ملاحظة : تحتوي هذه المطبوعة على محاور مكملة تتضمن المحتويات الاساسية لبرنامج الاحصاء، الموجهة إلى طلبة السداسي الثاني – ماستر تنظيم وعمل.

المحور الأول: اختبار كاي– تربيع X^2 لجودة التوفيق والاستقلال

يستخدم توزيع كاي– تربيع X^2 للاختبارات التالية:

1. إذا كانت التكرارات المشاهدة تختلف "معنويا" عن التكرارات المتوقعة عندما يكون عدد النواتج الممكنة أكثر من أثنين؛
2. إذا كان التوزيع الذي أخذت منه العينة ذا الحدين، أو الطبيعي، أو أي توزيع أخر؛
3. إذا كان المتغيران مستقلين أم لا.

ولأجل ذلك نعتمد على إحصائية X^2 المحسوبة من بيانات العينة المعطاة بالصيغة:

$$X^2 = \sum \frac{(f_o - f_e)2}{f_e}$$

f_o = التكرارات المشاهدة

f_e = التكرارات المتوقعة

فإذا كانت X^2 المحسوبة أكبر من القيمة الجدولية X^2 عند مستوى المعنوية و درجات الحرية المحددة (من الملحق الخاص بتوزيع X^2)، يُرفض الفرض العدمي H_0 لصالح الفرض البديل H_1.

أولا. اختبار X^2 لجودة التوفيق:

أ. لاختبار ما إذا كانت التكرارات المشاهدة تختلف "معنويا" عن التكرارات المتوقعة عندما يكون عدد النواتج الممكنة أكثر من أثنين، نمضي كمايلي:

- هناك فرضيتان H_0 ، H_1

H_0 : ليس هناك فرقا معنويا بين التكرارات المشاهدة والمتوقعة(f_o، f_e)

H_1 : هناك فرقا معنويا بين التكرارات المشاهدة والمتوقعة(f_o، f_e)

- درجات الحرية لاختبار جودة التوفيق: $df=c-m-1$

c= عدد الفئات

m= عدد معالم المجتمع التي يجري تقديرها من إحصائيات العينة.

n = حجم العينة الإجمالي

مثال: وجد محل تجاري من خبرته الماضية أن 30% من التلفزيونات المباعة من الحجم الصغير، 40% من الحجم المتوسط، 30% من الحجم الكبير، لتحديد حجم المخزون الواجب الاحتفاظ به من كل نوع، أخذ المدير عينة عشوائية من 100 من المبيعات الحديثة للتلفزيون فوجد أن منها 20 من النوع الصغير، 40 من النوع المتوسط، 40 من النوع الكبير. باستخدام مستوى معنوية 5%، يختبر المدير الفرض أن نمط المبيعات الماضي لازال سائدا (أي لا فرق بين التكرارات المشاهدة، المبيعات الحديثة والتكرارات المتوقعة، نمط المبيعات الماضي):

H_0 : لا فرق بين التكرارات المشاهدة، المبيعات الحديثة والتكرارات المتوقعة، نمط المبيعات الماضي

H_1 : هناك لا فرق بين التكرارات المشاهدة، المبيعات الحديثة والتكرارات المتوقعة، نمط المبيعات الماضي

$$X^2 = \sum \frac{(f_o - f_e)2}{f_e} = \frac{(20-30)2}{30} + \frac{(40-40)2}{40} + \frac{(40-30)2}{30}$$
$$= \frac{100}{30} + \frac{100}{40} = 5.83$$

$$Df=c-m-1=3-0-1=2$$

وحيث أنه لم يتم حساب أي من معالم المجتمع من البيانات فإن **m=0** .**df=2** تعني أننا إذا علمنا فئتين من الثلاث والمجموع، فإن الفئة الثالثة لا تكون "حرة" التغير. وحيث أن القيمة المحسوبة **X^2=5.83** أصغر من القيمة الجدولية **X^2=5.99** بمستوى معنوية **α =0,05** ودرجة حرية 2 (أنظر الملحق)، فإنه يجب قبول الفرضية الصفرية **H_0** بأن نمط المبيعات في الماضي ما زال سائدا.

ملاحظة: عندما يكون التكرار المتوقع في أي فئة أقل من 5 فإنه يجب ضمها لفئة مجاورة.

جدول: المشتريات المشاهدة والمتوقعة لأجهزة التلفزيون حسب حجم الشاشة.

	جحم الشاشة			الإجمالي
	صغير	متوسط	كبير	
النمط المشاهد f_o	40	40	20	100
النمط الماضي(المتوقع)f_e	30	40	30	100

ثالثا. اختبار X^2 لمعرفة مإذا كان المتغيران مستقلين أم لا

- **هناك فرضيتان** H_0 ، H_1

H_0 : ليس هناك فرقا معنويا بين التكرارات المشاهدة والمتوقعة(f_o، f_e) أي أن المتغيران مستقلان؛

H_1 : هناك فرقا معنويا بين التكرارات المشاهدة والمتوقعة(f_o، f_e) **أي أن المتغيران غير مستقلان**

- درجات الحرية لاختبارات الاستقلال أو اختبارات الاقتران معطاة بالصيغة التالية:

$$df=(r-1)(c-1)$$

r= عدد الصفوف في جدول الاقتران

c= عدد الاعمدة

$$f_e=\frac{\sum r \sum c}{n}$$

n= الحجم الاجمال للعينة

مثال: جمع تاجر سيارات البيانات الموضحة في الجدول1 عن عدد السيارات الأجنبية والمحلية التي يشتريها عملاء أعمارهم تحت سن 30 سنة، والتي يشتريها عملاء اعمارهم في سن 30 سنة فأكثر. لاختبار ما إذا كان نوع السيارة المشتراة (أجنبية، محلية) مستقلا عن سن المشتري عند مستوى 1%، ننشئ جدول التكرارات المتوقعة في الجدول 2

ولحساب قيمة التكرار المتوقع في الخلية الاولى صف 1 وعمود1 (طبعا بالاعتماد على جدول التكرارات المشاهدة):

$$f_e=\frac{\sum r \sum c}{n}=\frac{(70)(50)}{170}=21$$

الجدول1: جدول الاقتران لمشتري السيارات (التكرارات المشاهدة)

الاجمالي	نوع السيارات		السن
	أجنبية	محلية	
70	40	30	تحت سن 30
100	80	20	30 سنة فأكثر
170	120	50	إجمالي

ويمكن الحصول على التكرارات المتوقعة الثلاثة الباقية بالطرح من مجموع الصفوف ومجموع الأعمدة أو باستعمال طريقة مجاميع الصفوف والاعمدة.

درجة الحرية : df=(r−1)(c−1)=(2−1)(2−1)=1

الجدول 2:جدول التكرارات المتوقعة المناظرة للتكرارات المشاهدة في الجدول 2

الاجمالي	نوع السيارات		السن
	أجنبية	محلية	
70 100	49 71	**21** 29	تحت سن 30 30 سنة فأكثر
170	120	50	إجمالي

$$X^2 = \sum \frac{(f_o - f_e)2}{f_e} = \frac{(30-21)2}{21} + \frac{(20-29)2}{49} + \frac{(80-71)2}{71} = 9{,}44$$

وحيث أن قيمة ^{2} **المحسوبة تتجاوز قيمة** X^2 **عند 0,01= α** و df=1 (انظر الملحق) ، نرفض H_0 القائل بأن السن ليس عاملا في تحديد نوع السيارة المشتراة، أي أن هناك اقتران بين العمر ونوع السيارة المشتراة.

تمارين ومسائل محلولة

1. أختبار X^2 **لجودة التوفيق**

التمرين الاول:

اخذ مدير مصنع عينة عشوائية من 100 يوم من الاجازات المرجعية، ووجد أن 30% من القوة العاملة في المصنع في فئة العمر 20−29 سنة قد أخذوا أجازة مرضية 26 يوماً من الاجمالي 100يوم، و أن 40% من القوة العاملة في فئة العمر 30−39 قد أخذوا 37 يوماً، وأن 20% من فئة العمر 40−49 قد أخذوا 24 يوماً، وأن 10% في فئة العمر 50 فأكثر قد أخذوا 13 يوماً أجازة مرضية. كيف يمكن للمدير عند مستوى 5% أن يختبر الفرض أن العمر ليس عاملا في أخذ أجازة مرضية؟

الحل:

إذا كان العمر ليس عاملا، في أخذ أجازة مرضية، فإن العدد المتوقع للإيام المرضية التي يأخذها العاملون في كل فئة عمر يجب أن يكون بنفس عدد العاملين في كل فئة عمر إلى العدد الإجمالي للعاملين بالمصنع (كما هو موضح في الجدول)

الجدول: الأجازات المرضية المشاهدة و المتوقعة

فئة العمر	20−29	30−39	40−49	50 فأكثر	إجمالي
f_o	26	37	24	13	100
f_e	30	40	20	10	100

$$X^2 = \sum \frac{(f_o - f_e)2}{f_e}$$

$$= \frac{(26-30)2}{30} + \frac{(37-40)2}{40} + \frac{(24-20)2}{20} + \frac{(13-10)2}{10} = 2,46$$

درجات الحرية : df= c−m−1=3 ، وحيث أنه لم يتم تقدير أي معلمة من المجتمع، m=0,df= 3 بمعنى أننا إذا عرفنا ثلاث قيم من الفئات الأربع، فإن القيمة الرابعة ليست "حرة" أن تتغير. و حيث أن القيمة المحسوبة $X^2 = 2{,}46$ أصغر من القيمة الجدولية $X^2 = 7{,}81$ عند α=0,05 ودرجات حرية df=3(أنظر ملحق توزيع X^2) ، فإننا لا نستطيع أن نرفض H_0، بأن العمر ليس عاملا في أخذ أجازة مرضية. لاحظ أنه كما في حالة توزيع t فإن توزيع X^2 مختلفا لكل من درجات الحرية المختلفة. ولكن، اختبار ^{2} مختلفا لكل من درجات الحرية . ولكن، اختبار X^2 يستخدم هنا كاختبار الذيل الايمن فقط.

التمرين الثاني:

ألقيت نردة 60 مرة فأعطت النتائج التالية: ظهر العدد1، 12 مرة، ظهر العدد 2، 8 مرات، ظهر العدد 3، 13 مرة، ظهر العدد 4، 12 مرة، ظهر العدد 5، 7 مرات، وظهر العدد 6، 8 مرات. هل النردة متوازنة عند مستوى معنوية 5%.

الحل:

إن احتمال الحصول على عدد من الاعداد على 6 أوجه لنردة(1،2،3،4،5،6)هو:

$\frac{1}{6}$ وبالتالي فإن العدد المتوقع لعدد مرات ظهور العدد 5 يساوي :

60=10* $\frac{1}{6}$ ، أن العدد المتوقع لظهور أي عدد من الاعداد الستة هو 10

الجدول: التكرارات المشاهدة والتكرارات المتوقعة عند رمي نردة 60 مرة

	1	2	3"	4	5	6	إجمالي
f_o	12	8	13	12	7	8	60
f_e	10	10	10	10	10	10	60

$$X^2 = \sum \frac{(f_o - f_e)2}{f_e}$$
$$= \frac{(12-10)2}{10} + \frac{(8-10)2}{10} + \frac{(13-10)2}{10} + \frac{(12-10)2}{10} + \frac{(7-10)2}{10} + \frac{(8-10)2}{10} = 3,4$$

حيث أن قيمة X^2 المحسوبة أصغر من القيمة الجدولية $X^2 = 11,07$ عند α=0,05 و **df=5** فإننا نقبل H_0 و نرفض H_1 ، وبالتالي ليس هناك فرقا معنويا بين التكرارات المشاهدة والتكرارات المتوقعة ، وبالتالي النردة متوازنة.

2. أختبار X^2 للاستقلال والاقتران

التمرين الاول:

يوضح الجدول الاتي للاقتران عدد النوبات القلبية التي تعرض لها الذكور والاناث في فئات العمر المختلفة في مدينة ما. باستخدام مستوى معنوية 1% اختبر الفرض أن العمر و الجنس مستقلان فيما يتعلق بحدوث النوبات القلبية.

الجدول: عدد النوبات القلبية للذكور والاناث في فئات العمر المختلفة في إحدى المدن

فئات العمر	ذكور	إناث	إجمالي
أقل من 30	10	10	20
من 30 إلى 60	50	30	80
أكثر 60	30	20	50
إجمالي	90	60	150

لأختبار هذا الفرض، يجب تقدير التكرارات المتوقعة f_e (انظر الجدول التالي):

للخلية في الصف الاول**r** والعمود الأول **c** $f_e = \frac{\sum r \sum c}{n} = \frac{(20)(90)}{150} = 12$

للخلية في الصف الثاني **r** والعمود الاول **c** $f_e = \frac{\sum r \sum c}{n} = \frac{(80)(90)}{150} = 48$

ويمكن الحصول على باقي التكرارات المتوقعة بالطرح من مجموع الصف أو العمود المناظر، وعليه:

$$X^2 = \sum \frac{(f_o - f_e)2}{f_e}$$
$$= \frac{(10-12)2}{12} + \frac{(10-8)2}{8} + \frac{(50-48)2}{48} + \frac{(30-32)2}{32} + \frac{(30-30)2}{30} + \frac{(20-20)2}{20} = 1{,}04$$

درجات الحرية **$df=(r-1)(c-1)=(3-1)(2-1)=2$** , **$\alpha = 0{,}01$** ، وحيث أن X^2 **المحسوبة أصغر من X^2 الجدولية** $X^2 = 9.21$ ، **نقبل الفرض العدمي،** H_0 ، **أن العمر مستقل عن الجنس في حدوث النوبات القلبية.**

الجدول: التكرارات المتوقعة للنوبات القلبية

فئات العمر	ذكور	إناث	إجمالي
أقل من 30	12	8	20
من 30 إلى 60	48	32	80
أكثر 60	30	20	50
إجمالي	90	60	150

التمرين الثاني:

أعطت عينة عشوائية مكونة من 37 عاملا فوق سن 65 في مدينة ما النتائج الواردة في جدول الاقتران الاتي . باستخدام مستوى المعنوية 10% اختبر الفرض بأن عدد الإناث

والذكور من العاملين، في مجموعات السن 66−70 و 71 فأكثر، في المدينة مستقل عن الجنس.

جدول الاقتران1: العاملون من الذكور والإناث فوق سن 65 في المدينة

فئة العمر	ذكور	إناث	إجمالي
66− 70 سنة	9	17	26
71 فأكثر	8	3	11
إجمالي	17	20	37

ملاحظة: **50 >n** ، فيجب استخدام معامل التصحيح عند حساب X^2

جدول الاقتران 2 يعطي التكرارات المتوقعة، أما بالنسبة للخلية الأولى:

$$\mathbf{f_e}=\frac{\sum r \sum c}{n} = \frac{(26)(20)}{37} = \mathbf{14}$$

بالنسبة لباقي الخلايا، يمكن إيجاد $\mathbf{f_e}$ (أي باقي التكرارات المتوقعة) بالطرح من مجموع الصف العمود df=(r−1)(c−1)=1 وحيث أن df=1 , **50 >n** ، فيجب استخدام معامل التصحيح لحساب X^2، **كمايلي:**

$$X^2 = \sum \frac{(f_o - f_e)2}{f_e}$$
$$= \frac{(|17-14|-0,5)2}{12} + \frac{(|9-12|-0,5)2}{8} + \frac{(|3-6|-0,5)2}{48} + \frac{(|8-5|-0,5)2}{32} = 3,25$$

وحيث أن قيمة X^2 **المحسوبة أكبر من قيمة الجدولية عند , 0,01=α** ودرجات حرية **df=1**، فإننا ترفض الفرض $\mathbf{H_0}$ ، بأن الذكور والإناث فوق سن 65 يستمرون في العمل بصورة مستقلة عما إذا كانوا فوق أو تحت سن 70 في هذه المدينة. أن نسبة العاملين أعلى بدرجة جوهرية للذكور في فئة السن 66−70 وللإناث في فئة السن 71 فأكثر. لاحظ أن

نفس التعديل المشار إليه في معادلة حساب X^2 **يجب اجراؤه أيضا عند اختيار جودة التوفيق في حالة 50 >n،** df=1 ,

جدول الاقتران 2: العدد المتوقع للعاملين من الذكور والاناث فوق سن 65.

فئة العمر	ذكور	إناث	إجمالي
66- 70 سنة	12	14	26
71 فأكثر	5	6	11
إجمالي	17	20	37

المحور الثالث: مقاييس النزعة المركزية

يهدف هذا الفصل إلى التعريف بأهم المقاييس التي تسمح بتلخيص سلسلة مشاهدات إحصائية، إلى جانب العمل على توضيح مفهوم المركز في السلسلة . وتسمى هذه المقاييس خصائص الوضعية (caractéristiques de position) أو خصائص النزعة المركزية للسلسلة الإحصائية للمتغير. و سنتطرق في هذا الفصل إلى: المنوال، المتوسط و الوسيط.
و قد عرف الإحصائي الإنجليزي جورج يول G Yule (1871-1951) سنة 1911 الشروط المناسبة المطلوبة بالنسبة لقيمة مركزية:

- أن تكون معرفة بصفة موضوعية إنطلاقا من السلسلة؛
- أن تعتمد على كامل حدود السلسلة؛
- أن تكون مفهومة من طرف حتى غير الأخصائيين؛
- أن تكون بسيطة الحساب؛
- أن لا تتأثر بتغيرات المعاينة؛

- أن تكون قابلة للحسابات الجبرية.

لكن فيما يأتي من تعاريف، حسب الإحصائي يول، لا توجد أي قيمة من القيم المركزية " مثالية "(Antoine,1998, 75).

1. المنوال

عند ملاحظة التمثيل البياني لأحد التوزيعات الإحصائية (مخطط أعمدة أو مدرج تكراري)، يشتد تركيز العين على العمود أو المستطيل الأكثر ارتفاعا. من القيم النموذجية لسلسلة احصائية هو المنوال (القيمة المسيطرة).

و يبدو أن الشكل الجرسي للتوزيع الطبيعي، يعطينا نظرة جيدة عن هذا المعنى. فتعريفا المنوال هو القيمة للمتغير التي لها أكبر تكرار (أو التكرار النسبي) و يرمز له بـ Mo .

في بعض الميادين مثل الإقتصاد، في مشاكل التغذية، الدخل، السكن، و إلخ.، الفئة التي لها أكبر وزن هي المنوال. هذا الأخير يموقع جيدا مكان القيم الأكثر تواجدا (Droesbeke, 2001,198).

و ليس هناك مشكلة معقدة في تعيين المنول، لكن يجب أولا تمييز ما إذا كان المتغير كيفيا أو كميا منفصلا ، أم حالة المتغير المتصل.

و توجد سلاسل أحادية المنوال(unimodales) و سلاسل متعددة المنوال(plurimodales).

1. المتغيرات الكيفية و المتغيرات الكمية المنفصلة

إذا كان المتغير كيفيا أو كميا منفصلا، فيمكن تعريف المنوال بصفة مباشرة بالتعرف على تكرار المتغير الذي يناسب العدد الأكبر(أو النسبة الكبرى).

أما المنوال بالنسبة لسلسلة منفصلة فهو واحدا من قيم السلسلة. بيانيا، يشار إلى العمود الأطول(إلى الأعمدة الأطول في حال سلسلة متعددة المنوال).

2. المتغيرات الكمية المتصلة

إذا كان المتغير كميا متصلا، فيجب المضي وفق مرحلتين:

أ. تحديد الفئة المنوالية، أي الفئة مثلة في المدرج التكراري بأطول مستطيل: وهي الفئة ذات أكبر كثافة d_i (densité). أما إذا كانت الفئات متساوية الطول، فإن الفئة المنوالية فهي ذات أكبر تكرار (أو أكبر نسبة).

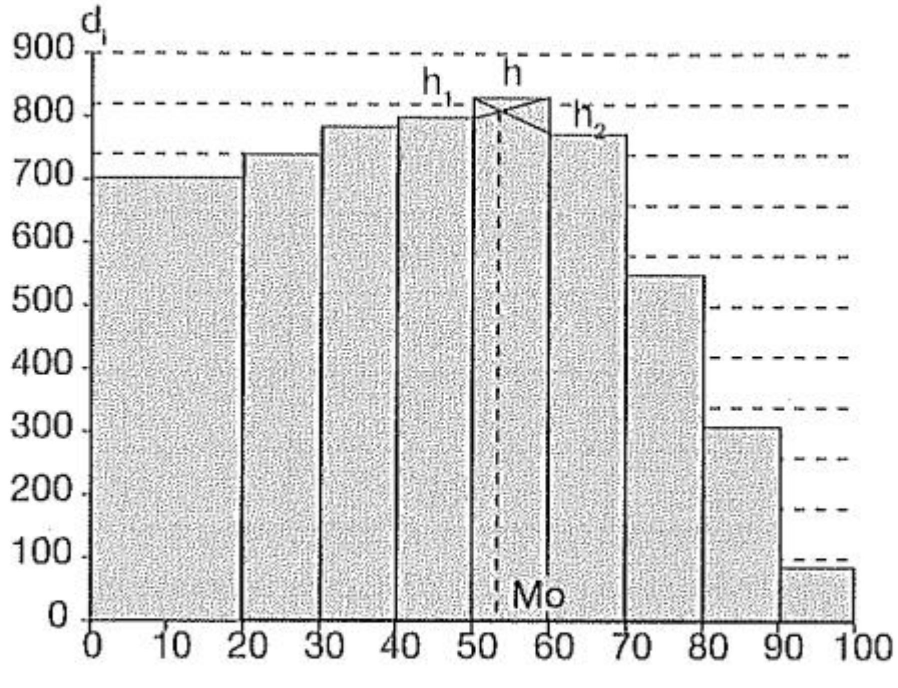
و يجب أن نذكر أن كثافة عدد في فئة i، برمز d_i، هو العلاقة $d_i=n_i/A_i$، بـ n_i التكرار و A_i طول الفئة. تمثل هذه الكثافة عدد المفردات حسب وحدة الطول(unité d'amplitude).

ب. تحديد المنوال داخل هذه الفئة.

من أجل تقدير أولي، يمكن اعتبار مركز الفئة المنوالية كمنوال. و تطبيقيا المنوال هو ترتيب أعلى نقطة يصلها منحنى كثافة التكرار النسبي(Bressoud, 2010,55).

تطبيقيا، لا نملك إلا المدرج التكراري. يمكن تقدير المنوال بطريقة la méthode des diagonales الشاقوليات .

شكل: المدرج التكراري للأعداد و تحديد للمنوال: الهيكلة العمرية في 2020



بيانيا، تمثل الفئة المنوالية قمة إرتفاع المدرج المصحح و المنوال يناسب عندئذ، ترتيب لنقطة تقاطع Deux diagonales. فيتم حسابه إذن باعتماد التكرار المصحح n_ic . (و هو المؤشر الوحيد الذي يحسب باعتماد التكرارات المصححة).

في حالة الفئات متساوية الطول، الفئة المنوالية هي الفئة ذات أكبر تكرار (أو أكبر نسبة). اما طريقة تحديد المنوال فيه نفسها.

لتكن] x_1 ; x_2 [الفئة المنوالية، h_1 وh_2 هي ارتفاعات (التكرار المصحح أو الكثافة) للمستطلين المجاورين لمستطيل المنوالي، h ارتفاع المستطيل المنوالي و Mo المنوال.

لحساب المنوال، يجب التخلي عن فرضية التوزيع المتساوي داخل الفئة المنوالية، التي تأخذ متوسط الفئة. فالفرضية المفضلة هنا تتمثل في وجود توزيع متأثر بالقيم h_1 وh_2 ، و أن المنوال يكون "منجذب" إتجاه المستطيل المجاور ذو أكبر كثافة. فيفترض أن الكثافة تتزايد من القيمة h_1 إلى اقصاها h و تتناقص من h إلى h_2 بنفس السرعة ، مما يعطي، بمعدلات زيادة:

$$\frac{h-h1}{Mo-X1}=\frac{h-h2}{X2-Mo}$$

يعني:

$\frac{K1}{Mo-X1}=\frac{K2}{X2-Mo}$, بـــ k_1 و k_2 الفروق : $k_1 = h-h_1$ et $k_2 = h-h_2$.

عند القيام بجداء الطرفين والوسطين بالزيادة: $Mo=\frac{K2X1+K1X2}{K1+K2}$

المنوال يظهر كوسط مرجح لـ x_1 وx_2 اللذين يتأثران على التوالي بالمعاملين h_1 وh_2 .

و توجد هناك معادلة مساوية للمنوال معطاة بـ:$Mo= x_1+\frac{K1}{K1+K2}(x_2-x_1)$. تبين هذه المعادلة تماما،مثلا، تحرك المنوال إلى في حال كان$k_1 < k_2$ ، بمعنى أين $0{,}5 > \frac{K1}{K1+K2}$.

مثال :حساب المنوال على متغير كمي متصل

لنعتبر أفاق البنية الديمغرافية لفرنسا في 2020 :

العمر	n_i	a_i	d_i
0- 19	14100	20	705
20−29	7400	10	740
30−39	7800	10	780
40−49	8000	10	800
50−59	8300	10	830
60−69	7700	10	770
70−79	5500	10	550
80−89	3100	10	310
90−99	900	10	90

Source : Insee, projections des ménages à l'horizon 2020 pour la France métropolitaine, juillet 2006.

بمأن أطوال الفئات ليست متساوية، فإننا سنعمد إلى استعمال الكثافات لتحديد الفئة المنوالية و تمثيل المدرج التكراري. فتكون الفئة المنوالية إذن الفئة 50−59 سنة، أي]50؛60 [بكثافة 830.

x_1 =50 ; x_2=60 ; h= 830 ; h_1=800 ; h_2 =770 ; k_1= 830-800=30 ; k_2=830-770=60.

بتطبيق معادلة المنوال:

Mo= $\frac{K2X1+K1X2}{K1+K2}$= $\frac{60*50+30*60}{30+60}$= 53,33 ans (أي 53 سنة و 4 أشهر)

ثانيا: المتوسطات

يمكن تعريف في هذا المجال أربعة نماذج من المتوسطات: المتوسطات الحسابية، ومعها المتوسطات الأخرى الأقل استعمالا، وهي المتوسطات الهندسية، التوافقية والرُبيعية. ويحتفظ

المتوسط الحسابي بدور أساسي باعتبار سهولة حسابه، وكذلك بسبب موضعه المهم في إطار نظرية أخطاء الملاحظة (loi de Laplace–Gauss)[1] وفي نظرية الانحدار.

إن الفكرة الأساسية لمفهوم المتوسط أنه يرمي إلى تمثيل معالمي غير متساوية بمعلمة واحدة لا تغير من الوضعية الإجمالية. فثلا في مؤسسة أين يتقاضى العمال أجورا مختلفة، تبقى كتلة الأجور ثابتة حين يتقاضى كل العمال نفس الأجر المتوسط(Schlacter, 1986, 70).

1. المتوسط الحسابي

لقد استتب المتوسط الحسابي، لأول مرة، في ميدان علم الفلك مع تيشو براي[2] Tycho Brahe لكن جوهان برنولي[3] J Bernoulli يعرفها في موسوعته على أنها" الوسط الواجب أخذه بين جملة الملاحظات".

هذا المتوسط، المتعلق بالجمع، هو المتوسط الأكثر شيوعا و استعمالا. فتمثل جيدا فكرة الوسط و التوازن.

بعض التعريفات

الوسط الحسابي يعبر عن مجموع القيم الملاحظة، قسمة على عدد الملاحظات، يرمز له بـ $\bar{X}$.

يحسب المتوسط الحسابي البسيط لـ n عدد حقيقي(أي معطيات في جدول خام) بقسمة مجموع القيم على عددها. لتكن $x_1, x_2, \ldots, x_n$ مشاهدات للمتغير X : المتوسط الحسابي يكتب :

$$\mathrm{X} = \frac{1}{n}\sum_{i=1}^{n} Xi$$

[1]Pierre Simon de Laplace (1749-1827), mathématicien, astronome français. Carl Friedrich Gauss (1777-1855), astronome, mathématicien allemand
[2]Tycho Brahe (1546-1601), astronome danois.
[3] Johann Bernoulli (1667-1748), mathématicien suisse.

هذه المعادلة تستلزم أن : $\sum_{i=1}^{n} Xi = nX$

المتوسط الحسابي المرجح لـ r حقيقي $x_1, x_2, \ldots, x_r$ (المعطيات في جدول إحصائي)، المعطاة على التوالي معاملات n_i، مثل $\sum_{i=1}^{r} ni = n$ و يكتب : $x = \frac{1}{n}\sum_{i=1}^{r} nixi = n$

أو كذلك $x = \frac{1}{n}\sum_{i=1}^{r} fixi$

بالنسبة للاحتمالات نتحدث عن أمل متغير عشوائي و نكتب متوسطا : $E(x) = \sum_{i=1}^{r} pixi$ أين تعوض الاحتمالات p_i النسب f_i .

2-حساب المتوسط في حالة المتغير المنفصل

مثال: حساب متوسط حساب مرجح

توزيع أعداد أطفال ما قبل الابتدائي في المؤسسات العمومية حسب السن خلال سنوات 2005-2006 كالتالي:

العمر	n_i	f_i (%)
2 سنة	154141	0,0702
3 سنوات	667328	0,3038
4 سنوات	685158	0,3119
5 سنوات	680202	0,3097
6 سنوات	9683	0,0044

Source :ministère de l'Education nationale, 2007.

يجب التنويه، مسبقا، أن العمر متغير متصل، لكن كثير من الإدارات تقدمه كمتغير منفصل وسنعالجه هنا على هذا الأساس. من ناحية أخرى سنأخذ 6 سنوات بالنسبة للترتيب الأخير.

لأجل حساب المتوسط، يجب حساب كل n_ix_i قبل حساب المجموع (كما في الجدول اللاحق).

يتم الحصول على المتوسط بقسمة مجموع لـ n_ix_i على عدد المفردات الإجمالي.

العمر المتوسط(الوَسطي) للمؤسسات العمومية ماقبل الإبتدائي هو: $x = \frac{8510006}{2196512} =$

سنة 3,87 أي 3 سنوات و10 أشهر. كما يمكن إيجاد هذه القيمة بحساب كل واحدة من f_ix_i ثم جمعها مع بعضها.

جدول: توزيع عدد الاطفال ما قبل الابتدائي

العمر(x_i)	n_i	$n_i\ x_i$	f_i	$f_i\ x_i$
2	154 141	308 282	0,0702	0,1404
3	667 328	2001 984	0,3038	0,9114
4	685 158	2 740 632	0,3119	1,2477
5	680 202	3 401 010	0,3097	1,5484
6	9683	58 098	0,0044	0,0265
إجمالي	2 196 512	8 510 006	1,00	3,87

بنفس الطريقة، عند حساب كل واحدة من $f_i\ x_i$ ثم جمعها مع بعضها، نحصل العمر الوسطَي لمدارس ما قبل الابتدائي هو $x = 3,8$ سنة.

3-حساب المتوسط في حالة المتغير المتصل

ليس هناك فرق بين مجمل المعادلات، القوانين والتعاريف المتعلقة بالوسط الحسابي البسيط والمرجح و تلك المتعلقة بالمتغير المنفصل. فتبقى الطريقة متماثل فيما عدا استعمال فرضية التوزيع المنتظم داخل الفئات والتركيز في مراكز الفئات، ما يسمح بحساب المتوسط انطلاقا من مراكز الفئات(موساوي،2004، 83).

مثال: حساب وسط حسابي مرجح على متغير متصل داخل فئة

لنأخذ البنية العمرية لسكان الجزائر سنة 2015، ونحسب العمر الوسطي. من أجل حساب $n_i\ x_i$ ، يجب أولا حساب مراكز الفئات x_i.إذا كان a_i و b_i يمثلان على الترتيب الحدود الدنيا و العليا للفئات، تكون إذن مراكز الفئات تساوي : $x_i = \frac{a_i+b_i}{2}$. حين تتوفر x_i ، يتم حساب كل واحد من $f_i\ x_i$ ، قبل جمعها كلها في نتيجة واحدة كما هو الحال في الجدول اللاحق.

جدول: حساب الوسط الحساب باستعمال نظام Excel

العمر	n_i	f_i	x_i	$n_i x_i$	$n_i\|x_i - \bar{x}\|$
15 −0	5915607	0,29234235	7	41409249	131093213
24−15	3397975	0,16792393	19,5	66260512,5	32826369,6
34−25	3755219	0,18557851	29,5	110778961	1274640,33
44−35	2764699	0,13662818	39,5	109205611	109205611
54−45	1952676	0,09649895	49,5	96657462	39716320,1

$\bar{X} = 29,91$ سنة

4−خصائص الوسط الحسابي

يتصف الوسط الحسابي بخاصية الخطية Propriété de linéarité:

$\overline{x+y} = \bar{x} + \bar{y}\ et\ \overline{ax} = a\bar{x}$ a قيمة ثابتة.

في مثال عن مؤسسة ما، دخل الموظفين يتكون من x الأجر و من المنحة y ، الوسطي الشهري يساوي 35000 دينار و المنحة المتوسطة الشهرية 8000 دينار. الدخل الوسطي الشهري يكون 43000 دينار . و من ذلك، إذا كانت كل الأجور ترتفع بـ 5 %، يكون الأجر الوسطي: 1,05 x35000= 36750 دينار.

إذا كانت كل قيم المشاهدات متماثلة، متوسط هذه المشاهدات تساوي هذه القيمة المشتركة. بمعنى، متوسط قيمة احصائية ثابتة تساوي نفسها.

ومنه : $\overline{ax+b} = a\bar{x}+b$ ، a و b كقيمتين ثابتتين. يسمح هذا لاسيما بتغيير الوحدة، أو أصلا، أي تغيير خطي على المتغير يؤثر المتوسط.

يكون متوسط الانحرافات عن المتوسط معدوما.

$$\sum_{i=1}^{r} ni(xi-\bar{x}) = \sum_{i=1}^{r} nixi - \sum_{i=1}^{r} ni\bar{x} = \sum_{i=1}^{r} nixi - n\bar{x} = 0$$

لأنه حسب قانون المتوسط :

$$\frac{1}{n}\sum_{i=1}^{r} nixi = \bar{x}, \text{أي } \sum_{i=1}^{r} nixi = n\bar{x}.$$

هذا يفسر لماذا نفضل متوسط الانحرافات المربعة لأجل قياس التشتت، المعروف أيضا بالتباين.

و يعتمد المتوسط الحسابي على كل حدود التوزيع(السلسلة)، ولأنه سهل الحساب، وهو مؤشر جيد لقياس النزعة المركزية حسب مفهوم جول Yule. بالمقابل، يعاب على المتوسط الحسابي أنه شديد التأثر بالقيم المتطرفة، ما جعل البعض يصنفه على أنه مؤشر غير صلب(موساوي،2004، 102).

ثالثا: الوسيط La médiane

لا تبدو دائما فكرة تقسيم السلسلة إلى مجموعتين لهما نفس العدد ممكنة، لذلك يجب التدقيق في تعريف الوسيط.

الوسيط *Me*، هي أصغر قيمة من السلسلة لأجلها يكون عدد المشاهدات الاصغر أو تساوي هذه القيمة، تمثل *50*% من العدد الإجمالي للسلسلة.

هذا يعنى أن هناك على الأقل 50 *%* من المشاهدات التي لها قيمة اصغر أو تساوي الوسيط، و على الاقل 50 *%* من المشاهدات التي لها قيمة أكبر أو تساوي الوسيط.

و يُحدد الوسيط عبر التكرارات المجمعة الصاعدة، انطلاقا من سلسلة من القيم المرتبة في اتجاه تصاعدي. كما يجب أن نميز بين حالة متغير ممثل بمعطيات غير مبوبة (معطيات خام) و حالة متغير ممثل في جدول إحصائي (بيانات مبوبة). بالنسبة للحالة الأخيرة نميز بين حال المتغير المنفصل و حال المتغير المتصل(Liorzou, 1984, 33).

1-الوسيط لبيانات غير مبوبة (خامة)

يجب أن تكون السلسلة مرتبة ترتيبا تصاعديا للقيم.

يتم تحديد بصفة مباشرة أو غير مباشرة للوسيط بحسب عدد المعطيات غير المبوبة.

1. إذا كان العدد فرديا، يمكن حينها تحديد الوسيط بصفة مباشرة.
2. إذا كان العدد زوجيا، يتم استنتاج الوسيط من المجال الوسيطي intervalle médian

طريقة الحساب 1: إذا احتوت السلسلة الخامة على عدد فردي للمشاهدات بـ $n= 2p +1$، الوسيط هو القيمة المركزية للسلسلة(مرتبة تصاعديا)، أي المشاهدة الـ $(p+1)$.

مثال: حساب الوسيط، عدد فردي لبيانات خام (غير مبوبة)

يمثل الجدول التالي معدل التشغيل (بـ %) للسكان بعمر 15 إلى 24 سنة، في 2005، في 7 بلدان من الاتحاد الأوروبي (ذات أعلى معدلات تشغيل).

نرتب أولا كل القيم ترتيبا تصاعديا. في مثالنا هذا، هذه القيم عددها $n= 7$، بمعنى عددا فرديا، و $p=3$، إذن الوسيط هو القيمة المركزية للسلسلة المرتبة، أي المشاهدة الرابعة: 40,5- 42 – 48,7 – 53,1 – 54 – 62,3 – 65,2. الوسيط هو $Me= 53,1$.

البلد	معدل التشغيل
ألمانيا	42
هولندا	65,2
النمسا	53,1
إيرلندا	48,7
المملكة المتحدة	54
الدانمارك	62,3
فلندا	40,5

Source : Insee, juillet 2007

طريقة الحساب 2: إذا احتوت السلسلة الخامة عددا زوجيا للمشاهدات، $n = 2p$، يجب تحديد المجال الوسيطي، المكون من مشاهدات من صف p و p+1 للسلسلة المرتبة. وتعريفا، يكون الوسيط هو وسط المجال الوسيطي.

مثال: حساب الوسيط، عدد زوجي للبيانات الخام

إذا أعدنا إعتبار المثال 5.2 مع اضافة فرنسا بمعدل 30,1 %، يصبح عدد القيم $n= 8$ ، إذن $p= 4$. المجال الوسيطي مكون من المشاهدتين 4 و 5، إذن المجال الوسيطي [53,1؛48,7]. النتيجة هي،

$$Me=\frac{48,7+53,1}{2} = 50,9$$

2-الوسيط في الجدول الاحصائي (البيانات المبوبة)

لأجل حساب الوسيط انطلاقا من جدول احصائي، يجب أن نميز بين حالتين:

أ. إما أن يكون المتغير منفصلا

ب. إما أن يكون المتغير متصلا

طريقة الحساب 1: في الحالة الأولى، تكون قيم المتغير (modalités de la variable) عبارة عن قيم مفصولة. يتم تحديد الوسيط بطريقة مباشرة باعتماد التكرارات المجمعة الصاعدة (الشكل 4.2)

مثال :حساب الوسيط لمتغير معطى كصفة منفصلة caractère discret

يعطي الجدول التالي عدد الاطفال أقل من 25 سنة في الاسرة، في فرنسا سنة 2005:

عدد الاطفال في الاسرة	n_i بالألاف
1	3700
2	3400
3	1200
4 فأكثر	400

Source :Insee, enquêtes de recensement,2004-2006

الشكل: التكرارات المجمعة الصاعدة

x_i	n_i	n_icc
1	3700	3700
2	3400	7100
3	1200	8300
4 أو أكثر	400	8700
إجمالي	8700	

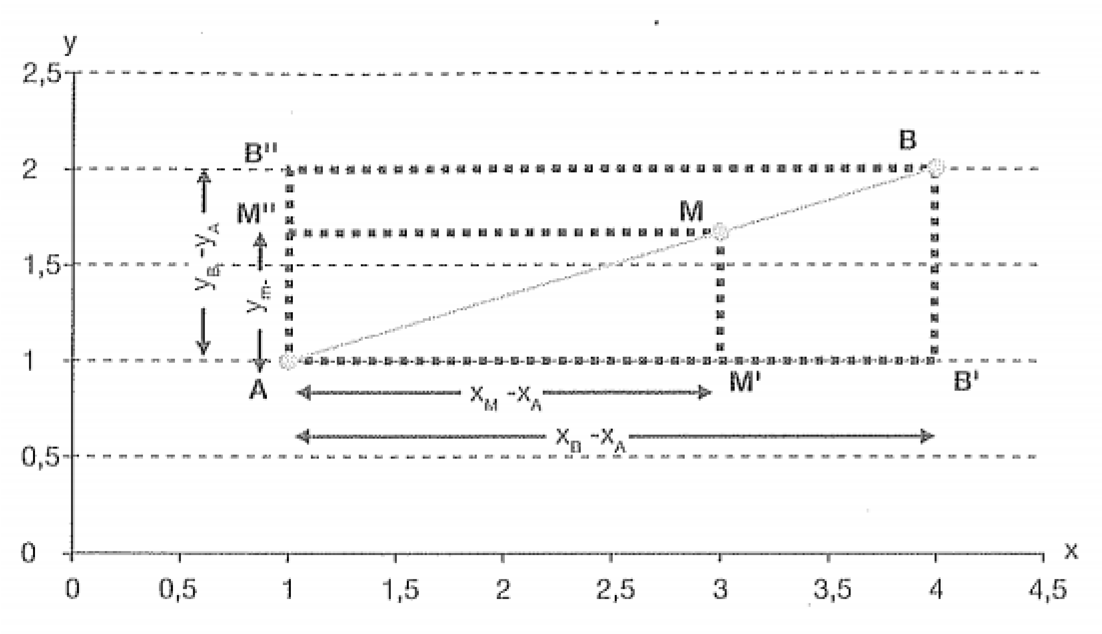
بمأن عدد المشاهدات زوجي، يتكون المجال الوسيطي إذن من المشاهدتين المركزيتين، أي برتبتين متتاليتين $p= \frac{n}{2} = \frac{8700}{2} = 4350$ و $p+1= 4351$. و تبين التكرارات المجمعة الصاعدة أن هذه المشاهدات هي في الترتيب أو القيمة modalité 2، إذن الوسيط، و وسطهم الحسابي، هو 2. فهناك 50 % من الأسر على الأقل لها عدد من الاطفال أقل أو يساوي 2 و 50 % من الأسر لها عدد من الاطفال أكبر أو يساوي 2.

طريقة حساب 2: في الحالة الثانية، عندما تكون قيم المتغير عبارة عن فئات. يعتمد تحديد الوسيط على فرضية أن المشاهدات تتوزع بصفة منتظمة و خطية (uniforme) داخل كل فئة. يُعرف الوسيط إذن بـ $F(Me)= 0,50$ أين F تمثل دالة التوزيع. و لحسابه نتبع حطوتين:

1. تعيين الفئة الوسيطية عبر التكرارات المجمعة الصاعدة أو التكرارات النسبة المجمعة الصاعدة؛

2. حساب الوسيط بطريقة خطية (interpolation linéaire) (Bressoud, 2010, 60)

شكل: توضيحي لتحديد قيمة ثالثة بطريقة الحساب الخطي



مثال: حساب الوسيط لمتغير ممثل n cc كصفة متصلة

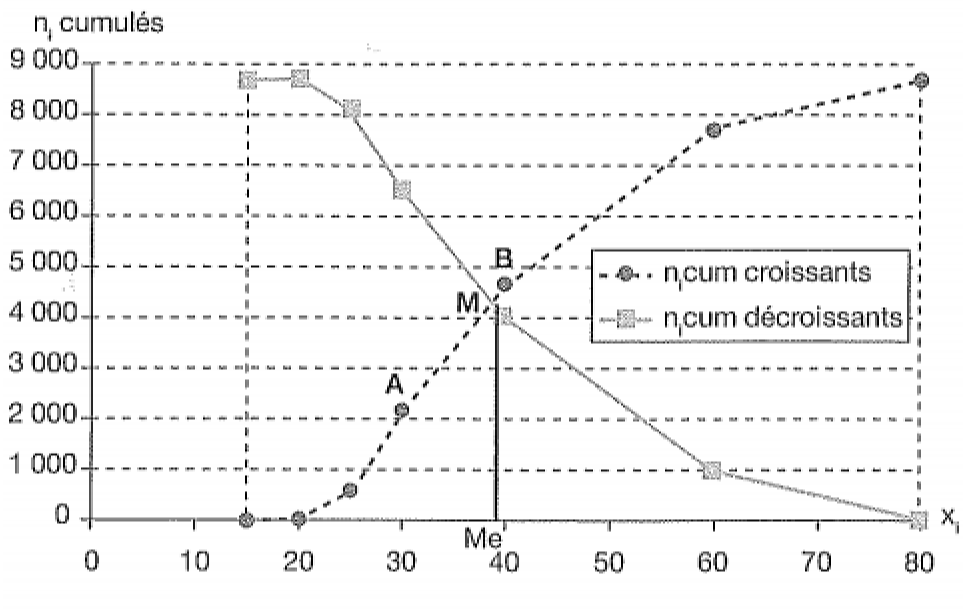
a_i	b_i	n_i	$n_i cc$	$n_i cd$
15	20	8	8	8691
20	25	563	571	8683
25	30	1594	2165	8120
30	40	2488	4653	6526
40	60	3063	7716	4038
60	80	975	8691	975
إجمالي	8691			

Source :Insee, recensement de la population, 1999

بحساب نصف العدد الاجمالي، 4346، و التكرارات المجمعة الصاعدة تسمح بتعيين الوسيط في المجال 30-40 سنة.

أما المنحنى المتجمع الصاعد فيسمح برؤية(حسيا) الوسيط على المخطط (مخطط...). أي النقاط الثلاث (A(30 ;2165),B(40 ;4653 و M(Me ; 4346)

شكل: الوسيط و التكرارات المتجمعة حسب العمر(15 سنة فأكثر) و المستوى التعليمي الجامعي



يمكن كتابة التتابع للنقاط الثلاث بالتساوي مع المعاملات الرأسية *coefficients directeurs(interpolation linéaire)*

$$\frac{4653+2165}{40-30}=\frac{4346-2165}{Me-30} \quad \text{أي} \quad \frac{2488}{10}=\frac{2181}{Me-30}$$

ما يعطي ، عن طريق جداء الطرفين و الوسطين: $Me = \frac{2181}{248,8}+30 = 38,76$

أو 38 سنة و 9 أشهر.

المحور الثالث: مقاييس التشتت

لا يخبرنا الإحصاء الوصفي المتعلق بخصائص النزعة المركزية أي المتعلق بالتموضع وخصائصه ،حول البنية الداخلية للتوزيع بالنسبة لسلسلة أو أخرى أو عن تغيراتها حول الوسط. لذلك يجب تكملة هذا العمل بإدراج خصائص التشتت. سندرس البعض منها مثل: المدى، الانحراف المعياري (المتعلق بالتباين) و أخيرا معامل الاختلاف.

المدى و المجال الربيعي و الانحراف المتوسط المطلق، تعرف أنها خصائص بسيطة للتشتت، لأنها خصائص محدودة في بنائها و استعمالتها، بالنسبة لمفهوم التباين الذي سنتطرق إليه لاحقا.

أولا: المدى L'étendue

إن أول قياس للتشتت لتوزيع معين هو المدى. هذا الأخير هو ابسط خصائص التشتت؛ فنتكلم عادة عن النطاق، أو مجال لتغيرات السلسلة. و يعرف المدى بالنسبة لسلسلة معينة على أنه: الفرق بين أكبر وأصغر قيمة مشاهدة. و يعبر عنه: $E=Max(x_i)-Min(x_i)$ (Lavallée,2005, 235)

ويسمح مفهوم المدى من تناول التشتت لمتغير بصفة سهلة ومبسطة، لكن معناه يبقى محدودا جدا، لأنه لا يأخذ في الحسبان إلا القمتين العظميين للسلسلة. في حين أنه يمكن إغفالهما أو حتى خطؤهما. من زاوية أخرى، لا يستقل المدى عن التكرار المشاهد ويمكن ان يعطي نظره خاطئة عن التشتت.

في الاخير، وفي حالة السلسلة المتصلة، لا يمكن ضبط المدى بدقة، لأن ضياع معلومات بسبب التجميع في فئات لا يسمح بمعرفة القيم الادنى و الاقصى المأخوذة من طرف المتغير حقيقة(Grenon, 1999, 267).

ثانيا: الانحراف المتوسط

يعتبر الانحراف المتوسط أكثر المقاييس بساطة للتشتت، وهو يقيس تغيرات السلسلة بالنسبة للمتوسط.

الانحراف المتوسط لـ n من المشاهدات هو المتوسط الحسابي للقيم المطلقة للانحرافات عن المتوسط:

$$Ea = \frac{1}{n}\sum_{i=1}^{n}|xi - \bar{X}|$$

الانحراف المتوسط لـ n من المشاهدات، مرتبة في جدول إحصائي $(x_i ; n_i)$، ممثلة بـ r من القيم، هو المتوسط الحسابي المرجح للقيم المطلقة للإنحرافات عن المتوسط:

$$Ea = \frac{1}{n}\sum_{i=1}^{n}|xi - \bar{x}|$$

و يعين r عدد القيم modalités، بـ $n = \sum_{i=1}^{r} ni$

وتستعمل القيم المطلقة للانحرافات بالنسبة للمتوسط من أجل منع تعويض الانحرافات الموجبة بتلك السالبة، حيث أنه عن طريق هذا التعويض مجموع الانحرافات عن المتوسط تكون سالبة:

$$\sum_{\mathrm{i}=0}^{r} ni(xi - \bar{x}) = 0$$

إن الانحراف المتوسط يمتاز بكونه يأخذ في الحسبان كل قيم السلسلة. فقد استعمله لابلاس Laplace قبل التباين، لاسيما في طريقة التقدير L1 ، كطريقة بديلة عن طريقة المربعات الصغرى.(Ferreol,1995, 179)

ثالثا. التباين و الانحراف المعياري

يعتبر الانحراف المعياري أكثر مؤشرات التشتت استعمالا. و قد تم استخدامه لأول مرة سنة 1893 من طرف عالم الرياضيات، الاحصائي و الفيلسوف الانجليزي Karl Pearson تحت اسم Standard deviation. أما التباين، أو مربع الانحراف المعياريي، فقد ادخله في الاحصاء عالم الاحصاء والوراثة الانجليزي Ronald Fisher (Bressoud,2010,75).

يُرمز للانحراف المعياري بـ σ_x و هو الانحراف المعياري للتباين. ويعطى كل من التباين و الانحراف المعياري، في حالة n مشاهدات:

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

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

في حالة n من المشاهدات، المرتبة في جدول إحصائي(x_i ; n_i)، ممثلة في r من القيم:

$$V(x) = \sigma_x^2 = \frac{1}{n}\sum_{i=1}^{i=r} n_i(x_i - \bar{x})^2, \text{ و } Q(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{i=r} n_i(x_i - \bar{x})^2}$$

التباين هو المتوسط الحسابي لمربعات الانحرافات بالنسبة للمتوسط، و يرمز له بـ V(x).

رابعا. مسائل وحلول

مثال: تعطي السلسلة التالية الاجر الادنى للنمو بالنسبة لـ 169 ساعة من العمل في 9 بلدان أوروبية سنة 2006. قيمة هذا الأجر معطاة بين قوسين:

بلجيكا (1234)؛ إيرلندا (1293)؛ اليونان (667)؛ إسبانيا (631)؛ فرنسا (1218)؛ لوكسمبورغ (1503)؛ هولندا (1273)؛ البرتغال (450)؛ المملكة المتحدة (1269).

البلد	x_i	$n_i(x_i - \bar{x})^2$
بلجيكا	1234	30353
إيرلندا	1293	54393
اليونان	667	154274
إسبانيا	631	183850
فرنسا	1218	25034
لوكسمبورغ	1503	196446
هولندا	1273	45464
البرتغال	450	351829
المملكة المتحدة	1269	43774
إجمالي	9538	1105418

حساب المتوسط الحسابي:

$$\bar{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = \frac{9538}{9} = 1059{,}78$$

من هذه النتيجة، وبعد حساب كل الفروق مع هذا المتوسط، ثم تربيعها ،

$$V(x) = \frac{1}{9}\sum_{i=1}^{9} n_i\,(x_i - \bar{x})^2 = \frac{1105418}{9} = 122824$$

إذن التباين لقيمة الاجر الأدنى لدول أوروبا المعنية هو 122824، و منه الانحراف المعياري:

$$Q_x = \sqrt{V(x)} = \sqrt{122824} = 350{,}46$$

مثال: حول الخصائص البسيطة للتشتت

f_i %	n_i	العمر
29,2	5915607	0- 15 سنة
16,8	3397975	15-24 سنة
18,5	3755219	25-34 سنة
13,5	2764699	35-44 سنة
9,6	1952676	45-54 سنة
6,4	1298940	55-64 سنة
3,3	675632	65-74 سنة
2,3	474456	75 - 100سنة
100	20235204	إجمالي

المطلوب:

1. حساب المدى

2. الانحراف المتوسط

الحل

1. المدى هو الفرق بين العمر الأقصى و العمر الأدنى

المدى= $\text{Max}\{x_i\} - Min\{x_i\} = 100 - 0$

المدى = 100. البنية العمرية في الجزائر تتوزع على 100 عام.

2. لأجل حساب الانحراف المتوسط، نحتاج لمعرفة المتوسط الحسابي. مراكز الفئات(x_i) ، و (n_ix_i) و المجاميع الأخرى كما هو موضح في الجدول.

المتوسط يساوي:

$$\bar{x} = \frac{1}{20235204}\sum_{i=1}^{8} n_i x_i = \frac{590070049}{20235204}$$

, يعني $\bar{x} = 29{,}16$ سنة

متوسط العمر للسكان الجزائريين الذكور هو تقريبا 29 و شهرين.

عندما نتعرف على المتوسط الحسابي، يتم حساب $|n_i - \bar{x}|$ n_i و مجاميعها كما هو في الجدول.

الانحراف المتوسط يساوي :

$$e_x = \frac{1}{20235204}\sum_{i=1}^{8} n_i|x_i - \bar{x}| = \frac{408459359}{20235204}$$

أي $e_x = 20{,}185$

متوسط انحرافات الاعمار عن المتوسط هو تقريبا 20 سنة و شهرين و أكثر من 3 أسابيع.

العمر	n_i	f_i	x_i	$n_i x$	$n_i \lvert n_i - \bar{x} \rvert$
15 -0	5915607	0,29234235	7	41409249	131093213
24-15	3397975	0,16792393	19,5	66260512,5	32826369,6
34-25	3755219	0,18557851	29,5	110778961	1274640,33
44-35	2764699	0,13662818	39,5	109205611	109205611
54-45	1952676	0,09649895	49,5	96657462	39716320,1
64-55	1298940	0,06419209	59,5	77286930	39409101,4
74-65	675632	0,03338894	69,5	46956424	27254610,9
100 - 75	474456	0,02344706	87,5	41514900	27679493,4
إجمالي	20235204			590070049	408459359

Source :ONS,Démographiealgérienne,2015

in :http://www.ons.dz/img/pdf/demographie2015.pdf

مثال: التباين والانحراف المعياري بالنسبة لمتغير منفصل

طلب أستاذ الإحصاء من طلبته عدد الافلام التي شاهدوها خلال الشهرين الأخيرين. و كانت النتائج على النحو التالي، كما هو موضح في الجدول:

عدد الافلام المشاهدة	عدد الطلبة
0	6
1	4
2	9
3	7
4	3

2	5

1. أحسب متوسط عدد الافلام المشاهدة.
2. أحسب:

أ. التباين بالنسبة لعدد الافلام المشاهدة؛

ب. الانحراف المعياري لعدد الافلام المشاهدة.

3. أحسب معامل الأختلاف.

الحل

1. المتوسط الحسابي هو:

$$\bar{x} = \frac{1}{31}\sum_{i=1}^{6} n_i x_i = \frac{65}{31} \text{ أي } \bar{x} = 2{,}096$$

إذن متوسط عدد الافلام المشاهدة عند كل طالب خلال الشهرين الاخيرين هو 2,1 فيلم.

2. أ. التباين هو:

$$V(x) = Q_x^2 = 1{,}444^2 \text{ أي } V(x) = 2{,}087.$$

أو أيضا بمعادلة أكثر تطورا:

$$\mathrm{V(x)} = \frac{1}{31}\sum_{I=1}^{6} n_i x_i^2 - \bar{X}^2 = \frac{201}{31} - 2{,}1^2 = 2{,}087 (\text{مقربة})$$

التباين لعدد الافلام المشاهدة حسب كل طالب خلال الشهرين الاخيرين هو 2,1 فيلم.

ب. الانحراف المعياري يساوي: $Q_x = \sqrt{V(x)} = 1{,}44$

الانحراف المعياري لعدد الافلام المشاهدة خلال الشهرين الاخيرين هو 1,44 فيلم.

3. معامل الاختلاف يساوي: $CV(x) = 0,69$ أي , $CV(x) = \frac{x}{\bar{x}} = \frac{1,44}{2,1}$

الانحراف المعياري أقل من المتوسط.

مثال: مقارنة توزيعات بمتغير متصل

يحصي الجدول التالي عدد سكان الاناث و الذكور في الجزائر حسب العمر في 1جويلية 2015

العمر	n_i	x_i
0−15	5915607	5599653
24−15	3397975	3267899
34−25	3755219	3690332
44−35	2764699	2742298
54−45	1952676	1966460
64−55	1298940	1273097
74−65	675632	687550
100 − 75	474456	500754
إجمالي	20235204	19728043

Source :ONS,Démographiealgérienne,2015

in :http://www.ons.dz/img/pdf/demographie2015.pdf

1. بالنسبة للنساء، أحسب:
   أ. المتوسط
   ب. الانحراف المعياري
   ج. معامل الاختلاف
2. بالنسبة للرجال، احسب:
   أ. المتوسط
   ب. الانحراف المعياري

ج. معامل الاختلاف

3. قارن بين التوزيعين

الحـــل

1. بالنسبة للنساء

العمر	n_i	x_i	$n_i\,x_i$	$n_i x_i^2$
0-15	5599653	7	39197571	274382997
24-15	3267899	19,5	63724031	1242618595
34-25	3690332	29,5	1,09E+08	3211511423
44-35	2742298	39,5	1,08E+08	4278670455
54-45	1966460	49,5	97339770	4818318615
64-55	1273097	59,5	75749272	4507081654
74-65	687550	69,5	47784725	3321038388
100 - 75	500754	87,5	43815975	3833897813
إجمالي	19728043	361,5	E+085,85	2,5213E+10

Source :ONS,Démographiealgérienne,2015

in :http://www.ons.dz/img/pdf/demographie2015.pdf

أ. المتوسط هو :

$$\bar{x} = \frac{1}{19728043}\sum_{i=1}^{8} n_i x_i = \frac{584796908}{19728043}$$ أي، $\bar{x} = 29,6429255$

العمر الوسطي هو تقريبا 29 سنة و 7 أشهر.

ب. الصيغة المركبة للتباين هي:

$$V(x) = \frac{1}{19728043}\sum_{i=1}^{8} n_i x_i^2 - \bar{x}^2 = \frac{2,5213E+10}{19728043} - 29,64^2$$
$$= 399,33$$

التباين بالنسبة لأعمار النساء هو 399,33.

ج. الانحراف المعياري يساوي: $Q_x = \sqrt{V(x)} = \sqrt{399,33}\ , soit\ Q_x = 19,98$

الانحراف المعياري لأعمار النساء هو 19,98 سنة، مقربة إلى 20 سنة.

معامل الاختلاف بالنسبة للنساء يساوي: $CV(x) = \frac{x}{\bar{x}} = \frac{19,98}{29,64}, soit\ CV(x) =$
0,674

2. و باتباع نفس الخطوات بالنسبة للرجال نحصل:

العمر	n_i	x_i	$n_i\,x_i$	$n_i x_i^2$
0−15	5915607	7	41409249	289864743
24−15	3397975	19,5	66260513	1292079994
34−25	3755219	29,5	E+08 1,11	3267979335
44−35	2764699	39,5	E+08 1,09	4313621615
54−45	1952676	49,5	96657462	4784544369
64−55	1298940	59,5	77286930	4598572335
74−65	675632	69,5	46956424	3263471468
100 − 75	474456	87,5	41514900	3632553750
إجمالي	20235204	361,5	E+08 5,9	2,5443E+10

أ. المتوسط هو :

$$\bar{x} = \frac{1}{20235204}\sum_{i=1}^{8} n_i x_i = \frac{5,9E+08}{20235204}$$ ، أي $\bar{x} = 29,16$

العمر الوسطي هو تقريبا 29 سنة و شهرين تقريبا.

ب. الصيغة المركبة للتباين هي:

$$V(x)=\frac{1}{20235204}\sum_{i=1}^{8}n_i x_i^2-\ \bar{x}^2=\frac{25442687608}{20235204}-29{,}16^2$$
$$=407{,}0089756$$

التباين بالنسبة لأعمار الرجال هو *407*.

ج. الانحراف المعياري يساوي: $Q_x=\sqrt{V(x)}=\sqrt{407}\ ,soit\ Q_x=20{,}17$

الانحراف المعياري لأعمار النساء هو *20,17* سنة، 20 سنة و شهرين.

معامل الاختلاف بالنسبة للنساء يساوي: $CV(x)=\frac{x}{=}=\frac{20{,}17}{29{,}16},soit\ CV(x)=$ 0,691

3. يبدو أن الرجال، في المتوسط، اكثر شبابية من النساء (العمر الوسطي: 29,2 مقابل 29,7).

في العموم، هناك تقارب بين انحراف أعمار الرجال و النساء (انحراف معياري: 20,17 مقابل 19,98 على التوالي).

ملحق توزيع X^2 :

df	نسب مساحة توزيع كاي – تربيع										
	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00004	0.00016	0.00098	0.00393	0.0158	0.455	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	1.386	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	2.366	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	3.357	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	4.251	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.83	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.43	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	51.17	60.39	64.28	79.33	98.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	89.33	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	99.33	118.5	124.3	129.6	135.8	140.2

مثال * : بالنسبة للمساحة المظللة والتي تمثل 0.05 من المساحة الكلية 1.0 تحت دالة كثافة الاحتمال ، قيمة X^2 هي 18.31 عند درجات حرية df = 10

المصدر : من جدول 4

Fisher and Yates, *Statistical Tables for Biological, Agriculturaland Medical Research*, 6th ed. 197
published by Longman Group Ltd., London (previously published by Oliver & Boyd, Edinburgh).
بتصريح من المؤلفين والناشرين