كلية العلوم

مقدمة
فى
الاحصاء الحيوي

قسم الرياضيات

الباب الاول

مقدمة

مكانة علم الإحصاء
Status of Statistics

إن علم الإحصاء هو علم قديم، وقد بقى جزءً من الرياضيات خلال القرون الأخيرة، حيث إن العمل الأصلي به قد تم بواسطة علماء الرياضيات مثل باسكال -1623) (1623، جيمس برنولي (1654-1705)، ديموفر (1667-1754)، لابلاس (1749-1827)، جاوس (1777-1855)، لاجرانش، بابيز، ماركوف، ...إلخ. هؤلاء الرياضيين كانوا مهتمين أساسًا بتطوير نظرية الاحتمالات وتطبيقاتها على نظرية المباريات والظواهر الطبيعية الأخرى. وحتى القرن 19، كانت الإحصاء مرتبطة أساسًا بالإحصائيات الرسمية المطلوبة لجمع المعلومات عن الدخل الحكومي، السكان، مساحات الأراضي، ...إلخ للولاية أو المملكة أو الدولة.

وتطور علم الإحصاء تدريجيًا وتعاظم مجال تطبيقه يومًا بعد يوم. ولذلك فمن الصعب إعطاء تعريف دقيق للإحصاء. وتغير التعريف من زمن لآخر اعتمادًا على استخدامه وتطبيقه. وقد صدرت تعريفات متعددة من أناس مختلفين. هذه التعريفات تعكس الزاوية الإحصائية ومجال النشاط. وسنقدم فيما يلي بعض هذه التعريفات:

١- تعريف الإحصاء Definition of Statistics

استخدمت كلمة "إحصاء" لأول مرة في كتاب Elements of Universal Erudition بواسطة البارون "بايفيلد" J.F. von Bielfeld، والمترجم بواسطة "هيبر" W. Hooper M.D. (لندن، 1970). وتُعرف الإحصاء هنا كالتالي:

«العلم الذي يُعلمنا ما هو النظام السياسي لكل الحالات الحديثة للعالم المعروف».

"ويبستر" Webster «الحقائق المصنفة التي تمثل حالة الناس في ولاية، خاصة تلك الحقائق التي يمكن أن تُذكر في أعداد أو في جداول من الأرقام أو أي نظام جدولي أو مصنف».

"هورس سيكرست" Horace Secrist «الإحصاء هي تجميع للحقائق التي تؤثر على مدى معلوم بواسطة أسباب متعددة -معبر عنها عدديًا- مسرودة أو مقدرة طبقًا لمقياس معقول من الدقة، مجمع بأسلوب نظامي لغرض مُسبق التحديد وموضوعة في علاقة مع بعضها البعض».

وأعطى البروفيسور "بولي" A.L. Bowley عدة تعريفات للإحصاء كالتالي:

(١) علم الحساب. (٢) علم المتوسطات.

(٣) علم قياس الظاهرة الاجتماعية. (٤) موضوع لا يقتصر على علم واحد فقط.

"بودنجتن" A.L. Boddington «الإحصاء هي علم التقدير والاحتمالات».

"كروكستن" و"كاودن" Croxton & Cowden «يمكن تعريف الإحصاء الإحصاء بأنها جمع وعرض وتحليل وتفسير البيانات العددية».

"وولس" و"روبرتس" Wallis & Roberts «يمكن اعتبار الإحصاء كجسم لطرق اتخاذ قرار حكيم في مواجهة عدم اليقين».

"فيشر" R.A. Fisher «علم الإحصاء هو أساسًا فرع من الرياضيات التطبيقية، ويمكن اعتباره كالرياضيات المطبقة على بيانات الملاحظة».

من كل التعريفات، يعتبر التعريف المعطى بواسطة فيشر الأكثر صحة حيث يغطي كل جوانب ومجالات الإحصاء.

٢. وظائف الإحصاء Functions of Statistics

من وجهة نظر التطورات الأخيرة في مجال الإحصاء، فهو يعتبر كعلم اتخاذ القرارات، تحت عدم اليقين والريبة، مع أو بدون بيانات. ويرجع التوسع في استخدام الإحصاء أساسًا لإحصائي إنجلترا. فقد تغيرت تمامًا فكرة ومجال الإحصاء في القرن العشرين.

وعلاوة على ذلك، فإن نظرية الاستدلال theory of inference، وتصميم التجارب، ونظرية العينات sampling theory أثبتت أنها نقط تحول في تطور الإحصاء. وهناك بعض المساهمات الأخرى دعمت وجهة النظر هذه. اخترع "فرانسيس جالتون" Francis Galton نظرية الانحدار regression theory، واخترع "كارل بيرسون" Karl Pearson نظرية التوزيع وتحليل الارتباط المتبادل theory of distribution and correlation analysis. اخترع "جوسّت" W.S. Gosset اختبار t في 1908. وقد أنجز فيشر كثير من الأعمال في

اتجاهات مختلفة مثل نظرية التقديرات، الاستدلال الإسنادي، توزيع العينات، نظرية تصميم التجارب. وقد استخدم فيشر طرق إحصائية في عدد من العلوم مثل الزراعة، والجينات، وعلم الاجتماع، والتعليم. وكانت أفكاره أساسية وقائمة على الجانب التطبيقي. ومنذ ذلك الوقت توسع مجال الإحصاء يومًا بعد يوم وازداد استخدامه في كثير من العلوم.

الطرق الإحصائية أو التقنيات تكون قابلة للتطبيق عند إتاحة بعض البيانات فقط بغض النظر عن طريقة جمع البيانات. البيانات يمكن أن تكون كمية أو نوعية. فإذا كانت البيانات نوعية فإنها تُكمم باستخدام تقنيات مثل الترتيب حسب الفئة، التحزز، التشفير، ...إلخ. وتُجمع البيانات إما بالتجارب أو بطرق الاستقصاء (المباشر أو غير المباشر) ثم تُجدول وتُحلل إحصائيًا. ومهما كانت القيم الناتجة المتحصل عليها من التحليل استدلالات صحيحة فإنه يجب أن تُرسم من تلك القيم العددية. وتؤدي هذه الاستدلالات إلى القرار النهائي. وعلى أساس هذه الأفكار، نستطيع إعطاء الوظائف التالية للإحصاء:

(١) تجميع البيانات. (٢) جدولة البيانات.
(٣) تحليل البيانات. (٤) تفسير النتائج.

الأربع وظائف السابقة سيتم وصفها لاحقًا. وسيكون تطبيقها ونفعيتها واضحة من مناقشة الموضوعات المعطاة في متن هذا الكتاب.

تجميع البيانات Collection of Data

بمجرد تحديد نوع الدراسة المطلوب إجراؤها، يصبح من الضروري جمع معلومات عن تلك الدراسة، وغالبًا على هيئة بيانات. ولذلك، يجب أن تُجمع المعلومات من مصادر معينة مباشرة أو غير مباشرة. وهذه التقنية تعرف بـ"طريقة الاستقصاء". وهي تستخدم عادة في العلوم الاجتماعية أي المشاكل المرتبطة بالمجتمع، والسياسة، وعلم النفس، والدراسات الاقتصادية المختلفة. في الاستقصاءات يتم الإمداد بالمعلومات المطلوبة بواسطة الشخص أو الفرد تحت الدراسة أو تعتمد على قياسات وحدات معينة. وعامة يتم اختيار المستجيبين أو الوحدات من مجتمع الدراسة باستخدام بعض تقنيات العينة القياسية. وهناك طريقة أخرى لجمع البيانات بواسطة إجراء التجارب، أي تنفيذ تجربة عملية على شخص معين أو الوحدات التي سيتم عليهم الاستدلال. ويتم أخذ ملاحظات عن الموضوع الذي تحت الدراسة. مثل هذه الدراسات التجريبية تكون شائعة في الزراعة، والبيولوجي، والطب، والكيمياء، والصناعة، ...إلخ.

كيفية أختيار العينة :

لاختيار العينة موضع البحث ، لابد من تحديد المجتمع الذى سوف تؤخذ منه هذه العينة ثم تحديد وحدة المعاينة هل هى الفرد أو الأسرة مثلا وبعد ذلك تحديد نوع العينة وحجمها وذلك يتوقف على مدى الدقة المطلوبة فى النتائج وأنواع العينات المستخدمة هى العينات العشوائية البسيطة والعينات العشوائية الطبقية ، العينات المنتظمة ثم العينات متعددة المراحل وسوف ندرس أنواع العينات بالتفصيل فيما بعد .

ثالثا : عرض البيانات الاحصائية جدوليا وبيانيا .

رابعا : تحليل وتفسير البيانات بأستخدام الاسلوب الاحصائى مع ملاحظة أن المرحلة الثالثة والرابعة سوف نتناولها بالتفصيل باذن الله فيما بعد .

طرق عرض البيانات

قبل البدء فى معرفة الطرق المختلفة لعرض البيانات الاحصائية لابد من معرفة أنواع البيانات كما يلى :

أنواع البيانات :

تصنف البيانات الاحصائية الى بيانات وصفية وبيانات رقمية (كمية) .

١ ـ البيانات الوصفية : هذه البيانات تشمل البيانات الناتجة عندما تصنف الوحدة التجريبية طبقا لنهج معين فمثلا : الحالة التعليمية لمجموعة من الاشخاص تصنف الى أمى ، يقرأ فقط ، يقرأ ويكتب ، مؤهل متوسط ، مؤهل جامعى ، مؤهل فوق جامعى ، وكذلك تصنيف الاشخاص حسب الجنسية . . . . وهكذا بالنسبة لبعض الظواهر الاخرى .

٢ ـ البيانات الرقمية : وهذه البيانات الرقمية تصنف الى بيانات متقطعة (أو منفصلة) وبيانات مستمرة (أو متصلة) .

أ ـ البيانات الرقمية المتقطعة (المنفصلة) .

وتمثل البيانات الرقمية المتقطعة البيانات التى تأخذ أرقاما منفصلة محددة فى حدود مدى معين ممكن أن تنتقل من رقم الى رقم آخر دون الاخذ فى الاعتبار ما بينهما من أرقام كسرية أى أن هذه البيانات تأخذ أرقاما صحيحة موجبة .

فمثلا عدد الحجرات فى شقة معينة ، عدد الاسرة فى مستشفى معين ، عدد أفراد الاسرة ، ..... كلها بيانات رقمية متقطعة .

ب ـ البيانات الرقمية المستمرة (المتصلة) .

البيانات الرقمية المتصلة هى البيانات التى تأخذ أى قيمة فى حدود معينة وسواء كانت هذه القيمة كسرية أم صحيحة أى لا توجد فجوات بين جميع القيم الممكنة فمثلا : الوزن ، الطول ، الدخل العمر ، المسافة ، ..... كلها تمثل ظواهر رقمية متصلة حيث كل منها يتغير بصورة متصلة ولا يقفز من قيمة معينة الى قيمة أخرى .

وسوف ندرس فى هذا الباب الطرق المختلفة لعرض وتقديم البيانات الاحصائية فى الصورة المنتظمة التى يمكن بواسطتها التوصل الى أستنتاجات هامة خاصة بالمشكلة موضوع الدراسة ، من هذه الطرق ما يلى :

أولا : عرض البيانات الاحصائية جدوليا .

ثانيا : عرض البيانات الاحصائية بيانيا .

أولا : عرض البيانات الاحصائية جدوليا .

تعتبر الجداول الإحصائية من أفضل الطرق لعرض البيانات الاحصائية فى صورة منظمة وواضحة وسهلة اذ عن طريق هذه الجداول من الممكن الاستغناء عن أى تفسير لها . هذا وتوجد جداول أحصائية لها أغراض عامة مثل جداول تعدادات السكان وتشمل بيانات عن النوع ، العمر ، الوظيفة الحالة التعليمية ، الحالة الاجتماعية ، ..... والجداول الصناعية وتشمل بيانات عن أنواع الصناعات وعدد العمال ورأس المال والارباح والخسائر ، المواقع المختلفة للمصانع بالاضافة الى

ذلك توجد جداول أحصائية ذات أغراض خاصة ، والجداول ذات الاغراض الخاصة قد تكون جداول بسيطة أى تدرس ظاهرة واحدة فقط أو مزدوجة أى تدرس ظاهرتين أو مركبة وتدرس أكثر من ظاهرتين .

الشروط الواجب توافرها فى الجداول الاحصائية .

١ ــ الترقيم لابد أن يكون لكل جدول احصائى رقم معين سواء أكان هذا الجدول وارد فى تقرير أو مجلد أو نشرة وذلك حتى يمكن تحديده من بين الجداول الاحصائية الاخرى المنشورة وحتى يسهل الرجوع اليه عند الحاجة .

٢ ــ العنوان : يجب أن يكون عنوان الجدول الاحصائى واضح يذكر ماهية البيانات ، وأين جمعت وكيف صنعت ، الفترة الزمنية التى تخصها .

٣ ــ تقسيم الجدول الاحصائى : يقسم رأس الجدول الاحصائى على شكل خانات مربعة أو مستطيلة حسب حجم البيانات المكتوبة فى كل خانة من الخانات .

٤ ــ الايضاحات السفلية : وهذه تستخدم اذا لزم الامر ذلك ولتفسير احدى المصطلحات العلمية أو الفنية والتى لا يدرك معناها غير المتخصصين .

٥ ــ مصادر البيانات : وتذكر أسم المصدر الذى أخذت منه بيانات الجدول وذلك تيسيرا لمهمة من يريد الرجوع الى المصدر الاصلى للتأكد أو لاستكمال بعض البيانات .

٦ ــ وحدات قياس البيانات التى يحويها أى جدول أحصائى وذلك اما أسفل الجدول فى حالة ما تكون وحدات القياس واحدة أما اذا أختلفت وحدات القياس فلابد من توضيح ذلك فى مدلول السطر وعنوان العمود .

وفيما يلى طرق عرض البيانات الاحصائية جدوليا منفصلة ومستمرة كل على حدة :

(١) ــ العرض الجدولى لبيانات وصفية :

توجد مراحل يجب أتباعها عندما يراد عرض مجموعة من البيانات الوصفية فى صورة جداول وهذه المراحل سوف نتناولها مستخدمين تطبيقها على المثال التالى :

مثال (١) : فيما يلى بيان بالحالة التعليمية لعدد ٢٠ شخصا . متوسط ــ جامعى ــ متوسط

أمى ـ جامعى ـ يقرأ فقط ـ يقرأ ويكتب ـ متوسط ـ فوق الجامعى ـ أمى ـ يقرأ ويكتب ـ يقرأ فقط ـ متوسط ـ جامعى ـ أمى ـ فوق الجامعى ـ يقرأ ويكتب ـ يقرأ فقط ـ جامعى ـ متوسط
والمطلوب : عرض هذه البيانات فى جدول تكرارى .

الحل :

بالنسبة لمثالنا هنا نجد أن الظاهرة التعليمية تشتمل على صفات وهى : أمى ، يقرأ ويكتب ، يقرأ فقط ، مؤهل متوسط ، مؤهل جامعى ، مؤهل فوق الجامعى .

ثم نكون الجدول لتفريغ البيانات ويتكون من ثلاثة أعمدة ، العمود الاول يمثل صفات (أو أقسام) الظاهرة ، العمود الثانى هى خانة العلامات وفيها يتم وضع علامة بعد الاخرى أمام كل صفة تنتمى اليها المشاهدة ونضع علامة عبارة عن شرطة مائلة (/) أمام تلك الصفة ..... وهكذا بالنسبة لبقية المشاهدات ويتم تكوين مايسمى بالحزم والحزمة تحتوى على خمس علامات تكون شكلها كالتالى (~~////~~) أى أربعة شرط مائلة فى أتجاه واحد والخامسة تكون بصورة عكسية وذلك لتسهيل العد ، والعمود الثالث هى خانة التكرارات وفيها يتم ترجمة رقم كل صفه من صفات الظاهرة التى ندرسها .

جدول (١) جدول تفريغ الحالة التعليمية لـ ٢٠ شخصا .

subjects الصفات	Marks (العلامات)	freguency (عدد الاشخاص
أمى (a)	///	3
يقرأ فقط (b)	///	3
يقرأ ويكتب (c)	///	3
مؤهل متوسط (d)	~~////~~	5
مؤهل جامعى (e)	////	4
مؤهل فوق جامعى (f)	//	2
Σ		20

أما الجدول التكرارى يتكون من جدول التفريغ بعد أهمال عمود العلامات (الحـــزم) أى أنه يتكون من العمود من الاول والثالث فقط كالاتى ،

جدول (٢) التوزيع التكرارى للحالة التعليمية لـ ٢٠ شخص .

Subjects	a	b	c	d	e	f	Σ
frequency	3	3	3	5	4	2	20

(٢) العرض الجدولى لبيانات رقمية متقطعة ومستمرة :

أ ــ العرض الجدولى لبيانات رقمية متقطعة (منفصلة) .

كما ذكرنا البيانات الرقمية المتقطعة هى أرقام منفصلة محددة فى حدود مدى معين ويمكن الانتقال من رقم الى رقم آخر أى أنها تأخذ أرقاما أو قيما صحيحة موجبة .

وفى دراستنا أيضا للعرض الجدولى للبيانات المتقطعة سوف نتبع نفس فكرة العرض الجدولى للبيانات الوصفية والتى تتمثل فى ثلاثة مـراحل هى : تحديد صفات الظاهرة ، تكوين جـــدول التفريغ ، وأخيرا تكوين الجدول التكرارى ، مع فارق بسيط وهو أن الصفات سوف تستبدل بأرقـــام هى أقسام الظاهرة التى ندرسها .

مثال (٢) :

البيانات التالية تمثل عدد الحجرات (Rooms) المتواجدة فى شقق 30 أسرة :

3, 5, 4, 2, 3, 1, 6, 5, 4, 2
1, 3, 6, 3, 2, 5, 4, 3, 6, 1
2, 3, 2, 6, 5, 4, 3, 2, 5, 2

والمطلوب عرض هذه البيانات فى جدول تكرارى .

الحل :

لتحديد أقسام الظاهرة نلاحظ أنها عدد حجرات فى الشقق والارقام التى تأخذها هــــذه

الظاهرة هى : 6، 5، 4، 3، 2، 1 ولتكوين جدول التفريغ كما سبق أن أوضحنا فى حالة المتغيرات الوصفية يتكون من ثلاثة أعمدة هى نفسها نستخدمها فى حالة المتغيرات المتقطعة وحيث العمود الاول يمثل أقسام الظاهرة (أو المتغير) الذى ندرسه وهنا تمثل عدد الحجرات ثم العمود الثانى عمود الحزم أو العلامات وفى هذا العمود يتم قراءة قيم المتغير موضوع الدراسة قيمة بعد الاخرى وفى كل مرة يتم تحديد القسم الذى تنتمى اليه تلك القيمه ونضع علامة أمام ذلك القسم ... . وهكذا وتكون ما يسمى بالحزم أو العلامات لتسهيل عملية العد ، أما العمود الثالث يتم فيه كتابة عدد العلامات أمام كل قسم من أقسام الظاهرة أو المتغير الذى ندرسه

جدول (٣) جدول تفريغ عدد حجرات 30 شقة

Rooms	Marks	Freguency
1	///	3
2	//// //	7
3	//// //	7
4	////	4
5	////	5
6	////	4
Sum		30

لاحظ فى جدول التفريغ وحتى تتأكد أن جميع المفردات المعطاه قد تم تبويبها لابد أن يكون المجموع (30 فى المثال الحالى) يساوى للعدد الاجمالى للمفردات المراد تبويبها فى جدول تكرارى .

ومن جدول التفريغ السابق يمكن تكوين جدول التوزيع التكرارى المطلوب كالتالى :

جدول (٤) التوزيع التكرارى لعدد حجرات 30 شقة .

Rooms	1	2	3	4	5	6	Sum.
families	3	7	7	4	5	4	30

وهذا الجدول تقرأ بياناته كالتالى : من بين 30 أسرة توجد 3 أسر تتكون شققهم من حجرة واحدة ، توجد 7 أسر تتكون شققهم من حجرتين ، .... وهكذا بالنسبة للبقية .

ب ــ العرض الجدولى لبيانات رقمية متصلة :

١ ــ العرض الجدولى لبيانات ظاهرة واحدة (جدول تكرارى بسيط) :

وفى دراستنا للعرض الجدولى للبيانات المستمرة فاننا سوف نتبع نفس فكرة العرض للبيانات المتقطعة والتى تتمثل فى ثلاثة مراحل : تحديد أقسام الظاهرة ، ثم تكوين جدول التفريغ وأخيرا تكوين جدول التوزيع التكرارى . ولكن أقسام الظاهرة بالنسبة للبيانات المتصلة تستبدل بما يسمى بالفئات وتشمل كل فئة على أرقام متتالية تكون محصورة بين حدى الفئة .

مثال (٣) :

البيانات التالية توضح الانفاق اليومى بالجنيهات لعدد 100 عامل :

70	55	58	60	88	92	75	105	68	115
92	94	90	54	56	62	65	74	66	76
110	112	106	103	101	89	83	53	57	68
65	67	111	116	117	102	107	63	66	76
100	101	108	64	69	83	86	92	97	90
83	86	91	96	103	114	107	116	76	87
81	93	118	119	72	71	99	98	74	82
80	82	87	88	92	94	102	83	85	81
83	85	91	103	107	78	70	85	89	87
74	75	96	97	112	82	83	72	93	78

والمطلوب تبويب البيانات السابقة فى جدول توزيع تكرارى متساوى الفئات .

الحـل :

أ ــ تحديد أقسام الظاهرة أى الفئات :

بالنسبة للبيانات المتصلة فان أقسام الظاهرة ماهى الا الفئات ولتحديد هذه الفئات لابد من

معرفة ثلاث مؤشرات وهـــى :

١ ــ عدد الفئات

٢ ــ تحديد أكبر وأصغر قيمة فى البيانات .

٣ ــ تحديد طول الفئة .

ولتحديد عدد الفئات (الفترات Intervals) يتوقف على عدد مفردات الظاهـــرة موضوع الدراسة والتى هى مجموع التكرارات والذى نرمز لها بالرمز n وتوجد قاعدة لتحديد عـــدد الفئات k وهى :

$$k = 1 + 3.3 \log n$$

وهذه الطريقة قد تكون غير دقيقة ، فى بعض الاحيان ، فى تحديد عدد الفئات وقد يقوم الباحث بنفسه بتحديد عدد الفئات وعادة من الفضل الا يقل عن 6 فئات وألا يزيد عن 20 فئة وبالنسبـــة للمثال نجد أن :

$$n = 100$$

$$\therefore k = 1 + 3.3 \log 100 = 7.6$$

هذا وسوف نختار عدد الفئات فى هذا المثال = 7 فئات . كما أننا نجد أن أصغر قيمة = 53 وأكبر قيمة = 119 .

ولتحديد طول الفئة اذا كان الجدول المطلوب تكوينه متساوى الفئات فان طول الفئـة (C) يعطى بالصورة :

$$\text{طول الفئة } (C) = \frac{\text{أكبر قيمة} - \text{أصغر قيمـــة}}{\text{عـدد الفئـــات}} + 1 = \frac{\text{المـدى}}{\text{عدد الفئات}} + 1$$

وبتطبيق هذا القانون على المثال الحالى نجد أن :

$$C = \frac{119 - 53}{7} + 1 = 10 \quad \text{تقريبا}$$

ممكن أن يكون طول الفئة 10 . ولسهولة العرض الجدولى يجب أن يكون طول الفئة عـــدد صحيح وأيضا يأخذ قيما عبارة عن 5 أو مضاعفاتها .

أما اذا كانت طبيعة البيانات تستلزم أن يكون الجدول غير متساوى الفئات فاننا لا نستخدم قانون طول الفئة السابق ذكره ، ويحدد طول كل فئة على حدة وكما يريده الباحث ولكن بشرط ألا تكون أطوال بعض الفئات كبيرا جدا وبعضها الآخر صغيرا جدا .

ولتحديد بداية ونهاية كل فئة :

عندما نحدد بداية ونهاية كل فئة ، لابد أن تكون بداية الفئة الأولى أصغر من أقل قيمة في البيانات المعطاه وذلك حتى تدخل أقل قيمة في البيانات في نطاق الفئة الأولى . وكذلك لابد أن تكون نهاية الفئة الأخيرة أكبر من أكبر قيمة في البيانات وذلك حتى تكون أكبر قيمة في البيانات داخل نطاق الفئة الأخيرة وبذلك يتكون لدينا جدول بسيط مغلق أي بداية الفئة الأولى ونهاية الفئة الأخيرة معلومتين .

وعلى هذا تكون فئات الانفاق للمثال السابق كما يلي :

50 - 60 لـ تَقرأ 50 but less than 60

أى 50 الى أقل من 60 جنية .

60 - 70 ===> from 60 but less than 70

70 - 80 ===> from 70 but less than 80

.... وهكذا .

ويمكن كتابة الفئات على النحو الآتي :

50- , 60 - , 70 - , ... وهكذا

بعد تحديد فئات البيانات المعطاة نكون جدول التفريغ وهو كما سبق أن ذكرنا يتكون من ثلاث أعمدة : العمود الأول ـ يمثل أقسام الظاهرة والتي هي الفئات في البيانات المتصلة ثم العمود الثاني خاصة بالعلامات أو الحزم حيث يتم قراءة قيم المتغير موضوع الدراسة قيمة بعد الأخرى وفي كل مرة يتم تحديد الفئة التي تنتمي اليها تلك القيمة وتضع علامة أمام تلك الفئة ..... وهكذا ، وتكون ما يسمى بالحزم ، والحزمة تتكون من خمس أشرطة أربعة مائلة في أتجاه والخامسة عكسية عليها وذلك لتسهيل عملية العد .

أما العمود الثالث فيه عدد العلامات أمام كل فئة من فئات البيانات المعطاه وتسمى

التكرارات · F regiencies

جدول (5) جدول تفريغ الانفاق اليومى بالجنيه لعدد 100 عامل .

Intervals	Marks	frequencies
50 - 60	~~||||~~ |	6
60 - 70	~~||||~~ ~~||||~~ ||	12
70 - 80	~~||||~~ ~~||||~~ ~~||||~~	15
80 - 90	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ||||	24
90 - 100	~~||||~~ ~~||||~~ ~~||||~~ |||	18
100 - 110	~~||||~~ ~~||||~~ ||||	14
110 - 120	~~||||~~ ~~||||~~ |	11
Σ		100

وعلى ذلك فان الجدول التكرارى للتوزيع السابق يكون كالتالى :

جدول (6) : جدول التوزيع التكرارى للانفاق اليومى بالجنيه لعدد 100 عامل .

Intervals	Freq.
50 - 60	6
60 - 70	12
70 - 80	15
80 - 90	24
90 - 100	18
100 - 110	14
110 - 120	11
Sum	100

بيانات الجدول السابق عرضاً كالتالى :

من بين 100 عامل يوجد 6 عمال يتراوح أنفاقهم اليومى بين 50 جنيه وأقل من 60 جنيـــة

يوجد 12 عامل يتراوح أنفاقهم اليومى ما بين 60 جنية وأقل من 70 جنية ، ..... وهكذا

٢ ــ العرض الجدولى لبيانات ظاهرتين (الجدول التكرارى المزدوج) :

مما سبق درسنا العرض الجدولى أو الجداول التكرارية البسيطة والتى تدرس ظاهرة واحـــدة فقط أو تمثل متغيرا واحدا فقط وأنتهينا الى أن جدول التوزيع التكرارى يشتمل على عمودين فقط وهما الفئات (الفترات) والتكرارات وذلك كما هو موضح فى جدول (٦) وهذا النوع من الجداول أطلقنا عليه أسم جداول تكرارية بسيطة .

ولكن اذا كنا نهتم بدراسة ظاهرتين أو متغيرين قد تكون بينهما علاقة ما تربطهما فاننا نحتاج ليس الى جداول تكرارية بسيطة ولكن الى ما يسمى بالجداول التكرارية المزدوجة . وعند تكوين الجدول التكرارى المزدوج نجد أنه يتكون من تقسيم أفقى أو صفوف وتقسيم رأسى أو أعمدة .

ولعرض البيانات الخاصة بظاهرتين أو متغيرين فى جدول تكرارى مزدوج نتبع نفس الخطـــوات أو المراحل السابقة كما فى حالة ظاهرة واحدة وهذه المراحل هى : تحديد الفئات ، تكوين جدول التفريغ المزدوج ، وأخيرا أشتقاق جدول التوزيع التكرارى المزدوج وذلك من جدول التفريغ المزدوج وخطوات تكوين جدول التوزيع التكرارى المزدوج سوف نوضحها بالمثال التالى : ــ

مثال (٤) :

البيانات التالية تمثل الصادرات (X) والواردات (Y) بآلاف الجنيهات لعدد 50 شركة فى دولة معينة :

والمطلوب : عـرض تلك البيانات فى جدول تكرارى مزدوج .

x	y	x	y	x	y	x	y
45	21	87	75	56	44	55	48
83	75	52	32	80	74	42	46
68	59	64	54	62	62	67	64
99	89	68	72	95	80	72	67
32	28	86	80	82	63	74	69
65	69	44	47	37	38	58	54
93	72	72	68	54	56	36	37
53	63	76	73	77	52	74	58
84	60	50	52	73	71	55	32
72	62	66	59	62	65	62	53
64	61	45	37	62	61	69	57
92	73	92	80	73	64		
60	42	51	42	54	45		

الحل :

لتكوين الجدول التكراري المزدوج لبيانات الصادرات والواردات المعطاه نتبع الفروض الثلاث وهي :

أ ـ تحديد الفئات (الفترات) :

وسوف نحدد فترات الصادرات وفترات الواردات كل على حدة كما يلي :ـ

حيث أن عدد أزواج المفردات المعطاه $n = 50$

عدد الفترات $k = 1 + 3.3 \log 50 = 7$

وهذا الرقم 7 لعدد الفترات خاص بكل من المتغيرين x ، y ، حيث x تمثل الصادرات ، y تمثل الواردات .

اى اعـدد فترات المتغير $x = 7$

• عـدد فترات المتغير $y = 7$

1_ بيانات الصادرات X :

أصغر قيمه = 32 واكبر قيمه = 99

$$\text{طول الفئه } C_1 = \frac{\text{أكبر قيمه – اصغر قيمه}}{\text{عدد الفئات}} + 1$$

$$C_1 = \frac{99 - 32}{7} + 1 \cong 10$$

لاحظ اننا سوف نستخدم جدول منتظم (الفترات متساويه) وذلك للمتغير x (الصادرات).

فان فئات المتغير x تكون كالتالى :

30 – , 40 – , 50 – , 60 – , 70 – , 80 – , 90 – 100 .

2_ بيانات الواردات y :

أصغر قيمه = 21 ، أكبر قيمه = 84

$$\therefore \text{طول الفئه } C_2 = \frac{\text{أكبر قيمه – اصغر قيمه}}{\text{عدد الفئات}} + 1$$

$$C_2 = \frac{84 - 21}{7} + 1 = 10$$

وحيث ان أصغر قيمه للمتغير y هو 21 فأننا نختار بدايه الفئه الاولى 20 وكذلك اكبر قيمه 84

وكذلك نختار بداية الفئه الاخيرة 80 وتنتهى عند اقل من 90 وتكون (80 – 90).

فان فئات المتغير y (الواردات) كما يلى :–

20 – , 30 – , 40 – , 50 – , 60 – , 70 – , 80 – 90 .

ب_ تكوين جدول التفريغ المزدوج :–

بعد تحديد فئات المتغير x وفئات المتغير y تخصص الأعمده للمتغير x ، الصفوف

للمتغير y ونبدأ فى غريغ البيانات وذلك بقراءة أزواج القيم المعطاه واحد تلو الاخر ونبحث عن الخليه التى يتبعها كل زوج من أزواج القيم المعطاه واحد تلو الاخر ونبحث عن الخلية التى يتبعها كل زوج من أزواج القيم وتكون الحزم كما هو موضح فى جدول (7).

جدول (7) : جدول تفريغ مزدوج للصادرات والواردات بالاف الجنيهات

لعدد 50 شركه فى دوله معينه

y \ X	30 -	40 -	50 -	60 -	70 -	80 -	90 - 100	Sum
20 - 30	/	/						2
30 - 40	//	/	//					5
40 - 50		//	////	/				7
50 - 60			///	~~////~~	//			10
60 - 70			/	~~////~~ /	~~////~~	//		14
70 - 80				/	//	///	//	8
80 - 90						/	///	4
Sum	3	4	10	13	9	6	5	50

فمثلا البيان الاول (21 و 45) ان الشركة الاولى كانت قيمه الصادرات لها 45 الف جنيه وقيمه الواردات فيها 21 الف جنيه.

لذلك فان الرقم 45 نجده فى الفئه الثانية للصادرات ، الرقم 21 نجده فى الفئه الاولى للواردات والخليه التى ينتمى اليها هذا البيان هى تقاطع العمود الثالث مع الصف الأول ونضع شرطه عند هذا التقاطع. كذلك البيان السادس (69 , 65) يكون موقعه فى تقاطع العمود الخامس مع الصف السادس أى فى فئه الصادرات (- 60) فئه الواردات (- 60) .....وهكذا

بالنسبة لبقيه البيانات .

تكوين الجدول التكراري المزدوج :ـ

من جدول التفريغ للمزدوج (7) فاننا نترجم العلامات الموجوده في الخلايا الى أرقام والجدول الناتج هو جدول التوزيع التكراري المزدوج للصادرات (x) ، الواردات (y) لعدد 50 شركه كما في جدول (8) .

جدول (8) : جدول التوزيع التكراري المزدوج للصادرات والواردات بالالف الجنيهات لعدد 50 شركة في دوله معينه .

y \ x	30 -	40 -	50 -	60 -	70 -	80 -	90 - 100	Sum
20 -	1	1						2
30 -	2	1	2					5
40 -		2	4	1				7
50 -			3	5	2			10
60 -			1	6	5	2		14
70 -				1	2	3	2	8
80 - 90						1	3	4
Sum	3	4	10	13	9	6	5	50

وهذا الجدول يقرأ كما يلي : مثلا توجد 4 شركات تصدر ما قيمته 50 الف جنيه حتى أقل من 60 ألف جنيه وتستورد ما قيمته 40 ألف جنيه حتى أقل من 50 ألف جنيه ، ... وهكذا .

هذا ومن جدول (8) يمكن استنتاج ثلاث جداول تخص ثلاثه توزيعات وهى :ـ

١ ـ جدول التوزيع التكراري المزدوج للصادرات والواردات وهو جدول (8) ويسمى الجدول

أو خلايا عبارة عن التكرارات المزدوجة للمتغيرين x ، y .

٢ ـ جدول التوزيع التكرارى البسيط للمتغير X ويشتمل على فئات الصادرات (x) والتكرارات التى هى مجموع التكرارات المزدوجة فى جدول (8) والتى تخص كل فئة من فئات المتغير x .
وهذا الجدول يسمى جدول التوزيع الهامشى للمتغير x . وذلك كما هو موضح فى جدول (9) .

٣ ـ جدول التوزيع التكرارى البسيط للمتغير الثانى y ويشتمل على فئات الواردات y والتكرارات التى هى مجموع التكرارات المزدوجة فى جدول (8) والتى تخص كل فئة من فئات المتغير .
وهذا التوزيع التكرارى البسيط يسمى بالتوزيع الهامشى للمتغير y وذلك كما هو موضح فى جدول (10) .

جدول (9)

Intervals (x)	f
30 - 40	3
40 - 50	4
50 - 60	10
60 - 70	13
70 - 80	9
80 - 90	6
90 - 100	5
Sum	50

جدول (10)

Intervals (y)	f
20 - 30	2
30 - 40	5
40 - 50	7
50 - 60	10
60 - 70	14
70 - 80	8
80 - 90	4
Sum	50

لاحظ انه فى الجداول التكرارية المزدوجة لا يشترط ان تكون عدد فترات المتغيرين متساوى (كما فى المثال (٩)) لان هذا يتوقف على المدى الذى يتحرك داخله بيانات المتغيرين والذى غالبا ما يكون مختلفا . وهذا يجعلنا نختار عدد من الفئات لكل من المتغيرين وهذا العدد من الفئات غير متساوى .

التوزيعات التكرارية المتجمعه الصاعدة والهابطة :-

تلزم بعض الدراسات والتحليلات الاحصائيه ان تحول أرقام جداول التكرارية البسيطه الى أرقام متجمعه اما تصاعديا أو تنازليا بمعنى آخر يلزمنا ايضا عدد المفردات أعلى أو أسفل حدود الفئه.

فمثلا من بيانات مثال (٣) قد يكون المراد هو معرفه عدد العمال الذين ينفقون أقل من مبلغ معين أو اكبر من مبلغ معين. أو قد يكون المراد هو معرفة عدد العمال الذين ينفقون أقل من 80 جنيه أو عدد العمال الذين ينفقون أكثر من 90 جنيه مثلا.

وحيث ان الجداول التكرارية البسيطه لا يمكن ان تعطى اجابه مباشرة لهذه التطلبات فانه يوجد نوعان من الجداول تسمى بالجداول التجمعيه تستطيع ان تعطى أو تقدم لنا الاجابه على هذه الاسئلة مباشرة وهذه الجداول هى الجداول التكرارية المتجمعه الصاعدة والتى تستخدم فى تحديد الشركات أو المفردات التى تأخذ قيمه اقل من قيمه معينه وكذلك تحديد القيمة التى يقل عنها عدد معين من المفردات ، والجداول التكرارية المتجمعه الهابطه التى تستخدم فى تحديد الشركات أو المفردات التى تأخذ قيما أكبر من أو تساوى قيمه معينه وكذلك تستخدم فى الحصول على القيمة التى يزيد عنها أو يساوى عدد معين من المفردات.

وفيما يلى نوضح كيفية تكوين الجداول التكرارية المتجمعه من الجداول التكرارية البسيطه.

أولا : جدول التوزيع التكرارى المتجمع الصاعد :-

وسوف نستخدم المثال التالى لشرح تكوين الجدول المتجمع الصاعد كما يلى :-

مثال (٥) :

بأستخدام بيانات جدول التوزيع التكرارى للانفاق اليومى لعدد ١٥٥ عامل والموضح فى جدول (٦) المطلوب انشاء جدول التوزيع التكرارى المتجمع الصاعد للبيانات المعطاه.

الحل :- كما فى جدول التوزيع التكرارى البسيط للانفاق اليومى يتكون من عمودين هى الفئات والتكرارات فان الجدول التكرارى المتجمع الصاعد يتكون من عمودين يتم اشتقاقهما من العمودين

الســابقين وهذين العمودين همــا :ـ

الحدود العليــا للفئــات والتكــرارات المتجمعــة الصــاعــده ٠

١ ـ يتكون العمود الاول من بيانات عمود الفئات فى الجدول الاصلى وذلك بأخذ الحد الاعلى لكــل فئه ونكتب أمام الفئه التى لها هذا الحد الاعلى " أقل من يأخذ الحد الاعلى " وذلك لجميــع الفئات وبالنسبة لبيانات المثال (٣) والموجوده فى جدول (6) نجد ان الحدود العليــا للفئات كما هو موضحه فى جدول (١١) تكون كما يلى أقل من 50 ، أقل من 60 ٠٠٠ وهكذا ٠

٢ ـ يتكون العمود الثانى من عمود التكرارات فى جدول التوزيع التكرارى البسيط حيث :

التكرار المتجمع الصاعد لاى فئه = التكرار الاصلى لهذه الفئه + مجموع تكرارات الفئات السابقة لهــا مباشــرة ٠

أو التكرار المتجمع الصاعد لاى فئه = التكرار الاصلى لهذه الفئه + التكرار المتجمع الصاعد للفئه السابقة لها مباشــره٠

وبالنســبة للمثال (٣) فى جدول (٦) نجــد ان :ـ

التكرار المتجمع الصاعد للفئه قبل الاولى = صــفر (التى فرضنــا وجودها) ٠

التكرار المتجمع الصاعد للفئه الاولى = $6 = 0 + 6$

التكرار الاصلى لهــا

التكرار المتجمع الصاعد للفئه الثانيه = $12 + 6 = 18$

وهكذا بالنسبة لبقيه الفئات كما فى الجدول التالى :ـ

ـ جدول (١١) التوزيع التكرارى المتجمع الصاعد للانفاق اليومى لعدد 100 عامــل

التكرارات المتجمع الصاعد	الحدود العليا للفئات	التكرارات	الفترات
6 + 0 = 6	less than 50 / less then 60	0 / 6	50 - 60
6 + 12 = 18	" " 70	12	60 - 70
18 + 15 = 33	" " 80	15	70 - 80
24 + 33 = 57	" " 90	24	80 - 90
18 + 57 = 75	" " 100	18	90 - 100
14 + 75 = 89	" " 110	14	100 - 110
11 + 89 = 100	" " 120	11	110-120

ثانيا : جدول التوزيع التكرارى المتجمع الهابط :ـ

مثال (٦) : من بيانات المثال (٣) فى جدول (٦) المطلوب انشاء جدول التوزيع التكرارى المتجمع الهابط من البيانات المعطاه .

الحل : كما ان جدول التوزيع التكرارى البسيط (٦) يتكون من عمودين هما الفئات والتكرارات فانه كذلك جدول التوزيع التكرارى المتجمع الهابط يتكون من عمودين يتم اشتقاقها عن عمودى الفئات والتكرارات السابقين ، وهذين العمودين اللذين يكونا جدول التوزيع التكرارى المتجمع الهابط هما : الحدود الدنيا للفئات ـ التكرارات المتجمعه والهابطه .

١ ـ يتكون العمود الاول من بيانات عمود الفئات فى الجدول الاصلى ، وذلك باخذ الحد الادنى لكل فئه من الفئات ونكتب امام الفئه التى لها هذا الحد الادنى " قيمه الحد الادنى فأكثر بالنسبة للبيانات الموجودة فى جدول (٦) نجد ان الحدود الدنيا للفئات ، وكما هى موضحة فى جدول (١٢) تكون كما يلى :ـ

50 فأكثر ، 60 فأكثر ،...... ، 110 فأكثر ، 120 فأكثر .

٢ ـ ويتكون العمود الثانى من عمود التكرارات فى الجدول الاصلى (٦) وذلك كما يلى :

سوف نوجد التكرارات المتجمعة الهابطة للفئات المختلفه ، وذلك ابتداء من الفئه الاخيرة ثم الفئه قبل الاخيرة ، ... وهكذا حتى التكرارات المتجمعه الهابطة للفئه الاولى وحيث نستخدم القاعدة التاليه :ـ

تكرار متجمع هابط للفئه المفروضه وهى بعد الفئه الاخيرة = صفر .

تكرار متجمع هابط للفئه الاخيرة = التكرار الاصلى لهذه الفئه = ١١

تكرار متجمع هابط للفئه قبل الاخيره = التكرار الاصلى لهذه الفئه +

+ تكرار متجمع هابط للفئه التاليه وهى الاخيرة .

... وهكذا . وبصورة عامه يمكن القول :ـ

التكرار المتجمع الهابط لاى فئه = التكرار الاصلى لهذه الفئه +

+ التكرار المتجمع الهابط للفئه التاليه لها مباشرة .

وذلك تحت شرط بدء الحصول على هذه التكرارات المتجمعه الهابطه من نهايه الجدول كما فى جدول (١٢) .

جدول (12) التوزيع التكرارى المتجمع الهابط للانفاق اليومى لعدد 100 عامل .

Intervals	freq.	lower limits	decreasing cumulative freq.
50 - 60	6	50 or more	6 + 94 = 100
60 - 70	12	60 " "	12 + 82 = 94
70 - 80	15	70 " "	15 + 67 = 82
80 - 90	24	80 " "	24 + 43 = 67
90 - 100	18	90 " "	18 + 25 = 43
100 - 110	14	100 " "	14 + 11 = 25
110 - 120	11	110 " "	11 + 0 = = 11
120 - 130	0	120 " "	0
	100		

ومن جدول (١٢) يمكننا الحصول على معلومات غير متاحه مباشرة فى جدول التوزيع التكرارى البسيط (6) فمثلا نجد انه يوجد 82 عامل انفاقهم اليومى يبلغ 70 فأكثر و 25 عامل من بين 100عامل انفاقهم اليومى يبلغ 100 جنيه فأكثر .

أنواع الجداول التكراريه لبيانات مستمره

أولا : جداول تكرارية بسيطه : وهى نوعان :-

أ - جداول تكرارية بسيطه منتظمه اى فئات الجدول تكون متساوية ومثال ذلك جدول توزيع الانفاق اليومى (6) .

ب - جداول تكرارية بسيطه غير منتظمة اى فئات الجدول تكون غير متساوية ويكفى فى هذه الجداول الغير منتظمة ان توجد فئه واحدة من بين فئات الجدول غير مساوية لاطوال بقية الفئات الاخرى حتى تجعل الجدول غير منتظم .

أيضا الجداول التكرارية البسيطة سواء متساوية الفئات أو غير متساوية الفئات قد تكون جداول مغلقة أو جداول مفتوحة.

والجداول المغلقة هي الجداول التي يكون معلوم فيها بداية الفئة الاولى ومعلوم ايضا نهاية الفئة الاخيرة.

والجداول المفتوحة ثلاثة انواع :-

أ - جدول مفتوح من أعلى أي غير معلوم بداية الفئة الاولى . ولتفادى هذا العيب احيانا تأخذ طول الفئة الاولى (المفتوحة) مساويا لطول الفئة التالية لها مباشرة.

ب - جداول مفتوحة من أسفل . . أى نهاية الفئة الاخيرة غير معلوم . ويمكن ايضا تحديد طول الفئة الاخيرة بحيث يكون مساويا لطول الفئة السابقة لها مباشرة.

جـ - جداول مفتوحة من الطرفين أى من أسفل ومن أعلى وهى تشمل النوعين السابقين (أ) ، (ب) .

ثانيا : جداول تكرارية مزدوجة :-

الجداول التكرارية المزدوجة تبين التوزيع التكرارى لظاهرتين مثل . الدخل والانفاق أو الطول والوزن أو الصادرات والواردات ، . . . الخ والجداول التكرارية المزدوجة نوعان :-

أ - جداول تكرارية مزدوجة منتظمة اى أطوال فئات كل من الظاهرتين على حدة تكون متساوية.

مثلا : فى جدول الدخل والانفاق لكى يكون الجدول المزدوج منتظم لابد ان تكون فئات الدخل متساوية فيما بينها وايضا فئات الانفاق تكون متساوية فيما بينها .

ب - جداول تكرارية مزدوجة غير منتظمة ، أى اطوال فئات احدى الظاهرتين غير متساوية أو أطوال فئات كل من الظاهرتين على حدة غير متساوية.

وأيضا كما فى الجداول التكرارية البسيطة ، فأنه فى الجداول التكرارية المزدوجة سواء منتظمة أو غير منتظمة اى متساوية الفئات أو غير متساوية الفئات قد تكون هذه الجداول أما مغلقة أو جداول مفتوحة.

ثالثا : جداول تكرارية مركبة :-

اى انها تبين التوزيع التكرارى لاكثر من ظاهرتين مثل الطول والوزن والعمر . . . .

وهذه الجداول المركبة ايضا قد تكون منتظمة او غير منتظمة وايضا قد تكون مغلقة أو مفتوحة.

ثانيا : عرض البيانات الاحصائية بيانيا

درسنا فيما سبق الطرق المختلفة لعرض البيانات فى جداول يسهل استخدامها ولكن هذه الجداول الصماء لا تكفى وحدها لعرض البيانات الاحصائيه ولا يمكن الاستفادة منها وخاصه لغير المتخصصين

لذا فانه يلزم عرض البيانات الاحصائيه فى صورة رسوم بيانية تمكن للقارئ العادى فهم هذه البيانات وسرعة أدراك مغزاها ومضمونها بصورة ايسر مما لو كانت معروضه فى جداول. كذلك الرسوم البيانية تسهل عليه مقارنة البيانات بالظواهر ببعضها البعض.

فيما يلى طرق عرض البيانات الاحصائية بيانيا وذلك سواء أكانت هذه البيانات وصفيه أو رقميه متقطعه (منفصله) ومتصله كل على حده.

وفى عرض البيانات الاحصائيه بيانيا سوف نفصل بين نوعين من البيانات وهى البيانات غير المبوبه البيانات المبوبه.

١ ـ العرض البيانى للبيانات غير المبوبه :

والبيانات غير المبوبه هى البيانات التى لا تعرض فى جدول توزيع تكرارى وسوف ندرس البيانات غير المبوبه كل من البيانات الوصفية والبيانات الرقمية المتقطعه والمتصله.

وأنواع الرسوم البيانية لعرض البيانات غير المبوبه هى : ـ

١ ـ الاعمدة أو الاشرطه البيانيه ٢ ـ الدوائر.

١ ـ الأعمدة أو الاشرطه البيانية :ـ

وفى تمثيل البيانات بالاعمدة نرسم محورين ، محور افقى تمثل عليه أقسام الظاهره المختلفة ومحور رأسى تمثل عليه قيمه الظاهرة موضوع الدراسة وذلك بمقياس رسم مناسب. وفى حاله الانواع الثلاثة للاشرطه البيانية تكون هذه الاشرطه عباره عن مستطيلات قواعدها متساويه وتمثل على المحور الافقى والتى هى أقسام الظاهرة وأرتفاعات هذه المستطيلات تمثل قيم الظاهره المختلفه والتى تتناسب

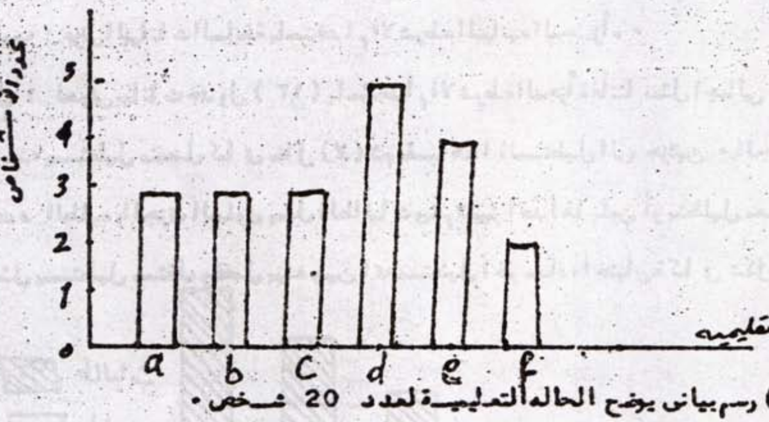
مع مساحات المستطيلات لان القواعد متساويه .

وفيما يلي انواع الاشرطة البيانية مع الامثلة التوضيحية .

أ ـ الاعمدة أو الاشرطه البيانية البسيطه أو المنفصله :ـ

مثال (٧) : باستخدام بيانات جدول (٢) . المطلوب عرض هذه البيانات باستخدام الاشرطه البيانية البسيطه (المنفصلة) .

الحل :ـ الاشرطة البيانية البسيطة عبارة عن مستطيلات متساوية القواعد كل منها يمثل قسم من اقسام الظاهرة لذلك فان ارتفاعاتها تتناسب مع مساحاتها .



شكل (١) رسم بياني يوضح الحالة التعليمية لعدد 20 شخص .

لاحظ ان هذه البيانات وصفية وان طريقه الاشرطه المنفصلة تعتبر أنسب عرض بياني لمثل هذه البيانات .

ب ـ الاعمدة أو الاشرطه البيانية المجزأة :ـ

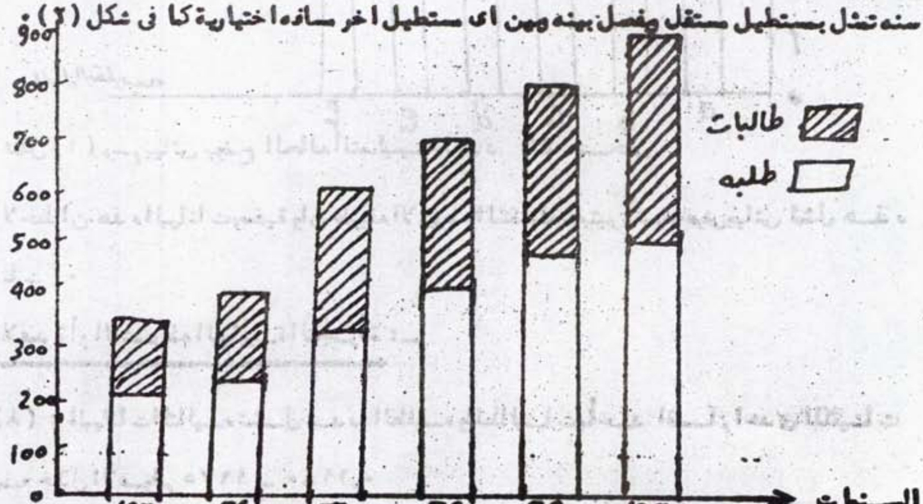
مثال (٨) : البيانات التاليه تمثل عدد الطلبه والطالبات بأحد اقسام احدى الكليات العلمية خلال الاعوام ١٩٧٥ ـ ١٩٨٠م .

جـــدول (١٣)

السنوات	الطلبــــه	الطالبـات	مجموع
1975	200	150	350
1976	230	170	400
1977	320	280	600
1978	400	300	700
1979	470	330	800
1980	500	400	900

والمطلوب : عرض البيانات السابقة باستخدام الاشرطه البيانيه المجـزأه .

<u>الحـل</u> : لعرض بيانات جدول (١٣) باستخدام الاشرطه المجزأة فأننا نمثل اجمالى عـدد الطلاب فى كل سنه بمستطيل منفصل كما فى شكل (٢) ثم نقسم هذا المستطيل الى جزئين . الجزء السـفلى يمثل عدد الطلبه والجزء العلوى يمثل الطالبات ويتم تمييز احداها بلون أو بتظليل معين . وكـل سنه تمثل بمستطيل مستقل ويفصل بينه وبين اى مستطيل اخر مسافه اختياريه كما فى شكل (٢) .

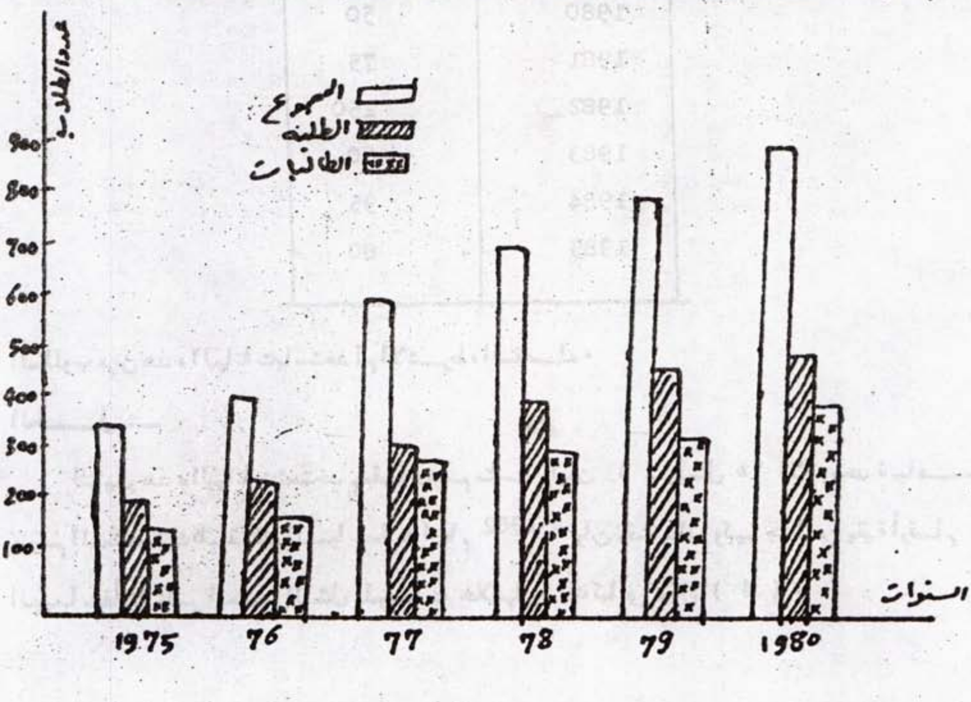


شكل (٢) اعداد الطلبة والطالبات بأحد اقسام الكلية العملية خلال الاعوام (1975-1980)

جـ- الاعمدة أو الاشرطة البيانية المتلاصقة :-

مثال (9) باستخدام البيانات جدول (١٣) المطلوب عرض هذه البيانات باستخدام الاعمدة المتلاصقة .

الحل : فى هذه الحالة يتم رسم ثلاثة مستطيلات لكل سنة على حدة وهذه المستطيلات الثلاثة يخصص منها الأول لمجموع الطلبة والطالبات معا ، والثانى لعدد الطلبة والثالث لعدد الطالبات ثم تميز هذه المستطيلات عن بعضها البعض ، وعمل دليل للرسم وذلك كما سنرى فى شكل (٣) :-



شكل (٣) اعداد الطلبة والطالبات باحد اقسام احدى الكليات العلمية

خلال الاعوام (1975 - 1980) .

مثال (10) :ـ

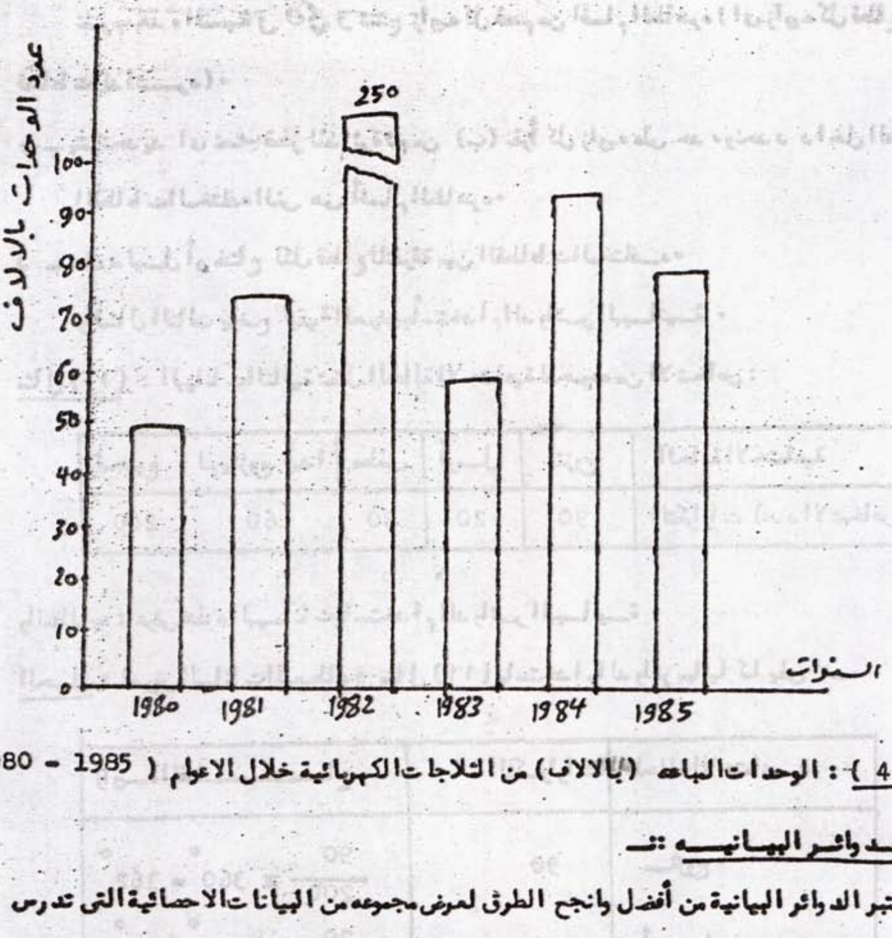
البيانات التالية تمثل عدد الوحدات المباعه بالآلاف من الثلاجات الكهربائية لاحدى الشركات خلال الاعوام (1980 - 1985) .

years	Units no.
1980	50
1981	75
1982	250
1983	60
1984	95
1985	80

المطلوب عرض هذه البيانات باستخدام الاشرطه المنفصله .

الحـــل :ـ

لتمثيل هذه البيانات نستخدم مقياس رسم مناسب وليكن ١ سم يمثل ١٠ آلاف وحدة مباعـــه ونرسم المستطيلات كاملة لكل السنوات ما عدا عام 1982 وان رقمة شاذ وكبير جدا عن بقية أرقـام المبيعات فأننا نكسر العمود الممثل للمبيعات خلالها وذلك كما فى شكل (4) .



شكل (4) : الوحدات المباعه (بالالاف) من الثلاجات الكهربائية خلال الاعوام (1985 - 1980)

٢ ـ الـدوائر البيانيــه :ـ

تعتبر الدوائر البيانية من أفضل وانجح الطرق لعرض مجموعه من البيانات الاحصائية التى تدرس ظاهرة معينه تحوى عدد من الاقسام. وعند تمثيل بيانات تلك الظاهرة بأستخدام الدوائر البيانيــه فأنة يتم تمثيل الظاهرة بدائرة ، حيث مساحة الدائرة تمثل اجمالى قيم اقسام الظاهره ، وبعد رسم الدائرة بمقايس رسم مناسب نقسم الى قطاعات كما يلى :ـ

أ ـ تحويل قيم كل قسم الى نسبة مئوية من اجمالى قيم اقسام الظاهره.

وذلك بقسمة كل قيمه على اجمالى القيم والضــرب × ١٠٠ .

ب ـ تقسيم مساحة الدائرة حسب النسب المئوية لاقسام الظاهرة والتى حصلنا عليها فى (أ) لــذا

تضرب هذه النسبة فى 360 ْ تنتج زاويه كل قسم من اقسام الظاهره (اى زاويه كل قطاع من قطاعات الدائـــره) .

جـ ـ يتم تحديد اى نصف قطر للدائرة ثم من (ب) نقرأ كل زاويه على حده ونحدد داخل الدائـــرة القطاعات المختلفه التى هى أقسام الظاهره.

د ـ عمل دليـل أو مفتاح لكل قطاع للتفرقة بين القطاعات المختلفــه.

والمثال التالى يوضح كيفية العرض بأستخدام الدوائـر البيـانيــة .

مثال (١١) : البيانات التالية تمثل الحالة الاجتماعية لمجموعه من الاشخاص :

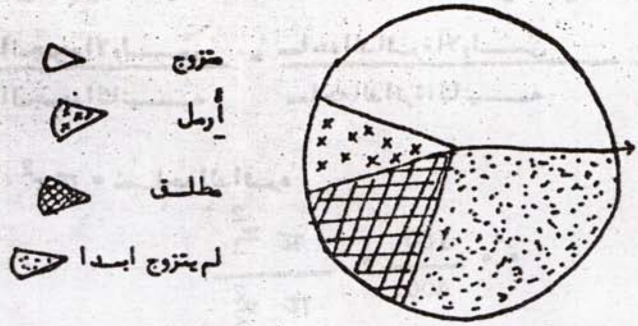
الحالة الاجتماعية	متزوج	أرمـل	مطلق	لم يتزوج ابدا	المجموع
التكرارات (عدد الاشخاص)	90	20	30	60	200

والمطلوب : عرض هذه البيـانات بأستخدام الدوائـر البيـانيـة .

الحـل : لعرض البيانات المعطاه فى مثال (١١) بإستخدام الدوائر بيانيا كما يلى :ـ

الحالة الاجتماعيـة	التكـرارات	زاويـه القطـــــــاع
متزوج	90	$\frac{90}{200} \times 360^\circ = 162^\circ$
أرمـــل	20	$\frac{20}{200} \times 360^\circ = 36^\circ$
مطلـــق	30	$\frac{30}{200} \times 360^\circ = 54^\circ$
لم يتزوج ابدا	60	$\frac{60}{200} \times 360^\circ = 108^\circ$
Σ	200	360°

ويمكن تمثيل بيانات الحالة الاجتماعية باستخدام الدائرة كما فى شكل (٥)٠



شكل (٥) الحالة الاجتماعية لعدد 200 شخص٠

هذا ويمكن استخدام الدوائر البيانية فى مقارنة اقسام الظاهرة المختلفة خلال فترتين زمنيتين او مكانين مختلفين والمثال التالى يوضح ذلك٠

مثال (١٢) :ـ البيانات التالية توضح توزيع أعداد مجموعتين من العمال على الأنشطة الاقتصاديه المختلفه:

المجموعه الثانيه	المجموعه الاولى	الانشطه الاقتصادية
140	30	صناعيه
120	40	زراعية
80	20	تجاره
60	10	خدمات
400	100	$\sum$

والمطلوب المقارنة بين اعداد العمال الذين يمارسون الانشطة الاقتصادية المختلفة فى المجموعتين باستخدام الدوائر البيانيه٠

الحل : لتمثيل البيانات المعطاه باستخدام الدوائر البيانية فانه يلزم رسم دائرة لكل مجموعه من

المجموعتين وحيث ان مساحه الدائرة تمثل احدى المجموعتين تتناسب مع مجموع افراد هذه المجموعه فان :

$$\frac{\text{مجموع افراد المجموعه الاولى}}{\text{مجموع افراد المجموعه الثانيه}} = \frac{\text{مساحه الدائرة الاولى}}{\text{مساحه الدائرة الثانيه}}$$

$$\text{مساحه الدائره} = \pi r^2$$

$$\therefore \frac{100}{400} = \frac{\pi r_1^2}{\pi r_2^2}$$

$$\therefore \frac{r_1^2}{r_2^2} = \frac{1}{4}$$

$$r_1 = \frac{1}{2} r_2$$

$$r_1 = 2 \text{ cm} \Rightarrow r_2 = 4 \text{ cm}$$

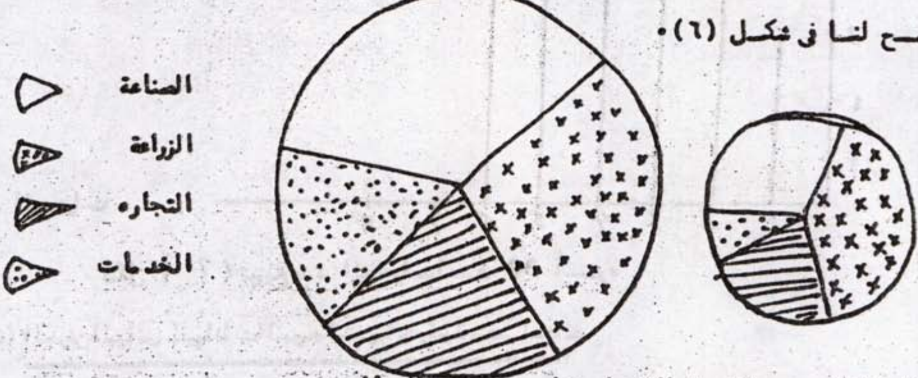
بعد تحديد نصف القطر لكل من الدائرتين اللتين تمثلا المجموعتين من العمال فانه يلزم تحديد زاويه القطاع لكل دائرة ولذلك نتبع نفس الطريقة السابق استخدامها فى مثال (11) ونكون الجدول التالى :-

الانشطه الاقتصادية	المجموعه الاولى		المجموعه الثانية	
	عدد العمال	زاويه القطاع	عدد العمال	زاويه القطاع
الصناعة	30	108°	140	126°
الزراعه	40	144°	120	108°
التجاره	20	72°	80	72°
الخدمات	10	36°	60	54°
	100	360°	400	360°

حيث زاويه القطاع كما ذكرنا يمكن الحصول عليها

$$= \frac{\text{التكرار الاصلي}}{\text{مجموع التكرارات}} \times 360^\circ$$

وبعد ذلك نرسم دائرتين ، دائرة تمثل المجموعة الاولى بنصف قطر 2 cm والثانية تمثل المجموعة الثانية بنصف قطر 4 cm وتقسم كل منها الى قطاعات كما في الجدول السابق وهذا يتضح لنا في شكل (٦).



شكل (٦) توزيع الانشطه الاقتصادية لمجموعتين من العمال .

هذا ويمكن استبدال التشيل الدائرة بنصفها أو بربعها ويقابل كل 1% من النسب زاويه قدرها 1.8° على نصف الدائرة ، 0.9° درجة على ربع دائرة.

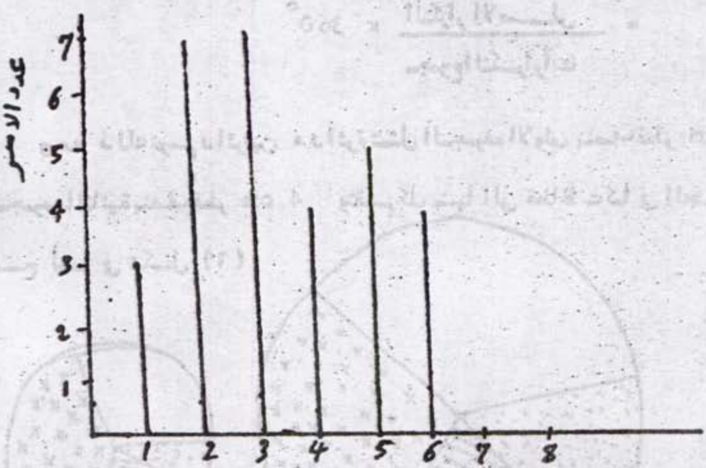
٢ — العرض البياني للبيانات المبوبه :—

البيانات المبوبه هي البيانات المعروضة في جداول التوزيع التكرارية وعرض البيانات المبوبه لمتغيرات رقمية سواء متقطعه او متصله بيانيا وذلك كما يلي :—

أ — العرض البياني للبيانات المبوبه لمتغيرات رقمية متقطعه :—

عند عرض البيانات الرقمية المتقطعة المبوبه بيانيا فاننا نرسم المحورين الافقي والتي تمثل قيم الظاهره ، ثم الرأسي والذى تمثل عليه تكرارات كل قيمه من قيم الظاهره ثم نقرأ كل قيمة والتكرار الناظر لها ومن نقطة التلاقي نرسم خط عمودى الاحداثي الافقي له هو قيمة الظاهره والاحداثي الرأسي له هو التكرار الناظر لهذه القيمه.

مثال (١٢): اذا كانت لدينا البيانات الواردة في جدول (4) والمطلوب عرض هذه البيانات بيانيا .



شكل (7) توزيع عدد الحجرات لعدد 30 أسره

(ب) العرض البياني للبيانات المبوبه لمتغيرات رقمية متصله :ـ

يمكننا عرض البيانات المبوبه بيانيا لمتغيرات رقمية متصله والتي يكون الجدول التكراري لها في صورة فئات (سواء متساوية او غير متساوية) وتكرارات وذلك باستخدام احدى الطرق التالية :ـ

١ ـ المدرج التكراري　　٢ ـ المضلع التكراري

٣ ـ المنحنى التكراري　　٤ ـ المنحنى التكراري المتجمع الصاعد

٥ ـ المنحنى التكراري المتجمع الهابط.

١ ـ المدرج التكراري :ـ

لاستخدام المدرج التكراري في عرض البيانات الرقمية المتصلة سوف ندرس ذلك في حالة الجداول التكرارية المنتظمة والغير منتظمه كما يلي :ـ

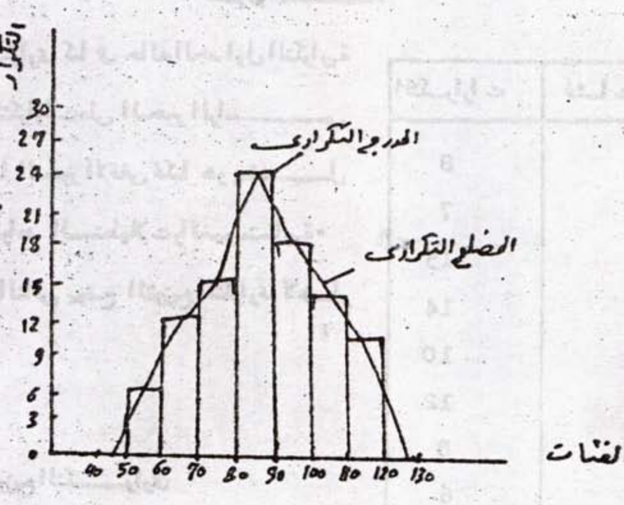
أ ـ المدرج التكراري لجداول تكرارية منتظمة (متساويه الفئات) :ـ

المدرج التكراري عباره عن مستطيلات متلاصقة قواعدها على المحور الافقي تمثل أطوال الفئات والتي تكون متساوية وأرتفاعاتها تمثل التكرارات الخاصه بهذه الفئات.

ولرسم المدرج التكرارى نرسم محورين متعامدين الافقى يقسم الى اجزاء متساوية هى طول الفئه مع ترك طول فئتين أو اكثر على يمين المحور الرأسى . أما المحور الراسى يمثل التكرارات ومقياس رسم مناسب على ان يبدأ من الصفر . ثم بعد ذلك نقرأ كل فئه والتكرار المناظر لها أبتداء من الفئـــه الاولى فى الجدول ونرسم مستطيل قاعدته هى طول الفئه وأرتفاعة هو تكرار هذه الفئه ثم بالنسبه للفئه الثانية . . . . . وهكذا حتى نهايه الجدول وبذلك يتكون لدينا شكل متدرج يطلق عليه اســـم المدرج التكرارى " .

مثال (١٣) : باستخدام البيانات الواردة فى جدول (6) والخاصه بالتوزيع التكرارى للانفـــاق اليومى بالجنيه لعدد 100 عامل . أعرض هذه البيانات فى شكل مدرج تكرارى .

الحـــل :-



شكل (8) المدرج والمضلع التكرارى للانفاق اليومى لعدد 100 عامـل .

(ب) المدرج التكراري لجداول تكرارية غير منتظمة (غير متساوية الفئات)

عند رسم المدرج التكراري لجداول تكرارية منتظمة نجد ان مساحة المستطيلات تتناسب مع ارتفاعات هذه المستطيلات والتي هي التكرارات وذلك لان اطوال الفئات متساوية والتي هي قواعد هذه المستطيلات.

ولكن اذا كانت الفئات غير متساوية فان قواعد المستطيلات تكون غير متساوية وبالتالي فان ارتفاعات المستطيلات لا تمثل التكرارات لذا فان المساحة الكلية للمستطيلات لا تتناسب مع مجموع التكرارات الكلية.

لتفادي هذا الخطأ فانه عند رسم المدرج التكراري لجدول تكراري فئاته غير متساوية يلزم تعديل تكرار كل فئة وذلك بالحصول على ما يسمى " بالتكرار المعدل " وذلك لكل فئة على حدة حيث :-

$$\text{التكرار المعدل لاي فئة} = \frac{\text{التكرار الاصلي لهذه الفئة}}{\text{طول الفئة}}$$

ثم نقوم برسم المدرج التكراري كما في حالة الجداول التكرارية المنتظمة مع استبدال التكرارات على المحور الراسي بالتكرارات المعدلة. اما المحور الافقي فكما هو يمثل الفئات التي تكون هي قواعد المستطيلات والغير متساوية.

__مثال (14)__: الجدول التالي يوضح التوزيع التكراري لاعمار مجموعة من الاشخاص.

__جدول (15) التوزيع التكراري__
لاعمار 80 شخص.

فئات العمر	التكرارات
20 -	8
25 -	7
30 -	15
35 -	14
40 -	10
45 -	12
55 -	8
65 -	6
	80

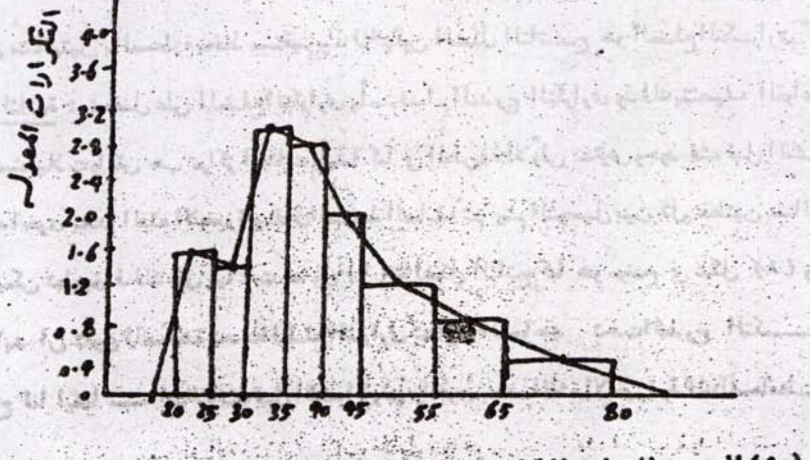
المطلوب : رسم المدرج التكراري لهذه البيانات

الحـل : حيث ان جدول (15) غير منتظم أى فئاته غير متساوية ، فانه لابد من ايجاد التكرارات المعدله وذلك كما يلى :ـ

جدول (16) : جدول التوزيع التكرارى والتكرارات المعدله لاعمار 80 شخص .

فئات العمر	التكرارات	طول الفئه	التكرارات المعدله
20 - 25	8	5	1.6
25 - 30	7	5	1.4
30 - 35	15	5	3.0
35 - 40	14	5	2.8
40 - 45	10	5	2.0
45 - 55	12	10	1.2
55 - 65	8	10	0.8
65 - 80	6	15	0.4
	80		

المدرج التكرارى الخاص ببيانات جدول (16) موضح فى شكل (9) .



شكل (٩) المدرج والمضلع التكرارى لاعمار 80 شخص .

لاحظ من شكل (٩) ان مجموع ساحات المستطيلات لابد ان يكون ساويا لمجموع التكرارات .

٢ ــ المضلع التكراري :ــ

سوف ندرس المضلع التكراري في حاله الجداول التكرارية المنتظمة والجداول التكرارية الغير منتظمة وقبل ان نبدأ لابد لنا ان نعرف ما هو مركز الفئه.

مركز الفئه = بدايه الفئه + $\frac{1}{2}$ طول الفئه

$$= \frac{\text{بداية الفئه + نهايه الفئه}}{2}$$

أ ــ المضلع التكرارى لجداول تكرارية منتظمة (متساوية الفئات) :ــ

توجد طريقتان للحصول على المضلع التكرارى في كلتا الحالتين تستخدم مراكز الفئات والتكرارات وذلك كما يلى :ــ

الطريقة الاولى :ــ نرسم المحور الافقى والذى يمثل الفئات وتحدد عليه مراكز الفئات ونرسم المحور الراسى ونمثل عليه التكرارات ثم يتم قراءة كل مركز فئه من الفئات والتكرار المناظر لها ونحدد نقطة التلاقى .... وهكذا ثم نتخيل أو نفترض وجود فئه قبل الفئه الاولى ونفس طول الفئه الاولى ولها تكرار يساوى صفر ونوجد مراكزها ونحدده بنقطة . كذلك نفترض وجود فئه بعد الفئه الاخيره طولها يساوى نفس طول الفئه الاخيرة وليس لها تكرار ايضا . ونحدد مركزها بنقطة. ثم بعد ذلك يتم التوصيل بين كل نقطتين متتاليتين بالمسطرة بخط مستقيم وبذلك يكون الشكل الناتج هو المضلع التكرارى.

الطريقة الثانية : نحصل على المضلع التكرارى بأستخدام المدرج التكرارى وذلك بتنصيف القواعد العليا للمستطيلات والتى هى مراكز الفئات وأيضا كما فى الطريقة الاولى نفترض وجود فئه قبل الفئه الاولى وفئه اخرى بعد الفئه الاخيرة وبنفس الشروط السابقة ثم يتم التوصيل بين كل نقطتين متتاليتين بالمسطرة ويمكن تطبيق ذلك على بيانات جدول (٦) والشكل الناتج كما هو موضح فى شكل (٨) ومنه نجد انه لابد ان تكون المساحة تحت المضلع التكرارى تساوى المساحه تحت المدرج التكرارى وهذا يوضح لنا ايضا سبب اضافة فئه قبل الفئه الاولى واخرى بعد الفئه الاخيرة وذلك للمحافظة

على ان المساحة تحت الضلع التكرارى تكون مساوية تماما للمساحة تحت المدرج التكرارى •

(ب) الضلع التكرارى لجداول تكرارية غير منتظمة (غير متساويه الفئات) :

هنا نستخدم المدرج التكرارى كما فى الشكل (٩) بان تنصف القواعد العليا للمستطيلات الممثلة لها ونتخيل وجود فئه قبل الفئه الاولى ومساويه لها فى الطول وتكرارها يساوى صفر وتنصفها ايضا ثم من نقط التنصيف نصل بين كل نقطتين متتاليتين بالمسطرة بخط مستقيم ٠٠٠٥ وهكذا حتى الفئه الاخيرة والتى فرضنا ان تكرارها يساوى صفر • وذلك حتى تكون المساحة تحت الضلع مساويه للمساحة تحت المدرج التكرارى تقريبا .

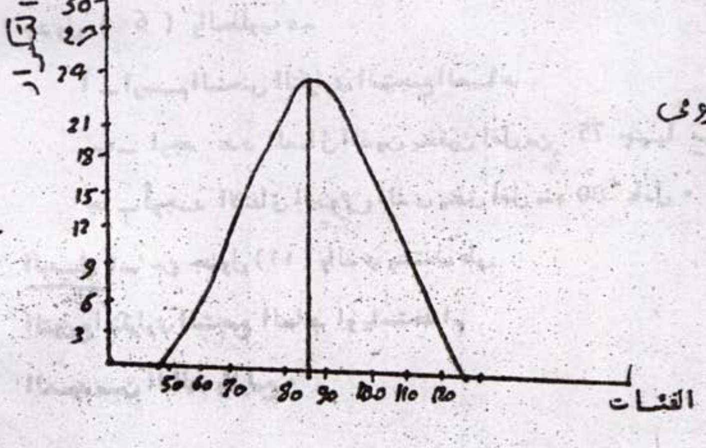
٣ ـ المنحنى التكرارى :ـ

يتم رسم المنحنى التكرارى وذلك باستخدام الضلع التكرارى وذلك عن طريق تمهيد خطوطه المستقيمة باليد • حيث ان الضلع التكرارى عباره عن مجموعه من الخطوط المستقيمة المتصلة ببعضها وكأنه خط مستقيم تم تكسيره فانه عند زيادة عدد الفئات وصغر حجمها فان حدة تكسير الخط المستقيم المكون للضلع التكرارى تقل وبالتالى عندما تكون أطوال الفئات متناهيه فى الصغر فان شكل الضلع التكرارى يكون انسيابى يقترب من شكل منحنى يسمى بالمنحنى التكرارى ومن ذلك نجد ان المساحة تحت المنحنى التكرارى غير مساوية للمساحة تحت الضلع أو المدرج التكرارى والتى تتناسب مع مجموع التكرارات الكلى •

شكل (١٠)

المنحنى التكرارى للانفاق اليومى

لعدد 100 عامل•



وايضا من شكل (٩) يمكن رسم المنحنى التكرارى وذلك باستخدام المضلع التكرارى وبدلا من التوصيل بين نقاطه المتتاليه بخط مستقيم فاننا نمهد المنحنى التكرارى باليد بحيث يمر بالغالبية العظمى من هذه النقط.

٤ـ المنحنى التكرارى المتجمع الصاعد :ـ

سبق ان ذكرنا ان الجدول التكرارى المتجمع الصاعد يستخدم فى تحديد عدد المفردات التى تأخذ قيمه اقل من قيمة معينه وكذلك تحديد القيمة التى يقل عنها عدد معين من المفردات . ويتم تمثيل بيانات الجدول التكرارى المتجمع الصاعد بيانيا باستخدام المنحنى التكرارى التجميع الصاعد ويتم رسم هذا المنحنى كما يلى :ـ

نرسم محورين متعامدين ، المحور الافقى يمثل الحدود العليا للفئات والمحور الراسى يمثل التكرارات المتجمعه الصاعدة ويتم توقيع نقاط الجدول التكرارى المتجمع الصاعد ، حيث كل نقطه عبارة عن الحد الاعلى لكل فئه ، ... وهكذا حتى نهاية الجدول ثم التوصيل بين كل نقطتين متتاليتين بالتمهيد باليد والشكل الناتج يسمى المنحنى المتجمع الصاعد .

مثال (15) :ـ

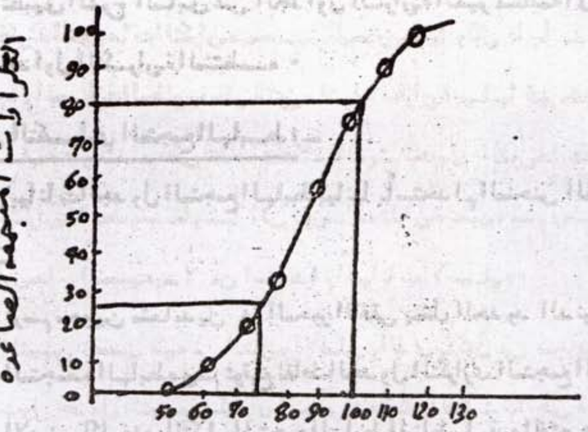
باستخدام بيانات جدول التوزيع التكرارى للانفاق اليومى لعدد 100 عامل والموضحة فى جدول (6) والمطلوب :ـ

أ ـ ارسم المنحنى التكرارى المتجمع الصاعد

ب ـ اوجد عدد العمال الذين ينفقون اقل من 75 جنيها يوميا .

جـ ـ أوجد الانفاق اليومى الذى ينفق اقل منه 80 عامل .

الحل :ـ من جدول (١١) والذى يشتمل على التوزيع التكرارى المتجمع الصاعد او باستخدام العمودين الثالث والرابع .



شكل (١١) المنحنى التكرارى المتجمع الصاعد للانفاق اليومى

لعدد 100 عامل

ب ـ عدد العمال الذين ينفقون أقل من 75 جنيه. يومياً

بعد رسم المنحنى المتجمع الصاعد كما فى شكل (١١) فانه للحصول على عدد العمال الذين ينفقون أقل من 75 جنيها يومياً فمن المحور الافقى ومن موقع الرقم 75 نرسم موازيا للمحور الرأسى يلاقى المنحنى المتجمع الصاعد فى نقطه ومنها نرسم موازيا للمحور الافقى يلاقى المحور الرأسى فى نقطة تكون هى عدد العمال المطلوب ايجادة .

ومن الرسم نجد ان عدد العمال المطلوب هو 25 عامل .

جـ ـ للحصول على الانفاق اليومى الذى ينفق اقل منه 80 عامل .

ايضا بعد الرسم كما فى الشكل ومن المحور الرأسى والذى يمثل التكرارات المتجمعه الصاعده والتى هى عدد العمال وعند موقع الرقم 80 نرسم من هذه النقطة موازى للمحور الافقى يقطع المنحنى المتجمع الصاعد فى نقطة ثم نسقط من هذه النقطة عمود مواز للمحور الرأسى يلاقى المحور الافقى عند نقطة تكون هى قيمة الانفاق المطلوب .

ومن الرسم نجد ان الانفاق اليومى الذى ينفق أقل منه 80 عامل هو 104 جنيها تقريبا .

لاحظ انه يمكن تطبيق الشرح السابق على الجداول التكرارية الغير منتظمة اى غير متساوية الفئـــات أيضـا كما فى الجداول التكـرارية المنتظمـه •

٥ ـ المنحـنى التكـرارى المتجمع الهابـط :ـ

يتم تمثيل بيانات الجدول المتجمع الهابط بيانيا بأستخدام المنحنى المتجمع الهابط وذلـــك كمـا يلـى :ـ

كما سبق نرسم محورين متعامدين ، المحور الافقى يمثل الحدود الدنيا للفئه والمحور الرأســى يمثل التكرارات المتجمعه الهابطه ويتم توقيع نقاط الجدول التكرارى المتجمع الهابط حيث كل نقطـــه عباره عن الحد الادنى لكل فئه والتكرار المتجمع الهابط المناظر لهذه الفئه ٠٠٠٠ وهكذا حتى نهايه الجدول ثم التوصيل بين كل نقطتين متتاليتين بالتمهيـد باليد والشكل الناتج يسمى المنحنى المتجمع الهابـط •

مثـال (١٦) :ـ

بأستخدام بيانات جدول التوزيع التكرارى للانفاق اليومى لعدد 100 عامل والموضحة فـــى جـدول (6) المطلوب :ـ

أ ـ رسـم المنحنى المتجمع الهابط •

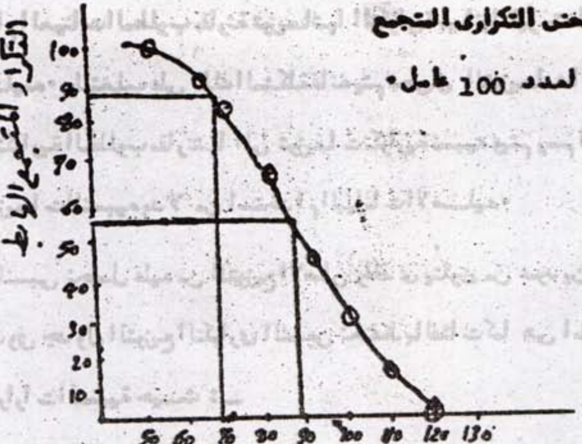
ب ـ عدد العمال الذين ينفقون 65 جنيها يوميا أو أكثـر •

جـ ـ المبلغ الذى ينفقـه (أو أكبر منه) 55 عامـل •

الحـــل :ـ

أ ـ رسم المنحنى المتجمع الهابط من جدول (12) الموضح به التوزيع التكرارى المتجمع الهابـــط فى العمودين الثالث والرابع. وبأستخدام بيانات هذين العمودين نرسم المنحنى المتجمـــع الهابط ، حيث يمثل المحور الافقى الحدود الدنيا للفئات والمحور الرأسى يمثل التكـــرارات المتجمعـه الهابطـه كما فى شكل (12) •

شكل (12) المنحنى التكراري المتجمع الهابط للانفاق اليومي لعدد 100 عامل.



الحدود الدنيا للفئات

ب- عدد العمال الذين ينفقون 65 جنيها يوميا أو اكثر.

بعد رسم المنحنى المتجمع الهابط كما في شكل (12) فانه عند 65 على المحور الافقي نرسم موازيا للمحور الرأسي يلاقي المنحنى المتجمع الهابط في نقطه ومنها نرسم موازيا للمحور الافقي يلاقي المحور الرأسي في نقطة تكون هي عدد العمال المطلوب ايجاده ونجده هو 88 عامل أي ان عدد العمال الذين ينفقون 65 جنيها يوميا أو اكثر هو 88 عامل .

جـ- المبلغ الذي ينفقة (أو اكبر منه) 55 عامل .

ايضا بعد الرسم المنحنى المتجمع الهابط كما في شكل (12) فمن المحور الرأسي وعند الرقم 55 نرسم من هذه النقطة مواز للمحور الافقي يقطع المنحنى المتجمع الهابط في نقطه ثم نسقط من هذه النقطة عمود مواز للمحور الرأسي يلاقي المحور الافقي عند نقطه هي قيمة الانفاق المطلوب ايجاده ونجده 86 جنيها تقريبا من الشكل .

الجدول والمنحنى التكراري النسبي

من أهم مميزات أستخدام المنحنيات التكرارية عند عرض التوزيعات التكرارية بيانيا هو أنه من الممكن رسم منحنيين تكراريين أو اكثر في نفس الشكل كل منهم يخص توزيع تكراري معين وذلك حتى

تسهل عملية مقارنة التوزيعات التكرارية ببعضها البعض بيانيا ولكن غالبا ما تكون مجموع التكرارات التى هى احجام العينات المطلوب مقارنة توزيعاتها التكرارية بيانيا غير متساوية وبالتالى تكون عمليه المقارنة غير سليمه. وللتغلب على تلك المشكلة فانه يتم تحويل التوزيعات التكرارية المطلوب مقارنتها الى توزيعات تكرارية المطلوب مقارنتها الى توزيعات تكرارية نسبيه ويتم رسم المنحنيات التكرارية بأستخدام التوزيعات النسبيه بدلا من استخدام البيانات الاصليه.

التوزيع النسبى نحصل عليه من التوزيع الاصلى والذى يتكون من عمودين هما عمود الفئات ثم عمود التكرارات وفى جدول التوزيع التكرارى النسبى نحتفظ بالفئات كما هى اما التكرارات فتستبدل بما يسمى بالتكرارات النسبية حيث :ـ

$$\text{التكرار النسبى لاى فئه} = \frac{\text{تكرار الفئه}}{\text{مجموع التكرارات}} \times 100$$

$$= \frac{f}{n} \times 100$$

مثال (17) : الجدول التالى يوضح انتاجيه مجموعتين من العمال على نفس النوع من الالات المستخدمه فى انتاج نوع معين من الاقمشة بالامتار كما فى جدول (17).

جدول (17)

المجموعه الثانيه

عدد العمال	انتاجيه العامل
27	10 - 20
48	20 - 30
75	30 - 40
72	40 - 50
42	50 - 60
24	60 - 70
12	70 - 80

المجموعة الاولى

عدد العمال	انتاجيه العامل
10	10 - 20
15	20 - 30
20	30 - 40
25	40 - 50
12	50 - 60
11	60 - 70
7	70 - 80

والمطلوب مقارنة هذين التوزيعين باستخدام المنحنيات التكرارية النسبيه .

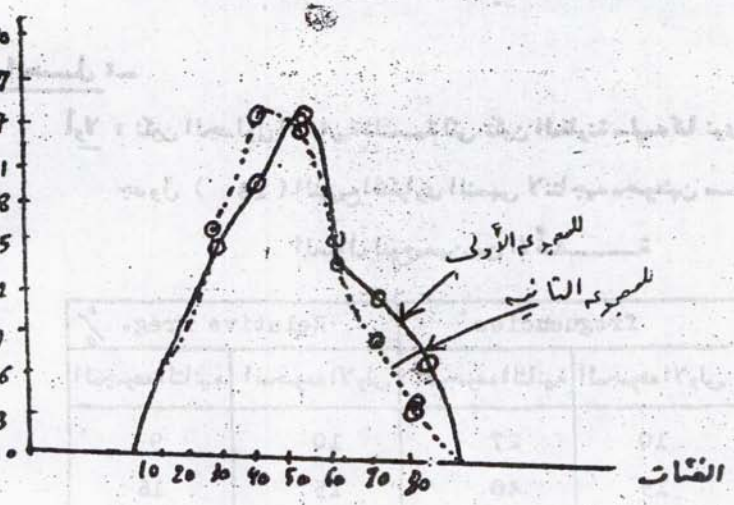
الحـــل :ـ

أولا : نكون الجداول التكرارية النسبية لكى تكون المقارنة سليمه كما نرى فى الجدول (18) .

جدول (18) التوزيع التكرارى النسبى لاتجاه مجموعتين من العمال لنوع معين من الانشطة

اتجاه العامل Intervals	frequencies		Relative freq. %	
	المجموعه الاولى	المجموعه الثانية	المجموعه الاولى	المجموعه الثانيه
10 - 20	10	27	10	9
20 - 30	15	48	15	16
30 - 40	20	75	20	25
40 - 50	25	72	25	24
50 - 60	12	42	12	14
60 - 70	11	24	11	8
70 - 80	7	12	7	4
	100	300	100 %	100 %

وباستخدام بيانات جدول (18) فأنه يمكن رسم المنحنى التكرارى النسبى للمجموعه الاولى والمنحنى التكرارى النسبى للمجموعه الثانيه .



شكل (13) المنحنيات التكرارية النسبية لانتاجيه مجموعتين من العمال لنوعين من الاقمشـــه .

من الشكل السابق نلاحظ ان انتاجيه العامل فى المجموعتين متساوية حتى الفئه 40 = 30تق
ثم بعد الفئه 50 – 40 نجد ان انتاجيه للعامل فى المجموعه الأولى اكبر من انتاجيه العامل
المجموعة الثانيـــة .

أنواع المنحنيات التكراريه :ـ

كما ذكرنا فيما سبق ان المنحنيات التكرارية تختلف باختلاف التوزيعات التكراريه . وكذ ل
المنحنى التكرارى يتغير بالنسبة للظاهره الواحد ة طبقا لاختلاف العينات المأخوذ ه لدراسة ال
ومن المنحنيات التكراريه ما هو ذو قمه واحد ة أو ذو قمتين أو متعدد القمم أو بدون قـ
اطلاقا . كذلك ما هو متماثل او غير متماثل .

١ ـ المنحنيات ذو القمه الواحد ه :ـ

من أشهر وأهم المنحنيات التكراريه ذو قمة واحد ة ومتماثلة هو المنحنى الطبيعى أو الـ
المعتد ل . ويأخذ هذا المنحنى شكل الجرس أو الناقوس اى انه له نهايه عظمى عند منتصفه
يأخذ هذا المنحنى فى الاقتراب من المحور الافقى تدريجيا بطريقه متماثله على جانبيه دون ا

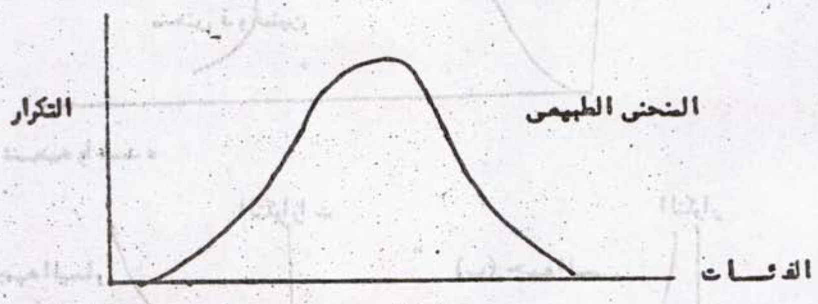
المحور الافقى . والمنحنى الطبيعى مهم جدا فى مجال التطبيقات الاحصائيه ويستخدم كأساس لمقارنة كثير من أنواع المنحنيات الاخرى .

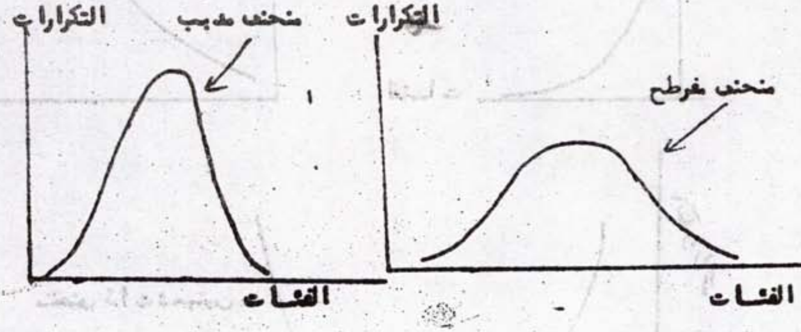
ومن المنحنيات المتماثله نجد منحنيات لها قمه ضيق واكثر أرتفاع من قمه المنحنى الطبيعى ويسمى هذا النوع بالمنحنى المدبب .

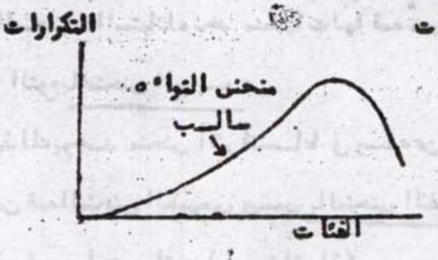
وعكس ذلك يوجد منحنى أكثر أتساعا فى وسطه عن المنحنى الطبيعى وله قمه أوسع أو اقل انخفاضا من قمه المنحنى الطبيعى ويسمى بالمنحنى المفرطح .

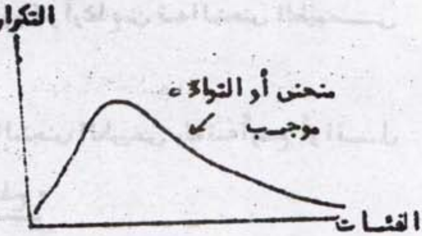
منحنيات ذو قمه واحده ملتويه (غير متماثلة)

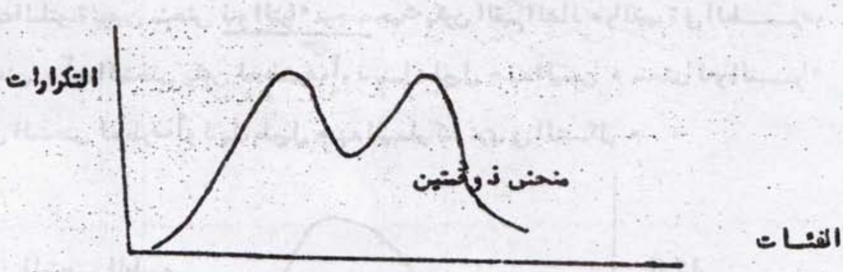
المنحنيات الملتوية نوعين منحنى ذو التواء موجب حيث يكون القيم الشاذه والكبيرة فى الطرف الايمن من المنحنى ، أى المنحنى يكون له طرف أو ذيل طويل جهه اليمين ، منحنى ذو التواء سالب حيث يكون المنحنى له طرف أو ذيل طويل جهه اليسار كما نرى فى الشكل .



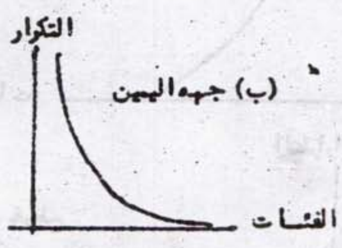


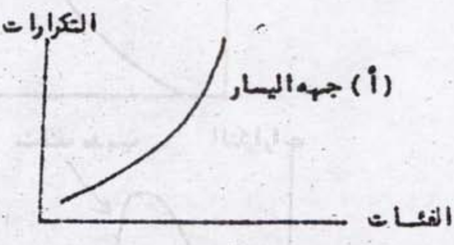


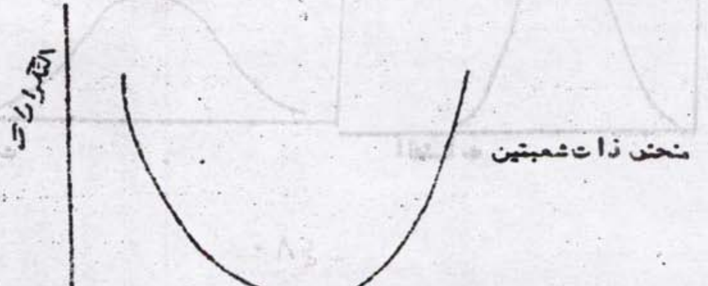




منحنى ذات شعبه واحده







منحنى ذات شعبتين

تمارين (١)

أ ـ والمطلوب عرض هذه البيانات باستخدام الشكل المناسب .
وهى تمثل عدد أفراد الأسرة لعدد ١٥٠ أسرة فى احدى المدن

عدد افراد الاسره	1	2	3	4	5	6	Σ
عدد الاسر	6	14	20	34	16	10	100

ج ـ البيانات التاليه توضح الدخل والانفاق اليومي لعدد 25 شخص بالجنيه :ـ

الدخل	الانفاق	الدخل	الانفاق	الدخل	الانفاق
26	22	33	22	50	20
35	29	28	21	20	14
44	13	17	16	48	17
34	11	14	12	40	25
29	21	32	18	30	24
16	15	59	22	12	10
52	35	29	35	25	22
44	32	21	118	15	12
10	8				

والمطلوب :ـ

أ ـ عرض هذه البيانات فى توزيع تكرارى مزدوج .

ب ـ أوجد التوزيع الهامشى لكل من الدخل والانفاق كل على حده .

جـ ـ قارن بين التوزيع التكرارى للدخل والتوزيع التكرارى للانفاق مستخدما المضلع التكرارى .

د ـ بأستخدام التوزيع المتجمع المناسب . أوجد مبلغ الانفاق الذى يقل عنه % 40 من اجمالى عدد الاشخاص .

هـ ـ عدد الاشخاص الذين يتقاضون 35 جنيه يوميا على الاقل مستخدما التوزيع المتجمع المناسب .

الباب الثاني

أولا : مقاييس الوضع أو النزعة المركزيه

ذكرنا فيما سبق بعض الطرق لعرض البيانات الاحصائيه في صورة سهله وبسيطه عن طريق أستخدام الرسوم البيانية والجداول والتي بواسطتها يمكننا الحصول على فكرة عابرة عن البيانات الرقمية الخاصه باحدى الظواهر . ولا شك ان هذه الطرق ما هي الا بعض الوسائل التي يمكننا بواسطتها وصف البيانات .

وحيث ان الهدف الاساسي من استخدام الاسلوب الاحصائي من تحليل البيانات الخاصه بظاهره ما هو الوصول الى خصائص المجتمعات التي جمعت عنها هذه البيانات ولذا فان استخدام بعض المقاييس الاحصائيه والتي فيها ما يطلق عليه مقاييس الوضع أو مقاييس المتوسط أو مقاييس النزعه المركزيه والتي بواسطتها يمكن وصف مجموعه من البيانات عن طريق تقدير قيمه من قيم المتغيرات ذات وضع معين .

ويقصد بالنزعه المركزيه ميل الغالبية من البيانات الرقمية نحو التمركز حول قيمة معينه ويطلق على هذه القيمة كلمة المتوسط ويعرف بالمتوسط بانه تلك القيمة التي يتجمع عندها او حولها الغالبية العظمى من البيانات .

ومن المقاييس الاحصائيه للوضع او للنزعه المركزيه التي سوف ندرسها ما يلي :-

١ - الوسط الحسابي ٢ - الوسيط

٣ - المنوال ٤ - الوسط الهندسي

٥ - الوسط التوافقي .

ومن اهم صفات القياس الملائم للوضع أو المتوسط كالتالي :-

١ - لابد ان تكون قيمه المتوسط محددة وليست مقدره .

٢ - يجب استخدام جميع المفردات للحصول على قيمه المتوسط .

٣ - ينبغي ان يكون للمتوسط الجيد خصائصه التي تميزه عن بقيه المقاييس .

٤ - ان يكون القياس سهل الحصول عليه بعمليات حسابية غير معقدة .

١ ـ الوسط الحسابي

تعريف : الوسط الحسابي لمجموعه من المفردات عبارة عن مجموع هذه المفردات مقسوما على عددها . للبيانات الغير مبوبة اذا كان لدينا n من المشاهدات لمتغير ما x حيث x تأخذ القيم $x_1 , x_2 , \ldots\ldots , x_n$ فان الوسط الحسابي يأخذ الشكل .

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1)$$

وبالتالي يمكن القول :

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

مثال (١) : اذا كان لدينا مجموعه من الدخول اليومية بالجنيه لمجموعه الاشخاص ولتكن :ـ

$$5 , 19 , 13 , 8 , 15 , 12$$

فان متوسط الحسابي للدخل اليومي هو.

$$\bar{x} = \frac{5 + 19 + 13 + 8 + 15 + 12}{6} = \frac{72}{6} = 12$$

خصائص الوسط الحسابي :ـ

١ ـ مجموع انحرافات قيم المشاهدات عن وسطها الحسابي يساوي صفر اي اذا كانت قيم المتغير x هي $x_1 , x_2 , \ldots\ldots , x_n$ ووسطها الحسابي $\bar{x}$ فان الانحرافات :

$$(x_1 - \bar{x}) , (x_2 - \bar{x}) , \ldots\ldots (x_n - \bar{x})$$

فان مجموع هذه الانحرافات = صفر لان :ـ

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

(ملحوظة: $\sum_{i=1}^{n} C = nC$ حيث C مقدار ثابت)

للسهوله سوف نكتب $\sum$ بدلا من $\sum_{i=1}^{n}$) .

٢ ـ اذا طرح (جمع) مقدار ثابت من كل قيمة من قيم x فان الوسط الحسابي $\bar{x}$ يكون مساويا

للوسط الحسابى للقيم الجديدة مضافا اليه (مطروحا منه) هذا المقدار الثابت.

نفرض ان A مقدار ثابت يسمى بالوسط الفرضى .

اذا كانت قيم x : $x_1, x_2, \ldots\ldots, x_n$.

انحرافات قيم x عن A هى : $(x_1 - A), (x_2 - A), \ldots, (x_n - A)$

اذا كانت $x_i' = x_i - A$ فان

$$\bar{x}' = \frac{\sum x_i'}{n} = \frac{\sum (x_i - A)}{n} = \bar{x} - A$$

$$\therefore \bar{x} = \bar{x}' + A$$

أى الوسط الحسابى للقيم الاصلية = الوسط الحسابى لانحرافات القيم عن A + المقدار الثابت) .

٣ـ اذا قسمنا (ضربنا) قيم المشاهدات على (فى) مقدار ثابت فان الوسط الحسابى $\bar{x}$ يكون مساويا للوسط الحسابى للقيم الجديد مضروبا فى (أو مقسوما على) هذا المقدار الثابت.

$$x : \quad x_1, x_2, \ldots\ldots, x_n$$

$$x' : \quad \frac{x_1}{c}, \frac{x_2}{c}, \ldots\ldots, \frac{x_n}{c}$$

$$\therefore \bar{x}' = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{c} = \frac{1}{c} \left(\frac{1}{n} \sum x_i \right) = \frac{\bar{x}}{c}$$

$$\therefore \bar{x} = \bar{x}' c$$

أى ان الوسط الحسابى للمجموعه من القيم = الوسط الحسابى للقيم الجديدة بعد القسمة على مقدار ثابت C مضروبا فى هذا المقدار الثابت C .

وباستخدام الخاصتين (٢)، (٣) يمكن حساب الوسط الحسابى كما يلى :ـ

$$\bar{x} = \frac{\sum_{i=1}^{n} \left(\frac{x_i - A}{C} \right)}{n} C + A \qquad (3)$$

مثال (٢) : احسب الوسط الحسابى لاعمار 10 أشخاص كما يلى :ـ

35 , 20 , 25 , 40 , 15 , 30 , 45 , 60 , 55 , 65

الحـل :ـ

أ ـ الطريقه المباشرة

$$\bar{X} = \frac{1}{n}\sum x = \frac{390}{10} = 39 \text{ years} .$$

ب ـ طريقه بالمعادله (3) :

$$\bar{X} = \frac{\sum \frac{x_i - 40}{5}}{10} .5 + 40$$

$$= \frac{-2}{10} .5 + 40 = 39 \text{ years} .$$

x	$x-40$	$\frac{x-40}{5}$
35	− 5	− 1
20	− 20	− 4
25	− 15	− 3
40	0	0
15	− 25	− 5
30	− 10	− 2
45	5	1
60	20	4
55	15	3
65	25	5
$\sum$ 390		− 2

مثال (٣) : البيانات التالية تمثل المبيعات اليومية بالجنيهات لشركة للادوات الكهربيه خلال خمسة ايام هى :

2000, 2500, 1500, 1750, 1250

والمطلوب ايجاد الوسط الحسابى للمبيعات اليوميه

الحل :ـ

أ ـ بالطريقه المباشرة :ـ

$$\bar{X} = \frac{\sum x_i}{n} = \frac{9000}{5}$$

$$= 1800 \text{ £}$$

ب ـ بالاستخدام المعادله (3) ينتج ان

$$\bar{X} = \frac{\sum \left(\frac{x - 1500}{250} \right)}{5} 250 + 1500$$

x	$x - 1500$	$\frac{x-1500}{250}$
2000	+ 500	2
2500	+ 1000	+ 4
1500	0	0
1750	250	1
1250	− 250	− 1
9000		$6 = \sum \frac{x-A}{C}$

$$\overline{X} = \frac{6}{5}(250) + 1500$$

$$= 1800 \; £ .$$

الوسط الحسابى للبيانات المبوبه (اى المعروضه فى جدول تكرارى)

فى حالة المتغيرات المتقطعه تأخذ قيم المتغير x المختلفه مباشرتا .

اما فى حالة المتغيرات المتصلة تأخذ قيم المتغير x بمراكز الفئات حيث

$$\text{مركز الفئه} = \frac{\text{الحد الادنى للفئه} + \text{الحد الاعلى لها}}{٢}$$

لايجاد الوسط الحسابى للبيانات المبوبه فاننا سوف نتبع نفس الطرق السابقة ونفس الخطوات الحصول عليه من كل منها ولكن مع فارق بسيط هو وجود عدد مرات تكرار كل فئه من الفئات ، وأستبدال قيم x الاصلية بمراكز الفئات فى حالة البيانات المتصله ، يمكن حساب المتوسط الحسابى بأحدى الطريقتين :ـ

أ ـ الطريقه المباشرة :

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{n}$$

حيث x_i علامة الفترة i ، f_i تكرار هذه الفترة ، k عدد الفترات.

ب ـ الطريقه المختصره :ـ

$$\bar{x} = \frac{\sum_{i=1}^{k} \left(\frac{x_i - A}{C}\right) f_i}{n} C + A$$

مثال ٤

: البيانات التالية توضح التوزيع التكرارى لعدد 46 أسره حسب عدد الحجرات داخل شقة كل منها . احسب المتوسط الحسابى ؟

عدد الحجرات x	عدد الاسر f	$x_i f_i$
1	5	5
2	5	10
3	15	45
4	10	40
5	6	30
6	5	30
٧	46	160

الحل :

$$\bar{x} = \frac{160}{46} = 3.47$$

مثال (٥) : احسب الوسط الحسابي لجدول التوزيع التكراري للانفاق اليومي لمائة موظف المبين بالجدول .

الانفاق اليومي	50 -	60 -	70 -	80-	90 -	100 -	110- 120	Σ
التكرارات	6	12	15	24	18	14	11	100

الحل :-

أ ـ الطريقه المباشره

فئـــات	التكرار f	مركز الفئات x	xf
50 - 60	6	55	330
60 - 70	12	65	780
70 - 80	15	75	1125
80 - 90	24	85	2040
90 - 100	18	95	1710
100 - 110	14	105	1470
110 - 120	11	115	1265
Σ	100		8720

$$\therefore \overline{X} = \frac{\sum x_i f_i}{\sum f_i}$$

$$= \frac{8720}{100} = 872$$

ب ـ الطريقه المختصره :

$$\bar{X} = \frac{\sum \left(\frac{x_i - A}{C}\right) f_i}{n} C + A$$

وعادة : نختار قيمة A بقيمة x المقابلة لاكبر تكرار او التى تقع فى الوسط، C بطول الفترة وذلك فى حالة الفئات المنتظمة فقط .

الوسط الحسابى للانفاق اليومى لعدد 100 موظف (طريقة المختصرة)

الفئــــــات	التكرار f	مركز الفئات x	$x - A$ ($A = 85$)	$(x - A)/C$ ($C = 10$)	$\frac{x - A}{C} f$
50 - 60	6	55	- 30	- 3	- 18
60 - 70	12	65	- 20	- 2	- 24
70 - 80	15	75	- 10	- 1	- 15
80 - 90	24	85	0	0	0
90 - 100	18	95	10	1	18
100 - 110	14	105	20	2	28
110 - 120	11	115	30	3	33
$\sum$	100				22

من الجدول نجد ان :ـ

$$A = 85 , \quad C = 10 \quad , n = 100 , \quad \sum\left(\frac{x - A}{C}\right)f = 22$$

$$\therefore \bar{X} = \frac{22}{100} \cdot 10 + 85 = 87.2 \ \mathcal{L}$$

الوسط الحسابى لمجموعه كبيرة مكونه من عده مجموعات صغيرة يساوى الوسط الحسابى المرجح لمتوسطات تلك المجموعات الصغيرة .

فاذا كان لدينا r من المجموعات وكانت اوساطها الحسابية لتلك المجموعات هـــــــــــ $\bar{x}_1 , \bar{x}_2 , \ldots , \bar{x}_r$ وان احجام هذه المجموعات هى $n_1 , n_2 , \ldots\ldots.., n_r$

فان الوسط الحسابى للمجموعه الكبيرة التى أخذت منها هذه المجموعات الصغيرة يكون كما يلى :ـ

$$\bar{X} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \ldots\ldots.. + n_r \bar{x}_r}{n_r + n_r + \ldots. + n_r} = \frac{\sum_{i=1}^{r} n_i \bar{x}_i}{\sum_{i=1} n_i}$$

كما انه لا يمكن حساب الوسط الحسابى من الجداول المفتوحه وذلك لعدم معرفه مركز الفئه المفتوحــة ولتفادى هذا العيب كما يلــى :ـ

١ ـ أهمال الفئات المفتوحة وخاصة اذا كانت تكرارها صغير نسبيا .
٢ ـ اننا نأخذ طول الفئه الاولى المفتوحه مساويا لطول الفئه الثانية . اما اذا كان الجدول مفتوح من اسفل فاننا نأخذ طول الفئه الاخيرة المفتوحه مساويا لطول الفئه قبل الاخيره.
٣ ـ امكانية ايجاد الوسط الحسابي باستخدام المقاييس الاخرى ومنها الوسيط والمنوال وذلك كما سنرى فيما بعد .

كما ان الوسط الحسابي يتأثر بالقيم الشاذة او المتطرفة . ويمكن تفادى ذلك بأهمال تلك القيم الشاذه او المتطرفه.

٢ ـ الوسيط Median

تعريف : من الممكن تعريف الوسيط لمجموعه من القيم بانه القيمة التى تقع فى منتصف هذه المجموعة من القيم وذلك بعد ترتيبها تصاعديا او تنازليا .

أ ـ الوسيط للبيانات غير المبوبه :ـ

للحصول على قيمة الوسيط من بيانات غير مبوبه ، فانه لا بد من ترتيب الوسيط فاذا كان عدد المفردات المراد ايجاد الوسيط لها هو n فان موقع أو ترتيب الوسيط نحصل عليه من العلاقة :ـ

$$\text{ترتيب الوسيط} = \frac{n+1}{2}$$

واذا كانت عدد المفردات n زوجيه فان قيمة الوسيط تساوى الوسط الحسابي للمفردتين المركزتيين أى الوسط الحسابي لقيمة المفردة $\frac{n}{2}$ وقيمة المفرده $\frac{n}{2}+1$.

مثال (٦) : أوجد الوسيط لاعمار تسعه أشخاص بالاعوام كالاتى :ـ

20, 29, 28, 25 , 24 ; 27, 22, 30, 26 ;

الحل :ـ لايجاد الوسيط للبيانات المعطاه فأننا نرتب هذه المفردات تصاعديا كما يلى :ـ

20 , 22, 24 , 25 , 26, 27, 28, 29 , 30

وحيث ان عدد المفردات $n = 9$ أى فردى فان ترتيب الوسيط هو :ـ

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

∴ الوسيط هو الفرد ءرقم خمسه بعد الترتيب :

$Me = 26$

مثال (٢) : البيانات التالية توضح اوزان ثمانى أشخاص بالكيلو جرام :-

54, 62, 70 , 83 , 58 , 76 , 87 , 68

والمطلوب : ايجاد قيمه الوسيط .

الحل : ترتيب المفردات المعطاه تصاعديا كما يلى :-

54, 58, 62, <u>68</u>, <u>70</u>, 76, 83 , 87

وعدد المفردات $n = 8$ أى زوجى .

الوسيط متوسط القيمتين اللتان رتبتهما : $\frac{n}{2}$, $\frac{n}{2}+1$ أى 4 , 5

$$Me = \tfrac{1}{2}(68 + 70) = 69 \text{ kg.}$$

ملاحظة هامة : اذا كان عدد المفردات المراد ايجاد الوسيط لها كبير كبرا كافيا فانه فى هذه الحاله يكون ترتيب الوسيط $\frac{n}{2}$ أى الوسيط هو الفرد . $X_{n/2}$ سواء كان عدد للمفردات فرديا أو زوجيا .

ب ـ الوسيط للبيانات المبوبه :-

عند ايجاد الوسيط من بيانات مبوبه فى جداول تكرارية فاننا نستخدم طريقتين هى الطريقة الرياضية ، الطريقه البيانيه .

أولا : ايجاد قيمه الوسيط للبيانات المبوبه رياضيا :-

اذا كانت البيانات معروضة فى جدول توزيع تكرارى فان الوسيط هو القيمة التى تتوسط قيم المفردات بعد ترتيب القيم تصاعديا أو تنازليا .

وعملية ترتيب المفردات سواء تصاعديا أو تنازليا فى حاله الجداول التكرارية ما هو الا تكوين الجداول التكرارية المتجمعه الصاعدة أو الهابطه .

الوسيط تعطى بالعلاقه : $Me = b + c \dfrac{d}{f_m}$

حيث m الوسيط ، I للفترة التى تحتوى على الوسيط ، l الحد الادنى للفترة I ،

c طول الفترة I ، (التكرار التجمع الصاعد قبل الفترة I) – $\frac{n}{2}$ = d ،

f_m تكرار الفترة I ·

مثال (٨) : بأستخدام بيانات الجدول الخاص بالتوزيع التكرارى للانفاق اليومى بالجنيه لعدد 100 عامل · المطلوب ايجاد قيمة الوسيط للانفاق اليومى ·

الحل :ـ

الفئـــات	تكرارات	الحدود العليـا للفئـــه	التكرار المتجمع الصاعد	الحدود الدنيـا للفئـــه	التكرار المتجمع الهابط
50 -60	6	less than 60	6	50 or more	100
60 - 70	12	,, ,, 70	18	60 ,, ,,	94
70 - 80	15	,, ,, 80	33 ← $\frac{n}{2}$	70 ,, ,,	82
80 - 90	24	,, ,, 90	57	80 ,, ,,	67
90- 100	18	,, ,, 100	75	90 ,, ,,	43
100-110	14	,, ,,, 110	89	100,, ,,	25
110-120	11	,, ,, 120	100	110,, ,,	11
$\sum$	100				

للبحث عن الفترة I ، نحسب $\frac{N}{2} = 50$ وهى تمثل نصف عدد الطلاب ويقع هذا الرقم 50 بين التكرارين 33 ، 57 ومن هذا يجب ان تكون الفترة I (الفترة التى تحتوى على الوسيط) وهى الفترة 80 -90 حدها الادنى b = 80 وطولها هو C=10 وتكرارها $f_m = 24$

وقيمة d هى

$$d = 50 - 33 = 17$$

واذن فان الوسيط هو :

$$Me = 80 + 10 \; \frac{17}{24} = 87.083 \; .$$

وتفسير هذه النتيجه يكون كالتالى من بين ال 100 عامل نجد ان 50 عامل من هولاء العمال ينفقون اقل من 87.083 جنيه شهريا ، 50 عامل ينفقون اكثر من أو تساوى نفس الرقم .

وبالمثل يمكن حساب الوسيط رياضيا من جدول التجمع النازل .

ثانيا : ايجاد قيمة الوسيط لبيانات مبوبه بيانيا :-

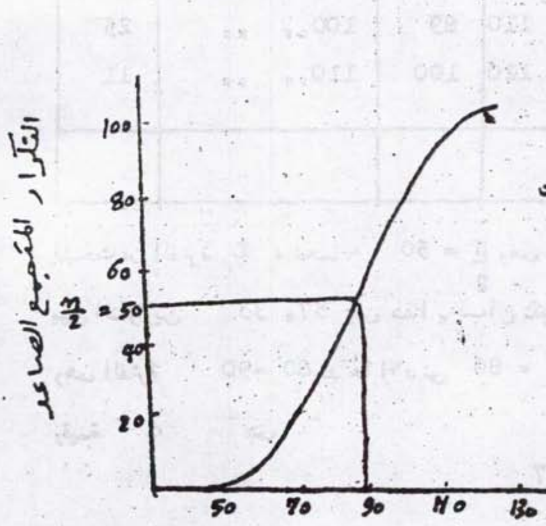
كما تم الحصول على قيمة الوسيط باستخدام التوزيع التكرارى التجمع الصاعد والتوزيع التكرارى التجمع الهابط ، فاننا سوف نجد قيمة الوسيط باستخدام المنحنى التكرار التجمع الصاعد والمنحنى التكرارى التجمع الهابط أو بأستخدام التكراريين الصاعد والهابط معا وذلك كما يلى بالمثال الاتى :-

مثال (٩) : من بيانات الجدول الخاص بالانفاق اليومى لعدد 100 عامل . أوجد قيمة الوسيط باستخدام المنحنى التكرارى التجمع .

الحــل :-

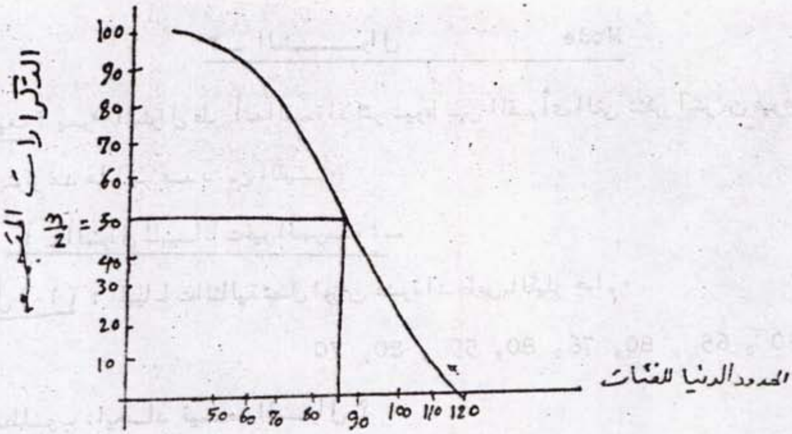
لرسم المنحنى التجمع الصاعد يلزم اولا تكوين الجدول التكرارى التجمع الصاعد وهذا ما نجده فى الجدول السابق (مثال ٨) من بيانات هذا الجدول نرسم المنحنى التكرارى التجمعى الصاعد .

من الشكل نجد ان قيمة الوسيط تساوى 87.083 جنيه تقريبا .



وبنفس الطريقه يمكن ايجاد الوسيط من المنحنى المتجمع والهابط .

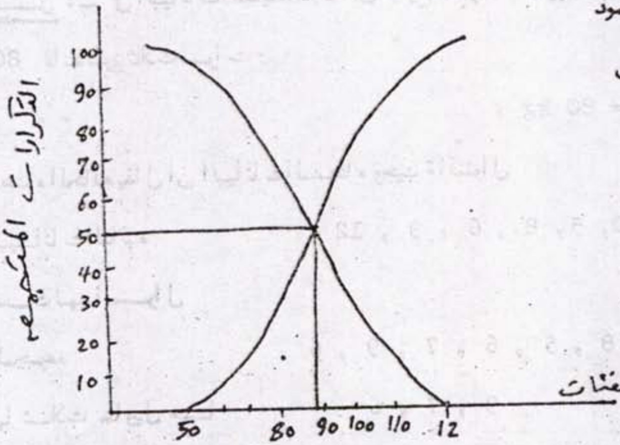
وكذلك يمكن ايجاد قيمة الوسيط باستخدام المنحنين المتجمعين الصاعد والهابط معا .



من نقطه التقاطع هذه نسقط عمود

على المحور الافقى فى نقطه تكون

قيمه الوسيط .



ويعتبر الوسيط من أهم المقاييس الاحصائيه للنزعــه المركزيه بعد الوسط الحسابى ومن مزايا الوسيط :ـ

١ ـ طريقه حساب قيمة الوسيط لا تتأثر سواء كان التوزيع التكرارى منتظم (فئاته متساويه ام غير متساوية) .

٢ ـ قيمة الوسيط لا تتأثر بالقيم الشاذة او المتطرفة بعكس حاله الوسط الحسابى .

٣ ـ يمكن ايجاد قيمة الوسيط من الجداول التكرارية المفتوحه سواء من أحد الطرفين او من كلا الطرفين .

٤ ـ يتميز الوسيط بسهوله حسابة سواء من بيانات مبوبه أو بيانات غير مبوبه وكذلك امكانية أيجاد قيمته رياضـــيا وبيـانيا .

٣ ـ المنـــــوال Mode

تعريف : يعرف المنوال على أنه القيمة الاكثر شيوعا بين القيم أى التى تتكرر أكثر من غيرها من القيـم اى يتركز عند ها اكبر عـدد من القيـم .

أ ـ المنوال للبيـانات غير المبوبـه :ـ

مثال (١٠) : البيانات التالية تمثل اوزان عشرة اشخاص بالكيلو جرام .

55, 48, 90, 65, 80, 76, 80, 59, 80, 70

والمطلـوب :ايجـاد قيمـه المنـوال ؟

الحـــل :ـ فى البيانات السابقة نجد ان جميع المفردات تكررت مرة واحدة فقط ما عدا عند الوزن 80 فانه تكرر ثلاث مرات .

$\therefore \quad Mo. = 80$ kg .

فى هذه الحاله يقال ان البيانات المعطاه وحيدة المنوال

وللبيـانات التاليه 7, 10, 5, 8, 6, 9, 12

ليسـت لهـا منـوال

والمجموعه 10, 12, 9, 7, 8, 5, 6, 7, 9, 6

لها ثـلاث منـاويل هما 6, 7, 9

ومن الامثلة السابقة نجد انه لايه مجموعه من البيانات قد يوجد منوال واحد او منوالين أو اكثـر وأيضا قد لا يوجد لها منوال على الاطلاق وذلك عكس المقاييس الاخرى للمتوسطات والتى سبق دراستها اذ وجدنا انه لابد من وجود مقياس وحيد للمتوسط (وسط حسابى واحد أو وسيط واحد) .

ب ـ المنـوال للبيـانات المبـوبـه :ـ

لايجاد قيمه المنوال لبيانات مبوبه فى جداول تكرارية فاننا سوف نستخدم طريقتين همـا

الطريقة الرياضية البيانية.

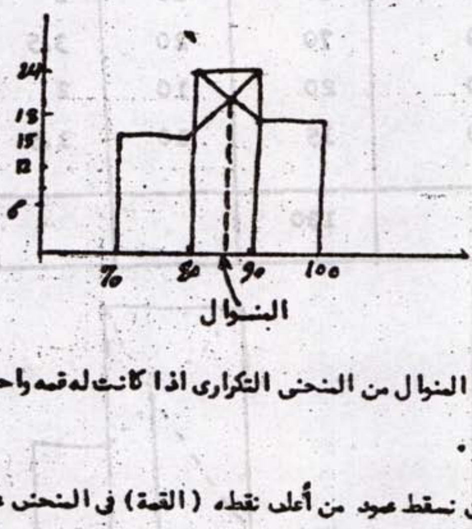
١ ـ ايجاد قيمة المنوال لبيانات مبوبه في جدول تكراري منتظم وهى علامه الفترة المقابلة لاكبر تكرار وهى قيمة تقريبيه.

فان المنوال لبيانات جدول التوزيع للانفاق اليومى لعدد 100 عامل من الجدول نلاحظ ان الفترة المحتويه على المنوال هى 90 ـ 80 لانها تقابل اكبر تكرار فان المنوال :

$$Mo. = \frac{80 + 90}{2} = 85$$

وكذلك يمكن ايجاد المنوال بيانيا باستخدام المدرج التكراري.

وذلك لنفس المثال للانفاق اليومى لعدد 100 عامل.



. ويمكن ايضا ايجاد المنوال من المنحنى التكراري اذا كانت له قمه واحدة فان هذه القمة تكون امام أكبر تكرار.

لذا فمن الممكن ان نسقط عمود من أعلى نقطه (القمة) في المنحنى على المحور الافقي الممثل للفئات فيلاقيه في نقطة تكون هى قيمة المنوال.

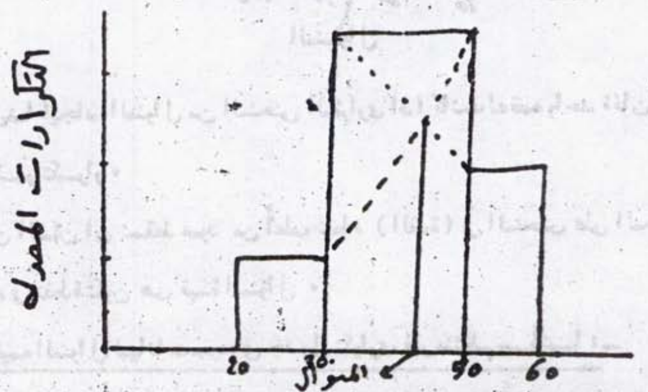
٢ ـ ايجاد قيمه المنوال لبيانات مبوبه في جدول تكراري غير منتظم بيانيا :ـ

لايجاد قيمة المنوال بيانيا سوف نستخدم طريقه المدرج التكراري فقط. وكما ذكرنا في حالة الجداول التكرارية المنتظمة سوف نرسم ثلاثة أعمدة فقط تمثل ثلاث فئات وهى الفئه السابقة للفئه المنواليه والفئه اللاحقه للفئه المنواليه مع أستخدام التكرارات المعدله وليس التكرارات الاصلية لاى الفئات

غير متساويه .

مثال : ايجاد المنوال لاعمار 120 شخص :-

الفئـــــات	التكرارات	طول الفئه	التكرار المعدل
15 - 20	5	5	-1
20 - 30	10	10	1
30 - 50	70	20	3.5
50 - 60	20	10	2
60 - 70	15	10	1.5
	120		



ايجاد المنوال لاعمار 120 شخص .

خصائص المنوال :-

١ ـ لا يتأثر بالقيم المتطرفه .

٢ ـ البساطه فى الفكرة التى يستند اليها حيث انه يعتبر كمثل لجميع المفردات .

٣ ـ امكانية الحصول على المنوال فى الجداول المفتوحه ..

٤ ـ امكانيه تقدير قيمه المنوال رياضيا وبيانيا .

العلاقة بين الوسط الحسابي والوسيط والمنوال :ــ

١ ــ اذا كان التوزيع التكراري الممثل للبيانات التي ندرسها متماثلا أي يأخذ شكل المنحنى الطبيعي أو يكون مدببا أو مفرطحا عن المنحنى الطبيعي فان :ــ

$$\bar{X} = Me = Mo$$

وبالتالي فان هذه القيم الثلاث المتساويه تقسم المنحنى الى جزأين منطبقين تماما .

٢ ــ اذا كان التوزيع التكراري الممثل للبيانات التي ندرسها ذو التواء موجب اي القيم الشاذه أو المتطرفة والكبيرة في الطرف الايمن من المنحنى ونجد ان :ــ

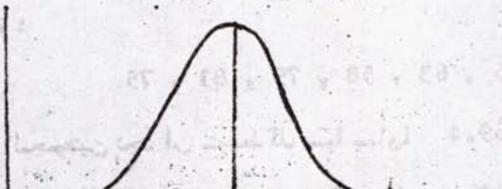
$$\bar{X} > Me > Mo .$$

٣ ــ اذا كان التوزيع التكراري الممثل بالبيانات التي ندرسها ذو التواء سالب اي ان القيم الشاذه أو المتطرفة والصغيرة تكون في الطرف الايسر من المنحنى اي انه في التوزيعات ذات التواء السالب نجد ان :ــ

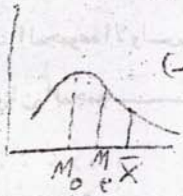
$$\bar{X} < Me < Mo$$

وفي الثلاث حالات الشكل الاتي يبين موضع كل من الثلاث قيم

متماثل



(أ)



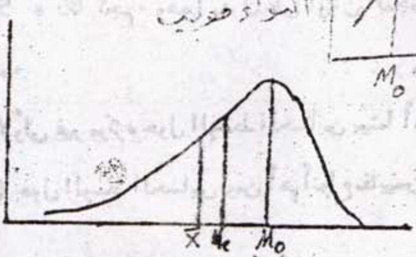
(جـ)

التواء موجب

ملاحظة علاقة اختيارية

$$2\bar{X} = 3M_e - M_o$$

ويطلق عليه ايجاد احراف؟ ... الالتيه الالطرفيه



(ب)

التواء يسير

ثانيا : مقاييس التشتت

ذكرنا سابقا انه عند اجراء بحث أحصائى لابد مشكله وبعد جمع البيانات اللازمه فانه يتم عرض هذه البيانات اما فى جداول تكراريه تحوى كميات هائله من البيانات فى عدد محدود من الفئات ثم عرضها فى شكل رسوم بيانيه، وايضا درسنا مقاييس المتوسطات والتى تمثل مجموعة البيانات وتعبر عنها فى صورة مختصرة جدا وهى قيمه اى من هذه المقاييس .

وهذه المقاييس لا تعطى فكره دقيقة او تصف وصف كامل لطبيعة وخصائص البيانات، لذا الامر يحتاج الى مقاييس اخرى لتوضيح كيفية اختلاف أو تفاوت البيانات عن المتوسط وهو التشتت ويقصد بالتشتت التغير او الانتشار لمجموعه من القيم.

ودراسة مقاييس التشتت يتيح لنا معرفه مدى تجانس أو عدم تجانس (أختلاف أو تباين) مفردات تلك الظاهره.

والمثال التالى يوضح ذلك :

البيانات التاليه توضح اوزان مجموعتين من الطلبه بالكيلو جرام :

أوزان المجموعه الأولى :

35, 54 ,58 , 63 ,67 ,115 , 69 , 73 , 84 , 76

أوزان المجموعة الثانيـة :

61,68, 55, 67, 85 , 63 , 58 , 79 , 83 , 75

بمقارنة اوزان الطلبة فى المجموعتين نجد ان متوسط كل منها مساويا 69.4 كجم. وعلى الرغم من تساوى الوسط الحسابى للمجموعتين الا ان وجه المقارنة غير سليم وغير كاف فنجد المجموعه الاولى تتراوح بين 35، 115 كجم والثانية بين 55 ، 85 كجم. وهذا يدل على اوزان المجموعه الأولى مشتتة فيما بينها اكثر من المجموعه الثانيه.

أو بمعنى آخر اوزان الطلبه فى المجموعه الأولى غير مركزه حول الوسط الحسابى بينما أوزان الطلبه فى المجموعه الثانية نجد انها مركزه (أو منتشره) حول الوسط الحسابى ومن أهم أنواع مقاييس التشتت

وأكثرها استخداما وهى :ـ

١ ـ المدى ٢ ـ نصف المدى الربيعى (الانحراف الربيعى) .

٣ ـ متوسط الانحرافات المطلقة (الانحراف المتوسط) .

٤ ـ الانحراف المعيارى ٥ ـ معامل الاختلاف.

١ ـ المـدى Range

تعريف : يعرف المدى لايه مجموعه من البيانات بأنه الفرق بين اكبر قيمة وأصغر قيمة من قيم هذه المجموعه من البيانات.

اذا كانت القيم غير مبوبه التى تأخذها مفردات اى ظاهره هى :

$x_1, x_2, \ldots\ldots, x_n$

وبالتالى فان المدى = أكبر قيمه ـ اصغر قيمه.

$$\text{range} = x_{(n)} - x_{(1)} \qquad -(1)$$

حيث $x_{(n)}$ اكبر قيمه من البيانات ، $x_{(1)}$ أصغرهم.

مثال (١) : البيانات التاليه تمثل الدخل الاسبوعى بالجنيه لثمانية اشخاص :

43,49, 35,56, 63,28 ,100, 92 والمطلوب ايجاد المدى .

الحل : المدى = اكبر قيمه ـ أصغر قيمه

$$\text{range} = 100 - 28 = 72$$

المدى للبيانات المبوبه فان المدى تأخذ الصوره

المدى = نهايه الفئه الاخيرة ـ بدايه الفئه الاولى .

مثال (٢) : من بيانات الجدول التى توضح الانفاق اليومى بالجنيهات لعدد 100 عامل أوجد المدى .

الحل : حيث ان نهاية الفئه الاخيرة = 120 وبداية الفئه الاولى = 50 فان

$$\text{range} = 120 - 50 = 70 .$$

خصائص المدى :-

١ - يعتبر المدى من أبسط مقاييس التشتت، الا انه أقلها تعبيرا عن تشتت أى مجموعة من البيانات لانه لا يأخذ فى الحسبان الا القيم الاكبر والقيم الاصغر بين مفردات •

٢ - لا يمكن استخدامه فى مقارنة تشتتى مجموعتين من البيانات اذا كانت وحدات قياسها مختلفه•

٣ - يتأثر المدى بالقيم الشاذه والمتطرفه •

٤ - لا يمكن ايجاده فى حالة الجداول المفتوحه ولكن اذا كان من الضرورى الحصول عليه فأننا نغلق الفئات المفتوحه•

٥ - يستخدم المدى بكثرة فى خرائط المراقبة الاحصائيه لجودة الانتاج حيث تؤخذ عينات من نفس الحجم وعلى فترات متقاربه وتحتاج الى حساب سريع لتشتت كل مجموعه•

٢ - نصف المدى الربيعى (الانحراف الربيعى)

تعريف : يعرف الانحراف الربيعى بانه يساوى نصف الفرق بين الربيعين الاعلى والادنى •

واذا رمزنا للربيع الادنى بالرمز Q_1 والربيع الاعلى Q_3 فان نصف المدى الربيعى بالرمز Q

المدى الربيعى = الربيع الاعلى - الربيع الادنى

الانحراف الربيعى (نصف المدى الربيعى) = $\frac{\text{الربيع الاعلى - الربيع الادنى}}{٢}$

$$Q = \frac{Q_3 - Q_1}{2} \qquad (2)$$

والفكرة الاساسية فى هذا المقياس مستمده من تقسيم اى توزيع الى اربع اقسام متساوية وذلك بعد ترتيب المفردات اما تصاعديا أو تنازليا ونجد ان :-

١ - قيمه الربيع الادنى الاول (Q_1) هى قيمة تلك المفردة التى تقع بعد 25% أى بعد ربع المفردات•

٢ - قيمه الربيع الاوسط (الثانى) Q_2 هى قيمة تلك المفرده التى تقع بعد 50% (نصف) من المفردات • وما هو الا الوسيط Q_2 or (Me)

٣ - قيمه الربيع الاعلى الثالث Q_3 هى قيمة تلك المفردة التى تقع بعد 75% (ثلاثه أرباع) من المفردات•

وعند الحصول على قيمه كل منهما تتبع نفس طريقه الحصول على قيمة الوسيط والتى ذكرناها سابقا مع اجراء التعديل اللازم والملائم .

أ ـ نصف المدى الربيعى للبيانات غير المبوبه :ـ

مثال (٣) : البيانات التاليه تمثل أوزان مجموعه من الاشخاص بالكيلو جرام عددها 13 شخصا .

70, 62, 83, 79 ,60 ,78, 93 , 76 ,66, 98, 102, 75, 100

والمطلوب : ايجاد نصف المدى الربيعى .

الحل : بداية لابد من ترتيب البيانات المعطاه تصاعديا نحصل على

60, 62, 66, 70, 75, 76, 78 , 79, 83, 93 , 98 , 100, 102

حيث $n = 13$

ترتيب الربيع الأول $(Q_1) = \frac{n+1}{4} = 3.5$

أى متوسط المفردة (الثالثه والرابعة) أى

$$Q_1 = \frac{66 + 70}{2} = 68 \text{ kg.}$$

ترتيب الربيع الثالث $(Q_3) = \frac{3}{4}(n+1) = 10.5$

فان الربيع الثالث (Q_3) يكون متوسط الحسابى للمفردة العاشرة والحاديه عشر .

$$Q_3 = \frac{93 + 98}{2} = 95.5$$

نصف المدى الربيعى

$$Q = \frac{Q_3 - Q_1}{2} = \frac{95.5 - 68}{2}$$

$$= \frac{27.5}{2} = 13.75 \text{ kg .}$$

ب- نصف المدى الربيعي للبيانات المبوبة :

ويمكن ايجاد الانحراف الربيعي بطريقتين : طريقة بيانية ، طريقة رياضيا .

مثال (٤) : من بيانات الجدول الخاص بالتوزيع التكراري للانفاق اليومي بالجنيه لعدد 100 عامل . المطلوب : ايجاد نصف المدى الربيعي للانفاق اليومي رياضيا .

الحل :

الفئات	f	الحدود العليا للفئة	التكرار المتجمع الصاعد	
50 - 60	6	less than 60	6	
60 - 70	12	,, ,, 70	18	$\rightarrow Q_1$
70 - 80	15	,, ,, 80	33	
80 - 90	24	,, ,, 90	57	
90 - 100	18	,, ,, 100	75	$\rightarrow Q_3$
100-110	14	,, ,, 110	89	
110 -120	11	,, ,, 120	100	
	100			

Q_1 ترتيب = $\frac{100}{4} = 25$

Q_3 ترتيب = $3 \ \frac{100}{4} = 75$

قيمة الربيع تأخذ الصوره :

قيمه الربيع = بدايه فئه الربيع + طول فئه الربيع × (ترتيب الربيع − السابق لترتيب الربيع التكرار المتجمع الصاعد) / تكرار فئه الربيع

فان

$$Q_1 = 70 + 10 \ \frac{25 - 18}{15} = 74.667 \ L$$

أما بالنسبه للربيع الثالث Q_3 نجد ان

$Q_3 = 100$

$$\therefore Q = \frac{Q_3 - Q_1}{2} = \frac{100 - 74.667}{2} = \frac{25.333}{2} = 12.667$$

ملاحظة هامة: اذا لم يكن ترتيب Q_1 أو ترتيب Q_3 معطى صراحه ى عمود التكرارات المتجمعه فاننا نستخدم القانون (3) للحصول على قيمة Q_1 او قيمه Q_3

مثال (٥) : البيانات التاليه توضح التوزيع التكرارى لسرعات عدد 200 سياره حسب فئات السرعة بالكيلو متر / ساعة. والمطلوب ايجاد نصف المدى الربيعى .

الحل :

فئات السرعه	عدد السيارات f	حدود عليا للفئات	تكرار متجمع صاعد
10 - 20	8	less than 20	8
20 - 30	15	,, ,, 30	23
30 - 40	20	,, ,, 40	43
40 - 50	25	,, ,, 50	68 ← Q_1
50 - 60	30	,, ,, 60	98
60 - 70	40	,, ,, 70	138
70 - 80	32	,, ,, 80	170 ← Q_2
80 - 90	20	,, ,, 90	190
90 - 100	10	,, ,, 100	200
	200		

أولا : تكون الجدول المتجمع الصاعد .

من بيانات الجدول السابق نجد ان :

$$\text{ترتيب } Q_1 = \frac{\sum f}{4} = \frac{200}{4} = 50$$

وبالتالى تكون فئـة الربيـع الأول Q_1 هـى $40 - 50$

ترتيب $Q_3 = \frac{3}{4} \sum f = 150$

فئه الربيع الاعلى Q_3 هى $70 - 80$

وبأستخدام القانون (3) نحصـل علـى

$$Q_1 = 40 + 10 \quad \frac{50 - 43}{25} = 42.8 \text{ km/ h}$$

$$Q_3 = 70 + 10 \quad \frac{150 - 138}{32} = 73.75 \text{ km/ h .}$$

$$\therefore Q = \frac{Q_3 - Q_1}{2} = \frac{73.75 - 42.8}{2} = 15.475 \text{ km / h}$$

ثانيـا : ايجـاد نصـف المدى الربيعى للبيـانات بيـانيـا :ـ

يتم الحصول على نصف المدى الربيعى وذلك عن طريق رسم المنحنى المتجمع الصاعد أو المتجمع الهابط ثم نحدد على المحور الرأسى نقطتى ترتيب الربيع الأول Q_1 والربيع الادنى Q_3 . ونسقط من كل من هذه النقط عمود أفقى على المنحنيات المتجمعه ومن نقطة التقاطع بين هذه الأعمـــد والمنحنيات المتجمعه نسقط أعمده رأسيه على المحور الافقى والممثل حدود الفئات كما يلى :ـ

مثال (٦) بأستخدام بيانات الجدول الخاص بالتوزيع التكرارى للانفاق اليومى بالجنيه لعدد 100 عامل . والمطلوب : ايجاد نصف المدي الربيعى للانفاق اليومى بيانيا .

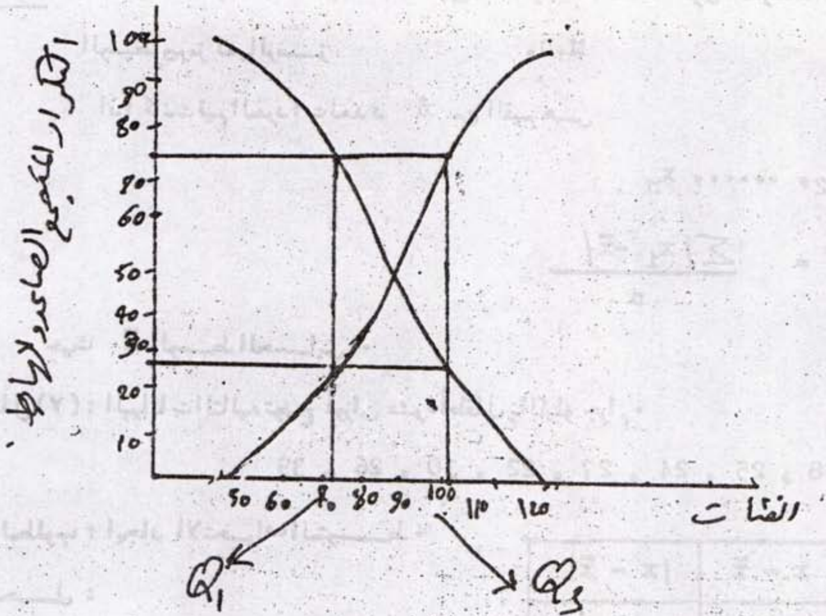
الحـل :ـ

سوف نستخدم المتجمع الصاعد والنازل كما يلـى :ـ

من الرسـم نجد ان :

$$Q_1 = 74.667 \qquad Q_3 = 100 \qquad \therefore Q = \frac{Q_3 - Q_1}{2}$$

$$Q = \frac{100 - 74.667}{2} = 12.667$$



خصائص نصف المدى الربيعى (الانحراف الربيعى)

١ ـ لا تتأثر قيمته بالقيم الشاذه أو المتطرفه .

٢ ـ من الممكن الحصول عليه من الجداول المفتوحه.

٣ ـ يمكن الحصول عليه رياضيا او بيانيا .

٤ ـ يعتبر مقياسا ملائما فى حاله التوزيعات شديده الالتواء.

٥ ـ فى حاله التوزيعات المماثله نجد ان الوسيط يقع فى منتصف المسافه بين Q_1 , Q_2

أى ان

$$Me\ (Q_2) = \frac{Q_1 + Q_2}{2}$$

٣ ـ الانحـراف المتوسـط   Mean Deviation

تعريف : الانحراف المتوسط بانه الوسط الحسابى للانحرافات المطلقة حول المتوسط الحسابى أو الوسيط ويرمز له بالرمـز   M.D.

اذا كانت قيم المفردات لعدد n من القيم هـى

$$x_1, x_2, \ldots\ldots, x_n$$

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

حيث $\bar{x}$ الوسـط الحسـابى .

مثال (٧) : البيانات التاليه توضح أوزان عشرة أطفال بالكيلو جرام .

20, 29 ,28 , 25 , 24 , 27 , 22 , 30 , 26 , 39

والمطلوب : ايجاد الانحـراف المتوسـط .

الحـل :

$$\bar{x} = \frac{\sum x_i}{n}$$

$$= \frac{270}{10} = 27$$

$$\therefore M.D. = \frac{36}{10} = 3.6 \text{ kg.}$$

x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
20	-7	7
29	2	2
28	1	1
25	-2	2
24	-3	3
27	0	0
22	-5	5
30	3	3
26	-1	1
39	12	12
270	0	36

الانحراف المتوسط للبيانات المبوبه تعطى بالعـلاقـة :

$$M.D. = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i}$$

وكما يمكن ايجاد الانحراف المتوسط حول الوسيط كما يلى :ـ

$$M.D. = \frac{\sum |x_i - Q_2| f_i}{\sum f_i} \quad \& \quad M.D. = \frac{\sum |x_i - Q_2|}{n}$$

مثال (٨) بأستخدام البيانات الخاصه بالتوزيع التكرارى للانفاق اليومى لعدد 100 عامل .

والمطلوب : ايجاد الانحراف المتوسط .

الحل : فى دراستنا لمقاييس النزعه المركزيا وجدنا ان المتوسط لهذا التوزيع هو :

$$\bar{x} = 87.2$$

ونحصل بيانات الانحرافات المتوسطه من الجدول التالى

فئات	f	x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$	$\lvert x - \bar{x} \rvert f$
50 – 60	6	55	-32.2	32.2	193.2
60 – 70	12	65	-22.2	22.2	266.4
70 – 80	15	75	-12.2	12.2	183.0
80 – 90	24	85	-2.2	2.2	52.8
90 – 100	18	95	7.8	7.8	140.4
100 – 110	14	105	17.8	17.8	249.2
110 – 120	11	115	27.8	27.8	305.8
$\sum$	100				1390.8

• الانحراف المتوسط باستخدام الوسط الحسابى :

$$M.D. = \frac{\sum |x - \bar{x}| f_i}{\sum f_i}$$

$$= \frac{1390.8}{100} = 13.908$$

٤ــ الانحــراف المعيــارى Standard Deviation

تعريف : يعرف الانحراف المعيارى بانه الجذر التربيعى الموجب لمتوسط المربعات انحرافات القيم عن وسـطها الحسـابى .

اذا كانت قيم المتغير x هى $x_1, x_2, \ldots, x_n$

فان الانحراف المعيارى لهذه القيم فى الصــوره :

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

وهذا وسوف نتعرض لكيفيه حساب الانحراف المعيارى بالطريقتين طريقه مباشرة ، طريقــــه الانحرافات المختصره (الوسط الفرضى ، الضرب فى أو القسمة على مقدار ثابت) .

ا ــ الطريقــه المباشــره : لايجــاد الانحــراف المعيــارى للمجمــوعه :

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

نحسب انحرافات القيم عن وسطها الحســابى $x_i - \bar{x}\ ,\ i=1,\ldots,n$

نوجد مربع هذه الانحرافات $(x_i - \bar{x})^2\ ,\ i=1,2,\ldots,n$

نقسم مجموع هذه المربعات على n يكون التباين مساويا لهذه القيمه

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

فان الانحراف المعيارى هو S أى الجذر التربيعى للتبــاين ويمكن كتابه التباين بالمعادله :

$$S^2 = \left(\frac{1}{n} \sum x_i^2 - \left(\frac{\sum x_i}{n} \right)^2 \right)$$

ولاثبات ذلك :

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

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

حيث : $\sum x_i = n\bar{x}$ ، $\sum \bar{x}^2 = n\bar{x}^2$ ،

$$S^2 = \frac{1}{n}\left(\sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2 \right)$$

$$= \frac{1}{n}\left(\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \right)$$

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

وهى أبسط الصيغ للحصول على الانحراف المعيارى .

$$S = \sqrt{\frac{1}{n} \sum x_i^2 - \left(\frac{\sum x_i}{n} \right)^2} \qquad (2)$$

مثال (١) : أحسب الانحراف المعيارى والتباين لأعمار عشرة أشخاص بالسنوات والتى بالقيم :

35 , 20 , 25 , 40 , 15 , 30 , 45 , 60 , 55 , 65

الحل : يمكن الحساب من الجدول الاتى :ـ

$$\bar{x} = \frac{390}{10}$$

بأستخدام العلاقة (١) :

$$S = \sqrt{\frac{1}{10} (2640)} = 16.248 \quad \text{year.}$$

بأستخدام العلاقة (٢) :

$$S = \sqrt{\frac{1}{10} \sum x^2 - \left(\frac{\sum x}{10} \right)^2} = \sqrt{\frac{17850}{10} - \left(\frac{390}{10} \right)^2} = \sqrt{264}$$

$$= 16.248 \quad \text{year.}$$

x	$x-\bar{x}$ ($\bar{x}=39$)	$(x-\bar{x})^2$	x^2
35	-4	16	1225
20	-19	361	400
25	-14	196	625
40	1	1	1600
15	-24	576	225
30	-9	81	900
45	6	36	2025
60	21	441	3600
55	16	256	3025
65	26	676	4225
390	0	2640	17850

أما في حالة البيانات المبوبة في جدول تكراري فانه يمكن ايجاد الانحراف المعياري بأستخدام أي من الطريقتين (١) ، (٢) وذلك بعد اجراء التعديلات ، وتأخذ الصورة :

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

$$S = \sqrt{\frac{\sum x_i^2 f_i}{n} - \left(\frac{\sum x_i f_i}{n}\right)^2} \qquad (2)$$

حيث $x_1, x_2, \ldots, x_k$ هي علامة كل فئة ، k عدد الفترات ، $\sum f_i = n$

مثال (١٠) : بأستخدام بيانات الجدول (6) أوجد الانحراف المعياري والتباين للانفاق اليومي لـ 100 عامل .

الحـل : للسهوله هنا نستعمل العلاقه (2) ونكتب للتعويض من الجدول الاتي :-

فئات الرتبـات	f	x	xf	$x_i^2 f_i$
50 - 60	6	55	330	18150
60 - 70	12	65	780	50700
70 - 80	15	75	1125	84375
80 - 90	24	85	2040	173400
90 - 100	18	95	1710	162450
100 - 110	14	105	1470	154350
110 - 120	11	115	1265	145475
	100		8720	788900

من الجدول نجد ان

$$S = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum x_i f_i}{n}\right)^2}$$

$$= \sqrt{\frac{788900}{100} - \left(\frac{8720}{10}\right)^2} = \sqrt{285.16} = 16.887$$

والتبـاين $S^2 = 285.16$

بعض الخواص الاساسيه للتبـاين :-

(١) بضرب (أو قسمة) علامه الفترات x_i في (أو على) مقدار ثابت فيجب ضرب أو قسمه التباين في (أو على) مربع نفس العـدد .

أعتبر المقدار الثابت c ونفرض ان $x_1, x_2, \ldots, x_k$ علامه الفترات .

التكرارات المناظره فان : $f_1, f_2, \ldots, f_k$

$$x_i' = \frac{x_i}{c}$$

$$\bar{x}' = \bar{x} / c$$

$$\text{variance } s'^2 = \frac{1}{n} \sum (x_i' - \bar{x}')^2 f_i$$

$$= \frac{1}{n} \sum \left(\frac{x_i}{c} - \frac{\bar{x}}{c} \right)^2 f_i = \frac{1}{c^2} \left(\frac{1}{n} \sum (x_i - \bar{x})^2 f_i \right)$$

$$= \frac{1}{c^2} \; s^2 \longrightarrow s^2 = c^2 s'^2$$

أى التباين يساوى مربع المقدار الثابت فى تباين العلامات بعد قسمه أو ضرب على أو فى مقدار ثابت.

٢ـ اذا جمع أو (طرح) على أو (من) علامة الفترات A مقدار ثابت فان قيمة التباين لـــن تتغير.

أخذ المقدار الثابت A فان

$$x_1, x_2, \ldots, x_n$$

بطرح مقدار ثابت فان

$$x_1 - A, \; x_2 - A, \; \ldots, \; x_n - A$$

$$x_i' = x_i - A$$

$$\bar{x}' = \frac{\sum (x_i - A)}{n} = \bar{x} - A$$

$$s'^2 = \frac{\sum (x_i' - \bar{x}')^2}{n} = \frac{1}{n} \sum ((x_i - A) - (\bar{x} - A))^2$$

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

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

أى أن قيمة التباين لن تتغير بطرح أو جمع مقدار ثابت من علامة الفترات.

واستخدام الخاصتين (١)، (٢) يمكن ايجاد التباين بصيغة مختصرة كما يلى :ـ

$$s^2 = c^2 \left(\frac{1}{n} \sum y_i^2 - \left(\frac{\sum y_i}{n} \right)^2 \right)$$

حيث $$y_i = \frac{x_i - A}{c}$$

في حالة البيانات المبوبة :

$$s^2 = c^2 \left(\frac{1}{n} \sum y_i^2 f_i - \left(\frac{\sum y_i f_i}{n} \right)^2 \right) \quad (3)$$

مثال (11) : للمثال السابق أحسب التباين والانحراف المعياري بالطريقة المختصرة .

الحل : باستخدام العلاقة (3) واستخدام الجدول نجد ان :

$$s^2 = 10^2 \left(\frac{1}{100} 290 - \left(\frac{22}{100} \right)^2 \right) = 100 \, (2.9 - 0.0484)$$
$$= 285.16 \ \mathcal{L}$$

فئات الانفاق	f	x	x- A (A=85)	$\frac{x-A}{c} = y$ (c =10)	yf	$y^2 f$
50 - 60	6	55	-30	-3	-18	54
60 - 70	12	65	-20	- 2	- 24	48
70 - 80	15	75	-10	-1	-15	15
80 - 90	24	85	0	0	0	0
90 - 100	18	95	10	1	18	18
100 -110	14	105	20	2	28	56
110 -120	11	115	30	3	33	99
$\sum$	100				22	290

ومنها الانحراف المعياري

$$s = \sqrt{285.16} = 16.887 \ \mathcal{L}$$

ملاحظات :-

١ - فى حالة الجداول التكرارية الغير منتظمة (غير متساوية الفئات) لا يكون فى الامكان دائماً أستخدام طريقة الانحرافات المختصرة (3) لايجاد الانحراف المعيارى ويمكن استخدام التبسيط بوسط فرضى فقط.

٢ - الانحراف المعيارى (أو التباين) والانحراف المتوسط يأخذان فى الاعتبار جميع المفردات عند قياس التشتت ، الا ان الانحراف المعيارى اكثر تأثرا بالقيم المتطرفة أو الشاذة . اما نصف المدى الربيعى فهو أقل هذه المقاييس تأثرا بالقيم المتطرفة ، ولذلك فانه يفضل استخدامه فى حالة التوزيعات شديدة الالتواء .

٣ - فى حالة التوزيعات المتماثلة أو بسيطة الالتواء ، هناك علاقة تجريبية تربط ما بين الانحراف المعيارى وكل من الانحراف المتوسط ونصف المدى الربيعى وهذه العلاقات هى :-

(i) $M.D. = \frac{4}{5} S$

(ii) $Q = \frac{2}{3} S$

(iii) $M.D. = 1.2 Q$

* معامل الأختلاف Coefficient of variation

اذا أردنا مقارنة تشتت مجموعتين تختلفان فى وحدة القياس أو مقارنة تشتت مجموعتين مختلفتين لها نفس وحدة القياس وللتخلص من الاختلاف فى الوسطين الحسابين فذلك بأستخدام مقياس نسبى يقيس التشتت بوحدات من الوسط الحسابى وهذا القياس هو ما يسمى " بمعامل الاختلاف " وان معامل الاختلاف يأخذ الصورة

$$\text{معامل الاختلاف} = \frac{\text{مقياس التشتت}}{\text{مقياس المتوسط}} \times 100$$

ومن اكثر معاملات الاختلاف استخداما يلى :-

(i) $V = \frac{S}{\bar{x}} \times 100$

$$(ii)\ V = \frac{Q}{Me} \times 100 =$$

$$= \frac{Q_3 - Q_1}{2 \times Me} \times 100$$

مثال (١٢) : البيانات التاليه توضح الأجر اليوس بالجنيه لمجموعتين من الاشخاص .

المجموعه الاولى : 5, 7 ,9 ,11, 13, 15

المجموعه الثانيه 75 , 77, 79 , 81 , 83 , 85

والمطلوب المقارنه بين تشتت الاجور اليوميه لكل من المجموعتين :ـ

الحـل :ـ

x_2	$x_2 - 80$	$(x_2 - 80)^2$
75	-5	25
77	-3	9
79	-1	1
81	1	1
83	3	9
85	5	25
480		70

x_1	x_1^2
5	25
7	49
9	81
11	121
13	169
15	225
60	670

واضح ان : $n = 6$

فان المتوسـط الحسابى لكل منهما : $\bar{x}_1 = \frac{60}{10} = 10$

$$\bar{x}_2 = \frac{480}{6} = 80$$

ولحسـاب الانحراف المعيـارى S_1, S_2 لكل منهمـا :

$$S_2 = \sqrt{\frac{70}{6}} = 3.416 = \sqrt{\frac{670}{6} - \left(\frac{60}{6}\right)^2} = S_1$$

∴ معامل الاختلاف للمجموعه الاولى

$$V = \frac{S_1}{\bar{x}_1} \times 100 = \frac{3.416}{10} \times 100 = 34.16 \ \%$$

ومعامل الاختلاف للمجموعة الثانيه :ــ

$$V = \frac{S_2}{\bar{x}_2} \times 100 = \frac{3.416}{80} \times 100 = 4.27 \ \%$$

ومن ذلك تستنتج ان التشتت النسبى للاجور اليومية للمجموعه الاولى اكبر من التشتت النسبى للأجـ اليوميه للمجموعة الثانيه .

(6)

مثال (١٣) : باستخدام بيانات الجدول الخاص بالانفاق الا لاجور العمال أوجد معامـل الاختلاف لهذا التوزيع .

الحل : سبق انَ أوجدنا المقاييس اللازمه للحصول على معاملات الاختلاف النسبيه وكانت كما يلى :

$\bar{x} = 87.20 \ \mathcal{L}$, $S = 16.887 \ \mathcal{L}$,

$Q_1 = 74.667 \ \mathcal{L}$, $Q_3 = 100 \ \mathcal{L}$

$Me = 87.083 \ \mathcal{L}$

$$\therefore V = \frac{S}{\bar{x}} \ 100 = \frac{16.887}{87.20} \times 100 = 19.366 \ \%$$

معامل الاختلاف الربيعى :

$$V = \frac{Q_3 - Q_1}{2 \ Me} \ 100 = \frac{100 - 74.667}{2\ (87.083)} = 14.545 \ \%$$

٦ ـ القيـم المعيـاريـه Standard value .

د رســنا فيما سبق بعض المقاييس الاحصائية التى يمكن بواسطتها المقارنة بين تشتتى مجموعتين وفى بعض الحالات يكون الباحث فى حاجه لان يقارن قيمتين فى مجموعتين مختلفتين • وفى هذه الحاله نحتاج الى مقياس احصائى •

فاذا فرضنا اننا نريد المقارنة بين وزن طالب وهو ٨٠ كيلو جرام ووزن طالب اخر هو ٧٠ كيلوجرام من مجموعتين مختلفتين من الطلبة فانه لا يصح هنا ان نقارن بين الرقمين المذكورين دون الوقــوف على شكل التوزيع المسحوبه منه القيمة الاولى والتوزيع المسحوبه منه القيمة الثانية • ولاجراء مثـل هـذه المقارنه يجب ان نقارن موضع كل قيمه من هاتين القيمتين على التوزيع الخاص بها • فنوجد بعد هـا عن متوسط التوزيع مقاسا بوحدات معياريه ، أى نوجد القيمه المعيارية لوزن الطالب الاول والقيمـه المعيارية لوزن الطالب الثانى ثم نقارن القيم المعيارية ببعضها البعض •

حيث تعرف القيمة المعيارية للفرد ، x التى وسطها الحسابى $\bar{x}$ للتوزيع والانحراف المعيارى وتعطى بالعلاقـة :ـ

$$Z = \frac{x - \bar{x}}{S}$$

<u>مثال (١٤)</u> : اذا كان لدينا نتيجه طالب فى مادتى الاحصاء والكيمياء وكانت 80, 70 واذا علمت ان متوسط درجات الطلبه فى الاحصاء والكيمياء هى 74 ، 67 درجه والانحراف المعيارى لهما على الترتيب 2 ، 1.5 • والمطلوب تحديد هل المستوى الافضل للطالب كان لمادة الرياضة ام لمادة الكيمياء •

الحـل : لتحديد المستوى الافضل للطالب بالدرجه المعيارية لكل مادة على حداه كما يلـي البيانات المعطاه :

$x_1 = 70\ ,\ \bar{x}_1 = 67\ ,\ S_1 = 1.5$

$x_2 = 80\ ,\ \bar{x}_2 = 74\ ,\ S_2 = 2$

$$z_1 = \frac{x_1 - \bar{x}_1}{S_1} = \frac{70 - 67}{1.5} = 2$$

$$z_2 = \frac{x_2 - \bar{x}_2}{S_2} = \frac{80 - 74}{2} = 3$$

حيث القيمه المعياريه للكيميا اكبر من القيمه المعياريه للاحصاء فانه يمكن القول ان مستوى اداء الطالب فى ماده الكيمياء أفضل من مستوى ادائه فى ماده الاحصاء .

ثالثا : مقاييس الالتواء

١ ـ مقياس الالتواء :ـ

ذكرنا فيما سبق ان المنحنيات التكراريه قد تكون متماثله أو غير متماثله (ملتويه) والالتواء هو بعد المنحنى عند التماثل وهو أما يكون التواء موجبا (الى اليمين) او التواء سالبا (الى اليسار) والتوزيع المتماثل هو الذى يمكن تقسيم المساحه تحت المنحنى التكرارى الذى يمثله الى نصفين متطابقين تماما . وان التكرارات تتزايد أو تتناقص بطريقه منتظمه على جانبى محور التقسيم . واما التوزيع موجب الالتواء فهو ذلك التوزيع الذى تتركز فيه التكرارات فى فئاته الدنيا ويكون ذيله (أو طرفه الايمن) أطول من ذيله الايسر .

والتوزيع سالب الالتواء هو ذلك التوزيع الذى تتركز فيه التكرارات فى فئاته العليا يكون ذيله (أو طرفه) الايسر أطول من ذيله الايمن .

ومن المعلوم ان الوسط الحسابى والوسيط والمنوال تتساوى جميعها من حيث القيمة فى حالة التوزيعات المتماثله .

ولكنا بعدت قيم هذه المتوسطات عن بعضها البعض كلما بعد التوزيع عن التماثل ويستخدم هذا الفرق فى ايجاد مقاييس الالتواء . ويسمى القياس الناتج بمعامل الالتواء . هذا ومعامل الالتواء لا يوضح نوع الالتواء فقط وانما يعطى ايضا قياسا لدرجة هذا الالتواء . وبصفه عامه فان معامل الالتواء يجب ان يحقق الشروط الاتيه :ـ

(١) ان تكون قيمته مساويه للصفر فى حالة المنحنيات المتماثله .

(٢) ان تكون قيمته مجرده من وحدات القياس (أى نسبه)

وفيما يلى بعض معاملات الالتواء الاكثر شيوعا :ـ

معامل الالتواء الاول (γ_1) = $\dfrac{\text{الوسط الحسابى ـ المنوال}}{\text{الانحراف المعيارى}}$

$$\gamma_1 = \frac{\bar{x} - Mo}{S}$$

معامل الالتواء الثانى (γ_2) = $\dfrac{3\,(\text{الوسط الحسابى ـ الوسيط})}{\text{الانحراف المعيارى}}$

$$\gamma_2 = \frac{3\,(\bar{x} - Me)}{S}$$

ويسميان معامل الالتواء الاول لبيرسون ومعامل الالتواء الثانى لبيرسون . وهذا ولا يمكن حساب هذين المعاملين فى حاله التوزيعات التكرارية المفتوحه كما انه يتأثر هذين المعاملين بالقيم المتطرفة او الشاذه .

وللتخلص من هذه المشكلة فى حالة التوزيعات التكرارية المفتوحه فقد أمكن استنتاج معامل ثالث للالتواء هو ما يسمى بمقياس بولى للالتواء او معامل الالتواء الربيعى الذى يأخذ الصوره التاليه :ـ

$$\gamma_3 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{(Q_3 + Q_1) - 2\,Q_2}{Q_3 - Q_1}$$

ويتميز هذا المعامل بانه معامل الالتواء الوحيد الذى يمكن ايجاده من الرسم . حيث يمكن حسابه باستخدام المنحنى التكرارى المتجمع الصاعد أو المنحنى التكرارى المتجمع الهابط .

كما ان الالتواء يمكن حسابه من العزوم المختلفه للتوزيع التى يمكن توضيحها فيما بعد .

مثال (١٤) : باستخدام البيانات الواردة فى الجدول (6) التوزيع الخاص باجور 100 عامل .

أوجد : معامل الالتواء الاول والثانى لبيرسون وكذلك معامل الالتواء الربيعى لبولى .

الحل : من الامثله السابقة أوجدنا قيم المقاييس الاحصائيه اللازمه لحساب هذه المعاملات الالتواء

الثلاثه وكانــت كما يلـى .

$\bar{x} = 87.20\,£$, $Me\ (Q_2) = 87.083\,£$, $Mo = 86\,£$

$Q_1 = 74.667\,£$, $Q_3 = 100\,£$, $S = 16.887\,£$.

$$\therefore \quad \gamma_1 = \frac{\bar{x} - Mo}{S} = \frac{87.20 - 86}{16.887} = 0.071 \quad ,$$

$$\gamma_2 = \frac{3(\bar{x} - Q_2)}{S} = \frac{3\ (87.20 - 87.083)}{16.887} = $$

$$= 0.021$$

$$\gamma_3 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

$$= \frac{(100 - 87.083) - (87.083 - 74.667)}{(100 - 74.667)}$$

$$= 0.020$$

يــلاحظ : اننا حصلنا تقريبا على نفس النتيجه وتفيد بان التوزيع موجب الالتواء وقيمته بسيطه جـدا مما يدل على ان التوزيع قريب جدا من التماثل أو معتدل الالتواء . ولا ختلاف هذه المعاملات فـانه يجب عند اجراء مقارنة بين درجات التواء التوزيعات المختلفه فانه يجب استخدام معامل واحد وهنـاك مقياس اخر للالتواء اكثر دقه يتوقف على ما يسمى عزم المنحنى .

تعريــف العــزوم : كلمه عزم مشتقه من علم الاستاتيكا حيث يقاس عزم القوه بحاصل ضرب مقدارها فــى ذراع عزمها (والذراع هو بعد عمل خط القوة عن مركز العزم) ويكون عزوم مجموعه من القوى = مجمـوع حاصل ضرب كل قوة فى ذراع عزمها .

تمارين (٥)

١ ـ التوزيع التكراري الاتي يوضح فئات الاعمار لمجموعه من الاشخاص :ـ

فئات الاعمار	5 -	20 -	35 -	50 -	65 -	70 -	85 - 100
عدد الاشخاص	2	7	10	16	25	12	8

أوجد : أ : ايجاد كل من المدى ، نصف المدى الربيعي والانحراف المتوسط ، والانحراف المعياري .

ب ـ ايجاد معاملات التشتت النسبيه (معامل الاختلاف) التى يمكن الحصول عليها .

جـ ـ ايجاد قيمة ونوع الالتواء وتفلطح التوزيع .

٢ ـ بالجدول التالى يبين توزيع عدد من العمال حسب الانفاق الشهرى .

فئات الانفاق	less than 20	20 -	40 -	60 -	80 -	100-110
عدد العمال	5	16	28	40	25	16

أوجد : أ ـ رسم المنحنيات المتجمع الصاعدة والهابطه واستنتج درجه التشتت والالتواء من الرسم

اذا كان طول الفئه الاولى مساويا لطول الفئه التاليه لها مباشرة فأوجد :

ب ـ الوسط الحسابى والوسيط والمنوال .

جـ ـ الانحراف المعيارى ـ الربيع الأول والربيع الثالث ـ المئين العاشر والمئين التسعون

د ـ أحسب المقاييس التاليه

معامل الاختلاف ـ معامل الالتواء ـ معامل التفلطح .

الباب الثالث

العلاقه بين متغيرين

أ : مقاييس الارتباط Measures of correlation

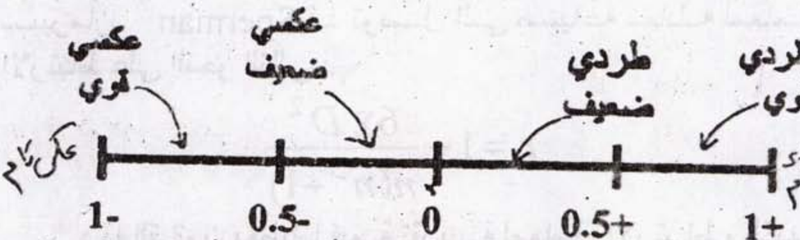
عرضنا فيما سبق بعض مقاييس الإحصائية التي تتناول متغير واحد بوصف نزعته المركزية أو متوسطة القيم التي يشملها و تشتت هذه القيم عن المتوسط و يهدف هذا الباب إلي معرفة و توضيح العلاقة بين متغيرين سواء في قيم مجموعة معينة موزعة حسب متغيرين كحالات فرديه أو موزعه فى جدول تكرارى مزدوج.

وقد يكون الارتباط بين المتغيرين المراد قياس العلاقه بينهما , فى نفس الاتجاه . بمعنى انه كلما زادت قيمه احد المتغيرين , زادت قيمه المتغير الاخر وهذا ما يسمى بالعلاقة الارتباطيه الموجبه . كما قد يكون الارتباط بين المتغيرين سالبا بمعنى انه كلما زادت قيمه احد المتغيرين , نقصت قيمه المتغير الاخر . وعلى هذا فأن مقاييس الارتباط توضح مدى التغير الذى يحدث فى ظاهره ما (متغير ما) وتصاحبه تغيرات فى ظاهره اخرى (متغير اخر) وفى نفس الاتجاه (موجب) او فى الاتجاه المضاد (سالب) . اى انه يمكن قياس الارتباط عن طريق التغيرات التى تحدث فى المتغيرين المراد دراستها .

وعلى هذا فان معامل الارتباط يلخص ارتباط البيانات العدديه لأى ظاهرتين أو متغيرين فى درجه واحده .

معامل الارتباط قيمته تتراوح بين (-1, 1) حيث تشير القيمة +1 إلى وجود علاقة طرديه أو موجبه تماما. وتشير القيمة -1 إلى وجود علاقة عكسية أو سلبية تماما و الملاحظ أن قيمة الارتباط تأخذ شكل كسر عشري أي جزء من الواحد الصحيح.

ويمكن القول أن الارتباط ضعيف إذا كانت قيمة معادل الارتباط بين -0.5،+0.5 وقوي إذا كانت بين (-0.5،-1) عكسي قوي ، بين (0.5، 1) طردي قوي كما يلي:



و يستخدم معامل الارتباط في حالة البيانات الكمية أو العددية سواء كانت لقيم فردية معروفة (غير مبوبة) أو لقيم موزعة في جدول تكراري مزدوج في حالة البيانات النوعية أو الوصفية فأننا نستخدم معاملات أخري من أهمها معامل الاقتران- معامل التوافق وسوف نعرض كل من هذه المقاييس كما يلي:

1- معامل ارتباط سبيرمان الرتب

Rank correlation coefficient

في كثير من الأحيان لا يستطيع الباحث تحديد قيم المتغير أثناء تغيره، و يصبح من السهل بالنسبة إليه ترتيب مراحل تغيره كأن

يحدد أيها الأول ، و أيها الثاني ، أيها الأخير . في هذه الحالة نضع ترتيب القيم المتعلقة لكل متغير أو ظاهرة .
و بحساب الفرق بين رتبتي كل قيمتين متناظرتين و هذه الفروق تتوقف قيمها علي شدة الاتفاق أو الاختلاف بين قيم المتغيرين فإذا أسمينا هذه الفروق D كانت مربعها D^2 . اذا كانت عدد القيم المعلومه لكل من المتغيرين n وكانت r ترمز لمعامل الارتباط فان سبيرمان Sperman قد توصل الى صياغه معادله لمعامل الارتباط على النحو التالى :

$$r = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

و هذا القانون يعطينا قيمة تقريبية لمعامل الارتباط و لكنها تمتاز بسهولة و سرعة حسابها . كما تمتاز هذه الطريقة بأنها تصلح لقياس الارتباط بين ظاهرتين من بيانات نوعيه غير كميه , ما دا فى الامكان ترتيب هذه البيانات النوعيه . كما فى الامثله الاتيه :

مثال (1):

اوجد معامل ارتباط الرتب بين المتغيرين التاليين للحالات التاليه:

x	15	12	18	14	10	34	40	60	44	50
y	40	30	43	35	14	25	32	50	37	48

الحل:

من الجدول التالى تكون رتب x ,رتب y والفرق D ومربع الفرق D^2 والترتيب من الاكبر الى الاصغر كما يلى

x	y	R(x)	R(y)	D	D^2
15	40	7	4	3	9
12	30	9	8	1	1
18	43	6	3	3	9
14	35	8	6	2	4
10	14	10	10	0	0
34	25	5	9	-4	16
40	32	4	7	-3	9
60	50	1	1	0	0
44	37	3	5	-2	4
50	48	2	2	0	0
					52

$$r = 1 - \frac{6\ (52)}{10\ (99)} = 0.685$$

ى أن المتغيرين يرتبطان معا ارتباط قويا نوعا .

مثال (2):

اوجد معامل ارتباط الرتب بين المتغيرين التاليين:

x	16	20	43	40	16	45	20	18	20	22
y	34	41	37	41	20	34	43	22	41	40

الحل:

نكون الجدول التالى لايجاد مربع الفرق للرتب D^2

x	y	R(x)	R(y)	D	D^2
16	34	9.5	7.5	2	4
20	41	6	3	3	9
43	37	2	6	-4	16
40	41	3	3	0	0
16	20	9.5	10	-0.5	0.25
45	34	1	7.5	-6.5	42.25
20	43	6	1	5	25
18	22	8	9	-1	1
20	41	6	3	3	9
22	40	4	5	-1	1
					107.5

$$r = 1 - \frac{6\ \ (107.5)}{10\ \ (99)} = 0.35$$

وهذا يعنى أن المتغيرين يرتبطان ارتباطا طردى ضعيفا .

مثال (3)

أحسب معامل ارتباط سبيرمان بين تقديرات عشر طلاب فى مادتى الجبر والاحصاء من البيانات التاليه :

الجبر x	B	G	P	Ex	VG	G	P	G	P	B
الاحصاء y	P	Ex	B	Ex	G	VG	G	P	P	G

(ضعيف B , مقبول P , جيد G , جيدجدا VG , ممتاز Ex)

x	y	R(x)	R(y)	D	D^2
B	P	9.5	8	1.5	2.25
G	Ex	4	1.5	2.5	6.25
P	B	7	10	-3	9
Ex	Ex	1	1.5	-0.5	0.25
VG	G	2	5	-3	9
G	VG	4	3	1	1
P	G	7	5	2	4
G	P	4	8	-4	16
P	P	7	8	-1	1
B	G	9.5	5	4.5	20.25
					69

$$r = 1 - \frac{6\ (69)}{10\ (99)} = 0.58$$

اى أن الارتباط متوسط وطردى بين تقديرات الطلاب فى هاتين المادتين .

2-معامل ارتباط بيرسون

معامل ارتباط بيرسون Pearson يهتم بالقيم التى تأخذها المتغيرات وليس برتبه كل منهم فان معامل ارتباط بيرسون ادق فى حسابه على أساس القيم ويتأثر بأى تغير فى القيم .

ولقد وضع بيرسون الصيغه كما يلى :

اذا كان لدينا المتغيرين x , y على النحو التالى

x	x_1	x_2	-------------------	x_n
y	y_1	y_2	-------------------	y_n

$$r(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y} \qquad (1)$$

حيث يسمى cov(x,y) بالتباين المتلازم بين المتغيرين

$$r(x,y)=\frac{\frac{1}{n}\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\Sigma(x_i-\bar{x})^2.\frac{1}{n}\Sigma(y_i-\bar{y})^2}} \qquad (2)$$

$$r(x,y)=\frac{\Sigma x_i y_i-n\bar{x}\bar{y}}{\sqrt{\left(\Sigma x_i^2-\frac{(\Sigma x_i)^2}{n}\right)\left(\Sigma y_i^2-\frac{(\Sigma y_i)^2}{n}\right)}} \qquad (3)$$

حيث $-1\le r(x,y)\le 1$

ولاثبات صحه هذه العلاقه (3)

$$\Sigma(x_i-\bar{x})(y_i-\bar{y})=\Sigma(x_i y_i-\bar{x}y_i-x_i\bar{y}+\bar{x}\bar{y})$$

$$=\Sigma x_i y_i-\bar{x}\Sigma y_i-\bar{y}\Sigma x_i+n\bar{x}\bar{y}$$

$$=\Sigma x_i y_i-n\bar{x}\bar{y}-n\bar{x}\bar{y}+n\bar{x}\bar{y}$$

$$\Sigma(x_i-\bar{x})(y_i-\bar{y})=\Sigma x_i y_i-n\bar{x}\bar{y}$$

حيث $\bar{y}=\frac{\Sigma y}{n}, \bar{x}=\frac{\Sigma x}{n}$

فى حالـه الارقـام الكبيـره يمكـن تسـهيل الحسـابات بطـرح ثـابتين c_1, c_2 لكل قيم x,y على الترتيب فتصبح المفردات x هى :

$$x_1-c_1, x_2-c_2, \ldots\ldots\ldots\ldots, x_n-c_1$$

فيكون الفرق

$$x_i-\bar{x}=(x_i'+c_1)-(\bar{x}'+c_1)=x_i'-\bar{x}'$$

بالمثل $y_i-\bar{y}=y_i'-\bar{y}'$

حيث $y_i'=y_i-c_2$

ولكن كما سبق وجدنا أن تباين كل من المتغير x,y لن يتأثر بطرح وسط فرضى وعلى ذلك فأن

$$\sigma_x=\sigma_{x'}, \sigma_y=\sigma_{y'}$$

فيكون معامل الارتباط فى هذه الحاله هو

$$r(x,y)=\frac{\Sigma x_i' y_i'-\frac{\Sigma x_i' \Sigma y_i'}{n}}{n\,\sigma_{x'}\sigma_{y'}}$$

$$=r(x',y') \qquad (4)$$

مثال (4)

أوجد معامل ارتباط بيرسون من التوزيعات التاليه

x يمثل الاجر الشهرى y يمثل الانفاق

X	100	101	102	102	100	99	97	98	96	95
y	98	99	99	97	95	92	95	94	90	91

الحل:

x	y	$x-98=x'$ $x-c_1$	$y-95=y'$ $y-c_2$	x'^2	y'^2	$y'x'$
100	98	2	3	4	9	6
101	99	3	4	9	16	12
102	99	4	4	16	16	16
102	97	4	2	16	4	8
100	95	2	0	4	0	0
99	92	1	-3	1	9	-3
97	95	-1	0	1	0	0
98	94	0	-1	0	1	0
96	90	-2	-5	4	25	10
95	91	-3	-4	9	16	25
		10	0	64	96	61

$$r(x,y)=\frac{61-\frac{(0)(10)}{10}}{\sqrt{(64-\frac{(10)^2}{10})(96-\frac{(0)^2}{10})}}=\frac{61}{\sqrt{54(96)}}$$

$= 0.85$

اى ارتباط طردى قوى

٣- معامل الارتباط بيرسون فى التوزيعات التكرارية المزدوجة

لإيجاد معامل الارتباط للتوزيع التكرارى المزدوج نتبع الخطوات التالية :

نتخذ وسطا فرضيا لكل من المتغيرين كلا على حسب قيمته كما هو متبع فى حالة إيجاد الوسط الحسابى بالطريقة المختصرة ثم نحدد انحرافات كل فئة عن المتوسط الفرضى.

ويمكن إيجاد معامل الارتباط من القانون التالى

$$r(x,y) = \frac{\sum x'_i y'_i f_{12} - \dfrac{(\sum x'_i f_1)(\sum y'_i f_2)}{n}}{\sqrt{\left(\sum x'^2_i f_1 - \dfrac{(\sum x'_i f_1)^2}{n}\right)\left(\sum y'^2_i f_2 - \dfrac{(\sum y'_i f_2)^2}{n}\right)}}$$

ونستخدم هذه الطريقة فى حالة الفئات المتساوية الطول .

مثال (٦)

أوجد معامل الارتباط بطريقة بيرسون من الجدول المزدوج التالى :

العمر / الأجر	10 -	20 -	30 -	40 -	50 -	60-70	$\sum$
6 -	2	1			3		6
10 -	3		5		2	1	11
14 -	6	1	9	4	3	2	25
18 -22			4	3		1	8
	11	2	18	7	8	4	50

ولحساب معامل الارتباط نتبع الخطوات التالية :-

أولاً : نوجد الجدول التوزيع الهامشى لكل من قيم x و قيم y ونوجد منها $\sum \acute{x} f_1$ ، $\sum \acute{x}^2 f_1$ ، $\sum \acute{y} f_2$ ، $\sum \acute{y}^2 f_2$:

فئات العمر	f_1	x	$\frac{X-35}{10} = \acute{x}$	$\acute{x} f_1$	$\acute{x}^2 f_1$
10 - 20	11	15	- 2	- 22	44
20 - 30	2	25	- 1	- 2	2
30 - 40	18	35	0	0	0
40 - 50	7	45	1	7	7
50 - 60	8	55	2	16	32
60 - 70	4	65	3	12	36
	50			11	121

فئات الأجر	f_2	y	$\acute{y} = \frac{y-16}{4}$	$\acute{y} f_2$	$\acute{y}^2 f_2$
6 - 10	6	8	- 2	- 12	24
10 - 14	11	12	- 1	- 11	11
14 - 18	25	16	0	0	0
18 - 22	8	20	1	8	8
	50			- 15	43

ثانياً لحساب $\sum x' y' f_{12}$ كما يلى :

		x' →	-2	-1	0	1	2	3	
y'	العمر / الأجر		10-20	20-30	30-40	40-50	50-60	60-70	ΣΣ
-2	6 - 10		2 (8)	1 (2)			3 (-12)		-2
-1	10 - 14		3 (6)				2 (-4)	1 (-3)	-1
0	14 - 18								
1	18 - 22					3 (3)		1 (3)	6
	ΣΣ		14	2		3	- 16	0	(3)

وعلى ذلك فإن $\sum x' y' f = 3$

$\sum x'^2 f_1 = 121$, $\sum x' f_1 = 11$,

$\sum y'^2 f_2 = 43$, $\sum y f_2 = -15$,

$$r(x,y) = \frac{3 - \frac{(11)(-15)}{50}}{\sqrt{\left(121 - \frac{(11)^2}{50}\right)\left(43 - \frac{(-15)^2}{50}\right)}}$$

$$= \frac{6.3}{\sqrt{118.58\ (38.5)}} = 0.09$$

وهو ارتباط طردى ضعيف أى الارتباط بين العمر والأجر ضعيف جداً

مثال (٧)

الجدول التكرارى المزدوج الآتى يبين حدة الخدمة والأجر لمجموعة 25 عاملاً والمطلوب حساب معامل الارتباط (بطريقة بيرسون)

الأجر \ مدة الخدمة	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10	$\sum$
50 - 60	1					1
60 - 70		1	2			3
70 - 80		1	2	5		8
80 - 90	1	1	2	3	2	9
90 - 100				1	3	4
$\sum$	2	3	6	9	5	25

الحل :ـ <u>**التوزيع الهامشى للأجر**</u>

sets	f_1	x	$\frac{X-75}{10} = x'$	$x' f_1$	$x'^2 f_1$
50 - 60	1	55	- 2	- 2	4
60 - 70	3	65	- 1	- 3	3
70 - 80	8	75	0	0	0
80 - 90	9	85	1	9	9
90 - 100	4	95	2	8	16
	25			12	32

التوزيع الهامشى لمدة الخدمة

sets	f_2	y	$y' = \frac{y-5}{2}$	$y' f_2$	$y'^2 f_2$
0- 2	2	$\cancel{1}$	- 2	- 4	8
2 - 4	3	3	- 1	- 3	3
4 - 6	6	5	0	0	0
6 - 8	9	7	1	9	9
8 - 10	5	9	2	10	20
$\sum$	25			12	40

وعلى الطالب ايجاد المطلوب (كتمرين)

II: توفيق المنحنيات Curve Fitting

كثيراً ما توجد في الحياة العملية علاقات بين متغيرين أو أكثر ونهتم بعد إجرائنا التجارب العملية بإيجاد العلاقات الرياضية التي تربط بين المتغيرات، كما يعنينا في كثير من الأحيان بعد الحصول على هذه العلاقات الرياضية التي تقرب البيانات التجريبية استخدامها في استخلاص معلومات عن المتغيرات تكون ذات فائدة في المستقبل أو يكون لها قيمتها على أساس أن استخدامها كان من خلال العلاقة الرياضية التي نحصل عليها ولم يكن من قيم البيانات التجريبية.

مثال :

يعتمد ذوبان قرص دواء معين على وزن القرص، فإذا رمزنا لوزن القرص بالرمز x، ولمعدل ذوبانه بالرمز y، وحصلنا على قيم للمتغير x (كأوزان لقرص لها نصف قطر واحد وسمك واحد)، وعلى قيم مقابلة للمتغير y (كمعدلات ذوبان هذه الأقراص) فإننا نسمي قيم المتغيرين x، y القيم التجريبية (أو البيانات) ولأننا نعلم أن هناك علاقة تربط بين المتغيرين x، y ولتكن $y = \Phi(x)$ فإننا نود الحصول على هذه العلاقة الرياضية $y = \Phi(x)$ التي توفق البيانات التجريبية التي نحصل عليها من إجراء التجربة الفعلية

على أقراص الدواء. وبعد الحصول على العلاقة الرياضية $y = \Phi(x)$ فإنه يسكننا الإجابة على السؤال الآتي :

إذا علمنا أن وزن القرص x_o (ولم تكن x_o ضمن البيانات التجريبية) فما هي القيمة المقابلة y_o لمعدل ذوبان القرص؟.

والإجابة هي $y_o = \Phi(x_o)$.

أي أننا نعوض في العلاقة الرياضية $y = \Phi(x)$ التي توفق البيانات التي حصلنا عليها من $x = x_o$ فنحصل على $y_o = \Phi(x_o)$.

<u>مثال :</u>

السلاسل الزمنية (وتستخدم تطبيقاتها في الاقتصاد والتخطيط) هي ظواهر تتغير فيها قيم الظاهرة تبعا لتغير الزمن.

فإذا رمزنا للزمن بالرمز x ولقيمه الظاهرة بالرمز y فإن $y = \Phi(x)$ تمثل سلسلة زمنية. ومن أمثلة السلاسل الزمنية أن تمثل الظاهرة فيها الدخل القومي العام لإحدى الدول على مدى السنين المتعاقبة (x) فإذا نظرت إلى قيم الدخل القومي العام لدولة ما (y) للأعوام المتعاقبة (x) ابتداء من 1925 حتى عام 1995 مثلا وسجلت الدخل العام المقابل لكل سنة من هذه السنين لنتجت سبعون قيمة للدخل العام تقابل السبعين عاما (x).

وتمثل هذه الظاهرة سلسلة زمنية نود نيها الحصول على الاتجاه العام $y = \Phi(x)$ الذي يمثل العلاقة الرياضية التي تربط بين المتغيرين x, y والتي توفق البيانات التي حصلنا عليها للظاهرة ولأسباب تتعلق بالسياسة التخطيطية للدول يهمنا معرفة الدخل العام لها في عام 2005 مثلا. ويمكننا

الحصول على هذه القيمة بالتعويض عن قيمــــة $x = 2005$ فـــي العلاقـــة الرياضية التي حصلنا عليها $y = \Phi(x)$ لتعطينا الدخل القومي المقابل لعـــام 2005.

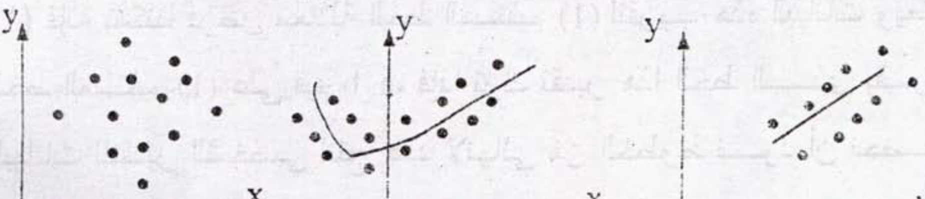
وهناك أمثلة كثيرة أخرى هامة متعددة للسلاسل الزمنية مثل كمية الإنتاج أو دخل القطن أو البصل أو الأرز ... الخ من المنتجات الصناعية أو المعـدن، والتنبؤ بكميات الإنتاج والدخل العام فيها مما يشكل الأساس التخطيطي الذي ترتكز عليه الدول في بناء إقتصادها.

العلاقة بين متغيرين :

إذا أجرينا تجربة معينة كان فيها المتغير المستقل هـــو x والمتغـــير التابع هو y فإن أول ما نفعل هو تجميع البيانـــات لـــهذه المتغـيرات مـــن التجارب التي نجريها فنحصل على الجدول المبيـــن حيـــث فيـــه (x_1, y_1)، (x_2, y_2)، ...، (x_n, y_n) تمثل n من القيم التجريبية التي يأخذها المتغيـر y التي تقابل n من القيم التجريبية التي يأخذها المتغير x، فإذا رســـمنا هـــذه النقط التجريبية على المستوى فإننا نحصل على ما نسميه (إنتشار البيانات).

x	x_1	x_2	...	...	x_n
y	y_1	y_2	...	...	y_n

ومن أمثلة انتشار البيانات الأشكال الآتية :



شكل (٣)　　شكل (٢)　　شكل (١)

في شكل (١) يبدو أن البيانات $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ يمكن تقريبها بخط مستقيم، وفي شكل (٢) بقطع مكافئ، بينما في شكل (٣) يبدو أنه لا يوجد علاقة بين المتغيرين (x, y).

وأهم ما في توفيق المنحنيات هو التوصل إلى المعادلة الرياضية المقترحة لتقريب البيانات، ويتم ذلك عن طريق الشكل الناتج من انتشار البيانات الذي يعطي صورة المعادلة الرياضية التي يمكن اقتراحها من خبرتنا عن المنحنيات التي تناظر المعادلات الرياضية المتعددة. فإذا لم يكن الانتشار بين البيانات خطي أو على صورة كثيرة حدود من درجة أعلى من الدرجة الأولى، فإننا نحاول رسم الانتشار بين $x, \log y$ أو بين $\log x$, $\log y$، أو بين $x, \frac{1}{y}$ فإن كان اتجاه النقط في أي من هذه الحالات خطيا فيمكننا فرض المعادلة الرياضية المناسبة. وإلا فيجب علينا اعتبار معادلات من أنواع أخرى.

<u>توفيق الخط المستقيم :</u>

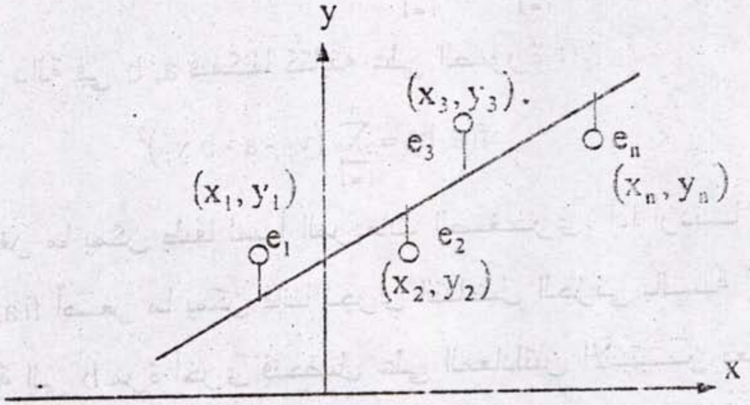
نعلم أن معادلة الخط المستقيم

$$y = a + b x \qquad (1)$$

فإذا رأينا أن البيانات $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ يأخذ انتشارها شكل (١) فإنه يمكننا فرض معادلة الخط المستقيم (1) لتقريب هذه البيانات ويعتمد الخط المستقيم (1) على قيم a, b، فإذا ترك تقدير هذا الخط الذي يقرب البيانات للتقدير الشخصي لنتج عدد لانهائي من الخطوط نود أن نحصل

على أفضلها معتمدين في ذلك على أساس رياضي معين حتى لا يترك تحديد الخط المستقيم للتقديرات الشخصية المتفاوتة.

مبدأ المربعات الصغرى :



نستخدم مبدأ المربعات الصغرى كأساس لتحديد قيم الثوابت a, b في الخط المستقيم، وسنسمي الخط الناتج $y = a + b\,x$ عن استخدام هذا المبدأ **خط المربعات الصغرى** أو أفضل خط يقرب البيانات.

وينص مبدأ المربعات الصغرى على أن يكون مجموع مربعات الأخطاء (أو الانحرافات e) للإحداثيات الصادية للنقط التجريبية عن النقط الافتراضية أصغر ما يمكن. فالإحداثي الصادي للنقطة التجريبية (x_i, y_i) يبعد عن الإحداثي الصادي للنقطة الافتراضية المقابلة (وهي على الخط المستقيم المفروض) بالمقدار $\pm e_i$ ، وهذا ما نسميه بالخطأ أو الانحراف.

فإذا تصورنا أن النقطة التجريبية (x_i, y_i) تقع على الخط (١) فلابد من إضافة الانحراف $\pm e_i$ الناتج من اعتبار أن النقطة تقع فعلا على الخط

المستقيم وعلى ذلك فإن :

$$y_i = a + b\,x_i \pm e_i \quad , i = 1, 2, \ldots, n$$
$$\therefore\ \pm e_i = (y_i - a - b\,x_i)$$

ويصبح مجموع مربعات الانحرافات على الصورة :

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

هذا المجموع هو دالة في a, b فيمكننا كتابته على الصورة :

$$f(a, b) = \sum_{i=1}^{n} (y_i - a - b\,x_i)^2$$

ويراد جعله أصغر ما يمكن طبقا لمبدأ المربعات الصغرى. إذا أردنا أن تكون الدالة $f(a, b)$ أصغر ما يمكن فإننا نجري التفاضل الجزئي بالنسبة إلى a مرة ثم بالنسبة إلى b مرة أخرى فنحصل على المعادلتين الآتيتين بعد مساواة التفاضل الجزئي الناتج من المعادلتين بالصفر.

$$\left.\begin{aligned} \sum_{i=1}^{n} y_i &= n\,a + b \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i\, y_i &= a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 \end{aligned}\right\} \qquad (2)$$

تسمى المعادلتين (2) بالمعادلتين الطبيعيتين وبحلهما آنيا نحصل على قيمتي a, b كما يلي :

$$b = \frac{n \sum x\,y - \left(\sum x\right)\left(\sum y\right)}{n \sum x^2 - \left(\sum x\right)^2}$$

$$a = \overline{y} - b\,\overline{x}$$

وبالتعويض عن a, b في معادلة الخط المستقيم (1) فإننا نحصل على خط المربعات الصغرى أو أفضل خط يترب البيانات.

تسمى المعادلة (1) خط انحدار y على x ، b معامل انحدار y على x

في حالة المنحنى يمثل قطع مكافئ معادلته:

$$y = a + bx + cx^2$$

والمطلوب ايجاد أنسب قطع مكافئ يمر خلال النقط المعطاه وهي $(x_1, y_1), (x_2, y_2), ---, (x_n, y_n)$ وفي هذه الحالة المعادلات القياسية باستخدام مبدأ المربعات الصغرى تأخذ الصورة:

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

وبحلها معا نحصل على قيم: a, b, c.

مثال(1) وفق أحسن خط يمثل البيانات الآتية:

x_i	1	3	4	6	8	9	11	14
y_i	1	2	4	4	5	7	8	9

(i) أوجد خط انحدار y على x (اذا كان x متغير مستقل)

(ii) أوجد خط انحدار x على y (اذا كان y متغير مستقل)

(iii) قدر قيمة y عندما $x=12$

(iv) قدر قيمة x عند $y=3$

(v) أوجد نقط تقاطع خطى الانحدار وما الذى تستنتجه ؟

الحل:

(i) معادلة الخط المستقيم $y = a + bx$

(ii) معادلة الخط المستقيم $X = \alpha + \beta y$

X	y	x^2	xy	y^2
1	1	1	1	1
3	2	9	6	4
4	4	16	16	16
6	4	36	24	16
8	5	64	40	25
9	7	81	63	49
11	8	121	88	64
14	9	196	126	81
$\Sigma x = 56$	$\Sigma y = 40$	$\Sigma x^2 = 524$	$\Sigma xy = 364$	$\Sigma y^2 = 256$

من مبدأ المربعات الصغرى يمكن ايجاد الثوابت كما يلي

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{8(364) - 56(40)}{8(524) - (56)^2} = 0.636$$

$$a = \bar{y} - b\bar{x} = \frac{40}{8} - (.636)\frac{56}{8} = 0.545$$

$$\therefore y = 0.545 + 0.636 X$$

عند $x = 12 \leftarrow$

$$y = 0.545 + 0.636(12) = 8.2$$

وكذلك

$$\beta = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2} = \frac{8(364) - 56(40)}{8(256) - (40)^2} = 1.5$$

$$\alpha = \bar{x} - \beta\bar{y} = \frac{56}{40} - 1.5\left(\frac{40}{8}\right) = -0.5$$

$$\therefore X = -0.5 + 1.5Y$$

قيمة y عند $X = 3 \leftarrow$

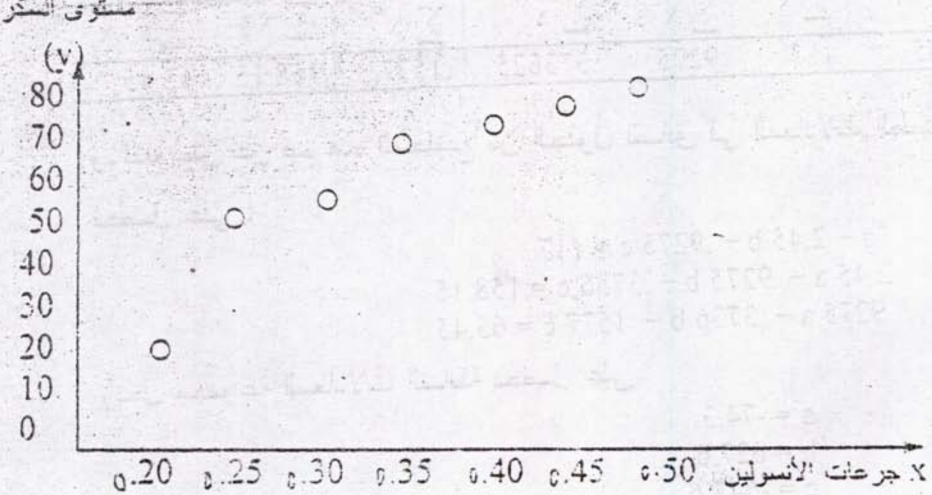
$$X = -0.5 + 1.5(3) = 4.0$$

مثال٢: في دراسة لمعرفة تأثير إحدى وصفات الأنســولين علــى تخفيــض مستوى السكر في دم مجموعة من الفئران قام باحث بإجراء هذه التجربــة على مجموعة من الفئران وحصل على النتائج التالية :

جرعات الأنسولين (x)	0.20	0.25	0.30	0.35	0.40	0.45	0.50
مستوى السكر (y)	20	50	58	63	65	68	73

(أ) ارسم شكل الانتشار واستنتج نوع العلاقة التي تربط بين المتغيرين x، y

(ب) باستخدام طريقة المربعات الصغرى أوجد معادلة النموذج المقترح في (أ)



واضح من شكل الانتشار أن العلاقة بين المتغيرين x، y هي علاقة قطــع مكافئ وليكن على الصورة

$$y = a + bx + cx^2$$

المعادلات الطبيعية هي :

$$\sum y = na + b\sum x + c\sum x^2$$
$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$
$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

x	y	x^2	x^3	x^4	xy	x^2y
0.20	20	.04	.008	.0016	4.0	0.80
0.25	50	.0625	.015625	.0039	12.5	3.125
0.30	58	.09	.027	.0081	17.4	5.22
0.35	63	.1225	.042875	.0150	22.05	7.7175
0.40	70	.16	.064	.0256	28.0	11.20
0.45	76	.2025	.091125	.0410	34.2	15.39
0.50	80	.25	.125	.0625	40.0	20.0
$\sum x$	$\sum y$	$\sum x^2$	$\sum x^3$	$\sum x^4$	$\sum xy$	$\sum x^2 y$
2.45	417	.9275	.373625	.1577	158.15	63.45

وبالتعويض عن قيم هذه المجاميع من الجدول السابق في المعادلات الطبيعية نحصل على :

$$7a + 2.45\,b + .9275\,c = 417$$
$$2.45\,a + .9275\,b + .3736\,c = 158.15$$
$$.9275\,a + .3736\,b + .1577\,c = 63.45$$

وبحل مجموعة المعادلات السابقة نحصل على

$$a = -74.3$$
$$b = 627.6$$
$$c = -647.6$$

وعليه تكون معادلة النموذج هي

$$y = -74.3 + 627.6x - 647.6x^2$$

مثال (٣)

البيانات التالية توضح العلاقة بين عدد الوحدات المنتجة (x) وتكلفة الوحدة (y) والمطلوب

١- إيجاد معادلة انحدار y على x على الصورة التربيعية

٢- تقدير تكلفة الوحدة فى حالة ما إذا كان عدد الوحدات المنتجة ٢٥٠٠ وحدة.

عدد الوحدات (ألف) X	1	2	3	4	5
تكلفة الوحدة Y	6	3	2	3	5

الحل

X	Y	X Y	X^2Y	X^2	X^3	X^4
1	6	6	6	1	1	1
2	3	6	12	4	8	16
3	2	6	18	9	27	81
4	3	12	48	16	64	256
5	5	25	125	25	125	625
15	19	55	209	55	225	979

معاملة خط الانحدار هى

$$Y = a + b\,x + c\,x^2$$

المعادلات القياسية هى

$$\sum y = n\,a + b \sum x + c \sum x^2 \;,$$

$$\sum xy = a\sum x + b \sum x^2 + c \sum x^3 \;,$$

$$\sum x^2y = a\sum x^2 + b \sum x^2 + c \sum x^4 \;,$$

بالتعويض بهذه المجاميع من الجدول السابق نجد أن

$$\therefore\ \left.\begin{aligned} 19 &= 5a + 15b + 55c \\ 55 &= 15a + 55b + 225c \\ 209 &= 55a + 225b + 979c \end{aligned}\right\} \Longrightarrow$$

$a = 10.4 \quad , \quad b = -5.343 \quad , \quad c = 0.857$

وتكون معادلة تقدير y بدلالة x (معادلة انحدار y على x) كما يلى :

$y_{est} = 10.4 - 5.343\,x + 0.857\,x^2$

ii- تقديـر تكلفة الوحدة فى حالة حجم انتاج قدرة 2500 وحدة أى y عند x=2.5 بالتعويض ينتج أن

$y_{est} = 10.4 - 5.343\,(2.5) + 0.857\,(2.5)^2$
$= 2.4$

مثال (٤)

الجدول التالى يبين سلسلة زمنية لإحدى الظواهر

الزمن السنة (t)	1953	1954	1955	1956	1957	1958	1959	1960
(y) القيمة	3	5	11	16	30	38	50	60

أوجد معادلة خط الانحدار العام على فرض أنها من الدرجة الثانية

الحل : نفرض أن معادلة الاتجاه العام هى

$Y = a + b\,x + c\,x^2$

لإيجاد قيم a ، b ، c نستخدم المعادلات القياسية السابق ذكرها فى المثال السابق .

t	y	X	X^2	X^3	X^4	X y	X^2y
1953	3	-7	49	-343	2401	-21	147
1954	5	-5	25	-125	625	-25	125
1955	11	-3	9	-27	81	-33	99
1956	16	-1	1	-1	1	-16	16
1957	30	1	1	1	1	30	30
1958	38	3	9	27	81	114	342
1959	50	5	25	125	625	250	1250
1960	60	7	49	343	2401	420	2940
$\sum$	213	0	168	.0	6216	719	4959

بالتعويض نجد أن

$$\left.\begin{array}{l} 213 = 8\,a + 168\,c \\ 719 = 168\,b \\ 4949 = 168\,a + 6216\,c \end{array}\right\} \Longrightarrow \qquad x = \frac{t - 1956.5}{0.5}$$

$$a = 22.9 \quad , \quad b = 4.3 \quad , \quad c = 0.2$$

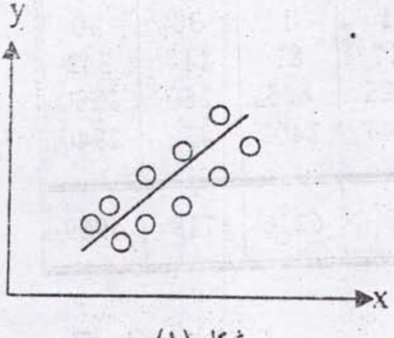
أى أن المعادلة العامة للانحدار هى

$$y = 22.9 + 4.3\,x + 0.2\,x^2$$

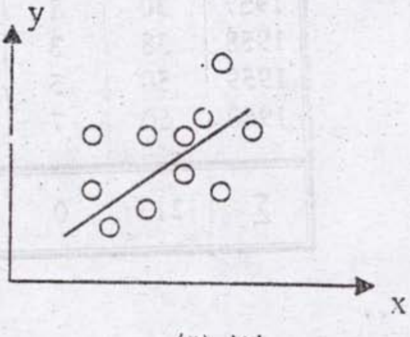
حيث نقطة الأصل هى منتصف سنة 1956 ،

الخطأ المعياري للتقدير Standard error of estimate

بعد أو وجدنا علاقة خطية رياضية في شكل خط انحدار بين المتغير المستقل x والمتغير التابع y يهمنا هنا قياس مدى اعتمادنا على المعادلة الرياضية التي وجدناها فمثلا بالنظر إلى الشكلين التاليين :



شكل (١)



شكل (٢)

نلاحظ أن خط الانحدار في شكل (١) يعطي تقديرا أكثر دقة، وذلك لكون البيانات متقاربة أكثر من الخط أي أنها تنتشر بشكل أقل حول خط الانحدار، بينما في شكل (٢) نلاحظ أن انتشار البيانات حول خط الانحدار أكثر تشتتا، ولهذا نتوقع أن تقديرات هذا الخط تكون أقل دقة.

هنا سوف نعرف كمية عددية نسميها الخطأ المعياري للتقدير ونرمز لها بالرمز $S_{y.x}$ وهي تقيس التغير (أو الانتشار) للمشاهدات المعطاة حول خط الانحدار، وتعطى بالعلاقة التالية : $S_{y.x} = \sqrt{\dfrac{\sum(y-\hat{y})^2}{n}}$

حيث y هي القيم المشاهدة المعطاة في الأزواج المرتبة (x, y)، $\hat{y}$ هي القيم المقدرة من المعادلة $y = a + b\,x$

ويمكن حساب $S_{y.x}$ من الصورة المبسطة التالية

$$S_{y.x} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n}}$$

وتفسير الخطأ المعياري للتقدير هو تماما كتفسير الانحراف المعياري فكلما كان الخطأ المعياري للتقدير كبيرا كلما كان تشتت المشاهدات حول خط الانحدار كبيرا.

معامل التحديد Coefficient of Determination

يعرف معامل التحديد بمقدار التغير في y الذي تفسره معادلة خط الانحدار. ويرمز له بالرمز r^2 حيث

$$r^2 = \frac{\sum\left(\hat{y} - \overline{y}\right)}{\sum\left(y - \overline{y}\right)^2}$$

أو من الصورة

$$r^2 = 1 - \frac{\sum\left(y - \hat{y}\right)^2}{\sum\left(y - \overline{y}\right)^2}$$

وفي حالة الانحدار الخطي يكون على الصورة

$$r^2 = \frac{b^2\left(\sum x^2 - n\overline{x}^2\right)}{\sum y^2 - n\overline{y}^2}$$

مثال : يريد باحث معرفة العلاقة بين عدد السجائر التي يدخنها مجموعة من العمال وعدد أيام الغياب بسبب المرض، فكانت النتائج التالية :

55	50	40	35	20	15	6	0	عدد السجائر اليومي (x)
16	9	12	10	6	4	3	2	عدد أيام الغياب (y)

(أ) مثل x، y بواسطة لوحة الانتشار واستنتج نوع العلاقة.

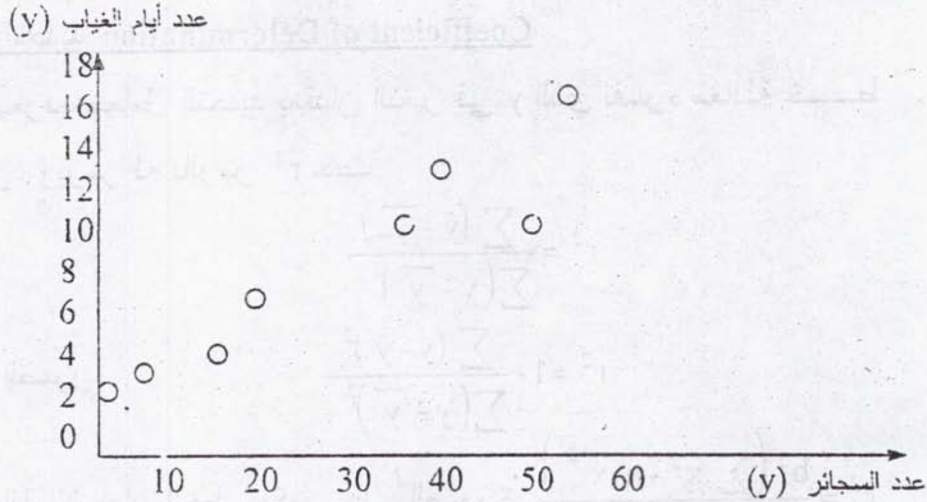
(ب) احسب r معامل الارتباط.

(جـ) أوجد معادلة انحدار y على x.

(د) قدر عدد أيام الغياب لعامل يدخن 25 سيجارة في اليوم.

(هــ) إحسب الخطأ المعياري للتقدير $S_{y.x}$

(و) أوجد معامل التحديد.



واضح أن شكل الانتشار هو وفق خط مستقيم.

x	y	xy	x^2	y^2
0	2	0	0	4
6	3	18	36	9
15	4	60	225	16
20	6	120	400	36
35	10	350	1225	100
40	12	480	1600	144
50	9	450	2500	81
55	16	880	3025	256
$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$
221	62	2358	9011	646

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

$$= \frac{8(2358) - 221(62)}{\sqrt{\left[8(9011) - (221)^2\right]\left[8(646) - (62)^2\right]}}$$

$$= 0.93$$

وهو ارتباط طردي قوي.

لإيجاد معادلة انحدار y على x، تعرف على أنها على الصورة :

$$y = a + b\,x$$

حيث

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

$$b = \frac{8(2358) - 221(62)}{8(9011) - (221)^2} = 0.222$$

ويكون

$$a = \overline{y} - b\,\overline{x}$$

حيث

$$\overline{x} = \frac{\sum x}{n} = \frac{221}{8} = 27.625$$

$$\overline{y} = \frac{\sum y}{n} = \frac{62}{8} = 7.75$$

$$a = 7.75 - (.222)(27.625) = 1.61725$$

ولهذا تكون معادلة إنحدار y على x هي

$$y = 1.617 + .222\,x$$

عدد أيام الغياب عندما تكون x = 25 هي

$$y(25) = 1.617 + .222\,(25) = 7.167$$

الخطأ المعياري للتقدير

$$S_{y.x} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n}}$$

$$= \sqrt{\frac{646 - 1.617(62) - .222(2358)}{8}} = 1.67$$

ويكون معامل التحديد
$$r^2 = \frac{b^2 \left(\sum x^2 - n \overline{x}^2 \right)}{\sum y^2 - n \overline{y}^2}$$

$$= \frac{(.222)^2 \left(9011 - 8(27.625)^2\right)}{646 - 8(7.75)^2} = 0.8657$$

أي أن 86.57% من التفسير يفسره الانحدار وحوالي 13.43% لا تفسير له.

مثال : يمثل الجدول التالي كمية الأدوية المستهلكة في إحـــدى المحافظـــات بملايين الجنيهات خلال عدة سنوات

السنوات (X)	1	2	3	4	5	6	7	8	9	10	11	12	1
كمية الأدوية (Y)	16.2	19.3	21.07	22	23.7	24.5	26	26.42	27.26	29	28.6	29.6	29

(أ) باستخدام طريقة المربعات الصغرى أوجد معادلة أفضل نموذج يمثل البيانات السابقة على الصورة : $y = a x^b$

(ب) أوجد الخطأ المعياري للتقدير.

(جـ) أوجد معامل التحديد واستنتج معامل الارتباط.

الحل: نفرض أن البيانات السابقة يمثلها النموذج $y = a x^b$ وبأخذ لوغاريتم الطرفين نحصل على :

$$\ln y = \ln a + b \ln x$$

وبفرض أن $u = \ln y ,\ v = \ln x ,\ \alpha = \ln a ,\ \beta = b$

فنحصل على $u = \alpha + \beta v$

وهي معادلة خط مستقيم يكون فيها
$$\beta = \frac{n \sum u v - \left(\sum u\right)\left(\sum v\right)}{n \sum v^2 - \left(\sum v\right)^2}$$

$$\alpha = \overline{u} - \beta \overline{v}$$

ويمكن أيضا الحصول على β ، α مباشرة من الآلة الحاسبة فنحصل على

$$\alpha = 2.7817 \ , \ \beta = 0.23897$$

وعليه يكون

$$a = e^{\alpha} = e^{2.7817} = 16.147$$

$$b = \beta = 0.24$$

ويكون النموذج المطلوب هو : $y = (16.147)x^{0.24}$

x	y	u = ln y	v = ln x
1	16.2	2.78501	0
2	19.3	2.96011	0.69315
3	21.07	3.04785	1.09861
4	22	3.09104	1.38629
5	23.7	3.16548	1.60438
6	24.5	3.17805	1.79176
7	26	3.25810	1.94591
8	26.42	3.27412	2.07944
9	27.26	3.30542	2.19773
10	29	3.36730	2.302585
11	28.6	3.35391	2.397895
12	29.2	3.37420	2.48491
13	29.7	3.39115	2.56495

(ب) الخطأ المعياري للتقدير هو $S_{y.x} = \sqrt{\dfrac{\sum (y - \hat{y})^2}{n}}$

ونحصل عليه بعد تكوين الجدول التالي :

x	y	$\hat{y}$	$(y-\hat{y})^2$	$(y-\overline{y})^2$
1	16.2	$(16.147)(1)^{.24} = 16.147$	.003	74.65
2	19.3	$(16.147)(2)^{.24} = 19.069$	.053	30.69
3	21.07	$(16.147)(3)^{.24} = 21.018$	.003	14.21

x	y	$\hat{y}$	$(y-\hat{y})^2$	$(y-\overline{y})^2$
4	22	$(16.147)(4)^{.24} = 22.521$	.271	8.07
5	23.7	$(16.147)(5)^{.24} = 16.147$	.004	1.30
6	24.5	$(16.147)(6)^{.24} = 24.823$	.104	0.12
7	26	$(16.147)(7)^{.24} = 25.758$	059	1.35
8	26.42	$(16.147)(8)^{.24} = 26.597$	.031	2.50
9	27.26	$(16.147)(9)^{.24} = 27.360$	.010	2.50
10	29	$(16.147)(10)^{.24} = 28.060$	.884	17.31
11	28.6	$(16.147)(11)^{.24} = 28.709$	.012	14.14
12	29.2	$(16.147)(12)^{.24} = 29.315$	.013	19.01
13	29.7	$(16.147)(13)^{.24} = 29.884$	.034	23.62
	322.95		1.49	21283

$$S_{y.x} = \sqrt{\frac{\sum (y-\hat{y})}{n}} = \sqrt{\frac{1.49}{11}} = 0.368$$

(جـ) يمكن حساب معامل التحديد كالآتي :

$$r^2 = 1 - \frac{\sum (y-\hat{y})^2}{\sum (y-\overline{y})^2}$$

$$r^2 = 1 - \frac{1.49}{212.83} = .993$$

أي أن نسبة التغير المفسر للانحدار هو 99.3% ويكون معامل الارتباط هو

$$r = \sqrt{.993} = .996$$

$$\bar{y}=\frac{\Sigma\, y_i}{n},\bar{x}=\frac{\Sigma\, x_i}{n}$$

2-معامل انحدار x على y

معادله انحدار x على y هي

$$x=\alpha+\beta y$$

المعادلتان القياسيتان هما :

$$\Sigma x_i=n\alpha+\beta\Sigma y_i$$

$$\Sigma x_i y_i=\alpha\Sigma y_i+\beta\Sigma y_i^2$$

كما سبق بحلها آنيا ينتج

$$\beta=\frac{\Sigma x_i y_i-\frac{\Sigma x_i\Sigma y_i}{n}}{\Sigma y_i^2-\frac{(\Sigma y_i)^2}{n}}\ldots\ldots\ldots\ldots(2)$$

وذلك بالقسمة على n للمعادلة الأولى :

$$\alpha=\bar{x}-\beta\bar{y}$$

وتسمى β معامل انحدار x علىy .

3-معامل الارتباط من معاملي الانحدار

من المعادلتين (1),(2) نجد أن

$$b\beta = \frac{\left[\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}\right]^2}{\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]\left[\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right]} = r^2(x,y)$$

أي أن حاصل ضرب معاملي انحدار y على x في معامل انحدار x على y نحصل على مربع معامل الارتباط بين المتغيرين.
وهذه الطريقة لحساب معامل الارتباط من معاملي الانحدار وذلك بأخذ الجذر التربيعى لحاصل ضرب معاملى الانحدار وتكون اشاره معامل الارتباط هى نفس اشاره معاملى الانحدار.

$$r(x,y) = \pm\sqrt{b\beta}$$

4-العلاقه بين الارتباط ومعامل انحدار y على x تعطى :-

$$b = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n})} = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{n\sigma_x^2},$$

$$r = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{n\sigma_x \sigma_y} = \frac{\frac{1}{n}(\Sigma xy - \Sigma x \Sigma y/n)}{\sigma_x \sigma_y}$$

بالقسمة ينتج أن $\frac{b}{r}=\frac{\sigma_y}{\sigma_x} \Rightarrow b=r\frac{\sigma_y}{\sigma_x}$ أ $r=b\frac{\sigma_x}{\sigma_y}$

وبالمثل يمكن إيجاد معامل انحدار x على y من معامل الارتباط كما يلي:

$$\beta=\frac{\Sigma x_i y_i-\frac{\Sigma x_i \Sigma y_i}{n}}{n\sigma_y^2}$$

$$\frac{\beta}{r}=\frac{\sigma_x}{\sigma_y} \Rightarrow \beta=r\frac{\sigma_x}{\sigma_y}$$ أ $r=\beta\frac{\sigma_y}{\sigma_x}$

<u>ملحوظة:</u> عند أيجاد معامل الارتباط عند طريق ضرب معاملي الانحدار نكر ! أن أشارة معاملي الارتباط هي اشاره معاملى الانحدار وذلك لأن $b=r\frac{\sigma_y}{\sigma_x}, \beta=r\frac{\sigma_x}{\sigma_y}$

وكلا من σ_x, σ_y كميات موجبه فان r تنتج نفس اشاره كلا من β, b:

ونلك لأنه إذا كان الانحدار سالب فان الارتباط بين المتغيرين عكسيا .

مثال (٥)

أوجد معامل الارتباط للبيانات المذكورة في المثال(٢) صـ١١٩

الحل:

y=0.545+0.636x

x=-0.5+1.5y

فنجد أن b=0.636 , β=1.5

فان معامل الارتباط المطلوب هو

$$r=\sqrt{b\beta}=\sqrt{0.636(1.5)}=0.9767$$

ارتباط طردي قوى .

5-نقطه تقاطع خطى الانحدار

تبين في السابق

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

$$\alpha=\bar{x}-\beta\bar{y}$$

بالتعويض في المعادلة (انحدار y على x) على الصورة

$$y=(\bar{y}-b\bar{x})+bx$$

أي

$$y-\bar{y}=b(x-\bar{x})$$

∴خط الانحدار y على x يمر بالنقطة $(\bar{x},\bar{y})$وبالمثل يمر أيضا بخط انحدار x على y .

وعلى ذلك فخطا الانحدار يتقاطعان في إحداثياتها هي $(\bar{x},\bar{y})$

ويمكن اختبار ذلك للمثال السابق (يترك كتمرين للطلاب)

6-الخطأ المعياري للتقدير :

إذا أخذنا y_{est} للتعبير عن قيمه y لقيمه x المقدرة من المعادلة

$$y = a + bx$$

لخط انحدار y على x فأننا نسمى الكميه

$$S_{Y,X} = \sqrt{\frac{\Sigma(y - y_{est})^2}{n}}$$

يسمى الخطأ المعياري لتقدير y على x

وبالمثل الخطأ المعياري x على y
$$S_{x,y} = \sqrt{\frac{\Sigma(x - x_{est})^2}{n}}$$

عموما نلاحظ أن $S_{x,y} \neq S_{y,x}$

ويمكن تقدير قيمه
$$S^2_{y,x} = \frac{\Sigma y_i^2 - a\Sigma y_i - b\Sigma x_i y_i}{n}$$

ولإثبات ذلك نجد أن
$$\Sigma(y - y_{est})^2 = \Sigma(y_i - a - bx_i)^2$$
$$= \Sigma y_i(y_i - a - bx_i) - a\Sigma(y_i - a - bx_i) - b\Sigma x_i(y_i - a - bx_i)$$
$$= \Sigma y_i^2 - a\Sigma y_i - b\Sigma x_i y_i$$

من المعادلات القياسية لخط انحدار y على x كانت

$$\sum y_i = na + b\sum x_i \Rightarrow \Sigma(y_i - a - bx_i) = 0$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 \Rightarrow \sum (x_i y_i - a x_i - b x_i^2) = 0$$

وهما الحدان الأخيرين في المتساوية السابقة

وبالمثل

$$S^2_{x,y} = \frac{\sum x_i^2 - \alpha\sum x_i - \beta\sum x_i y_i}{n}$$

<u>مثال (٦)</u>

أحسب الخطأ المعياري للمثال (١٠)

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9
y_{est}	1.18	2.453	3.1	4.361	5.633	6.27	7.54	9.45

فرق الإحداثيات الصادية التجريبي y_i والافتراض y_{est}

$$\left(\text{الخطأ} \qquad \pm e_i = y_i - y_{est}\right)$$

$$S_{y,x} = \sqrt{\frac{\sum y_i^2 - a\sum y_i - b\sum x_i y_i}{n}}$$

$$= \sqrt{\frac{256 - 0.545(40) + 0.636(364)}{8}}$$

$$= 0.5805$$

ونلاحظ أن الصف الثالث أمكن إيجاده بالتعويض فى المعادلة :

معادلات غير خطية يمكن تحويلها الى معادلات خطية :

من أهم ما يقابل الباحث معرفة نوع المعادلة الرياضية التي تربط بين المتغيرات حتى يمكنه على هداها استخدام التحليل المناسب لإيجاد قيم الثوابت لها. وكما ذكرنا سابقا فإن على الباحث أن يرسم العلاقة بين المتغيرات فيما أسميناه بالانتشار ومن هذا يلاحظ أن الاتجاه العام للعلاقة بين المتغيرات، فإذا كانت هذه العلاقة معروفة لديه استطاع فرض المعادلة، وإلا فإنه في المعتاد ما يستخدم تحويلات على إحدى المتغيرين أو كليهما لتصبح العلاقة بينهما معروفة لديه.

هناك الكثير من المعادلات الغير خطية يمكن تحويلها باستخدام تحويل مناسب إلى خط مستقيم أو قطع مكافئ وسنذكر منها على سبيل المثال القائمة الآتية :

المعادلة الغير خطية	لو أخذنا التحويل	فإننا نحصل على
$y = \frac{1}{a + b x^2}$	$u = \frac{1}{y}, v = x^2$	$u = a + b v$
$y = \frac{1}{a + b \log x}$	$u = \frac{1}{y}, v = \log x$	$u = a + b v$
$y = a + b (\log x)^2$	$v = (\log x)^2$	$y = a + b v$
$y = a + b \sqrt{x}$	$v = \sqrt{x}$	$y = a + b v$
$y = a + b(\log x) + c (\log x)^2$	$v = \log x$	$y = a + b v + c v^2$

المعادلة الغير خطية	لو أخذنا التحويل	فإننا نحصل على
$y=\frac{1}{a+b(\log x)^2}$	$u=\frac{1}{y},\ v=(\log x)^2$	$u=a+b\,v$
$y=\frac{1}{a+b\sqrt{x}}$	$u=\frac{1}{y},\ v=\sqrt{x}$	$u=a+b\,v$
$y=\frac{1}{a+b(\log x)+c(\log x)^2}$	$u=\frac{1}{y},\ v=\log x$	$u=a+b\,v+c\,v^2$

مثال9: يمثل الجدول الآتي القيم التجريبية للمتغير x المقابلة للمتغير y، فإذا ارتبط المتغيران x, y بعلاقة على الصورة $y=\frac{1}{a+b\log x}$ حيث a, b ثابتان.

مستخدما طريقة المربعات الصغرى أوجد أحسن قيم لكل من a, b ثم قدر قيمة y عندما $x=5$.

x	0.1	1	10	100
y	1	1/2	1/3	1/4

الحل:

$$y=\frac{1}{a+b\log x}$$

بأخذ مقلوب الطرفين نحصل على :

$$\frac{1}{y}=a+b\log x$$

بفرض أن $u=\frac{1}{y},\ v=\log x$ نحصل على : $u=a+b\,v$

وهذه معادلة خط مستقيم

المعادلات الطبيعية للخط المستقيم هي :

$$\sum u = n a + b \sum v$$
$$\sum u v = a \sum v + b \sum v^2$$

نكون الجدول الآتي :

x	y	$v = \log x$	$u = \frac{1}{y}$	u v	v^2
0.1	1	-1	1	-1	1
1	1/2	0	2	0	0
10	1/3	1	3	3	1
100	1/4	2	4	8	4
		$\sum v$	$\sum u$	$\sum u v$	$\sum v^2$
		2	10	10	6

وبالتعويض من الجدول السابق في المعادلات الطبيعية نحصل على :

$$10 = 4a + 2b$$
$$10 = 2a + 6b$$

وبحل المعادلتين السابقتين آنيا نحصل على : $a = 2 , b = 1$

وبالتالي تكون العلاقة المطلوبة على الصورة :

$$y = \frac{1}{2 + \log x}$$

وعندما $x = 5$ نحصل على : $y(5) = \frac{1}{2 + \log 5} = 0.37$

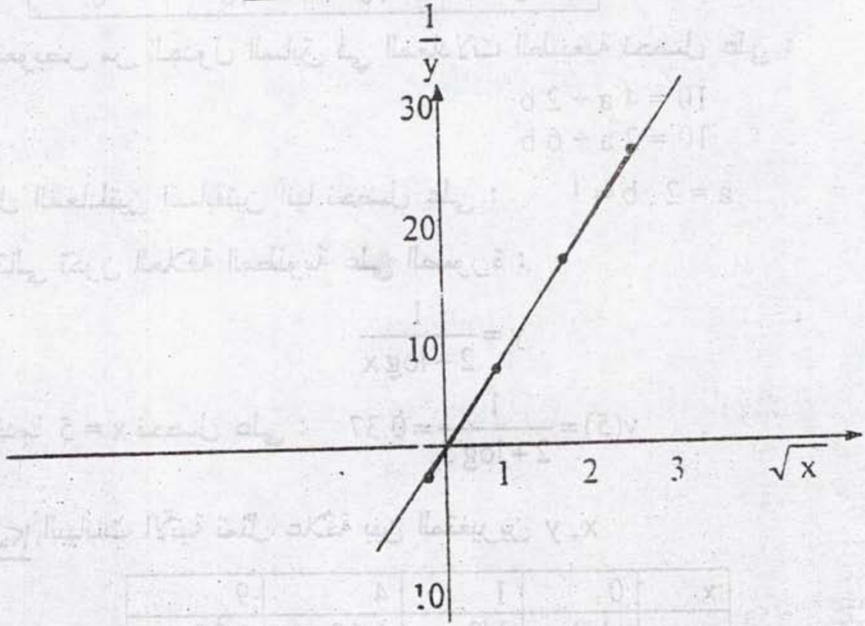
مثال٢: البيانات الآتية تمثل علاقة بين المتغيرين x, y

x	0	1	4	9
y	-1/2	1/8	1/18	1/28

والمطلوب : (أ) ارسم شكل الانتشار بين $\sqrt{x}$, $\frac{1}{y}$ ثم اقترح النموذج المناسب لمثل هذه البيانات.

(ب) مستخدما طريقة المربعات الصغرى أوجد النموذج المناسب المقترح في (أ).

x	y	$v=\sqrt{x}$	$u=\frac{1}{y}$	u v	v^2
0	-1/2	0	-2	0	0
1	1/8	1	8	8	1
4	1/18	2	18	36	4
9	1/28	3	28	84	9
		$\sum v$	$\sum u$	$\sum uv$	$\sum v^2$
		6	52	128	14



من شكل الانتشار واضح أن العلاقة بين $\sqrt{x}, \frac{1}{y}$ هي علاقة خط مستقيم

وبالتالي يكون النموذج المناسب هو : $y = \frac{1}{a + b\sqrt{x}}$

وبأخذ مقلوب الطرفين نحصل على :

$$\frac{1}{y} = a + b\sqrt{x}$$

وبأخذ التعويض $u = \frac{1}{y}, v = \sqrt{x}$ نحصل على : $u = a + b\,v$

وبالتالي تكون المعادلات الطبيعية هي :

$$\sum u = n\,a + b\sum v$$
$$\sum u\,v = a\sum v + b\sum v^2$$

وبالتعويض عن قيم المجاميع من الجدول السابق نحصل على :

$$52 = 4\,a + 6\,b$$
$$128 = 6\,a + 14\,b$$

وبحل المعادلتين السابقتين آنيا نحصل على $a = -2, b = 10$

ومنه يكون النموذج هو : $y = \frac{1}{-2 + 10\sqrt{x}}$

مثال ١٢: البيانات الآتية تمثل علاقة بين المتغيرين x, y على الصورة :

$$y = \frac{1}{a + b(\log x) + c(\log x)^2}$$

x	0.1	1	10	100
y	1	1	1/3	1/7

والمطلوب :

(أ) ارسم شكل الانتشار بين $\frac{1}{y}, \log x$.

(ب) مستخدما طريقة مبدأ المربعات الصغرى أوجد النموذج المقترح

تمارين (٢)

١- أوجد أنسب خط مستقيم يمثل البيانات الآتية :

x	78	65	63	55	53	24	43	46	48	55
y	55	61	50	53	38	31	35	35	47	45

٢- أوجد معادلة أنسب قطع مكافئ للبيانات الآتية التى تمثل سلسلة زمنية لإحدى الظواهر إذا علمت أن هذه البيانات يمكن تقريبها بقطع مكافئ ثم أوجد قيمة الظواهر فى سنة 1975 وفى سنة 2000. في سنة 2010 ؟

t	1952	1953	1954	1955	1956	1957	1958	1959	1960
y_i	2	3	4	9	16	30	38	50	60

٣- أ- أوجد خط المربعات الصغرى الذى يقرب البيانات الآتية

ب- أوجد قيمة y عندما x = 5 وعندما x = 12 ثم أوجد معامل الارتباط بينهما

X	3	5	6	8	9	11
y	2	3	4	6	5	8

٤- يمثل الجدول الآتى الدرجات النهائية فى الرياضيات والإحصاء لعشرة من الطلاب المختارين عشوائياً من عدد كبير من الطلاب

أ- ارسم الانتشار ب- أوجد خط المربعات الصغرى للبيانات

ج- إذا حصل طالب على 75 فى الرياضيات وأخر على 95 فيها فما هى درجاتهم فى الإحصاء

X_i	75	80	93	65	87	71	98	68	68	84
y_i	82	78	86	72	91	80	95	72	89	74

الباب الخامس

١ - المتغير العشوائي Random Variables

تعريف المتغير العشوائي: هو متغير يأخذ قيما معتمدا على الصدفة وبناءا على هذا التعريف يمكننا عرض الامثله التاليه للمتغير العشوائي

١- عدد مرات ظهور الصوره عند القاء قطعة نقود خمس مرات مثلا

٢- عدد المكالمات الى يتلقاها الشخص فى منزله

٣- وقت الانتظار الذي يقضيه الشخص في أحد البنوك لصرف شيك

٤- مده صلاحية ثلاجة أو كمبيوتر.

وبواسطة هذه الا مثله يمكن إعطاء التعريف الرياضي العشوائي التالي: هو داله X نطاقها فراغ العينه S، ونطاقها المصاحب الأعداد الحقيقية R

$$X : S \longrightarrow R$$

فمثلا فى تجربة رمى ثلاث قطع نقود وكان المتغير العشوائي X يمثل عدد الصور التى تظهر يكون

$X=3$: (HHH)

$X=2$: (HHT , HTH , THH)

$X=1$: (HTT ,THT , TTH)

$X=0$: (TTT)

فى هذا المثال نلاحظ ان المتغير العشوائي يأخذ قيما متقطعة $X=1, 2, 3, 0$ ولذلك يسمى متغير عشوائي متقطع Discrete Random Variable أما المثال الرابع فان المتغير العشوائي الذي يمثل مدة صلاحية الثلاجة او الكمبيوتر يأخذ كل القيم الحقيقية الممكنة حتى يتلف الجهاز ولذلك يسمى متغير عشوائي متصل Continuous Random Variables ويتحدد المتغير العشوائي تماما اذا عرفت القيم التي يأخذها بالاضافه لقيم الاحتمالات المناظرة . وتسمى الاحتمالات المناظرة بالوصف الاحتمالي للمتغير. ويمكن التعبير عن هذا الوصف باستخدام دوال او توزيعات احتمالية .

٢- دالة التوزيع للمتغير العشوائي Distribution Function

كل متغير عشوائى X يناظر دالة توزيع $F_X(x)$ بحيث . $F_X(x) = P[X \le x]$

وللسهوله نكتب $F(x) = P[X \le x]$ ودالة التوزيع معرفه سواء كان المتغير متقطع او مستمر

وهى دائما تحقق $0 \le F(x) \le 1$ لكونها احتمال الحادثه $[X \le x]$ وسوف نتعرض لكل من

نوعى المتغير العشوائى (المتقطع والمتصل)

١. متغير عشوائي متقطع Discrete Random Variable

بفرض ان المتغير العشوائي X يأخذ القيم المنفصلة التى على الصورة

$x_1 \; , \; x_2 \; , x_3 \; , \; , x_i ... \; .... \; x_n$

فان توزيع الاحتمالات: $P_k = P[X = x_k]$

حيث $\sum P_k = 1$ و $0 \le P_k \le 1$ ويمكن كتابة كل احتمال

مقرون بالمتغير العشوائي كما فى الجدول التالى

X	x_1	x_2	....	X_n	المجموع
P	P_1	P_2	....	P_n	1

حساب دالة التوزيع للمتغير العشوائى المتقطع باستخدام توزيع الاحتمال

$$F(x) = P[X \le x]$$

حيث ان :
$$= P[X = x_1] + P[X = x_2] + ... P[X = x_j]$$

$$[X \le x_1, x_2, ..., x_j]$$

حيث
$$F(x) = \sum_{x_1}^{x_j} P_k$$

٢- بعض التوزيعات الإحصائية

التوزيعات الاحتمالية :

إذا ما اعتبرنا زهرة نرد وهى تحتوى على ستة أوجه وعلى كل وجه عدد معين من النقاط من 1 إلى 6 فإن احتمالات ظهور النقط المختلفة وهى

X	1	2	3	4	5	6	Σ
P(X)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

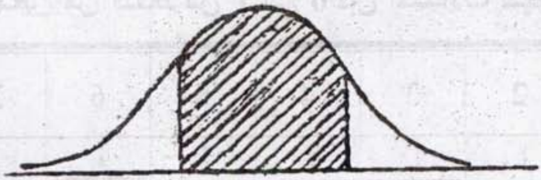
ويسمى هذا الجدول بجدول التوزيع الإحتمالى لعدد النقط X على زهرة نرد وإذا كانت لدينا زهرتى نرد فإن احتمالات ظهور النقط على الزهرتين وهى

X	2	3	4	5	6	7	8	9	10	11	12	Σ
P(X)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

وفى الحالتين السابقتين لنا نعتبر عدد النقط على زهرة النرد وهو عدد صحيح وكذلك فى حالة رمى قطعة العملة فكذا نحصل على عدد صحيح لظهور الصورة أو الكتابة فمثلاً من غير المعقول أن يكون عدد مرات ظهور الصورة 65 - 1 أو 7 - 5 ، ... إلخ وهذا دائماً صحيح فى حالة الأعداد .

وإذا كنا نتحدث عن ظاهرة معينة فإنه من الممكن حدوث أى قيمة مهما كانت كسرية ويسمى هذا التوزيع بالتوزيع المستمر (Continuous Distribution) .

وشكل التوزيع فى هذه الحالة يكون منحنى ممتداً محوره الأفقى يمثل المقاييس (بينما فى التوزيع فى التوزيع غير المستمر يمثل العدد) ويمثل المحور الرأسى الكثافة الاحتمالية (الاحتمال) فمثلاً أطوال مجموعة من الذكور فى نفس العمر يمثلها التوزيع الحالى



واحتمال اختيار شخص عشوائياً منحصر طوله بين 152.4 Cm إلى 182.9 Cm وهو عبارة عن النسبة بين المساحة المظللة والمساحة الكلية

وإذا اعتبرنا أن المساحة الكلية تحت المنحنى = 1 فإن احتمال اختيار شخص عشوائياً ينحصر بين 152.4 ، 182.9 سم هو عبارة عن المساحة تحت المنحنى المحصورة بين النقطتين .

التوزيعات الاحتمالية هامة للاحصائى إذا أنها تمكنه من استخدام عينه للحصول على بعض الاستنتاجات عن معالم المجتمع وهو ما سندرسه فيما بعد .

وسنعتبر التوزيعات الآتية :

١- التوزيع الطبيعى Normal Distribution

٢- توزيع ذى الحدين Binomial Distribution

٣- توزيع بواسون Poisson Distribution

وبالإضـــافة إلـــى الـــتوزيعات المتصـــلة ، ت ، مـــربع كا ى ، توزيع ف (T, X^2 - F-) وهـــى توزيعـــات كثـــيرة الاستخدام فى الإحصاء والتوزيع الأول توزيع متصل بينما التوزيعان الأخريان غير مستمران .

وســـنرى فـــيما بعد أن التوزيع الطبيعى هو أكثر التوزيعات أهمية فأكثرها شيوعاً واستخداماً وكما سنرى فيما بعد أنه لشروط معينة وإذا زاد حجم العينة فإن التوزيعين الأخريين يقتربان من التوزيع الطبيعى .

التوقع للمتغير العشوائي X Expected variables

المتغير العشوائي

X	X_1	X_2	----	X_n	Σ
P	P_1	P_2	---	P_n	1

حساب توقع المتغير العشوائي X.

$$EX = X_1P_1 + X_2P_2 + \cdots + X_nP_n$$
$$= \Sigma XP$$

وكذلك توقع المتغير X^2 هو

$$EX^2 = X_1^2P_1 + X_2^2P_2 + \cdots + X_n^2P_n$$
$$= \Sigma X^2P$$

مثال ٢ :

اوجد قيمه التوقع EX للتوزيع

X	50	-10	Σ
P	0.6	0.4	1

لحسب التوقع لمكسب ذلك البائع

الحـــل:

$P[X=50] = P[$ الجو ممطر $] = 0.6$

$P[X=-10] = P[$ الجو غير ممطر $] = 0.4$

X	50	-10	المجموع
P	0.6	0.4	1

$E(x) = 50(0.6) + (-10)(0.4) = 30 - 4 = 26$

مثـــــال ٣ : اذاكانت القراءه التى تظهر عند القاء زهرة النرد تمثل المتغير العشوائى X اوجد $E(x)$ و $E(x^2)$,

الحـــل:

X	1	2	3	4	5	6	المجموع
P	1/6	1/6	1/6	1/6	1/6	1/6	1

$$E(x) = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = \frac{21}{6}$$

x	P_x	$x P_x$	X^2P_x
1	0.166667	0.166667	0.166667
2	0.166667	0.333333	0.666667
3	0.166667	0.5	1.5
4	0.166667	0.666667	2.666667
5	0.166667	0.833333	4.166667
6	0.166667	1	6
	المجموع	3.5	15.16667

$E(x^2)=16.16667$

$E(x)$

Variance of random variables التباين للمتغير العشوائي

التباين للمتغير العشوائي X الذى له التوزيع

X	x_1	x_2	....	X_n	المجموع
P	P_1	P_2	....	P_n	1

يعطى بالعلاقه

$$\sigma^2 = E[(x-\mu)^2]$$

$$= E(x^2) - \mu^2$$

σ تسمى الانحراف المعيارى

مثـــــال ٣

اذا كان X متغير عشوائى داله التوزيع معرفه بالعلاقه

$$F(x) = \begin{cases} 0 & x < 0 \\ .13 & 0 \le x < 1 \\ .27 & 1 \le x < 2 \\ .53 & 2 \le x < 3 \\ .84 & 3 \le x < 4 \\ .92 & 4 \le x < 5 \\ 1 & x \ge 5 \end{cases}$$

١ ـ اوجد التوزيع الاحتمالى للمتغير X

٢ ـ اوجد قيم الاحتمالات التاليه $P[2<x\le 4]$, $P[x>3]$

٣ ـ اوجد قيمة التوقع الرياضى و التباين

الحـــــل :

من داله التوزيع يمكن معرفة القيم الممكنة و الاحتمالات المناظرة

X	0	1	2	3	4	5	المجموع
P	0.13	0.14	0.26	0.31	0.08	0.08	1

$P[x>3] = P[x=4]+P[x=5]$

$= 0.08 + 0.08 = 0.16$

$P[2<x\leq 4] = P[x=3]+P[x=4]$

$= 0.31 + 0.08 = 0.39$

x	P_x	$x P_x$	X^2P_x
0	0.13	0	0
1	0.14	0.14	0.14
2	0.26	0.52	1.4
3	0.31	0.93	2.79
4	0.08	0.24	0.96
5	0.08	0.4	2
	المجموع	2.31	7.25

$E(x^2)=7.25$

$\mu=2.31$

$\sigma^2 = E(x^2) - \mu^2$

$= 7.25-(2.31)^2 = 1.914$

بعض التوزيعات المتقطعة العامة

1- توزيع ذات الحدين Binomial Distribution

محاولات برنولى *Bernoulli Trials* هي سلسله من المحاولات المكررة والتي تحقق الشروط التالية :

- كل محاوله لها نتيجتان فقط هما نجاح (s) او فشل (f)
- كل المحاولات مستقله
- احتمال النجاح ثابت القيمه فى كل المحاولات و كذلك احتمال الفشل ثابت القيمه

حيث $P[f]=q$, $P[s]=p$ وبذلك يكون $p+q=1$

يتحدد تماما توزيع ذات الحدين اذا عرفنا متغير عشوائى X وعدد مرات النجاح خلال n من محاولات برنولى ولذلك تكون قيم المتغير العشوائى هى $1, 2, 3, \ldots, n$

ويعطى احتمال الحصول على k من مرات النجاح من ضمن n من محاولات برنولى بالعلاقه

$$P_k = P[x=k] = C_k^n \; p^k \; q^{n-k} \qquad , k = 0 , 1 , 2 , \dots , n$$

اما القيمه المتوقعه (المتوسط او التوقع الرياضى) فيعطى بالعلاقه:

$$\mu = E(x) = np$$

اما التباين فيعطى بالعلاقه:

$$\sigma^2 = npq$$

مثـــــــال ٤

اذا كان احتمال سحب عينه معيبه من نتاج احد المصانع 0.2 . فى اختبار ضبط الجوده يتم سحب 10 عينات

i- ما هو احتمال الحصول على ثلاث عينات معيبه

ii- ما هو احتمال الحصول على عينتين على الاقل معيبتان

iii- ما هو عدد العينات المعيبه المتوقع سحبها

iv- اوجد قيمة التباين والانحراف المعيارى

الحـــــــل:

$N=10 \quad , \quad p=0.2 \quad , \quad q=0.8$

$$(i) \qquad k=3 \quad P_k = P[x=3] = C_3^{10} \; (0.2)^3 \; (0.8)^7$$

$$= (120) \;\; (0.008) \; (0.209)$$

(ii)

$$P[x \geq 2] = p[x=2] + p[x=3] + p[x=4]$$
$$+ p[x=5] + p[x=6] + p[x=7] + p[x=8]$$
$$+ p[x=9] + p[x=10]$$
$$= 1 - (p[x=0] + p[x=1])$$

$$p[x \ge 2] = 1 - \{C_0^{10}(0.2)^0(0.8)^{10} + C_1^{10}(0.2)^1(0.8)^9\}$$
$$= 1 - \{(0.8)^{10} + 2(0.8)^9\}$$

(iii) $\mu = E(x) = np = 10 \times 0.2 = 2$

(iv) $\sigma^2 = npq \quad = 10(0.2)(0.8) \quad = 1.6 \rightarrow \quad \sigma = \sqrt{1.6}$

٣- توزيع بسواسون **Poisson Distribution**

وتوزيع بولسون هو توزيع احتمالات تحقق الحوادث النادره. والحوادث النادره هى التى تكون احتمالات تحققها ضئيله جداولمثلة هذا النوع من الحوادث :

١- عدد مرات النجاح فى عدد كبير من محلولات برنولى

٢- عدد حوادث السيارات فى الشهر داخل احدى المدن

٣-عدد الاخطاء للطبعيه فى احد الجرائد اليوميه.

وفى كل من حالات الحوادث النادره لذا كان عدد محلولات برنولى n والحتمال النجاح p

فان قيمة متوسط التوزيع يعطى بالعلاقه $\qquad E(x) = \lambda = np$

ويعرف التوزيع الاحتمالى على الصوره $\qquad P[X = k] = \dfrac{\lambda^k}{k!}e^{-\lambda}$

حيث $\qquad \lambda > 0 \quad , k = 0,1,2,\dots,n$

وايضا يعطى التباين لهذا التوزيع بالعلاقه $\qquad \sigma^2 = \lambda = np$

وهنا يلاحظ تساوى متوسط التوزيع مساوى للتباين . $\{e = 2.7183 \quad , 0! = 1 \quad , 1! = 1\}$

مثـــــال ١:

اذا كان من بين كل 100 وحده منتجه باحد مصانع الزجاج توجد وحده واحده معيبه .فما هو احتمال ان 30 وحده من انتاج هذا المصنع لا يكون من بينها اى وحده معيبه .ثم اوجد احتمال ان يكون من بين هذه الوحدات المنتجه وحده واحده معيبه.

الحـــــل:

$$E(x) = \lambda = np = 30 \times 0.01 = 0.3$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=0]=\frac{(0.3)^0}{0!}e^{-0.3}=e^{-0.3}=0.74$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=1]=\frac{(0.3)^1}{1!}e^{-0.3}=e^{-0.3}(0.3)=0.22$$

مثـــــال ٢:

تبين فى احد مصانع انتاج مسامير البرمه ان متوسط المسامير المعيبه هو 0.2 ان توزيع المسامير المعيبه يتبع توزيع بواسون . فما هو احتمال ان صندوقا من المسامير المنتجه يحتوى على مسمارين معيبين.

الحـــــل:

$$E(x)=\lambda=0.2$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=2]=\frac{(0.2)^2}{2!}e^{-0.2}$$

$$\Rightarrow P[X=2]=2.7183^{-0.2}\times\frac{(0.2)^2}{2!}=0.01637$$

مثـــــال ٣:

كانت المواصفات الموضوعه للانتاج بشركة لانتاج اللمبات الكهربائيه انه يوجد من بين كل 1000 لمبه منتجه 60لمبه معيبه . اخذت عينه مكونه من 50 لمبه احسب الاحتمالات التاليه :

أ- ان يكون من بين العينه لمبه واحد معيبه

أأ- ان يكون من بين العينه عدد 2لمبه معيبه

أأأ- ان يكون من بين العينه اكثر من 3لمبات معيبه.

الحـــــل :

$$E(x)=\lambda=np=50\times0.06=3$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=1]=\frac{(3)^1}{1!}e^{-3}=0.14961$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=2]=\frac{(3)^2}{2!}e^{-3}=0.22442$$

$$p[x>3]=1-(p[x=1]+p[x=2]+p[x=3]+P[X=0])$$

$$p[x=3]=\frac{3^3}{3!}\times 2.7183^{-3}=0.2242 \quad , \quad P[X=0]=\frac{3^0}{0!}e^{-3}=0.0498$$

$$\Rightarrow p[x>3]=1-(0.14961+0.22442+0.22442+0.0498)$$
$$=1-0.64825=0.35175$$

مثـــــال ٤ :

قامت إحدى دور النشر بدراسة الأخطاء الموجودة فى صفحات أحد الكتب تولت نشره فى 1000 صفحه وقد تبين ان توزيع صفحات الكتاب وفقا لعدد الأخطاء فى الصفحة الواحدة كما يلى :

عدد الأخطاء فى الصفحة	عدد الصفحات
0	672
1	166
2	60
3	40
4	30
5	20
6	10
7	2
المجمـــــــوع	1000

فاذا كان X التوزيع يتبع التوزيع البواسونى . فاحسب احتمال الحصول على خطئين على الاقل فى الصفحه الواحده فى كتاب آخر ستقوم بطباعته.

الحـــــل:

x	f	xf
0	672	0
1	166	166
2	60	120
3	40	120
4	30	120
5	20	100
6	10	60
7	2	14
	1000	700

$$\boxed{\lambda=\frac{700}{1000}=0.7}$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda}\Rightarrow P[X=0]=\frac{(0.7)^0}{0!}e^{-0.7}=0.49648$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda}\Rightarrow P[X=1]=\frac{(0.7)^1}{1!}e^{-0.7}=0.34754$$

$$p[x>2]=1-(p[x=0]+p[x=1])$$

$$\Rightarrow p[x>2]=1-(0.49648+0.34754)$$

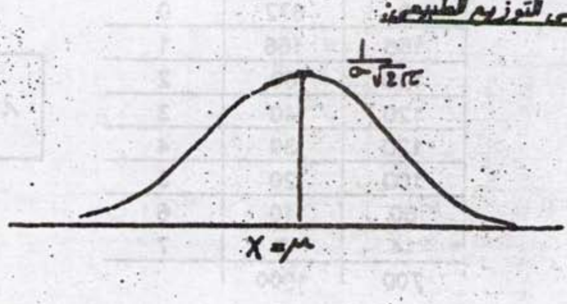
$$=1-0.84402=0.15598$$

بعض التوزيعات المتصلة الهامة

١- التوزيع الطبيعي

التوزيع الطبيعي هو الاداه الاحصائيه التي يمكن عن طريق صفاته تحليل بيانات المتغيرات المتصلة – والمتغير المتصل continuous variables هو المتغير الذي تمكن ان يأخذ جميع القيم بما فيها القيم ذات الكسور داخل المسافة التي يتحرك فيها وهذه المتغيرات ترتبط اكثر ما ترتبط بالظواهر الطبيعية كالأعمار والطوال والأوزان ... . وصفات وخواص التوزيع الطبيعي هى أساس النظرية الاحصائيه وتطبيقاتها فى المشروعات الصناعية. ويعد التوزيع الطبيعى من اهم التوزيعات النظريه . لان كثير من الظواهر الطبيعيه تتبع هذا التوزيع . كما ان توزيع متوسطات العينات لاى مجتمع من المجتمعات يتبع التوزيع الطبيعى مهما كان المجتمع المأخوذة منه العينه .

منحنى التوزيع الطبيعي:



شكل ١

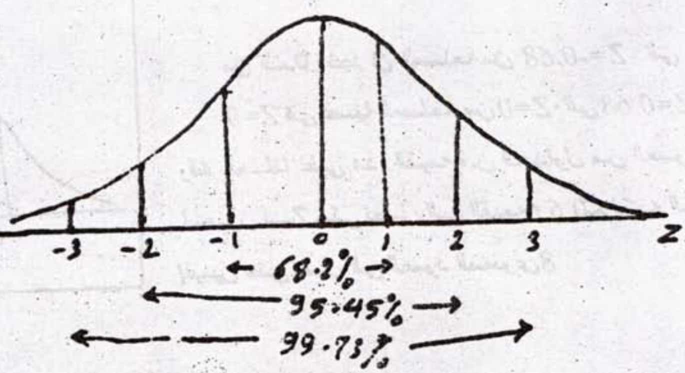
كما في شكل (١) فان منحنى التوزيع الطبيعي هو منحنى متماثل على شكل جرس مقلوب يمتد طرفاه الى ما لانهاية ويقتربان من القاعدة ولكن لا يلتقيان معها .

ودالة التوزيع الاحتمالي للتوزيع الطبيعي تعطى بالعلاقة
$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$
ويرمز له بالرمز : $X \sim N(\mu, \sigma^2)$

حيث μ =الوسط و σ = الانحراف المعياري و $e=2.71828$ & $\pi = 3.14159$

المساحه الكليه بين المنحنى والاحداثى السينى X تساوى الواحد الصحيح . وبهذا فان للمساحه تحت المنحنى بين الاحداثي $x=a$,$x=b$ حيث $a<b$ تمثل احتمال ان تقع X بين a ,b ويعبر عنها كما يلي: $p[a < X < b]$ وعندما نعبر عن بدلالة الوحدات المعيارية $Z = \frac{(X-\mu)}{\sigma}$ فان دالة
$$Z \sim N(0, 1)$$
التوزيع الطبيعى تصبح على الصوره :
$$Y = \frac{1}{\sqrt{2\pi}} e^{\frac{-Z^2}{2}}$$

وفى هذه الحالة يقال ان Z تتوزع توزيعا معتدلا متوسطة الصفر وتباينه الوحدة كما هو فى شكل (٢) .ويظهر فى هذا الشكل ان المساحه الواقعه بين $Z=-1$, $Z=1$ هى 68.27% وبين $Z=2$, $Z=-2$ تكون المساحة 95.45% وبين $Z=-3$, $Z=3$ هى 99.73% من المساحه الكلية والتي تساوى الوحدة .



شكل ٢

ونعتمد على الجدول I فى تحديد المساحات المطلوبه فى كل حاله حيث تعطى الجدولى المساحه تحت المنحنى بين الاحداثى $Z=a$, $Z=0$ حيث a عباره عن عدد موجب كما يمكن استخدام تماثل المنحنى فى تعيين المساحات المعرفة فى الجزء السالب لمحور Z.

مثـــــال ١:

اوجد المساحه تحت منحنى التوزيع الطبيعى فى كل من الحالات التاليه

١- من $Z=0$ الى $Z=1.2$

٢- من $Z=-0.68$ الى $Z=0$

٣- من $Z=-0.46$ الى $Z=2.21$

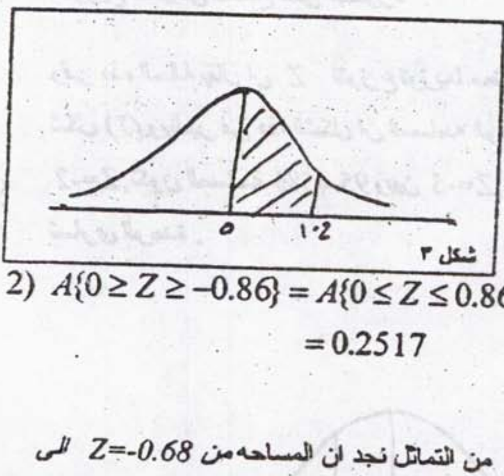
٤- من $Z=0.81$ الى $Z=1.49$

الحـــــــل:

اذا رمزنا للمساحه بالرمز A فان:

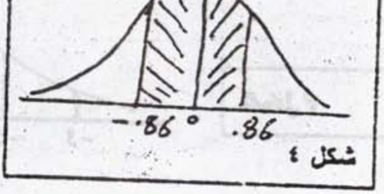
$$1)\ A\{0 \le Z \le 1.2\} = 0.3849$$

وقد حصلنا على هذه القيمه من الجداول من العمود المعنون بـZ حتى نصل الى القيمه 1.2 ثم نتجه الى اليمين حتى نصل الى العمود المعنوى 0



شكل ٣

$$2)\ A\{0 \ge Z \ge -0.86\} = A\{0 \le Z \le 0.86\}$$
$$= 0.2517$$



شكل ٤

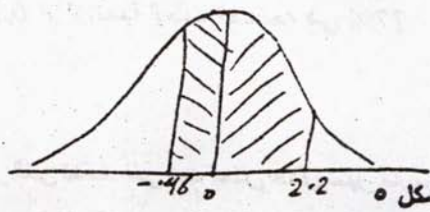
من التماثل نجد ان المساحه من $Z=-0.68$ الى $Z=0$ هى نفسها المساحه من $Z=0$ الى $Z=0.68$ وقد حصلنا على هذه القيمه من الجداول من العمود المعنون بـZ حتى نصل الى القيمه 0.6 ثم نتجه الى اليمين حتى نصل الى العمود المعنوى 8

3) $A\{-0.46 \le Z \le 2.2\} =$

$= A\{0 \le Z \le 0.46\}$

$+ A\{0 \le Z \le 2.2\}$

$= 0.1772 + 0.4864 = 0.6636$



شكل ٥

شكل ٦

4) $A\{0.81 \le Z \le 1.94\} =$

$= A\{0 \le Z \le 1.94\}$

$- A\{0 \le Z \le 0.81\}$

$= 0.4738 - 0.2910$

$= 0.1828$

مثـــــــال٢:

اذا كان متوسط القطر الداخلي في عينه من 200 جلبه مستديره من انتاج ماكينه معينه هو 5.02 mm وانحرافها المعيارى 0.05 mm وكان الهدف من استخدام هذه الجلب يسمح بانحراف في القطر اقصاه من 4.96 الى 5.08 mm وفيما عدا ذلك تعتبر الجلبه معيبه . اوجد النسبه المئويه للجلب التالفه من انتاج هذه الماكينه وذلك بفرض ان الاقطار تتوزع توزيعا طبيعيا .

$X \sim N(5.02,\ 0.05)$

الحـــــــل

$P[4.96 \le X \le 5.08] =$

$P[\frac{4.96 - 5.02}{0.05} \le Z \le \frac{5.08 - 5.02}{0.05}]$

$= P[-1.2 \le Z \le 1.2]$

$= 2P[0 \le Z \le 1.2] = 2 \times 0.3849$

$= 0.7689$

اى ان احتمال ان تكون الجلبه صالحه هو 0.77 او ان نسبة الجلب الصالحه هى 77%

وبالتالى تكون نسبة الجلب التالفه 23%

مثـــــال ٣ :

فى تقرير لهيئة الارصاد الجويه عن الامطار التى تسقط على احد المدن خلال شهر مارس ان كمية المطر تتوزع توزيعا طبيعيا بمتوسط 3.6 بوصه وانحراف معيارى 1.1 بوصه . احسب احتمال ان تكون كمية المطر التى تسقط فى شهر مارس من العام القادم :

١- اقل من 0.7 بوصه

٢- بين 2 و 3 بوصه.

٣- اكبر من 5.3 بوصه

$X \sim N(3.6, 1.1) \rightarrow Z = \frac{X-3.6}{1.1}$

الحـــــل

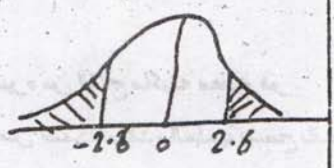
1) $P[x < 0.7] = P[z < \frac{0.7-3.6}{1.1}]$

$= P[Z < -2.6]$

$= P[z > 2.6] = 1 - P[z \le 2.6]$

$= 1 - 0.9953 = 0.0047$

$= 0.7689$



حاول ان ترسم المساحه

2)

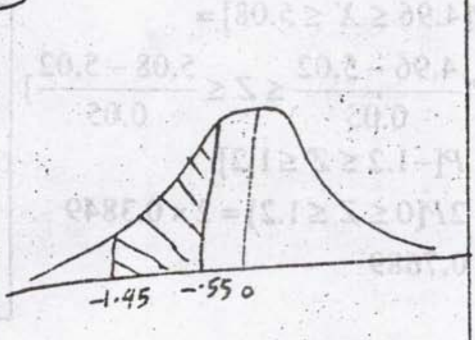
$P[2 \le X \le 3] =$

$P[\frac{2-3.6}{1.1} \le Z \le \frac{3-3.6}{1.1}]$

$= P[-1.45 \le Z \le -0.55]$

$= P[Z \le 1.45] - P[z \le 0.55]$

$= 0.9265 - 0.7088 = 0.2177$



3) $P[X > 5.3] =$

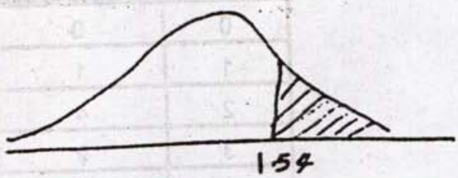
$$= P[Z > \frac{5.3-3.6}{1.1}]$$

$$= P[Z > 1.54]$$

$$= 1 - P[Z \leq 1.54]$$

$$= 1 - 0.9382 = 0.0618$$

حاول ان ترسم المساحة



العلاقه بين توزيع ذات الحدين والتوزيع الطبيعي:

اذا كانت n كبيره وكلا من p,q ليسا قريبين من الصفر فان توزيع ذات الحدين يمكن تقريبه بصوره جيده بالتوزيع الطبيعى ذى المتغير المعيارى $Z = \frac{x-np}{\sqrt{npq}}$ ويصير التقريب اكثر جوده كلما كانت $n \to \infty$ وعمليا اذا كانت $np>5$ يمكن استخدام التقريب .

العلاقه بين توزيع ذات الحدين والتوزيع البواسونى :

اذا كانت n كبيره وكانت $p \to 0$ وبالتالى $q \to 1$ فان توزيع ذات الحدين يمكن تقريبه باستخدام توزيع بواسون ويكون $\lambda = np$

مثـــــال :

الجدول التالى يوضح توزيع عدد الجزيئات فى معدن الذهب الملاحظ من اجهزه بصريه فى فترات زمنيه متساويه كل منها ثانيتين . اوجد التوزيع النظرى لعدد جزيئات الذهب

عدد الجزيئات x_i	0	1	2	3	4	5	6	7
عدد الفترات f_i	381	568	357	175	67	28	5	2

الحـــــــــل :

نحسب اولا الوسط الحسابى والانحراف المعيارى من الجدول كما سبق

x_i	x_i^2	f_i	$x_i f_i$	$X_i^2 f_i$
0	0	381	0	0
1	1	568	568	568
2	4	357	714	1428
3	9	175	525	1575
4	16	67	268	1072
5	25	28	140	700
6	36	5	30	180
7	49	2	14	98
المجمـــــوع		1583	2259	5621

$$\mu = \frac{2259}{1583} = 1.427 \qquad \sigma^2 = 1.514$$

ونظرا لان عدد الجزيئات لابد ان يكون عددا صحيحا غير سالب وان المتوسط والتباين كلاهما متقاربان فى القيمه .لذلك يمكن قانون التوزيع النظرى للمتغير العشوائى X المعبر عن عدد جزيئات الذهب الملاحظ بقانون بواسون حيث $\lambda = \mu = 1.427$ فيصبح قانون التوزيع على الصوره :

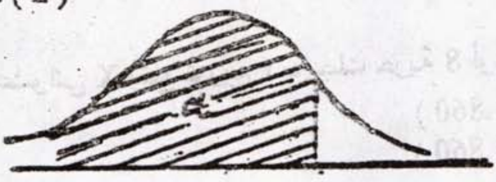
$$P[X=k] = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{(1.247)^k}{k!} e^{-1.427}$$

حيث k فى هذه الحاله تعبر عن عدد جزيئات الذهب المراد حساب احتمال وجودها.

٢ - توزيـــــع (ت)
T - distribution

" أنــه توزيـــع لمتغير عشوائى متصل يشبه إلى حد كبير التوزيع الطبيعى المعيارى " .

$$T = \frac{Z}{\sqrt{Y/n}} \sim t(n)$$



$t_{\alpha,\upsilon}$

$$Z \sim N(0,1)$$
$$Y \sim \chi^2(n)$$

الخواص :

١- له معلمة واحدة هى υ وتسمى درجات الحرية .

٢- التوزيع ليس وحيداً ولكنه عائله من التوزيعات ، ويتحدد شكل التوزيع بمجرد تحديد درجات الحرية υ .

٣- التوزيع متماثل حول المتوسط الحسابى الذى يساوى صفر .

٤- للتوزيع : المتوسط الحسابى = الوسيط = المنوال .

٥- مدى التوزيع يمتد من $-\infty$ إلى $+\infty$.

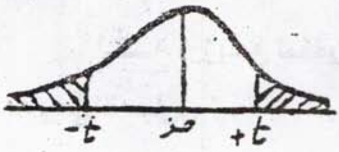
٦- بزيادة درجات الحرية يقترب التوزيع من التوزيع الطبيعى المعيارى

يوضح الجدول T بالملحق قيم المتغيرات والاحتمالات المناظرة لها بحيث أن :

$$p(X \le t_{\alpha,\upsilon}) = \alpha$$

وباعتبار أن التوزيع متماثل فإن

$$P(X < -t_{\alpha,v}) = 1 - p(X < t_{\alpha,v})$$



مثال (١)

متغير عشوائى X يتبع توزيع t بدرجات حرية 8 أوجد :-

a) $p(X < 1.860)$
b) $p\,X < -1.860)$

الحل :

بالرجوع للجدول . وأمام درجات الحرية 8 نجد أن .

a) $p(X < 1.860) = 0.95$

b) $p\,X < -1.860) = 1 - p(X < 1.860)$
$= 1 - 0.95 = 0.05$

مثال (٢) :

متغير X يتبع توزيع T بدرجات حرية 7 أوجد :

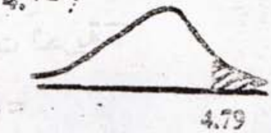
(i) $p(X < -2.36)$ (ii) $p(X > 4.79)$
(iii) $p(-2.36 < X < 4.79)$

الحل :

(i) $p(X < -2.36) = 1 - p(X < 2.36)$
$= 1 - 0.975 = 0.025$

$(ii)\ p(X > 4.79) = 1 - p(X < 4.79)$

$$= 1 - 0.999$$



$$= 0.001$$

$(iii)\ p(-2.36 < X < 4.79)$

$$= p(X < 4.79) - p(X < -2.36)$$

$$= p(X < 4.79) - 1 + p(x < 2.36)$$

$$= 0.999 + 0.975 - 1 = 0.974$$

٢ - توزيع كا٢ χ^2 - Distribution

للمتغير العشوائي X الذي يتبع توزيع طبيعي بمتوسط μ وانحراف معياري σ . أخذت عينة حجمها n وهي $(X_1, X_2, \ldots\ldots, X_n)$

متوسطها $\bar{X} = \frac{\sum X_i}{n}$ وانحرافها المعياري S

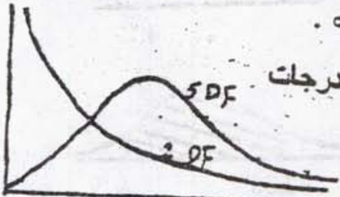
فإن المتغير العشوائي : $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

فإنها تعطي توزيع مربع كاي بدرجات حرية $\upsilon = n - 1$

وتوزيع مربع كاي هو توزيع مستمر وله استخدامات متعددة في الإحصاء

أهم خواصه :

١- له معلمة واحدة υ وتسمى درجات الحرية .



٢- حدى التوزيع يمتد من صفر إلى ∞ .

٣- التوزيع ملتو من اليمين . بزيادة درجات الحرية يميل إلى التماثل .

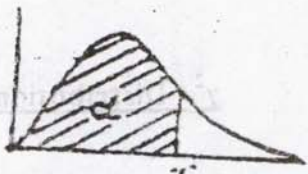
٤- متوسط التوزيع = $\upsilon = n - 1$

$= EX$

٥- تباين التوزيع .

$$V(X) = 2\upsilon = 2(n-1)$$

جدول بالملحق يعرض قيم $\chi^2_{\alpha,\upsilon}$ بحيث $P(X < \chi^2_{\alpha,\upsilon})$



$\chi_{\alpha,\upsilon}$

لدرجات الحرية أقرب من 30 نستخدم تقريب التوزيع الطبيعي

$$\chi^2_{\alpha} = \upsilon\left(1 - \frac{2}{9\upsilon} + Z_{\alpha}\sqrt{\frac{2}{9\upsilon}}\right)^3 \qquad (*)$$

حيث Z_{α} هي قيمة المتغير الطبيعي المعيارى .

مثال (٣) :

متغير X يتبع χ^2 بدرجات حرية 5 أوجد :

(i) $p(X > 11.1)$ (ii) $p(X < 2.67)$

(iii) $p(2.67 < X < 11.1)$.

الحل :

من جدول المساحات نجد أن :

عند $\upsilon = 5$

$$(i)\ p(X < 11.1) = 0.95$$
$$p(X > 11.1) = 1 - 0.95 = 0.05$$
$$(ii)\ p(X < 2.67) = 0.25$$
$$(iii)\ p(2.67 < X < 11.1) = 0.95 - 0.25 = 0.70$$

مثال (٤)

أوجد

$$\chi^2_{70, .975}$$

الحل :

نستخدم الصيغة (*)

$$\chi^2_{70\,.975} = 70\left[1 - \frac{2}{9(70)} + Z_\alpha\sqrt{\frac{2}{9(70)}}\right]^3 = 95.01$$

لاحظ أن القيمة من الجدول هى : 95.0

حيث أن $Z_{0.975} = 1.96$

ويمكن تكوين مربع كاى من العلاقة التالية :

المتغيرات العشوائية :

$X_i \sim N(\mu_i, \sigma_i)$ $\quad i = 1, 2, \ldots\ldots, n$

$$Z_i = \frac{X_i - \mu_i}{\sigma_i} \sim N(0, 1)$$

$$\sum_{i=1}^{\upsilon} Z_i = \sum_{1}^{\upsilon}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2(\upsilon)$$

٤- توزيع ف F - Distribution

إذا كانت S_1^2 , S_2^2 تباين لعينتين عشوائيتان حجم كل منهما n_1 , n_2 على الترتيب أخذتا من توزيعين طبيعيين له نفس التباين فإن المتغيرات العشوائية .

حيث

$$F = \frac{S_1^2}{S_2^2} \sim F_{\upsilon_1, \upsilon_2}$$

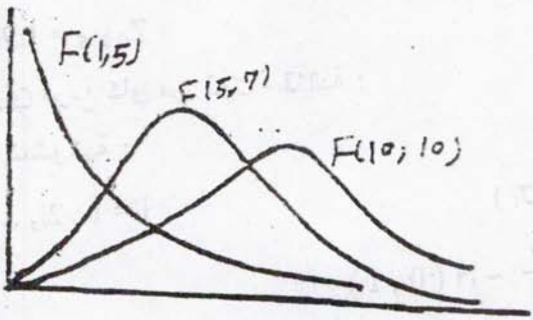
$\upsilon_1 = n_1 - 1$, $\upsilon_2 = n_2 - 1$ وتسمى بدرجات الحرية للتوزيع

فهو متصل وغير متماثل ويشبه إلى حد كبير توزيع مربع كاى (χ^2) أى

١- له معلمتان υ_1 , υ_2 كلاهما يسمى بدرجات الحرية .

٢- حدى التوزيع يمتد من صفر إلى ∞ .

٣- التوزيع ملتو من اليمين .



٤- إذا كان المتغير العشوائى X يتبع توزيع $F_{\upsilon_1, \upsilon_2}$ فإن $\frac{1}{X}$ يتبع توزيع $F_{\upsilon_1, \upsilon_2, 1-\alpha}$

(i) For $\upsilon_1 = 1$ and $\upsilon_2 = \upsilon$

فإن قانون F يتحول إلى $\chi^2(1) \equiv t^2$

(ii) $\upsilon_1 = \upsilon$ and $\upsilon_2 = \infty$

فإن قانون F يتحول إلى $\chi^2(\upsilon)$

ويمكن تكوين قانون F كما سبق

$$F_{\upsilon 1, \upsilon 2} = \frac{\chi_1^2 / \upsilon_1}{\chi_2^2 / \upsilon_2}$$

حيث χ_1^2 قانون مربع كاى بدرجة حرية υ_1 ،

χ_2^2 قانون مربع كاى بدرجة حرية υ_2 .

من (i) , (ii)

$$F_{1, \upsilon, \alpha} = t^2_{\upsilon, \alpha} , F_{\upsilon 1, \upsilon 2, \alpha} = \frac{1}{F_{\upsilon 1, \upsilon 2, 1-\alpha}}$$

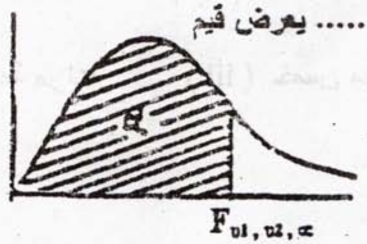
كما يلاحظ أن :

قانون t :
$$t = \frac{Z}{\sqrt{\chi^2/\upsilon}} \sim t(\upsilon)$$

$$\rightarrow N(0,1) \quad \text{if} \quad \upsilon \rightarrow \infty$$

بالجدول .... يعرض قيم $F_{\upsilon 1, \upsilon 2, \alpha}$ حيث

$$P(X < F_{\upsilon 1, \upsilon 2, \alpha}) = \alpha$$



تمارين (٤)

١ - أو جد العدد C بحيث يكون :

$p(z < c) = 0.2266$, $P(z < c) = 0.8643$

$p(z < c) = 0.05$, $P(z \geq -c) = 0.65541$

$p(-c < z < c) = 0.99$, $P(-c \leq z \leq c) = 0.95$

٢ - متغير عشوائى X يتبع $N(50 , 25)$ أحسب :

$p(x = 60)$, $P(|x - 40| > 5)$

$p(x > 62)$, $P(|x - 50| < 8)$

٣ - إذا كان :

$X \sim N(2,2)$, $Y \sim N(3,3)$, $Z \sim N(4,4)$

أحسب :

$(i)\ p(1 \leq x \leq 4)$, $(ii)\ p(x - 2 \leq 4)$

$(iii)\ p(2x + Y \geq 5)$, $(iV)\ p(z+2 < 4x - y \leq 3)$

$(V)\ p(x \geq y , z - 3 > 0)$

٤ - فى خمس رميات لزهر طاولة غير متميزة أوجد احتمال أن يظهر الرقم 3 ؟

(i) ثلاث مرات ، (ii) أربعة مرات ، (iii) خمس مرات

٥- مـن بـيانات الجدول الآتى أحسب المتوسط الحسابى والتباين ثم أوجد قانون التوزيع النظرى الذى يخضع إليه هذه البيانات ؟

X	0	1	2	3	4	5	6	7
n	367	376	218	89	33	13	2	1

٦- التوزيع الآتى يتبع توزيع بواسون . أحسب متوسطه وتباينه وبين أنهما متساويين ؟

X	0	1	2	3	4	5	6	7	8	9	10
P (x)	0.1354	0.2706	0.2708	0.1804	0.0902	0.0361	0.012	0.0034	0.0008	0.0002	0.0001

٧- إذا كـان احتمال إصابة الأشخاص اللذين يعانون من التأثير المضاد لمصل أعطى لهم هو 0.001 من بين 2000 شخص ما هو احتمال إصابة :

(i) ٣ فقط ، (ii) أكثر من فردين يعانون من التأثير المضاد .