جامعة الملك فيصل

عمادة التعلم الإلكتروني والتعليم عن بعد

كلية الآداب

اسم المقرر

الإحصاء الاجتماعي

أستاذ المقرر

د. سعيد سيف الدين

عمادة التعلم الإلكتروني والتعليم عن بُعد

المحاضرة الرابعة عشرة

التعريفات والعلاقات الأساسية لمقرر الإحصاء الاجتماعي

خطوات العملية الإحصائية :

- جمع البيانات : هي عملية الحصول على القياسات الخاصة بظاهرة معينة وعادةً ما تُسمى البيانات المجمعة بالبيانات الخام
- تنظيم وعرض البيانات : هي عملية وضع البيانات المجمعة في جداول خاصة وعرضها بطرق مناسبة
- تحليل البيانات : هي عملية إيجاد مقاييس تتحدد قيمها من البيانات وتعطي بعض الدلالات عن الظاهرة تحت الدراسة
- استقراء النتائج واتخاذ القرارات : هي الاستنتاجات التي يتوصل إليها الباحث من خلال تحليله للبيانات وعادةً ما تكون على شكل تقديرات أو تنبوءات أو تعميمات أو قرارات بالرفض أو القبول

أقسام علم الإحصاء :

- الإحصاء الوصفي : هو العلم الذي يهتم بجمع وتبويب وعرض ووصف البيانات وحساب بعض المقاييس الخاصة بها دون الوصول إلى نتائج أو استدلالات خاصة
- الإحصاء الاستقرائي أو الاستدلال الإحصائي أو الإحصاء الاستدلالي : هو العلم الذي يختص يبحث في استقراء النتائج واتخاذ القرارات

بعض المصطلحات الإحصائية :

- المفردة : وهى تعنى شخص أو مكان أو وقت أو تاريخ أو أى شئ محل الاهتمام فى الدراسة قابل للعد أو القياس ، وهى بمثابة العنصر أو الوحدة الواحدة .
- المتغير : وهو يعني أية خاصية كمية أو وصفية تأخذ مفرداتها قيماً مختلفة عند قياسها أو مشاهدتها .
- المجتمع : (والمقصود به المجتمع الإحصائي) وهو عبارة عن جميع المفردات المكونة لمجموعة البيانات محل الاهتمام .
- العينة : هي جزء فقط من مفردات المجتمع الإحصائي يُختار لتُجرى عليه الدراسة .
- المعلمة : هى قياس لأحد المتغيرات الخاصة بالمجتمع حيث يتم استخدام جميع مفردات المجتمع في قياسها .
- الإحصاءة : هى قياس لأحد المتغيرات الخاصة بالمجتمع حيث يتم استخدام مفردات عينة من المجتمع في قياسها والتي تكون تقديراً لمعلمة المجتمع .
- البيانات : هي مجموعة من "المشاهدات أو القياسات" التي تخص الظاهرة تحت الدراسة التي نقوم بها

مصادر البيانات : ثلاثة مصادر أساسية : مصادر تاريخية – الملاحظة – المصادر الميدانية .

أدوات جمع البيانات : الاختبارات والمقاييس – الاستبيانات – المقابلة – الملاحظة – استطلاعات الرأي .

أنواع البيانات الإحصائية :

- ❑ بيانات وصفية (أو نوعية) : وهي البيانات التي لا يُوصف فيها المتغير **بعدد**
  - ▪ وصفية ترتيبية : وهي البيانات الوصفية التي يمكن ترتيبها تصاعدياً أو تنازلياً مثل تقديرات الطلاب في مقرراتهم الدراسية ، ......
  - ▪ وصفية إسمية : وهي البيانات الوصفية غير الترتيبية مثل الجنس ، الحالة الاجتماعية ، ....
- ❑ بيانات كمية : وهي البيانات التي يُوصف فيها المتغير **بعدد**
  - ▪ كمية متقطعة : وهي البيانات التي يمكن فيها أن يأخذ المتغير قيمتين لكن لا يمكن أن يأخذ جميع القيم بين هاتين القيمتين ، مثل عدد أفراد الأسرة ، .....
  - ▪ كمية متصلة : وهي البيانات التي يمكن فيها أن يأخذ المتغير قيمتين وكذلك يمكنه أن يأخذ جميع القيم بين هاتين القيمتين ، مثل العمر ، الوزن ، الطول ، .....

تصنيف متغيرات الدراسة :

- ❑ متغير وصفي (أو نوعي) : أي لا يمكن التعبير عنه **بعدد** [مثل لون العين أو تقدير الطلاب]
- ❑ متغير كمي (متقطع أو متصل) : أي يمكن التعبير عنه **بعدد** مثل الطول والوزن أو عدد أفراد الأسرة

وسواء كان المتغير وصفياً أو كمياً فإنه قد يكون على أحد الأنواع التالية

مستقل	وهو المتغير الذي يقوم الباحث بتغييره ودراسة تأثير تغيره على متغير آخر (المتغير التابع)
تابع	وهو المتغير الذي يقوم الباحث بدراسة تغيره وهو يتأثر بالمتغير المستقل
دخيل	وهو المتغير الذي قد يؤثر على الدراسة لكنه غير مضبوط ويحاول الباحث ضبطه

أساليب إجراء البحث الميداني :

- ❑ أسلوب الحصر الشامل : يعتمد فيه البحث على دراسة جميع مفردات المجتمع الإحصائي
- ❑ أسلوب العينة : يعتمد فيه البحث على دراسة جزء فقط من مفردات المجتمع الإحصائي

طرق اختيار العينة :

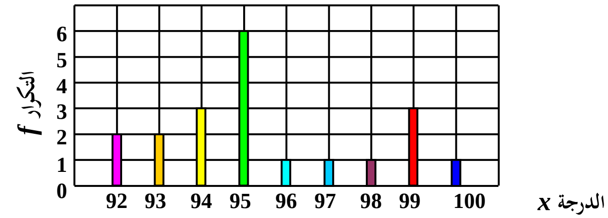
- ❑ الطريقة العشوائية (الاحتمالية) :
  - ▪ العينة العشوائية البسيطة .
  - ▪ العينة العشوائية المنتظمة .
  - ▪ العينة العنقودية .
  - ▪ العينة الطبقية .
- ❑ الطريقة غير العشوائية (غير الاحتمالية) :
  - ▪ العينة الصدفة
  - ▪ العينة العمدية (أو القصدية أو الغرضية)
  - ▪ العينة الحصية (أو الحصصية)

العرض البياني للبيانات المنفصلة (وصفية أو كمية متقطعة) :

❑ لظاهرة واحدة :

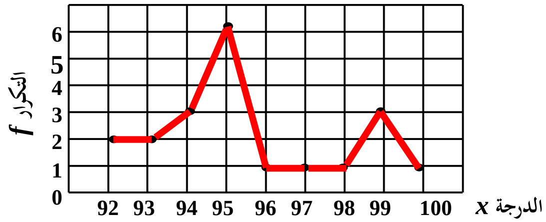
في هذه الطريقة تُمثل كل قيمة من قيم المتغير بعمود (أو خط رأسي) طوله يُعبر عن تكرار تلك القيمة

طريقة الأعمدة البيانية البسيطة



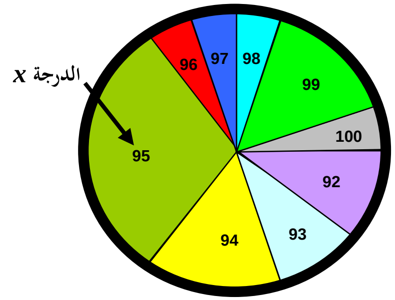
حيث تُمثل كل قيمة من قيم المتغير بنقطة إحداثيها الأفقي هو قيمة المتغير وإحداثيها الرأسي هو قيمة التكرار المناظر لتلك القيمة ثم تُوصل هذه النقاط بخط منكسر (بالمسطرة) .

طريقة الخط البياني المنكسر



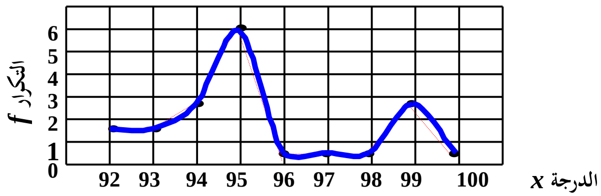
حيث تُمثل كل قيمة من قيم المتغير بقطاع من دائرة زاويته المركزية تتحدد طبقاً لتكرار تلك القيمة .

طريقة الدائرة البيانية



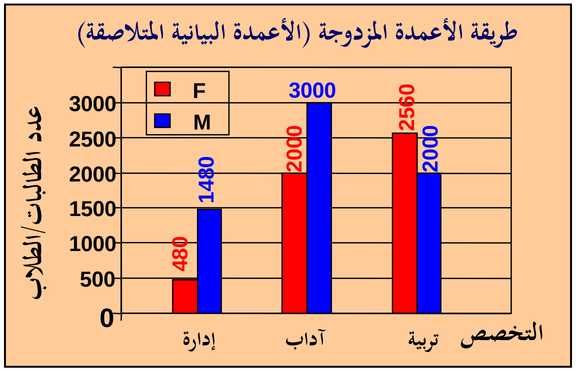
حيث تُمثل كل قيمة من قيم المتغير بنقطة إحداثيها الأفقي هو قيمة المتغير وإحداثيها الرأسي هو قيمة التكرار المناظر لتلك القيمة ثم تُوصل هذه النقاط بخط ممهد (باليد).

طريقة المنحنى البياني البسيط

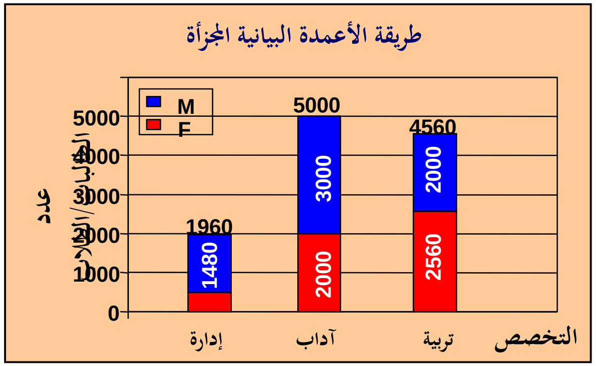


❑ لظاهرتين (أو أكثر) :

أي أن كل تخصص يُمثل بعمود مزدوج مكون من عمودين متلاصقين



أي أن كل تخصص يمثل بعمود طوله يعبر عن مجموع عدد طالباته وطلابه معاً ثم يتم تجزئته إلى عمودين كل منهما يمثل فئة من الفئات



عمادة التعلم الإلكتروني والتعليم عن بعد
Deanship of E-Learning and Distance Education

العرض البياني للبيانات المتصلة (كمية متصلة) :

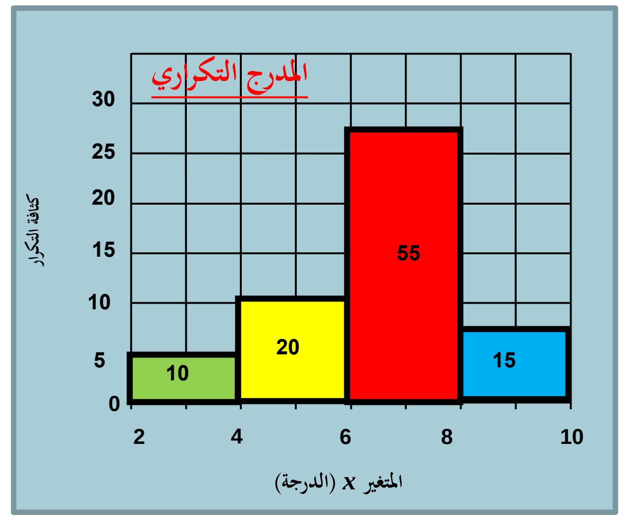
- ❑ **المدرج التكراري :** حيث تُمثل كل فئة بمستطيل عرضه يساوي طول الفئة ومساحته تساوي تكرار الفئة وارتفاعه يساوي كثافة التكرار ، والمستطيلات متلاصقة .
- ❑ **المضلع التكراري :** حيث تُمثل كل فئة بنقطة : إحداثيها الأفقي هو مركز الفئة ، وإحداثيها الرأسي هو كثافة تكرارها ، ثم نقوم بتوصيل هذه النقاط بخط منكسر (بالمسطرة) .
- ❑ **المنحنى التكراري :** حيث تُمثل كل فئة بنقطة : إحداثيها الأفقي هو مركز الفئة ، وإحداثيها الرأسي هو كثافة تكرارها ، ثم نقوم بتوصيل هذه النقاط بخط ممهد (باليد) .

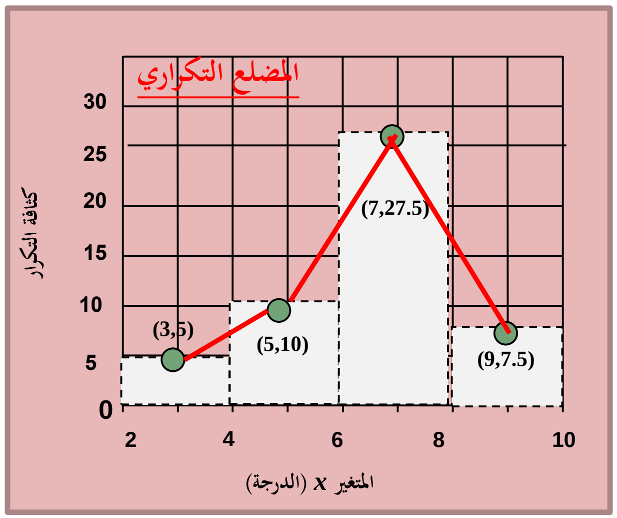
فمثلاً : إذا كان الجدول المقابل يبين التوزيع التكراري لدرجات 100 طالب في اختبار شهري لمادة اللغة الإنجليزية بأحد فصول مدرسة ثانوية (الدرجة العظمى 10) . هذه البيانات يمكن تمثيلها بطرق عديدة كالآتي :

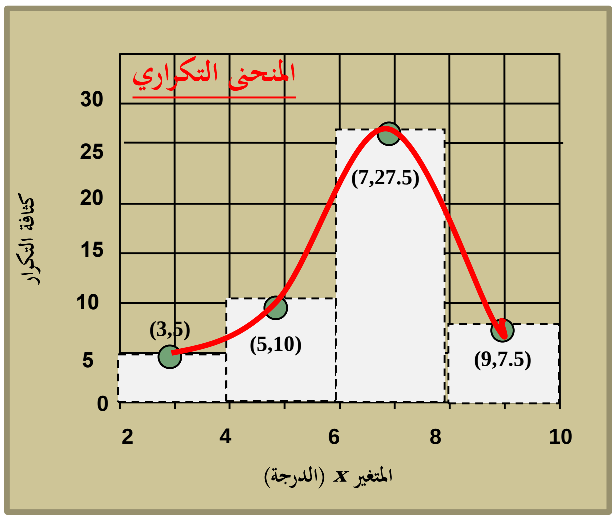
الجدول التكراري

النقطة الممثلة	مركز الفئة	كثافة التكرار f/c	طول الفئة c	التكرار f	المتغير x (الدرجة)
(3,5)	3	5	2	10	$2 \le x < 4$
(5,10)	5	10	2	20	$4 \le x < 6$
(7,27.5)	7	27.5	2	55	$6 \le x < 8$
(9,7.5)	9	7.5	2	15	$8 \le x < 10$

لا فراغات موجودة بين المستطيلات [حيث أن البيانات هنا بيانات متصلة] بخلاف طريقة الأعمدة في حالة البيانات غير المبوبة حيث يجب ألا تكون الأعمدة متلاصقة .







مقاييس النزعة المركزية : هي قيم نموذجية يمكن أن تمثل مجموعة من البيانات بحيث تعطي دلالة معينة لتلك البيانات . وحيث أن مثل هذه القيم تميل إلى الوقوع في المركز داخل مجموعة البيانات (عند ترتيبها حسب قيمها) ، فإن هذه القيم سميت بهذا الاسم . وهناك صور عديدة من هذه المقاييس وإن كان الأكثر شيوعاً :

الوسط الحسابي (أو باختصار الوسط أو المتوسط) – الوسيط – المنوال (الشائع)

(1) الوسط الحسابي

- **لبيانات غير متصلة :**

$$\text{الوسط الحسابي} = \frac{\text{مجموع قيم البيانات}}{\text{عددها}} = \frac{\sum x}{n} = \bar{x}$$

مثال : أوجد الوسط الحسابي للقيم

9 , 2 , 7 , 12 , 10

$$\bar{x} = \frac{\sum x}{n} = \frac{9+2+7+12+10}{5} = \frac{40}{5} = 8$$

مثال آخر : أوجد الوسط الحسابي لقيم x المبينة بالجدول التكراري المعطى.

التكرار f	المتغير x
6	5
2	3
2	6
5	4
2	2
3	8

$$\bar{x} = \frac{\sum fx}{\sum f}$$

المتغير x	التكرار f	fx
5	6	30
3	2	6
6	2	12
4	5	20
2	2	4
8	3	24
	20	96
	$\sum f = 20$	$\sum fx = 96$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{96}{20} = 4.8$$

- **لبيانات متصلة :**

مثال : أوجد الوسط الحسابي لقيم x المبينة بالجدول التكراري المعطى.

الجدول التكراري

الفئة	المتغير x	التكرار f
الأولى	$0 \le x < 20$	4
الثانية	$20 \le x < 30$	16
الثالثة	$30 \le x < 35$	12
الرابعة	$35 \le x < 40$	10
الخامسة	$40 \le x < 50$	6
السادسة	$50 \le x < 60$	2

$$\bar{x} = \frac{\sum f\,x_0}{\sum f}$$

الجدول التكراري

الفئة	المتغير x	التكرار f	مركز الفئة x_0	fx_0
1	$0 \le x < 20$	4	10	40
2	$20 \le x < 30$	16	25	400
3	$30 \le x < 35$	12	32.5	390
4	$35 \le x < 40$	10	37.5	375
5	$40 \le x < 50$	6	45	270
6	$50 \le x < 60$	2	55	110
		$\sum f = 50$		$\sum f\,x_0 = 1585$

$$\bar{x} = \frac{\sum f\,x_0}{\sum f} = \frac{1585}{50} = 31.7$$

خواص هامة للوسط الحسابي :

1. يمكن تحديد قيمة الوسط الحسابي بالضبط لأي بيانات كمية (ميزة) لكن لا يمكن تحديده للبيانات النوعية (عيب) ولا للفئات المفتوحة (عيب).
2. يأخذ في الاعتبار جميع البيانات (ميزة) ولا يتأثر بترتيب القيم (ميزة) .
3. يتأثر بالقيم المتطرفة في البيانات (عيب) .
4. إذا أضفنا (أو طرحنا) عدد ثابت c لكل قيمة من قيم البيانات ، فإن الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم القديمة + (أو −) هذا الثابت c
5. إذا ضربنا (أو قسمنا) كل قيمة من قيم البيانات في (على) عدد ثابت c ، فإن الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم القديمة مضروباً في (مقسوماً على) هذا الثابت c

(2) الوسيط :

(ببساطة) يُعرف الوسيط لمجموعة من القيم (المرتبة تصاعدياً أو تنازلياً حسب قيمها) على أنه القيمة التي تقسم مجموعة القيم إلى مجموعتين متساويتين في العدد ، أو بتعبير آخر هي القيمة التي في المنتصف . بتعبير آخر هي تلك القيمة التي تقسم مجموعة القيم بحيث تقع 50% من القيم تحتها ، 50% من القيم فوقها .

- لبيانات غير متصلة : مثال : مجموعة الأرقام (بعد ترتيبها تصاعدياً) : 9 7 6 6 5 4 3 3 2 وسيطها هو **5** [عدد القيم $n = 9$ (فردي)]

  في حين مجموعة الأرقام (بعد ترتيبها تصاعدياً) : 24 20 13 11 9 7 4 3 وسيطها هو $\frac{9+11}{2} = 10$ [عدد القيم $n = 8$ (زوجي)]

- لبيانات متصلة : لتحديد قيمة الوسيط لتوزيع تكراري كما هو مبين نتبع الخطوات التالية :

الخطوة الأولى : نحدد الفئة الوسيطية (أي الفئة التي يقع داخلها الوسيط)

- احسب $\frac{1}{2}\sum f$ ⟵ $\frac{1}{2}\sum f = \frac{70}{2} = 35$
- نبدأ بالصفر [في ذهننا] ، نزود على الصفر السابق تكرار الفئة الأولى [**14**] ينتج **14**

14 أقل من **35** ، يبقى الفئة الأولى ليست الفئة الوسيطية

- نزود على الـ **14** الأخيرة تكرار الفئة الثانية [**29**] ينتج **43**

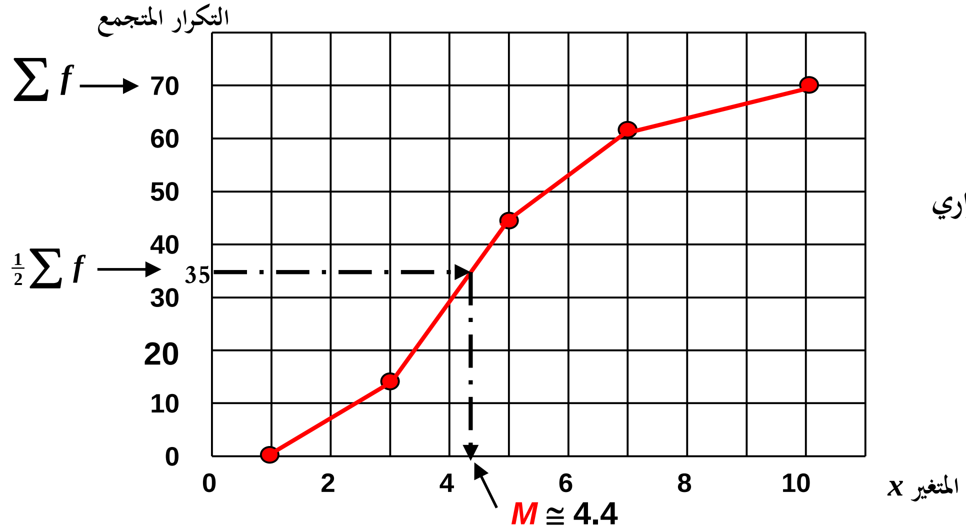
43 أكبر من **35** ، يبقى الفئة الثانية هي الفئة الوسيطية

الجدول التكراري

الفئة	المتغير x	التكرار f
1	$1 \le x < 3$	14
2	$3 \le x < 5$	29
3	$5 \le x < 7$	18
4	$7 \le x < 10$	9
		$\sum f = 70$

الخطوة الثانية : نقوم بحساب الوسيط من العلاقة :

الوسيط M = الحد الأدنى للفئة الوسيطية + [(نصف مجموع التكرارات – التكرار المتجمع السابق) / تكرار الفئة الوسيطية × طول الفئة الوسيطية]



أو نقوم بتحديده من المضلع التكراري المتجمع الصاعد (أو الهابط)

(3) المنوال (الشائع) : يُعرف المنوال لمجموعة من القيم على أنه القيمة التي تتكرر أكثر من غيرها أو القيمة الأكثر شيوعاً [لذا يُسمى في بعض الأحيان بالشائع.

فمثلاً مجموعة القيم : 18 12 11 10 10 9 9 9 7 5 2 2 لها منوال 9

ومجموعة القيم : 18 15 12 10 8 5 3 9 ليس لها منوال (أو عديمة المنوال)

ومجموعة القيم : 9 7 7 7 5 5 4 4 4 3 2 لها منوالان (4 , 7)

أي أن مجموعة القيم قد تكون وحيدة المنوال [لها منوال واحد] ، وقد تكون عديدة المنوال [منوالان أو أكثر] وقد تكون عديمة المنوال [لا يوجد لها منوال]

والمنوال [مقارنةً بالوسط الحسابي والوسيط] به العديد من العيوب منها :

- أنه لا يأخذ في الاعتبار جميع البيانات ولكنه يهتم فقط بالقيم الأكثر تكراراً .
- أنه قد لا يتواجد أو قد يكون هناك أكثر من منوال للبيانات .

سيارات في أحد المواقف	
عدد السيارات	لون السيارة
10	أحمر R
23	أزرق B
12	أبيض W
5	أصفر Y

بيانات نوعية
لها منوال وهو "اللون الأزرق"

إلا أنه أيضاً يتميز ببعض المزايا منها :

- أنه أسرع في تحديده من الوسط والوسيط
- من الممكن تحديده للتوزيعات التكرارية للبيانات غير المبوبة سواء كانت تلك البيانات كمية متقطعة أو نوعية [والبيانات الأخيرة (النوعية) لا يمكن حساب الوسط الحسابي لها أو الوسيط]

والمنحنيات التكرارية وحيدة المنوال والبسيطة الالتواء تحقق العلاقة الاعتبارية التالية :

الوسط – المنوال = 3 × (الوسط – الوسيط)

وللمنحنيات المتماثلة يكون : الوسط الحسابي = الوسيط = المنوال

وللمنحنيات الملتوية قليلاً لليمين يكون : الوسط الحسابي أكبر من الوسيط أكبر من المنوال

وللمنحنيات الملتوية قليلاً لليسار يكون : الوسط الحسابي أصغر من الوسيط أصغر من المنوال

مقاييس التشتت :

هي مقاييس تقيس الدرجة التي تتجه بها البيانات الكمية للانتشار حول قيمة متوسطة (أحد مقاييس النزعة المركزية) تُسمى تشتت أو تغير البيانات ، ومن أكثر المقاييس شيوعا لقياس هذا التشتت :
المدي – الانحراف المتوسط – الانحراف المعياري – الانحراف الربيعي [نصف المدى الربيعي] – معامل الاختلاف (مقياس للتشتت النسبي) .

أولاً : المدى R :

مدى مجموعة من البيانات الكمية هو الفرق بين أكبر قيمة في البيانات وأقل قيمة فيها

فمثلاً لمجموعة القيم : 15 3 7 6 12 18 5 3 13 15 ← المدى : $R = 18 - 3 = 15$

وللتوزيع التكراري المبين يكون

التكرار f	المتغير x
6	15
2	3
2	20
5	4

$R = 20 - 3 = 17$

أما للتوزيع التكراري المبين فإن

$R = 10 - 1 = 9$

الجدول التكراري

التكرار f	المتغير x	الفئة
14	$1 \le x < 3$	1
29	$3 \le x < 5$	2
18	$5 \le x < 7$	3
9	$7 \le x < 10$	4

وبالرغم من بساطة تحديده إلا أن له بعض العيوب مثل تأثره بالقيم المتطرفة ، كما أنه لا يمكن تحديده في حالة التوزيعات التكرارية المفتوحة .

العمر x	الفئة
$x < 6$	1
$6 \le x < 12$	2
$12 \le x < 15$	3
$15 \le x < 18$	4

مفتوح من أسفل

العمر x	الفئة
$6 \le x < 12$	1
$12 \le x < 15$	2
$15 \le x < 18$	3
$x \ge 18$	4

مفتوح من أعلى

العمر x	الفئة
$x < 6$	1
$6 \le x < 12$	2
$12 \le x < 15$	3
$x \ge 15$	4

مفتوح من الطرفين

لا يمكن تحديد مدى البيانات

ثانياً : الانحراف المتوسط [أو متوسط الانحرافات] : يُعرف الانحراف المتوسط (أو متوسط الانحرافات) [وسنرمز له بالرمز $M.D$] على أنه متوسط القيم المطلقة للانحرافات عن قيمة متوسطة للبيانات [عادةً تكون الوسط الحسابي] . أي أن :

$$M.D = \frac{\sum |d|}{n}$$

حيث $d = x - \bar{x}$ هي انحراف القيمة x عن الوسط الحسابي ، $|d|$ هي القيمة المطلقة للانحراف d .

فمثلاً لمجموعة القيم : 10 12 7 2 9

يمكننا حساب الانحراف المتوسط لها كالتالي :

x	d	$\|d\|$
9	1	1
2	-6	6
7	-1	1
12	4	4
10	2	2
40		14

$$\bar{x} = \frac{\sum x}{n} = \frac{40}{5} = 8$$

$$M.D = \frac{\sum |d|}{n} = \frac{14}{5} = 2.8$$

أما في حالة جدول تكراري كمل هو مبين ، فإن الانحراف المتوسط يتحدد من :

المتغير x	التكرار f
4	20
5	40
6	30
7	10

$$M.D = \frac{\sum f|d|}{\sum f}$$

الجدول التكراري

المتغير x	التكرار f	fx	$d = x - \bar{x}$	$\|d\|$	$f\|d\|$
4	20	80	4 – 5.3 = -1.3	1.3	20 × 1.3 = 26
5	40	200	5 – 5.3 = -0.3	0.3	40 × 0.3 = 12
6	30	180	6 – 5.3 = 0.7	0.7	30 × 0.7 = 21
7	10	70	7 – 5.3 = 1.7	1.7	10 × 1.7 = 17
	100	530			76
	$\sum f = 100$	$\sum fx = 530$			$\sum f\|d\|$

$$M.D = \frac{\sum f|d|}{\sum f} = \frac{76}{100} = 0.76$$

مزايا الانحراف المتوسط : من السهل حسابه – يأخذ في الاعتبار جميع البيانات – لا يحتاج لترتيب معين للبيانات

عيوب الانحراف المتوسط : يتأثر بشدة بالقيم المتطرفة – لا يمكن حسابه للتوزيعات التكرارية المفتوحة

أما في حالة البيانات الكمية المتصلة (كما في التوزيع التكراري المبين) ، تُستخدم نفس العلاقة السابقة لتحديد الانحراف المتوسط $M.D$ ، أي يكون :

$$M.D = \frac{\sum f \times |d|}{\sum f}$$

حيث $\boxed{d = x_0 - \bar{x}}$ ، x_0 تمثل مراكز الفئات

الفئة	المتغير x	التكرار f
الأولى	$0 \le x < 20$	4
الثانية	$20 \le x < 30$	16
الثالثة	$30 \le x < 35$	12
الرابعة	$35 \le x < 40$	10
الخامسة	$40 \le x < 50$	6
السادسة	$50 \le x < 60$	2

فمثلاً ، للتوزيع التكراري المبين ، يمكن حساب الانحراف المتوسط كالتالي :

إتجاه الحل

من هنا بداية الحل

الفئة	المتغير x	التكرار f	المركز x_0	fx_0
الأولى	$0 \le x < 20$	4	10	40
الثانية	$20 \le x < 30$	16	25	400
الثالثة	$30 \le x < 35$	12	32.5	390
الرابعة	$35 \le x < 40$	10	37.5	375
الخامسة	$40 \le x < 50$	6	45	270
السادسة	$50 \le x < 60$	2	55	110
		50		1585
		$\sum f$		$\sum fx_0$

$$\bar{x} = \frac{\sum f\, x_0}{\sum f} = \frac{1585}{50} = \underline{\underline{31.7}}$$

خاص بحساب الوسط الحسابي

$d = x_0 - \bar{x}$	$\|d\|$	$f \times \|d\|$
10 – 31.7 = -21.7	21.7	86.8
25 – 31.7 = -6.7	6.7	107.2
32.5 – 31.7 = 0.8	0.8	9.6
37.5 – 31.7 = 5.8	5.8	58
45 – 31.7 = 13.3	13.3	79.8
55 – 31.7 = 23.3	23.3	46.6
		388
		$\sum f\,\|d\|$

$$\therefore \quad M.D = \frac{\sum f\,|d|}{\sum f} = \frac{388}{50} = \underline{\underline{7.76}}$$

وهذا الجزء يُضاف إذا كان مطلوباً حساب الانحراف المتوسط

ثالثاً : التباين s^2 والانحراف المعياري s : يُعرف متوسط مربعات الانحرافات عن الوسط الحسابي على أنه تباين مجموعة البيانات [ويُرمز له بالرمز s^2] ، ويُعرف الجذر التربيعي للتباين على أنه الانحراف المعياري للبيانات [ويُرمز له بالرمز s] ، أي أن :

التباين $= s^2 = \dfrac{\sum d^2}{n}$

ومنه يكون

الانحراف المعياري $= s = \sqrt{s^2} = \sqrt{\dfrac{\sum d^2}{n}}$

فمثلاً لمجموعة القيم : 10 12 7 2 9

يمكننا حساب الانحراف المتوسط لها كالتالي :

x	d	d^2
9	1	1
2	-6	36
7	-1	1
12	4	16
10	2	4
40		58

$$\bar{x} = \frac{\sum x}{n} = \frac{40}{5} = \underline{\underline{8}}$$

$$s^2 = \frac{\sum d^2}{n} = \frac{58}{5} = 11.6$$

$$s = \sqrt{11.6} \cong \underline{\underline{3.4}}$$

أما في حالة جدول تكراري كمل هو مبين ، فإن التباين يتحدد من :

المتغير x	التكرار f
4	20
5	40
6	30
7	10

$$s^2 = \frac{\sum fd^2}{\sum f}$$

ومنه يمكن تحديد الانحراف المعياري s

الجدول التكراري

المتغير x	التكرار f	fx	$d = x - \bar{x}$	d^2	fd^2
4	20	80	4 – 5.3 = -1.3	1.69	20 × 1.69= 33.8
5	40	200	5 – 5.3 = -0.3	0.09	40 × 0.09= 3.6
6	30	180	6 – 5.3 = 0.7	0.49	30 × 0.49= 14.7
7	10	70	7 – 5.3 = 1.7	2.89	10 × 2.89= 28.9
	100	530			81
	$\sum f = 100$	$\sum fx = 530$			$\sum fd^2$

$$\therefore \quad s^2 = \frac{\sum fd^2}{\sum f} = \frac{81}{100} = 0.81 \Longrightarrow s = \sqrt{s^2} = \sqrt{0.81} = \underline{\underline{0.9}}$$

ومزايا وعيوب الانحراف المعياري هي نفسها مزايا وعيوب الانحراف المتوسط ؛ أي أن

مزايا الانحراف المعياري : من السهل حسابه – يأخذ في الاعتبار جميع البيانات – لا يحتاج لترتيب معين للبيانات

عيوب الانحراف المعياري : يتأثر بشدة بالقيم المتطرفة – لا يمكن حسابه للتوزيعات التكرارية المفتوحة

التكرار f	المتغير x	الفئة
4	$0 \le x < 20$	الأولى
16	$20 \le x < 30$	الثانية
12	$30 \le x < 35$	الثالثة
10	$35 \le x < 40$	الرابعة
6	$40 \le x < 50$	الخامسة
2	$50 \le x < 60$	السادسة

أما في حالة البيانات الكمية المتصلة (كما في التوزيع التكراري المبين) ، تُستخدم نفس العلاقات السابقة لتحديد التباين والانحراف المعياري ، أي يكون :

$$\text{التباين} = s^2 = \frac{\sum fd^2}{\sum f}$$

ومنه يكون

$$\text{الانحراف المعياري} = \sqrt{\frac{\sum fd^2}{\sum f}} = \sqrt{s^2} = s$$

حيث $\boxed{d = x_0 - \bar{x}}$ ، x_0 تمثل مراكز الفئات . فمثلاً ، للتوزيع التكراري المبين ، يكون :

إتجاه الحل

x

الفئة	المتغير x	التكرار f	المركز x_0	fx_0	$d = x_0 - \bar{x}$	d^2	$f \times d^2$
الأولى	$0 \le x < 20$	4	10	40	10 – 31.7 = -21.7	470.89	4 × 470.89 = 1883.56
الثانية	$20 \le x < 30$	16	25	400	25 – 31.7 = -6.7	44.89	16 × 44.89 = 718.24
الثالثة	$30 \le x < 35$	12	32.5	390	32.5 – 31.7 = 0.8	0.64	12 × 0.64 = 7.68
الرابعة	$35 \le x < 40$	10	37.5	375	37.5 – 31.7 = 5.8	33.64	10 × 33.64 = 336.4
الخامسة	$40 \le x < 50$	6	45	270	45 – 31.7 = 13.3	176.89	6 × 176.89 = 1061.34
السادسة	$50 \le x < 60$	2	55	110	55 – 31.7 = 23.3	542.89	2 × 542.89 = 1085.78
		50		1585			$\sum fd^2 = 5093$
		$\sum f$		$\sum fx_0$			

من هنا بداية الحل

$$\therefore \quad \bar{x} = \frac{\sum f\, x_0}{\sum f} = \frac{1585}{50} = \underline{\underline{31.7}}$$

$$\text{التباين} = 101.86 = \frac{5093}{50} = \frac{\sum fd^2}{\sum f} = s^2$$

$$\text{الانحراف المعياري} = \underline{\underline{10.09}} \cong \sqrt{101.86} = \sqrt{s^2} = s$$

رابعاً : الانحراف الربيعي [نصف المدى الربيعي] : لمجموعة من البيانات يُعرف الانحراف الربيعي [أو نصف المدى الربيعي وسنرمز له بالرمز Q] كالآتي :

$$Q \text{ [نصف المدى الربيعي] الانحراف الربيعي} = \frac{1}{2}(Q_3 - Q_1)$$

الربيع الثالث (Q_3) الربيع الأول (Q_1)

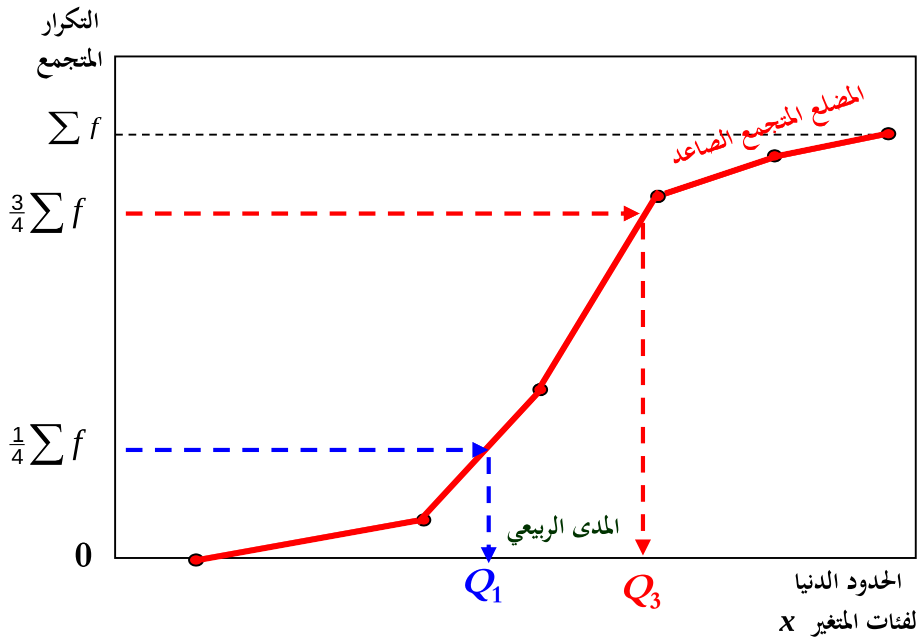
الربيعات :

Q_1 [الربيع الأول] هي تلك القيمة التي يقع تحتها 25% من القيم [وبالطبع وفوقها 75% من القيم]

Q_2 [الربيع الثاني] هي تلك القيمة التي يقع تحتها 50% من القيم [وبالطبع فوقها 50% من القيم] [أي الوسيط M]

Q_3 [الربيع الثالث] هي تلك القيمة التي يقع تحتها 75% من القيم [وبالطبع فوقها 25% من القيم]

ويكن تحديد الربيعين Q_1 (الأول) ، Q_3 (الثالث) بنفس الطريقة التي حددنا بها الوسيط M [الربيع الثاني Q_2] ، إلا أننا سنكتفي هنا بتحديدها [ومن ثم تحديد نصف المدى الربيعي Q] للبيانات الكمية المتصلة وبالطريقة التخطيطية (الرسم) كما هو مبين .



فيكون المدى الربيعي هو : $\boxed{Q_3 - Q_1}$ ونصف المدى الربيعي [أو الانحراف الربيعي] هو :

$$\boxed{Q = \tfrac{1}{2}(Q_3 - Q_1)}$$

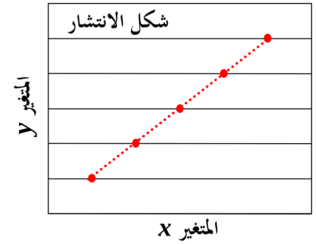
خامساً : معامل الاختلاف (معامل التشتت) :

$$\boxed{\text{معامل الاختلاف (كنسبة مئوية)} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100 = 100 \times \frac{s}{\bar{x}}}$$

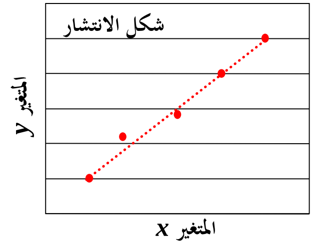
الارتباط الخطي :

يُعرف الارتباط الموجب (الطردي) بأنه علاقة بين متغيرين (x,y) بحيث إذا تغير أحد المتغيرين فإن المتغير الآخر يتبعه في نفس الاتجاه

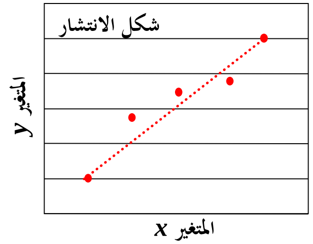
ويُعرف الارتباط السالب (العكسي) بأنه علاقة بين متغيرين (x,y) بحيث إذا تغير أحد المتغيرين فإن المتغير الآخر يتبعه في الاتجاه المضاد



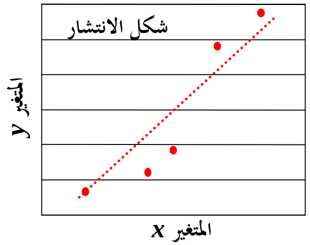
ارتباط طردي تام



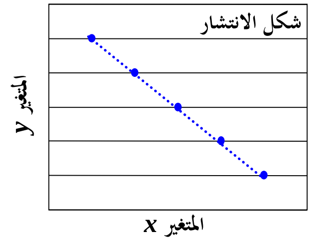
ارتباط طردي قوي



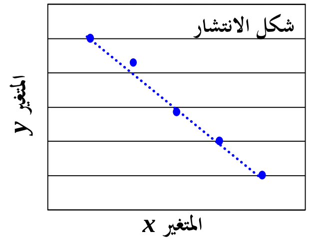
ارتباط طردي متوسط



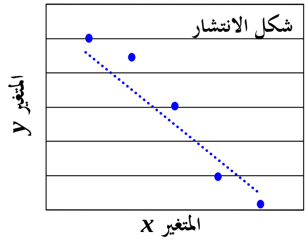
ارتباط طردي ضعيف



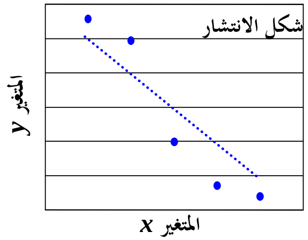
ارتباط عكسي تام



ارتباط عكسي قوي

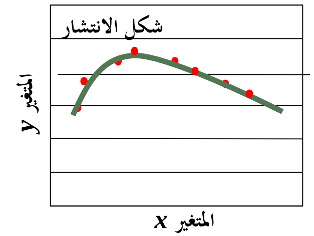


ارتباط عكسي متوسط

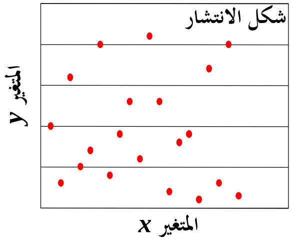


ارتباط عكسي ضعيف

أما إذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً وإنما يكون «ارتباط غير خطي»



وإذا لم يكن هناك ما يشير إلى وجود علاقة بين المتغيرين ، فإننا نقول إنه لا يوجد ارتباط بينهما أو أنهما غير مرتبطين .



ويُقاس الارتباط بين متغيرين x , y بما يُسمى بـ "معامل الارتباط" [وسنرمز له بالرمز r] وقيمته تكون محصورة بين -1 ، $+1$:

- فإذا كانت قيمته موجبه دل ذلك على أن الارتباط طردي
- وإذا كانت قيمته سالبة دل ذلك على أن الارتباط عكسي
- وإذا كانت قيمته صفراً دل ذلك على عدم وجود ارتباط

هذا بخصوص نوع الارتباط [طردي أم عكسي أم معدوم]

أما بخصوص قوة الارتباط فتحدده القيمة المطلقة لمعامل الارتباط كما يوضحه الجدول التالي :

القيمة المطلقة لمعامل الارتباط	قوة الارتباط
0	لا يوجد ارتباط خطي
$0 < r < 0.5$	ارتباط ضعيف
$0.5 \leq r < 0.7$	ارتباط متوسط
$0.7 \leq r < 1$	ارتباط قوي
1	ارتباط تام
> 1	كلام فارغ (خطأ في الحسابات)

أمثلة : إذا كان :

- $r = 0.45$ ← فهذا يعني ارتباط طردي ضعيف
- $r = -0.9$ ← فهذا يعني ارتباط عكسي قوي
- $r = 1.3$ ← فهذا يعني خطأ في الحسابات
- $r = 0.84$ ← فهذا يعني ارتباط طردي قوي
- $r = -0.22$ ← فهذا يعني ارتباط عكسي ضعيف
- $r = -1$ ← فهذا يعني ارتباط عكسي تام

ونذكر أن الإشارة الموجبة لمعامل الارتباط تعني أن الارتباط طردي (أو موجب) ، والإشارة السالبة تعني أنه عكسي (أو سالب)

عمادة التعلم الإلكتروني والتعليم عن بعد
Deanship of E-Learning and Distance Education

معامل بيرسون للارتباط الخطي البسيط : وهو يقيس قوة العلاقة واتجاهها ويُستخدم لمتغيرين كميين ، ويتم حسابه من العلاقة :

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

معامل ارتباط الرتب (معامل سبيرمان لارتباط الرتب) : وهو يقيس قوة العلاقة واتجاهها ويُستخدم لمتغيرين كميين أو متغيرين وصفيين ترتيبيين أو متغيرين أحدهما كمي والآخر وصفي ترتيبي ، ويتم حسابه من العلاقة :

$$\text{معامل سبيرمان لارتباط الرتب} = r_s = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

معامل بوينت بايسيريال للارتباط :وهو يقيس قوة العلاقة (وليس اتجاهها) بين متغيرين أحدهما كمي والآخر إسمي مستويين . وإشارته لا معنى لها ، ويتم حسابه من العلاقة :

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_2}{s}\sqrt{\frac{n_1 n_2}{n(n-1)}}$$

معامل الاقتران (معامل فاي أو معامل فاي للاقتران) : وهو يقيس قوة العلاقة (وليس اتجاهها) بين متغيرين إسميين كل منهما ثنائي التقسيم كالنوع (ذكر/أنثى) والإصابة بالمرض (مصاب/غير مصاب) والتدخين (مدخن/غير مدخن) .... الخ . وإشارته لا معنى لها ، ويتم احتسابه من العلاقة :

$$r_\phi = \frac{ad - bc}{\sqrt{E \times F \times G \times H}}$$

التحليل الإحصائي للبيانات السكانية :

أسس إجراء التعداد :

- الأساس الفعلي (أو الواقعي) : ويتميز بالسهولة حيث يتم حصر الأشخاص في مكان وجودهم وقت التعداد بصرف النظر عن كونهم من سكان هذا المكان أصلاً أو زائرين بصفة مؤقتة .
- الأساس النظري (أو الحقيقي) : وهو يُعطي صورة صحيحة عن السكان الدائمين بكل منطقة حيث يتم حصر الأشخاص حسب محل إقامتهم بصرف النظر عن أماكن تواجدهم وقت التعداد

مقاييس هامة :

- كثافة السكان = عدد السكان في الدولة ÷ مساحة الدولة بالكيلو متر المربع — وهو مقياس يدل على درجة ازدحام الدولة بالسكان
- كثافة السكن = عدد السكان في الدولة ÷ عدد حجرات المساكن — وهو مقياس يوضح درجة الازدحام داخل المسكن
- معدل الزيادة السنوية في عدد السكان = (عدد السكان في سنة المقارنة – عدد السكان في سنة الأساس) ÷ عدد السنوات — وهو مقياس يساعد على تقدير عدد السكان في غير سنوات التعداد
- معدل المواليد الخام = (عدد المواليد الأحياء خلال عام ÷ عدد السكان منتصف العام) × 1000
- معدل الوفيات الخام = (عدد الوفيات خلال عام ÷ عدد السكان منتصف العام) × 1000
- معدل الزيادة الطبيعية الخام = معدل المواليد الخام – معدل الوفيات الخام

الأرقام القياسية للأسعار :

$$I_S = \frac{\sum P_1}{\sum P_0} \times 100$$ ⟵ الرقم القياسي البسيط

$$I_L = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$ ⟵ رقم لاسبير (الرقم القياسي المرجح بكميات سنة الأساس)

$$I_P = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$ ⟵ رقم باشي (الرقم القياسي المرجح بكميات سنة المقارنة)

$$I_F = \sqrt{I_L \times I_P}$$ ⟵ الرقم القياسي الأمثل

سنة الأساس		سنة المقارنة					
P_0	Q_0	P_1	Q_1	P_1Q_0	P_0Q_0	P_1Q_1	P_0Q_1
——	——	——	——	——	——	——	——
——	——	——	——	——	——	——	——
——	——	——	——	——	——	——	——
——		——		——	——	——	——
$\sum P_0$		$\sum P_1$		$\sum P_1Q_0$	$\sum P_0Q_0$	$\sum P_1Q_1$	$\sum P_0Q_1$

تمت
بحمد الله