الفصل الحادى عشر

تلخيص وعرض البيانات

Data Summarizing and Presenting

11.1. مقدمة Introduction

تعرفنا من خلال الأجزاء السابقة علي أنواع مختلفة من البيانات، تعرفنا على البيانات الوثابة Discrete والمستمرة Continuous وتعرفنا علي نوعية هذه البيانات هل هي اسمية Nominal أم ترتيبية Ordinal أم فترة Interval أم نسبة Ratio وأشرنا إلى أنه باختلاف نوع البيانات ستختلف طريقة معالجتها ودراستها والتعامل معها، وهذه البيانات هي ما نحصل عليه عند إجراء دراسة ما على موضوع محدد، وفى الغالب ما تكون بيانات كثيرة من سجلات أو من عينات، وإذا تم عرض وتقديم البيانات وخاصة الكثيرة فى صورة جداول أو بطريقة جافة وغير منظمة فغالبا ما تفقد هذه البيانات أهميتها، بل وقد يؤدي عدم فهمها إلي فقد الثقة فيها، ويمكن أن تصبح هذه البيانات عقبة أمام متخذ القرار بدلا من أن تكون أداة لاتخاذ قرار صائب، ومن هنا تبرز أهمية معالجة هذه البيانات بطريقة صحيحة.

وسنعرف فى هذا الجزء كيف سيمكننا تنظيم وتبويب هذه البيانات وتلخيصها وعرضها فى صورة مركزة و تحويلها من صورتها الجافة وغير المعبرة إلي صور أخري أكثر وضوحا وتعبيرا وأسهل فهما وتناولا، وهذه المعالجة ستجعل البيانات المتاحة أكثر طواعية للدراسة والتحليل وستساعد فى كشف الصفات البارزة والخصائص العامة التي لا تظهر فى البيانات قبل تنظيمها، وسنتناول طريقتين لمعالجة هذه البيانات كما يلي:

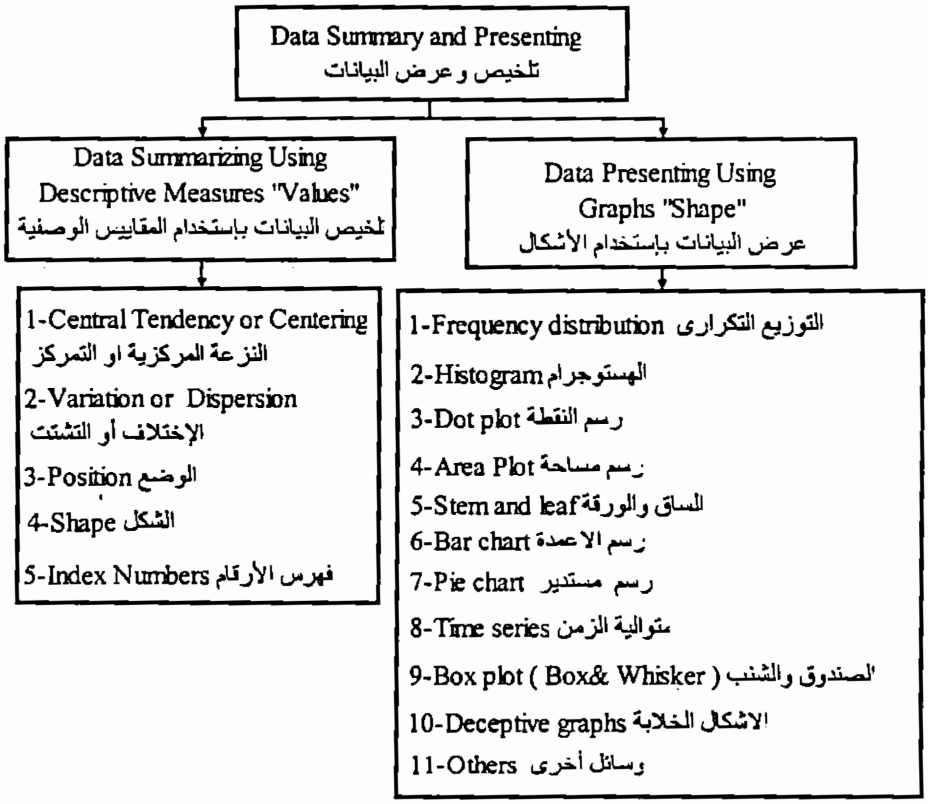
1. الأولى هي المقاييس الرقمية الوصفية Descriptive Measures:

وهى عملية اختزال وتحويل الأرقام الكثيرة إلي قيم ومعايير أقل عدداً ومتفق عليها وتعطي انطباعاً عن إجمالي الأرقام الأصلية مثل المتوسط Mean والمنوال Mode والانحراف المعيارى Standard Deviation و......

2. الثانية هي الأشكال و الرسوم البيانية Graphs:

وهى اختزال وتحويل الأرقام الكثيرة إلي رسومات بيانية أسهل فهماً وأكثر تعبيرا.

والشكل 11-1 ، يعرض ملخصا لهذه الطرق:



شكل رقم 11-1 تلخيص وعرض البيانات

11.2. عرض البيانات بالمقاييس الوصفية Data Summary Using Descriptive Measures

والجديد الذي نتعرض له الآن و سنحاول أن نعرفه، هو أنه يمكن تحويل هذه الأرقام الكثيرة إلي أرقام أو قياسات وصفية Descriptive Measures ذات دلالة محددة، ومفهومة، ومتفق عليها، وتعطي انطباعاً وتفسيراً للخصائص العامة لمجموعة الأرقام الأصلية.

والرقم الوصفى أو المقياس الوصفى Descriptive Measure هو رقم وحيد يتم حسابه من بيانات العينة ليعطي انطباعا وتفسيرا لخصائص المجتمع (لمعرفة المزيد عن العينة و المجتمع يرجى مراجعة الفصل الثالث عشر تحت عنوان العينات وطرق انتقائها فى هذا الكتاب)، و تنقسم هذه المقاييس الرقمية الوصفية إلي عدة أنواع، ويتوقف اختيار النوع الذي سنستخدمه على عدة أشياء منها وبصفة أساسية المعامل الذي نريد قياسه ودراسته وتحليله، ويمكن إجمال هذه الأنواع فيما يلي:-

1- مقاييس النزعة المركزية (التمركز/الدقة) Measures of Central Tendency (Centering / Accuracy).

وهي تجيب علي الأسئلة: أين يقع متوسط بياناتي ؟ وحول أى القيم تتركز؟ وما هي أكثر القيم تكرارا ؟

2- مقاييس الاختلاف/ التشتت Measures of Variations / Dispersion.

وهي تجيب علي: ما مدي تبعثر وتشتت بياناتي حول المتوسط ؟ وما قيمة و حدود هذا التبعثر ؟

3- مقاييس الوضع Measures of Position.

وهي تجيب علي: أين مكان قيمه معينة من باقي القيم المسجلة ؟ وأي القيم تزيد عن قيمة معينة؟ وأي القيم تقل عن قيمة معينة؟

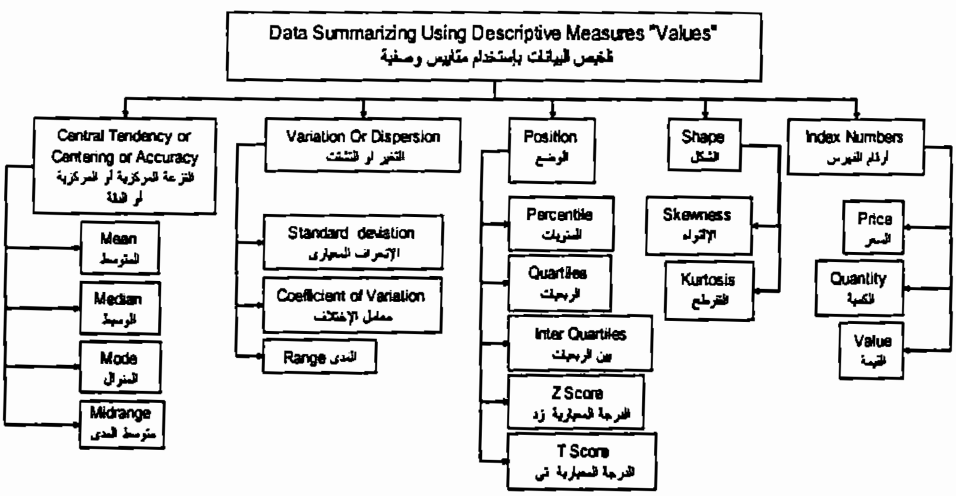
4- مقاييس الشكل Measures of Shape.

وهي تجيب علي:.هل قيم بياناتي (عند تمثيلها بيانيا على منحنى الهستوجرام مثلا) متماثلة ؟ أم أنها ملتوية ومتراكمة حول أحد الأطراف؟ وهل هي منبسطة و مفرطحة أم لا؟

5- فهرس الأرقام Index Numbers.

وهى وسيلة فعالة وطريق مختصر لوصف المتغيرات الاقتصادية، مثل التباين فى الأسعار أوالكميات أو القيمة مع الزمن.

والشكل 11-2 يعرض ما تم ذكره ، ويليه شرح مفصل لهذه المقاييس:-



شكل رقم 11- 2 المقاييس الوصفية

11.2.1. معايير النزعة المركزية Measures Of Central Tendency

فى كثير من التوزيعات نجد أن عددا كبيرا من البيانات يميل وينزع إلى التمركز والتراكم حول قيمة معينة، وهذه القيمة تسمى معيار النزعة المركزية، وهي مجموعة من المعايير التي تحدد مدى تمركز مجموعة البيانات وتنقسم لأربعة معايير سنقوم بشرحها علي مجموعة البيانات التالية 6،9،7،23،5

أ- المتوسط الحسابي $\overline{X}$ Arithmetic Mean μ &

المتوسط الحسابي هو متغير من المتغيرات الكمية، وهو أكثر معايير النزعة المركزية استخداما وشيوعا واستجابة للمعالجات الرياضية، وهو متوسط القيم المتاحة وله رمزان، ففي حالة العينة تكتب $\overline{X}$ وتقرأ (X–Bar)، وفي حالة المجتمع Population تكتب μ وتقرأ (MU).

$$\overline{X} = \frac{\sum X}{N} \qquad \mu = \frac{\sum X}{N-1}$$

وعليه فمتوسط القراءات السابقة هو

$$\bar{X} = \frac{6+9+7+23+5}{5} = \frac{50}{5} = 10$$

وللمتوسط الحسابي عدة خواص منها ما يلي:

1. المتوسط هو قيمة حسابية قد لا يكون لها وجود حقيقي فى العينة كما في هذا المثال، فالرقم 10 لا وجود له فى مجموعة الأرقام المعطاة.
2. يتأثر المتوسط الحسابي كثيرا بالقيم المتطرفة فى التوزيع، فهذه القيم حينما توجد تجعله غير صالح كمقياس من مقاييس النزعة المركزية، لأنه فى هذه الحالة يعطينا صورة خاطئة عن تجمع البيانات العددية، ولذا لا يفضل استخدامه عندما تكون البيانات تحت الدراسة شديدة الالتواء.
3. يزداد استقرار المتوسط الحسابي ويقل تذبذبه وتغيره كلما زاد حجم العينات.
4. تختلف قيمة المتوسط باختلاف العينة المأخوذة من المجتمع وهذا أحد عيوبه.
5. مجموع انحرافات القيم المتاحة عن المتوسط تساوى صفرا، ففي المثال المبين نجد أن المتوسط هو 10 وانحرافات القيم المتاحة بالترتيب هي $10-6=4$ و $10-9=1$ و $10-7=3$ و $10-23=-13$ و $10-5=5$ ، وبجمعها نجد أنها تساوى صفرًا.

6. يزداد الاطمئنان بأخذ المتوسط كتقدير لمعايير النزعة المركزية إذا كان نظام القياس دقيقا، ويقل هذا الاطمئنان كلما ضعف مستوى نظام القياس، ونقصد بنظام القياس: الموظف الذي يقيس، وآلة القياس ودقتها، وطريقة القياس والظروف المحيطة بعملية القياس.

ب- الوسيط M_d Median

الوسيط هو متغير من المتغيرات الترتيبية، و هو الرقم الأوسط من البيانات بعد ترتيبها من الأصغر إلي الأكبر ، وقيمته أكبر من كل الأرقام التي تسبقه وأقل من كل الأرقام التي تليه في الترتيب، وبمعرفته نتمكن من معرفة ما إذا كان رقم ما ينتمي إلى النصف الأصغر من البيانات أم إلى النصف الأكبر من تلك البيانات، ونحدد ترتيب الوسيط من العلاقة :

$$\mathrm{M}d = \left(\frac{\mathrm{N}+1}{2}\right)$$

وفي المثال نرتب القيم تصاعديا 5،6،7،9،23،

وتطبيق القانون $\mathrm{M}d = \frac{5+1}{2} = 3$ فيكون الوسيط هو الرقم الثالث من القيم التي تم ترتيبها أي يساوي 7 أما إذا كانت عدد زوجي فإنه بعد ترتيب هذه الأرقام تصاعديا، يكون الوسيط مساويا لمتوسط العددين الأوسطين من العينات فمثلاً لايجاد الوسيط لمجموعة الأرقام 7 1 2 5 3 9

- نرتب الأرقام 1 2 3 5 7 9
- جيث أن عدد الأعداد زوجى نحدد الرقمين الأوسطين $\mathrm{M}d = \frac{\mathrm{N}+1}{2} = \frac{6+1}{2} = 3.5$ فيكون الرقمين الأوسطين هما الثالث والرابع
- نعين الوسيط وهو متوسط الرقمين الأوسطين $\mathrm{M}d = \frac{3+5}{2} = 4$

وللوسيط عدة خصائص منها:

1. لا يتأثر بحجم العينة، ولذا يفضل استخدامه إذا كانت البيانات المتاحة قليلة.
2. قيمته لا تتأثر كثيرا بالقيم المتطرفة كما هو الحال مع المتوسط الحسابي، وإنما يتأثر بالقيم الوسطي، وهو بهذه الصفة على نقيض المتوسط الذي يتأثر بالقيم المتطرفة

أكثر من تأثره بالقيم الوسطي، ولذا يفضل استخدام الوسيط فى التوزيعات الملتوية مثل المرتبات وعدد ساعات العمل.

3. يصلح الوسيط لتمثيل البيانات النوعية غير الرقمية حيث يصعب القياس الكمي، وإنما نستطيع ترتيب البيانات بحسب نوعها أو حجمها (جيد – ممتاز- أو متوسط - عال- عال جدا -...).
4. تختلف قيمة الوسيط باختلاف العينة المأخوذة من المجتمع وهذا أحد عيوبه.
5. من عيوبه أيضا أنه قليل الحساسية، بمعنى أننا قد نستبدل قيما كثيرة من البيانات المتاحة بقيم أخرى دون أن يتأثر لأنه لا يعتمد فى حسابه على كل القيم، كما لا يسهل التعامل معه جبريا.
6. يرمز للوسيط فى بعض الحالات بالرمز $\tilde{X}$ يقرأ X Til-Dah.

ج- المنوال M_o Mode

المنوال هو متغير من المتغيرات الاسمية، و هو أكثر القيم تكرارا وشيوعا في العينة، وهو لا يستعمل كثيرا مع معايير المركزية لأنه لا يعبر عنها بصدق، ففي مثالنا لايوجد منوال Mode، ولذا نأخذ مثالا آخر 5، 1، 5، 4، 8، 7، 6، 9، 8، 10، 5، 8

وبدراسة هذا المثال وهذه المجموعة من الأرقام سنجد لها منوالين Two Modes وهي 5 & 8 وهو استثناء لكون معايير النزعة المركزية وحيدة القيمة، ومع ذلك فللمنوال فى هذا المثال قيمتان، وفى هذه الحالة يمكن حساب المنوال من المتوسط والوسيط من العلاقة التقريبية (صالحة ما دام المنحنى قريب من التوزيع الطبيعي) التي أوجدها بيرسون كما يلي:

$$Mo = 3 * M_d - 2\mu$$

إذن ففيم يستخدم المنوال؟

يستخدم مثلا في استقصاء مقاسات الملابس إشارة إلى أن أرقاما محددة تلقى قبولا أكثر من غيرها لدى العملاء، أو أن مقاسات معينة من الأحذية هي الأكثر مبيعا، أو عند حساب الدورة الزمنية Cycle Time كما سنرى فيما بعد، إذ أنه لمعرفة متوسط زمن عملية ما، فيمكن أن نقوم بتسجيل قيم كثيرة لأزمنة هذه العملية، ثم نختار أكثر القيم تكرارا ليكون هو الزمن المطلوب، وهكذا.

وللمنوال عدة خصائص منها :

1. لا يتأثر بحجم العينة، ولذا يفضل استخدامه إذا كانت البيانات المتاحة قليلة.

2. قيمته لا تتأثر بالقيم المتطرفة ولا بالقيم الوسطي، وهو بهذه الصفة أكثر ثباتا من المتوسط والوسيط.
3. يستخدم للظواهر غير الرقمية حتى التي لا يمكن ترتيبها مثل الحالة الاجتماعية (أعزب- متزوج-....)، و فصائل الدم (أ، ب، أب..).

د- متوسط المدى (M_r) Midrange

ومتوسط المدى هو أقل استخداما من المتوسط Mean والوسيط Median وأقل تعبيرا عن مركز البيانات لأنه يأخذ أكبر وأقل قراءة $Mr = \frac{L+H}{2}$ حيث L هي أقل قيمة في العينة و H هي أكبر قيمة في العينة

وفي المثال $Mr = \frac{5+23}{2} = 14$

مثال رقم 11- 1 على معايير النزعة المركزية Measures of Central Tendency

إذا كان لدينا عشرة قراءات لفترة تنفيذ مشروعات بناء منازل سكنية بالأشهر كما يلي:

4.1 &3.2 &2.8 & 2.6 & 3.7 & 3.1 & 9.4 & 2.5 & 3.5 & 3.8

والمطلوب المقارنة بين معايير النزعة المركزية & Mode & Midrange &Median Mean

الحل : سوف نعطى حلا لهذا المثال بطريقتين، الطريقة الأولى هى الطريقة النمطية بإستخدام الآلة الحاسبة والطريقة الثانية بإستخدام برنامج مينيتاب:

ولحساب المتوسط Mean

$$Mean = \frac{\sum X}{N} = \frac{4.1+3.2+............+3.8}{10} = 3.87 \ \ Months$$

ولحساب الوسيط Median نرتب الأرقام تصاعديا. 2.5،2.6، 2.8، 3.1، 3.2، 3.5، 3.7،3.8، 4.1، 9.4

و حيث إن الأعداد زوجية، إذن ترتيب الوسيط Medianهو الرقم الخامس والسادس

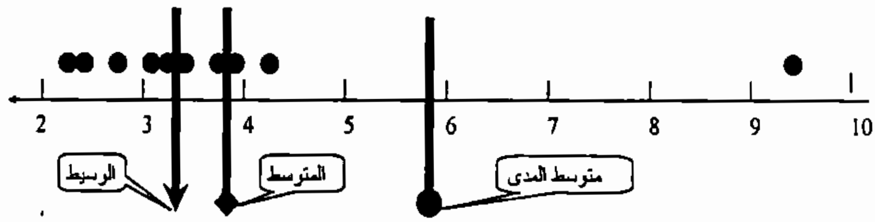
$$Md = \frac{5th+6th}{2} = \frac{3.2+3.5}{2} = 3.35$$

ولحساب متوسط المدى Midrange $Mr = \frac{L+H}{2} = \frac{2.5+9.4}{2} = 5.95$

ولإيجاد السنوال Mode نجد أنه لا يوجد منوال صريح، فنحسبه من العلاقة:

$$Mo = 3*M_d - 2\mu = 3*3.35 - 2*3.87 = 2.31$$

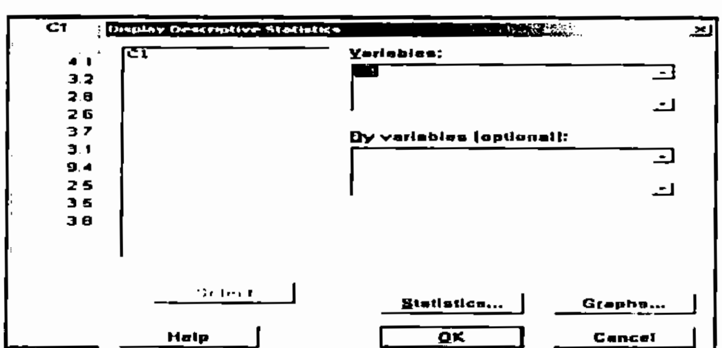
وبرسم هذه القيم كما فى شكل 11-3 علي محور الزمن سيتضح لنا وجود أكثر من معيار للنزعة للمركزية.



شكل رقم 11- 3 تمثيل لقيم المثال 11-1

الطريقة الثانية للحل بإستخدام المينيتاب Minitab: برنامج المينيتاب Minitab هو أحد البرامج الإحصائية التى تقوم بمعالجة البيانات وحساب القيم الإحصائية المختلفة بدقة عالية، كما تقوم كذلك برسم الأشكال البيانية والهندسية المختلفة ويتميز هذا البرنامج بسهولته وشموليته.

من المينيتاب Minitab نختار القوائم Stat >Basic Statistics > Display Descriptive Statistics، فيظهر الصندوق الحوارى الموضح فى شكل 11-4



شكل رقم 11- 4 الصندوق الحوارى من المينيتاب

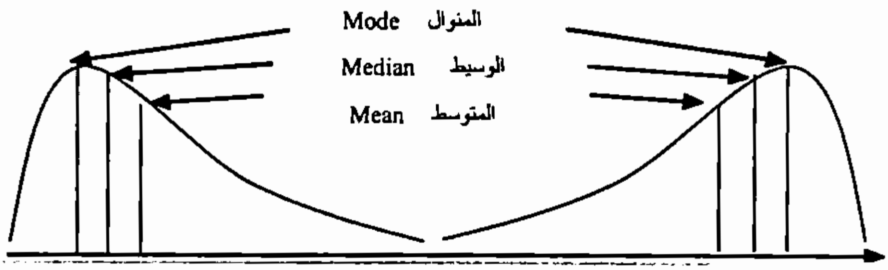
ومن نافذة النتائج Session Window نجد أن:

Descriptive Statistics: C1

Variable	Mean	Median
C1	3.870	3.350

ملاحظات هامة على الوسط ، الوسيط والمنوال:

- بالنظر إلي الأرقام المعطاة سنجد أن معظمها يدور حول القيم من 2 حتي 4 وأن هناك قيمة غريبة/شاذة Outlier وهي 9.4 وفي هذه الحالة يجب التأكد من صحة هذه القيمة، هل هي صحيحة أم أنها سجلت بطريقة الخطأ؟ (وهذا يقودنا لقضية هامة وهى دقة تسجيل البيانات ودقة القياس، و سيتم مناقشتها في الفصل السادس تحت عنوان تحليل نظام القياس Measurement System Analysis MSA).
- القيمة الغريبة/الشاذة Outlier هي قيمة موجودة في العينة تكون قيمتها كبيرة جداً أو صغيرة جداً مقارنة بباقى قيم العينة وإذا وجدت يجب التأكد من صحتها وأنها ليست مسجلة بطريقة الخطأ ومثال ذلك رقم 30 فى مجموعة الأرقام 1 11 9 7 5.5 30 4
- في المثال11-1 السابق سنجد أن قيم الوسيط والوسط ومتوسط المدى Median & Mean& Midrange هي قيم مختلفة لا تعبر بصدق عن متوسط القيم، وهو ما دفع الإحصائيين للبحث عن أكثر من معيار وليس معياراً واحداً، وبالتالي يكون لدينا أكثر من طريقة لوصف البيانات، حتى يكون انطباعنا صادقا ومعبرا عن مدى تمركز البيانات.
- يعتمد اختيار أى من مقاييس النزعة المركزية السابقة بدرجة كبيرة على نوع البيانات المتاحة، فيفضل استخدام المنوال Mode إذا كانت البيانات المتاحة من النوع الاسمى Nominal، ويمكن استخدام المنوال Mode أو الوسيط Median إذا كانت البيانات المتاحة من النوع الرتبى Ordinal، كما يمكن استخدام المنوال Mode أو الوسيط Median أو المتوسط الحسابي Mean إذا كانت البيانات المتاحة من النوع الفترى Interval أو النوع النسبي Ratio.
- عادة ما تقع قيمة الوسيط بين قيمتى المتوسط الحسابي والمنوال.
- تكون المتوسطات الثلاثة متساوية فى حالة التوزيع الاعتدالى، بينما يكون المنوال هو أكبر المتوسطات ويليه الوسيط ثم المتوسط الحسابى.
- الشكل 11-5 يوضح الموضع النسبى لكل من الوسيط والمنوال والمتوسط فى منحنى التوزيع الطبيعى.



شكل رقم 11- 5 موضع المنوال و الوسيط والمتوسط فى منحنى التوزيع الطبيعى

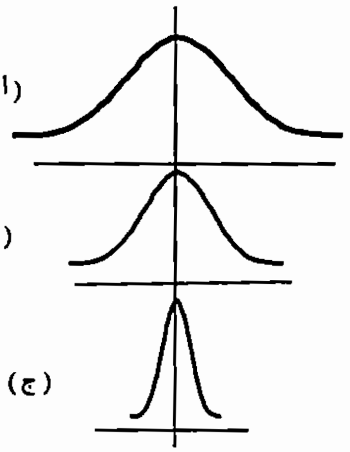
- والجدول 11-1 يعرض كذلك بعض أوجه الاختلاف والتشابه بين الوسط Mean والوسيط Median والمنوال Mode:

جدول رقم 11- 1 أوجه الشبه والاختلاف بين الوسط والوسيط والمنوال

	Mean الوسط	Median الوسيط	Mode المنوال
الرمز	μ	M_d	M_o
الاستخدام	البيانات الفترية Interval أو البيانات النسبية Ratio	البيانات الفترية Interval، و الترتيبية Ordinal أو النسبية Ratio	البيانات الفترية Interval، و الاسمية Nominal، و الترتيبية Ordinal، أو النسبية Ratio
يعتمد على	جميع الدرجات	الدرجات الوسطى فى التوزيع	الدرجة التى تكررت أكثر من غيرها
المكان	موقع العمود المار بمركز ثقل المنحنى، وأكثر ميلا نحو الجانب الملتوي	موقع العمود الذي يقسم مساحة المنحنى إلى قسمين متساويين	موقع العمود النازل من قمة المنحنى على المحور الأفقى
تأثره بالقيم المتطرفة	يتأثر كثيرا	لا يتأثر	لا يتأثر
تأثره بالقيم الوسطي	يتأثر قليلا	يتأثر كثيرا	لا يتأثر
تأثرة بزيادة حجم العينة	يزداد ثباتا	لا يتأثر كثيرا	لا يتأثر
تأثره بتغير العينة	لا يتغير كثيرا	يتغير كثيرا	يتغير كثيرا
المعالجة الجبرية	يقبل	لا يقبل	لا يقبل
العلاقة بينها	$Mo-\mu \cong 3(M_d-\mu)$ $Mo \cong 3M_d-2\mu$		

11.2.2.مقاييس التشتت Measures of Variation Or Dispersion

بالنظر للشكل 11-6 (أ، ب، ج) سيتضح جليا أن معايير المركزية – والسابق شرحه - وحدها لا تكفى للتعبير عن مدى تشتت البيانات، إذ أن المتوسط Mean للأشكال الثلاث متساو، وبالرغم من ذلك فإن البيانات الممثلة بالمنحنى (أ) أكثر تشتتا وأسوأ من الممثلة بالمنحنى (ب)، وأكثر تشتتا وأسوأ من تلك الممثلة بالمنحنى (ج)، وهذا لم تخبرنا به معايير المركزية لأنها تجيب عن سؤال واحد فقط وهو: أين يقع متوسط بياناتي ؟ و حول أى القيم تتركز باقى البيانات؟.



شكل رقم 11- 6 تساوى المتوسطات

ومن هنا كانت الحاجة لمعايير أخرى توضح مدي التباين أو التشتت Variation or Dispersion الحادث في البيانات، ولذا سنقوم بدراسة كل من معامل التشتت والانحراف المعياري والمدى Coefficient of Variation & Standard Deviation & Range ، والتي سيجيب على السؤال الذى يقول ما مدي تشتت بياناتي ؟ وما قيمة و حدود هذا التشتت ؟

أ- المدى Range

وهو الفرق بين أكبر قيمة وأصغر قيمة في القيم المتاحة للدراسة ففي مثالنا

$$(\text{Range}) = H - L = 23 - 5 = 5 = 18$$

وبالرغم من سهولة حسابه إلا أنه يعطي انطباعات مهمة جدا عن مدى التشتت والتباين الحادث فى مجموعة من البيانات.

ب- التباين والانحراف المعياري Variance and Standard Deviation

أحد مقاييس التشتت هو مجموع انحرافات البيانات عن وسطها الحسابي، لكن وكما سنعرف فيما بعد أن هذا المجموع دائما يساوى صفرا، ولذلك لابد من حذف الإشارة السالبة لنحصل على مقياس ذي معنى، وإحدى هذه الطرق التي تزيل بها الإشارة السالبة هي بتربيع الانحرافات وهو ما يطلق علية التباين Variance ويرمز له بالرمز "S^2" وهو يساوى

$S^2 = \sum_{i=1}^{n}(X^i - \bar{X})^2 / (n-1)$ - ويلاحظ أننا نقسم على (n-1) لكي يكون للتباين S^2

بعض الصفات الإحصائية المرغوب فيها على أساس أن البيانات التي لدينا هي مجرد عينة أخذت لدراستها، وهذا يساعدنا على تعميم النتائج على المجتمع الذي أخذت منه هذه العينات
- ووحدة التباين هي مربع وحدة البيانات قيد الدراسة.

أما الانحراف القياسي أو المعيارى Standard Deviation هو الجذر التربيعي للقيمة العددية للتباين Variance ووحدة الانحراف المعياري هي نفس وحدة البيانات محل الدراسة.

وهذان المعياران هما الأكثر شيوعا للتعبير عن التباين Variability في البيانات

وتتلخص خطوات حساب الانحراف المعياري يدويا (لن يتم الحساب فيما بعد يدويا ولكن هنا للشرح فقط) فى الخطوات التالية:

1. نكون الجدول المكون من الأعمدة

Sample العينة – *Value* *القيمة* - *Mean* المتوسط - $(X-X)^2$

2. نحسب المتوسط الحسابى $Mean = \bar{X} = \frac{\sum X}{N}$ ، ونحسب $(\bar{X}-X)$ ، ونحسب $(\bar{X}-X)^2$ ، ونحسب "S^2" وهو إجمالي التباين Variance

3. نحسب القيمة العددية للانحراف المعيارى للعينة S من العلاقة

$$S = \sqrt{\frac{\sum(\bar{X}-X)^2}{N-1}}$$

كما يجب الانتباه إلى أنه يمكننا حساب قيمة الانحراف المعيارى Standard Deviation بطريقة أخرى باستخدام خرائط المراقبة و التحكم Control Charts كما سبق شرحه فى الفصل الخامس من هذا الكتاب تحت عنوان خرائط التحكم والمراقبة، وسيلاحظ أن فرق القيم المستنتجة بين الطريقتين بسيطا ومقبولا.

مثال رقم 11- 2:

كيف يمكن حساب كل من المتوسط Mean والمنوال Mode و المدى Range و الوسيط Median وكذلك الانحراف المعيارى Standard Deviation للأوزان التالية بالكيلو جرام: 170 & 190 &160& 150 & 150 & 140 & 200& 160& 170

أولا الحل اليدوي

حيث S2 = التباين، N= عدد أفراد العينة، N-1= درجات الحرية، X = المتوسط نقوم بترتيب النتائج كما فى جدول 11-2

ولحساب المتوسط

$$\bar{X} = Mean = \frac{170+160+200+140+150+150+160+190+170}{9} = \frac{1490}{9} = 165.55$$

ولحساب الوسيط Median نرتب الأرقام تصاعديا 140 & 160 & 150 & 150 & 160 & 170 & 170 & 190 & 200

و حيث إن الأعداد فردية $Median = \frac{9}{2} = 4.5$ فيكون الوسيط هو الرقم الخامس فى الترتيب، أى أن

$$Median = 160$$

- أما المنوال Mode فهو أكثر الأرقام تكرارا وهى150 ،160، 170
- ولحساب المدى Range   Range = H-L =200-140 = 60
- وثانيا لحساب التباين والانحراف المعيارى Variance & SD نحسب ابتعاد قيمة متوسط مربع الفروق عن قيمة المتوسط العام للعينات وهو ما يطلق عليه التباين Variance

$= \frac{\sum(\bar{X}-X)^2}{N-1}$ كما بالجدول 11-2:

جدول رقم 11 -2- العمليات الحسابية المطلوبة لحساب التباين والانحراف

Point	Mean	$\bar{X}-X$	$(\bar{X}-X)^2$
170	165.6	-4.4	19.75
160		5.5	30.86
200		-34.4	1186.4
140		25.56	653.1
150		15.56	243.9
150		15.56	241.9
160		5.56	30.86
190		-24.4	597.5
170		-4.4	19.75
			3022

$$S^2 = \frac{\sum (X-\bar{X})^2}{n-1} = \frac{3022}{8} = 337.75$$

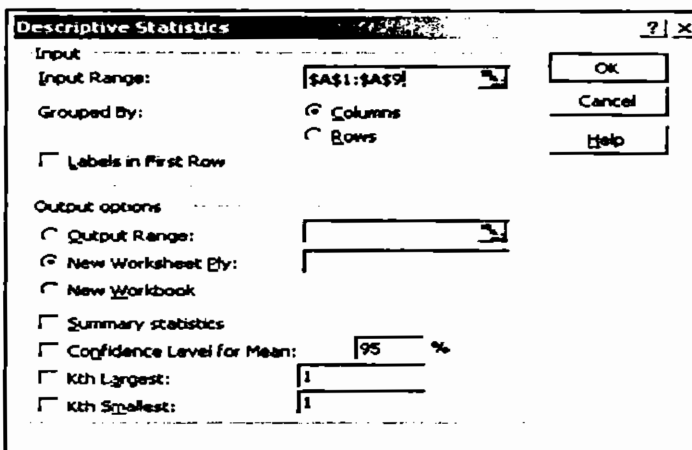
$$S = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}} = \sqrt{\frac{3022}{8}} = \sqrt{337.75} = 19.44$$

<u>ثانيا الحل باستخدام المينيتاب **Minitab**</u> ومن القوائم Minitab> Stat> Descriptive Statistics، فتظهر النتيجة التالية:

Descriptive Statistics: C1

Variable	N	N*	Mean	StDev	Variance	Median	Range
C1	9	0	165.56	19.44	377.78	160.00	60.00

<u>ثالثا الحل باستخدام الإكسل **Excel**</u> ومن القوائم Tools>Data Analysis> Descriptive Statistics فيظهر الصندوق الحواري كما فى شكل 11-7 .



شكل رقم 11- 7 الشكل الذى يظهر من برنامج Excel

ثم نختار مدى البيانات المتاحة ، فتظهر النتيجة كما بالشكل 11-8

Mean	165.5555556
Standard Error	6.478835439
Median	160
Mode	170
Standard Deviation	19.43650632
Sample Variance	377.7777778
Kurtosis	-0.234120119
Skewness	0.680409828
Range	60
Minimum	140
Maximum	200
Sum	1490
Count	9

شكل رقم 11- 8 نتائج إدخال البيانات فى برنامج Excel

ويلاحظ تساوى القيم عند حسابها يدويا أو بالكمبيوتر سواء ببرنامج Excel أو ببرنامج مينيتاب Minitab.

مثال رقم 11- 3:

احسب $\bar{X}$ & S & Range & Mo & M_d& Sigma Level للبيانات التالية

20، 35، 17، 22، 14، 10، 27، 16، 26، 16، 24، 16، اذاعلمت أن LSL = 10 ، USL = 34 .

- يتم بناء جدول بيانات مشابه لجدول 11-2
- المنوال Mode هو أكثر الأرقام تكرارا Mode = 16
- لحساب الوسيط Median نرتب الأرقام تصاعديا 10،14، 16، 16، 16، 17، 20،22، 24، 26، 27، 35

$$Md = \frac{12}{2} = 6 \rightarrow 6TH$$

وحيث إن الأرقام عددها زوجي فنحسب متوسط الرقم السادس والسابع

$$Md = \frac{6th + 7th}{2} = \frac{17 + 20}{2} = 18.5$$

- ولحساب المدى Range = 35 -10 = 25
- ولحساب المتوسط

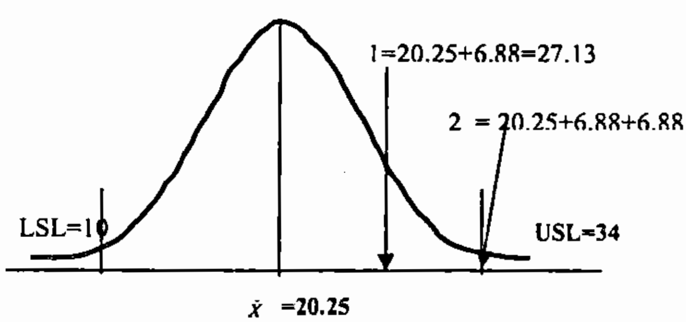
$$\bar{X} = \frac{10+14+16+16+16+17+20+22+24+26+27+35}{12} = 20.25$$

جدول رقم 11- 3 خطوة لحل المثال 11-3

Sample	$\bar{X}$	$\bar{X}-X$	$(\bar{X}-X)^2$
10		10.25	105
14		6.25	39
16		4.25	18
16		4.25	18
16		4.25	18
17	20.52	3.25	10.5
20		0.25	0.0625
22		1.75	3
24		3.75	14
26		5.75	33
27		6.75	45.5
35		14.75	217.6
			521.6

$$S = \sqrt{\frac{(\bar{X}-X)^2}{n-1}} = \sqrt{\frac{521.6}{11}} = = 6.88$$

ولحساب مستوى السيجما Sigma Level نقوم برسم منحنى التوزيع الطبيعي ونوقع عليه القيم و الحدود الدنيا والقصوى USL و LSL، ومن الرسم 11-9 يتضح أن هذه العملية تعمل على مستوى جودة يساوى انحرافين معياريين أى 2σ تقريبا وهى قيمة سيئة جدا.



شكل رقم 11- 9 منحنى التوزيع الطبيعي لحساب مستوى السيجما

مثال رقم 11- 4: قام بائع أجهزة كمبيوتر بتسجيل مبيعات أسبوع وكانت كالتالي: 16& 10

20 & 25 & 15 & 17 & 13 &، و المطلوب حساب $S \& M_d \& \bar{X}$

الحل :

$$\bar{X} = \frac{\sum X}{N} = \frac{10+16+13+17+15+25+20}{7} = \frac{116}{7} = 16.57 \quad \text{Set / day}$$

ولحساب الوسيط Median فلا بد من ترتيب الأرقام تصاعديا.

10 &13 & 15 &16 & 17 & 20 & 25 وحيث إن الأرقام عدد فردي يكون الوسيط Median هو الرقم الأوسط Median = 16

لحساب S نكون الجدول 11-4:

جدول رقم 11- 4 حساب قيم S, Md & X

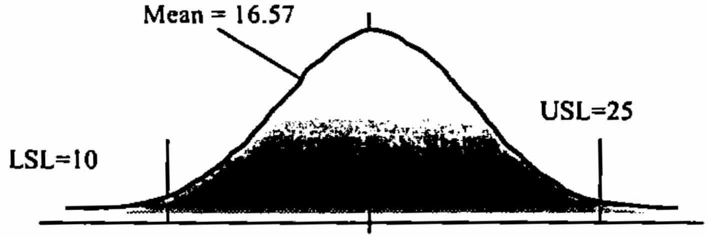
Sample	Mean	$\bar{X}-X$	$(\bar{X}-X)^2$
10		6.57	43.18
13		3.57	12.76
15		1.57	2.47
16	16.57	0.57	0..326
17		0.43	0.183
20		3.43	11.75
25		8.43	71.04
			141.7

$$S = \sqrt{\frac{(\bar{X}-X)^2}{N-1}} = \sqrt{\frac{141.7}{6}} = 4.86 \quad Set$$

فى المثال11-4 السابق: ما معني أن الانحراف المعيارى يساوى 4.86 ؟

برسم القيمة السابقة وبفرض أن لدينا عددا كبيرا من العينات يمكننا من رسم الشكل 11-10 ، وبتوقيع قيمة المتوسط Mean التي تساوي 16.57، وبفرض أن LSL تساوي 10 وأن قيمة USL تساوى 25، فإنه يمكننا حساب المدى

Range = USL –LSL =25-10 = 15



شكل رقم 11- 10 منحنى التوزيع لطبيعى للقيم المعطاة فى المثال

$$Sigma\ \ level = \frac{Range}{2*S} = \frac{15}{9.72} \approx 1.54$$

إذن فهذه العملية تعمل على مستوى سيجما يساوى 1.5 وهي قيمة صغيرة وسيئة، أي أنها عملية ضعيفة (ضعيفة لأن كثير من مخرجاتها في المدي الغير مسموح به) ولو كانت قيمة الانحراف المعيارى لهذه العملية مثلا 7 جهاز فإنه بحساب Sigma Level سنجد أنها تساوي $\frac{15}{2*7} \approx 1$ أي أنها أسوأ من السابقة، لماذا أسوأ ؟ لأن خطوة الانحراف تساوي في هذه الحالة 7 وهي أكبر من السابقة التى قيمتها 4.86 ، أي أن المساحة تحت المنحنى و المحصورة بالمسافة بين المتوسط Mean والحد الأعلى للماصفات Upper Specification Limit USL، والحد الأدنى للمواصفات Lower Specification Limit LSL ، ستحتوي علي عدد أقل من سيجما أي سيكون هناك مساحة كبيرة خارج حدود المسموح أي ستقل مستوى السيجما Sigma Level.

ومعني أن 4.86 = S لهذه المجموعة من الأرقام التي متوسطها $\bar{X} = 16.57$ هو أن المساحة المحصورة بين $\bar{X} - \sigma \& \bar{X} + \sigma$ أي بين القيمة $(16.5 + 4.8)$ والقيمة $(16.5 - 4.8)$ أي بين القيم 11.77&21.37 تمثل 68% من مساحة المنحني ككل،

وكذلك المساحة المحدودة بين $\bar{X} - 2\sigma \& \bar{X} + 2\sigma$ أي بين القيم + 16.57 & 16.57-9.7 9.7 أي بين القيم 26.27& 6.87 تمثل 95% وهكذا، وبالتالي لو تمكننا من معرفة نسبة المنتجات السليمة فيمكننا بالتناظر معرفة قيمة مستوى السيجما Sigma Level.

مثال رقم 11- 5: لو كان لدينا عملية تعطي النتائج المبينة بالجدول 11-5

جدول رقم 11- 5 نتائج إحدى العمليات

(1)	(2)	(3)	(4)	(5)
4	7	6	15	8

ولو حسبنا الإنحراف المعيارى S. Dev كالتالي :

$$Mean = \frac{40}{5} = 8$$

ثم نكون الجدول 11-6:

جدول رقم 11- 6 حساب قيم S, Md & X

No	Point	Mean	$\bar{X}-X$	$(\bar{X}-X)^2$
1	4	8	8-4=4	16
2	7	8	8-7=1	1
3	6	8	8-6=2	4
4	15	8	8-15=-7	49
5	8	8	8-8=0	0
			0	70

$$S = \sqrt{\frac{(\bar{X}-X)^2}{n-1}} = \sqrt{\frac{70}{4}} = 4.18 = 4.18$$

ولقياس مستوى الجودة Sigma Level لهذه العملية (بفرض القيم USL = 15 وأن LSL=4)

$$Sigma\ Level = \frac{11}{2 * 4.18} = 1.3$$ ويكون Range = USL– LSL= 15 – 4 =11

وحيث إن العملية تعمل علي مستوى Sigma 3 وهي قيمة سيئة وضعيفة، فيمكن القول بأن العملية سيئة جداً، ولو أننا حاولنا تعديل نفس العملية بحيث جعلنا القراءة رقم 4 لتصبح 10 بدلا من 15 مع الاحتفاظ بقيم LSL & USL وهي 4 & 10 فإن :

$$S.D = \sqrt{\frac{20}{5}} = 2.2$$

$$Sigma\ Level = \frac{15-4}{2 * 2.2} = 2.5$$ & Mean =7 & Variance = 0

أي أن مستوى السيجما Sigma Level زادت، أي أن العملية تحسنت، وهذا منطقي لأننا قللنا التباين ، فقل الانحراف المعيارى فزاد مستوى السيجما فتحسنت العملية.

وبتعميم هذا الطرح على أي مجموعة بيانات، يصبح من الممكن التعبير عن حجم المخرجات السليمة لأي عملية بدلالة قيمة مستوى السيجما Sigma level.

بمعنى أننا لو قلنا أن لدينا عملية تعمل عند One Sigma (أي واحد يمين وواحد يسار المتوسط Mean) فإن معنى ذلك أن حجم المخرجات السليمة سوف يكون 68.27% من إجمالي المخرجات (لتوضيح الفرق بين الانحراف المعيارى قيمة السيجما ومستوى السيجما، يرجى مراجعة الفصل الأول من هذا الكتاب تحت عنوان العلاقة بين قيمة ومستوى السيجما Relationship between Sigma Value and Level)، كذلك لو قلنا أن لدينا عملية تعمل عند Three Sigma (أي ثلاثة يمين وثلاثة شمال المتوسط Mean) فإن معنى ذلك أن حجم المخرجات السليمة سوف يكون 99.97% من إجمالي المخرجات.

وبالمثل لو قلنا أن شركتنا تعمل عند مستوى 2.5 Sigma وهناك شركة منافسة تعمل عند مستوى 3 Sigma فإن ذلك يعني أن تلك الشركة أفضل من شركتنا، وذلك لأن حجم المخرجات السليمة لها أكبر من حجم المخرجات السليمة لشركتنا.

وجدير بالذكر أننا سنتعرض للطرق المختلفة لحساب مستوى السيجما Sigma Level بالتفصيل عند شرح موضوع "مقدرة العملية و مستوى الجودة Process Capability and Quality Level" فى الفصل السابع من هذا الكتاب.

ج- معامل التباين The Coefficient of Variation

وهو معامل يستخدم للمقارنة بين مقدار التباين Variability لمجموعتين من البيانات الكمية Quantitative Data التي ليس لها نفس الوحدات أو متوسطاتها مختلفة اختلافا كبيرا، فمثلا لو كان لدينا مجموعتين من الأرقام كما يلي:

$$5 \quad -6 \quad -7 \quad -9 \quad -23 \quad \rightarrow \bar{X} = 10, S = 7.416$$

$$500 - 600 - 700 - 900 - 2300 \quad \rightarrow \bar{X} = 1000, S = 741.6$$

بالنظر لهاتين المجموعتين السابقتين من الأرقام، هل يمكننا تحديد أى من المجموعتين يحتوى على تباين Variability أكبر من الأخرى؟ ولأول وهلة سنقول إن مجموعة البيانات الثانية هي الأرقام التى تحتوى على تغير أو تباين أكبر، إذ أن التباين فيها بالمئات، ولكن لو دققنا النظر سنجد أنها نسبة من البيانات الأولى، وبالتالي فالتباين الموجود في كلا المجموعتين متساو، ولقد قبلنا هذا الاستنتاج نظرًا لسهولة التناظر بين الأرقام، ولكن ماذا لو كانت الأرقام مختلفة ونريد طرح نفس التساؤل ؟ عندئذ سنحتكم إلي معامل التباين

CV Coefficient of Variation وهو معامل كلما زادت قيمته دل ذلك علي حدوث تباين أكبر، ويحسب هذا المعامل بقسمة الانحراف المعياري على المتوسط الحسابى أى

$$CV = \frac{S}{\bar{X}} * 100$$

وبتطبيق ذلك على مثالنا 11-5 نجد أن:

$$CV_2 = \frac{741.6}{1000} * 100 = 74.16\% \quad \& \quad CV_1 = \frac{S}{\bar{X}} * 100 = \frac{7.416}{10} * 100 = 74.16\%$$

وحيث إن $CV_2 = CV_1$ فإن التباين الموجود في العينة الأولى يناظر ويساوى ويكافئ التباين الموجود في العينة الثانية.

كذلك لو كان لدينا مجموعتين من البيانات كما فى جدول 11-7 التالي:

جدول رقم 11- 7 قيم لعينتين أ & ب

A	6.19	6.2	6.2	6.21	6.21	6.22	6.24	6.24
B	163	186	220	250	278	283	318	334

$$CV_A = \frac{S}{\bar{X}} * 100 = \frac{0.0185}{6.2138} * 100 = 0.03\%$$

$$CV_B = \frac{S}{\bar{X}} * 100 = \frac{60.96}{254} * 100 = 23.99\%$$

ويتضح أن المجموعة A أكثر انسجاما واتساقا من المجموعة B أي يقال أنها أكثر تجانسا More Homogenous لأن معامل الاختلاف فيها ليس كبيرا (نسيج واحد)، وهو أقل من معامل الاختلاف فى المجموعة B.

ولا يقف استخدام معامل التباين CV فى مقارنة التشتت أو التباين بين العمليات المتشابهة بل يمتد استخدامه أيضا إلى مقارنة التشتت أو التباين بين الظواهر المختلفة، حتى وإن اختلفت وحدات القياس، وذلك لأن معامل التباين CV هو خارج قسمة متغيرين لهما نفس وحدات القياس، فلو كنا بصدد قياس أوزان وأطوال طلبة المرحلة الابتدائية وكانت النتائج كما فى جدول 11-8.

جدول رقم 11- 8 بيانات تعرض أوزان وأطوال الطلاب

طلبة المرحلة الابتدائية	الانحراف المعياري (σ)	المتوسط الحسابي $\bar{X}$
الأوزان	10 كجم	40 كجم
الأطوال	10 سم	140 سم

فإننا – طبقا للبيانات المتاحة- لا نستطيع القول استنادا إلى الانحراف المعياري وحده بأن التباين فى الأطوال يساوى التباين فى الأوزان، ويرجع ذلك إلى عاملين أولهما اختلاف وحدات القياس، وثانيهما اختلاف قيمة المتوسطات، ولمقارنة التباين فى الظاهرتين نحسب معامل التباين CV لكل منهما كما يلي:

$$CV_{weights} = \frac{S}{\bar{X}} * 100 = \frac{10}{40} * 100 = 25\%$$

$$CV_{Lengths} = \frac{S}{\bar{X}} * 100 = \frac{10}{140} * 100 = 7\%$$

وعلى ذلك وبالرغم من أن الانحراف المعيارى Standard Deviation متساو إلا أن التباين Variability فى الأوزان أكبر من التباين Variability فى الأطوال (مصطفى زايد 1989 صفحة 166).

11.2.3.مقاييس/معايير الموضع Measures of Position

لو تخيلنا شخصا يشعر بأنه مظلوم في عمله فأول ما سيقوم به هذا الشخص أنه سيقارن مرتبه بالآخرين الذين لهم نفس سنوات الخبرة، ونفس الأداء، ولمقارنة مرتب هذا الشخص بمرتب الآخرين، نستخدم معيار الموضع Position وهو مؤشر لموضع قراءة بالنسبة لباقي القراءات ويتم استخدام نوعين من المعيار لتحديد الموضع Position وهما معيار المئويات Percentile أو الربعيات Quartiles وكذلك الدرجات المعيارية Score و....

مثال رقم 11- 6:

ولمزيد من التوضيح نفترض أن مدير شؤون الأفراد يشرف علي 50 فرداً وأجري لهم اختبار أداء وكانت درجاتهم كما يظهر فى جدول 11-9 بعد ترتيب هذه الدرجات.

جدول رقم 11- 9 الدرجات المرتبة التى حصل عليها العاملون فى اختبار الأداء

الدرجات التى حصل عليها العاملون	22	44	56	68	78
	25	44	57	68	78
	28	46	59	69	80
	31	48	60	71	82
	34	49	61	72	83
	35	51	63	72	85
	39	53	63	74	88
	39	53	63	75	90
	40	55	65	75	92
	42	55	66	76	96

الحل: وبحساب المتوسط نجد أن قيمته $\bar{X} = 60.36$، وبحساب الانحراف المعياري نجد أن قيمته $S = 18.61$.

فإذا حصل أحد الأفراد علي نتيجة 83 درجة، فكيف نقيم أداء هذا الفرد بالنسبة لباقي أفراد المجموعة؟ أي أين مكان أو ترتيب هدا الفرد بالنسبة لباقي الأفراد؟ والإجابة على ذلك سوف تكون من خلال حساب المئينات، والربعيات والدرجات المعيارية كما سيتضح من الشرح التالى.

أ- المئويات **Percentiles**

فى كثير من الأحيان نحتاج لمعرفة نسبة البيانات التي تقل عن قيمة معينة أو تساويها، فمثلا ما هي نسبة العاملين الحاصلين على نسبة 85% أو أقل فى التقييم السنوي لهؤلاء العاملين، ولتحديد ذلك نرتب القيم المتاحة ترتيبا تصاعديا ونوزعها على محور أفقى مقسم إلى مائة جزء ثم نحدد النسبة المطلوبة فمثلا المئين الأول ورمزه P1 هو القيمة التي تساوى أو تكون أكبر من واحد على مائة من القيم، وكذلك المئين التسعون P90 هو القيمة التي يسبقها أو يساويها 90% من البيانات ويليها أو يساويها 10% منها على فرض أن البيانات مرتبة ترتيبا تصاعديا.

وبعض المئينات له أهمية خاصة مثل P25& P50 & P75 حيث P25 يسمى الربيع الأول Q1 وهو القيمة التي يسبقها ربع البيانات ويليها ثلاثة أرباع البيانات، و حيث P50 يسمى الربيع الثاني Q2 وهو القيمة التي يسبقها نصف البيانات ويليها النصف الأخر من البيانات وهو الوسيط فى هذه الحالة، وحيث P75 يسمى الربيع الثالث Q3 وهو القيمة التي يسبقها ثلاثة أرباع البيانات ويليها ربع البيانات .

و المنويات Percentiles هو أكثر مقاييس الموضع Position شيوعا.

$$Position = N * \frac{p_n}{100}$$

$$Percentile = p_n = \frac{Position}{N} * 100$$

حيث Pn المنوية Percentile، وحيث الموضع Position هو ترتيب الرقم في مجموعة الأرقام الكلية، وحيث N هي عدد الأرقام الكلي.

وإذا كانت نتيجة المعادلة السابقة رقم كسري، نقوم بأخذ القيمة التالية للكسر فمثلا $N * \frac{P}{100} = 30.5$ نأخذ القيمة الواحدة والثلاثين ويكون الترتيب هو " الواحد والثلاثون"، وإذا كان ناتج المعادلة رقم صحيح نقوم بأخذ متوسط هذه القيمة والقيمة التي تليها فمثلا $N * \frac{P}{100} = 30$ فنقوم بأخذ القيمة الثلاثين والقيمة الواحد والثلاثين ونحسب متوسطهم

$$P = \frac{30ThValue + 31ThValue}{2}$$

وبالعودة إلى المثال 11-6 الذي تحدثنا عنه للتو، ولمعرفة درجة الشخص الذي ترتيبه هو $P83$ نقوم بالآتي:

- نرتب الأرقام ترتيبا تصاعديا.
- نحدد العدد الإجمالي للدرجات فنجد أنه 50 درجة (50 قراءة).
- نحدد مكان ال$P83$ فنجد أنه رقم $Pn = N * \frac{P}{100} = 50 * \frac{83}{100} = 41.5$ وحيث إن الرقم 41.5 هو رقم كسرى، فإن مكان ال$P83$ هو الثاني والأربعون من الدرجات المرتبة، ومن الجدول المرتب نجد أن قيمته $P83 = 78$.

مثال رقم 11- 7:

لنفس الجدول 11-3 السابق المطلوب معرفة ترتيب الشخص الحاصل علي 83 درجة.

الحل:

بالرجوع إلي المثال 11-6 ولمعرفة ترتيب الشخص الحاصل علي 83 درجة، نحدد الموضع Position يدويا فنجد أن موقع القيمة 83 يكون 45، و نطبق العلاقة

$$Percentile = Pn = \frac{Position}{N} * 100 = \frac{45}{50} * 100 = 90 \rightarrow 90\%$$

مثال رقم 11- 8:

علي نفس البيانات السابقة ما هى القيمـ$P50$ أو 50^{th} Percentile، ويلاحظ أن هذه هي القيمة التي تتوسط الأرقام الكلية وهى الوسيط Median.

الحل:

نطبق العلاقة

$$\text{Position} = N * \frac{p_n}{100} = 50 * \frac{50}{100} = 25$$

وحيث إن القيمة 25 عدد صحيح فإن

$$p_{50} = \frac{25Th Value + 26Th Value}{2} = \frac{61+63}{2} = 62$$

ب- المعيار الربعي (Q) Quartiles

وهو مقياس يقسم البيانات المتاحة إلى أربعة أجزاء متساوية، وهو بذلك يعتبر حالة خاصة من المنينات Percentile وذلك بعد ترتيب الأرقام تصاعديا

Q_1 = Its Quartile =25 TH Percentile = (P25) القيمة التي قبلها ربع القيم، وكذلك

Q_2 =2 nd Quartile = 50th Percentile = Median =P50 القيمة التي قبلها نصف القيم، وأيضا Q_3 =3rd Quartile = 75th Percentile = P75 القيمة التي قبلها ثلاثة أرباع القيم، ويستخدم هذا المقياس في عمليات التصنيف مثل A، B، C، D لدرجات امتحان مجموعة من الأشخاص.

وكما حسبنا المنويات أو المنينات Percentile في الجزء السابق نحسب هنا P 75 & P 50 & P 25

فإذا كان المطلوب معرفة قيم الأرقام التي ترتيبها الخامس والعشرون، والخمسون، والخامس والسبعون وبتطبيق ذلك علي أرقام المثال11-6 السابق نجد:

$$Q_1 = P25 = \frac{N * P25}{100} = \frac{50 * 25}{100} = 12.5 \cong \text{No } \#13$$

أي أن Q_1 تساوي الرقم الذي ترتيبه 13 ومن الجدول المرتب نجد أنه (46)

$$Q_2 = Median = P50 = \frac{N * P50}{100} = \frac{50 * 50}{100} = 25$$

وحيث إن هذه القيمة عدد صحيح نأخذ متوسط قيمة العدد الخامس والعشرين والسادس والعشرين.

$$Q_2 = \frac{25rh + 26rh}{2} = \frac{61 + 63}{2} = 62$$

ولاحظ أنها هي قيمة الوسيط Median

$$Q_3 = P75 \frac{N * P75}{100} = \frac{50 * 75}{100} = 37.5 \rightarrow 38$$

أي أن Q_3 تساوي الرقم الذي ترتيبه 38 ومن الجدول المرتب نجد أنه (75)

ج- المدى بين الربعى Inter Quartile Range (IQR)

وهو مقياس لتوضيح مدي التشتت أو التباين المحصورفي 50% الوسطي من البيانات، وكلما كان كبيرا كان التباين فى البيانات Data Variability كبيرا، وعيبه أنه يبحث هذا التباين في جزء من البيانات فقط ، وليس فيها كلها كما في المدى، وسوف يتم تمثيل هذه القيمة بيانيا فيما بعد بمخطط الصندوق Box Plot أو الصندوق والشنب Box & Whisker.

د- الدرجة المعيارية زد Z Scores

والدرجة المعيارية مقياس آخر لتحديد الموضع Position (مثل موضع كل طالب بين أقرانه فى الامتحان) و تعتمد فى حسابها علي قيمة كل من المتوسط Meanوالانحراف المعياري Standard Deviation، وهو يقيس بعد مكان نقطة معينة عن المتوسط Mean بعدد من الانحرافات المعيارية، وتوجد جداول لهذا المعامل تعطى المساحة تحت المنحنى لقيمه المختلفة، وسنرى كيف يمكن استخدام هذا المعامل فى قياس قدرة العملية Process Capability كما سيتضح فى الفصل السابع "مقدرة العملية ومستوى الجودة Process Capability and Quality Level" من هذا الكتاب

وتحسب قيمة الدرجة المعيارية بالعلاقة (Z)

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

حيث X هى الدرجة الخام، و X- Bar هى المتوسط، و S هى الانحراف المعيارى، ويتميز هذا المقياس أو المعيار على قدرته الفائقة للمقارنة النسبية آخذا فى الاعتبار كافة المؤثرات التى يمكن أن تؤثر فى القيم قيد المقارنة وهذا هو مفهوم مقارنة تفاحة بتفاحة أى Apple to Apple.

لنفرض أن لدينا طالبين، أحدهما فى الفصل "أ" وحصل على 55 درجة من 60 فى مادة الرياضيات، والثاني فى الفصل "ب" وحصل على 43 درجة من 60 فى امتحان مختلف لنفس المادة، فهل يمكننا أن نجزم بأن الطالب الأول أفضل من الطالب الثاني؟ أرجو ألا نتسرع فى الإجابة فهي لا تتم بصورة مباشرة، إذ أن الدرجات المعطاة تسمى الدرجات الخام Raw Scores وهى درجات ليس لها دلالة قاطعة دون أن نحولها إلى درجات معيارية "زد" Z Scores تأخذ فى الاعتبار موضع وترتيب كل طالب بين زملاء فصله، وهذه الدرجات تسمى أيضا درجات الميزان Scaled Scores ، فالظروف والملابسات التى صاحبت كلا الامتحانين بالقطع لا تكون متشابهه مثل المهلة التى أعطيت لكلا الطالبين قبل الامتحان، وعدد الأبواب التى تم فيها الامتحان، والوقت المتاح لأداء الامتحان، والتوقيت الذى أجرى فيها الامتحان والجو النفسى والتعليمات التى أعطيت قبل الامتحان، والتوضيحات التى صاحبته، والمكان التى عقد فيه الامتحان، و......وهذه الطريقة تطبق فعليا فى العديد من جامعات العالم وفى الإمتحانات النهائية للدراسة الثانوية فى الولايات المتحدة ودول أخرى.

وبالعودة للمثال الذي بين أيدينا، فإن الموظف الذي حصل علي 83 في اختبار الأداء و اللياقة لمجموع الدرجات المعطاة، والذي كان متوسطها هو 60.36 وانحرافها المعياري هو 18.61 وكان ترتيب هذا الموظف هو ال90 فإنه بحساب Z

$$Z = \frac{83 - 60.36}{18.61} = 1.22$$

وهذا يعني أن درجة هذا الموظف تبعد بمقدار 1.22 مرة من الانحراف المعياري عن متوسط قيمته صفرا، فإذا كان لدينا موظفا ثانيا يبعد عن المتوسط بدرجة 1.1 فإن ذلك يعنى أن الموظف الأول أفضل منه لأن الدرجة Z له أكبر.

ويلاحظ أنه إذا كانت قيمة Z موجبة دل ذلك علي أن النقطة محل الاختبار تقع يمين المتوسط (أعلى من المتوسط)، وأما إذا كانت قيمة Z سالبة دل ذلك علي أن النقطة محل الاختبار تقع يسار المتوسط (أقل من المتوسط).

$$\frac{X-\bar{X}}{S} = \text{بعد النقطة عن المتوسط بعدد من الـ } Z$$

أي نقول أن النقطة محل الدراسة تبعد عن المتوسط العام لباقي النقط بمقدار كذا انحراف معياري وكلما كانت هذه القيمة كبيرة كلما كان موضع هذه النقطة بعيدا عن المتوسط، وكلما قلت هذه القيمة كلما قل بعد هذه النقطة عن المتوسط، مع التأكيد على أن البعد عن المتوسط يمكن أن يكون إيجابا (أعلى من المتوسط) أو سلبا (أقل من المتوسط).

مثال رقم 11- 9:

إذا حصل أحد الطلاب فى أحد الفصول على درجة 57 فى امتحان الرياضيات وكان المتوسط الحسابي لهذا الامتحان هو 50 درجة وكان الانحراف المعياري 3.5 درجة، وإذا حصل طالب آخر فى فصل آخر على 100 درجة فى امتحان الرياضيات وكان المتوسط الحسابي لهذا الامتحان هو 80 درجة وكان الانحراف المعياري 20 درجة فأى الطالبين أفضل فى الأداء؟

الحل :

$$Z_1 = \frac{X-\bar{X}}{S} = \frac{57-50}{3.5} = 2$$

$$Z_2 = \frac{X-\bar{X}}{S} = \frac{80-100}{20} = 1$$

وحيث إن الدرجة المعيارية للطالب الأول أكبر من درجة الطالب الثاني، فإن الطالب الأول أفضل أداءا من الثاني.

ويمكن استخدام هذا المفهوم عند تقييم أداء الموظفين الذين يعملون فى أكثر من إدارة أو أكثر من فرع لذات الشركة بشكل عادل وموضوعى تماما،، فإذا حصل أحد الموظفين فى إدارة ما أو فى فرع ما على درجة 68 من مائة، وحصل موظف ثان فى إدارة أخرى أو فى فرع آخر على نفس الدرجة، فليس معنى ذلك أن أداء كلا الموظفين متساو (اذ أن ظروف وضغوط وبيئة العمل وحجم ودرجة تعقد وأهمية هذا العمل تختلف من وظيفة لأخرى ومن مكان لآخر لنفس الوظيفة)، ذلك لأننا لم نأخذ فى الاعتبار المتوسط العام والانحراف المعياري لدرجات باقي الموظفين فى كلا الإدارتين أو الفرعين.

مثال رقم 11- 10:

إذا تصادف تساوى درجات التقييم الشهري لموظفين أ، ب انضما حديثا لإحدى الشركات ويعملان فى قسمين مختلفين وحصلا على درجة 82 من مائة، وأرادت إدارة الشركة اختيار

أحدهما لتكريمه، وكانت درجات الإدارتين كما يلي فى جدول 11-10 فمن هو الموظف الأكثر جدارة بالتكريم :

جدول رقم 11- 10 درجات التقييم الشهري لموظفين أ، ب

الإدارة الأولى	75	71	80	82	86	77
الإدارة الثانية	80	79	76	82	85	81

الحل:

- نحسب المتوسط والانحراف المعياري لكلا الإدارتين فنجد أنه على التوالي للإدارة الأولى 78.5 و 2.17 بينما للإدارة الأخرى 80.5 و 1.23.
- نحسب الدرجة المعيارية لكلا الموظفين.

$$Z_1 = \frac{X - \bar{X}}{S} = \frac{82 - 78.5}{2.17} \cong 1.6$$

$$Z_2 = \frac{X - \bar{X}}{S} = \frac{82 - 80.5}{1.23} \cong 1.2$$

- نقارن الدرجة المعيارية لهذين الموظفين وحيث إن الدرجة المعيارية للموظف الأول (أ) أكبر من الدرجة المعيارية للموظف الثاني (ب) فالموظف الأول أفضل أداء من الثاني وأحق بالتكريم.

هـ- الدرجة المعيارية تى T Scores

رأينا عند شرح القيمة المعيارية "زد" Z أن هذه القيم تحتوى على قيم موجبة وأخرى سالبة، كما تحتوى على قيم كسرية وهذا قد يكون غير محقق للرضا التام من الناحية العملية، وبالتالي فقد لجأ الإحصائيون إلى إيجاد مقياس آخر لا يحتوى على هذه العيوب وهو القيمة المعيارية التائية أو درجات "تى" أى T Scores، وفى هذا النوع يضاف مقدار ثابت للتخلص من الإشارات السالبة فى المعيار زد، ويزداد حجم المقياس للتخلص من الكسور العشرية.

وتحسب القيمة المعيارية التائية من العلاقة:

$$T = 50 + 10Z = 50 + 10 * \frac{X - \bar{X}}{S}$$

والدرجة التائية هى درجات متوسطها يساوى 50 وانحرافها المعيارى يساوى 10، وكذلك الدرجة التائية المعدلة هى درجة معيارية متوسطها يساوى 100 وانحرافها المعيارى يساوى 15، وهى نسبة الذكاء فى اختبارات وكسلر.

مثال رقم 11- 11:

أوجد الدرجة المعيارية "زد"، والدرجة المعيارية "تى" إذا حصل أحد الطلاب على درجة 40 فى أحد الامتحانات، إذا علم أن متوسط الدرجات فى هذا الاختبار هو 64، والانحراف المعياري هو 15.

الحل:

$$Z = \frac{X - \bar{X}}{S} = \frac{40 - 64}{15} = -1.6$$

$$T = 50 + 10Z = 50 + 10 * (-1.6) = 34$$

وعلى ذلك يمكن تطبيق واستخدام هذا المعيار لتقييم مستويات أداء الموظفين أو مستوى توزيع منتجات معينة.... إلى آخره

مثال رقم 11- 12:

إذا تصادف تساوى درجات التقييم الشهري للاعبى كرة قدم أ، ب انضما حديثا لأحد الأندية، ويلعبان فى خطين مختلفين وحصلا على درجة 82 من مائة، وأرادت إدارة النادى اختيار أحدهما لتكريمه، وكانت الدرجات التى حصل عليها أقران هذين اللاعبين كما يلي فى جدول 11-11 فمن هو اللاعب الأكثر جدارة بالتكريم :

جدول رقم 11- 11 درجات التقييم الشهرى للاعبين أ، ب

المجموعة الأولى	75	71	80	82	86	77
المجموعة الثانية	80	79	76	82	85	81

الحل:

- حيث إن $T = 50 + 10Z$ ، وبما أن $Z = \frac{X - \bar{X}}{S}$ فيمكن استنتاج ان:

$$T = 50 + 10 * \frac{X - \bar{X}}{S}$$

- نحسب المتوسط والانحراف المعياري لأقران اللاعبين فنجد أنه على التوالي للمجموعة الأولى 78.5 و 2.17 بينما للمجموعة الثانية 80.5 و 1.23.
- نحسب T لكلا اللاعبين من العلاقة السابقة:

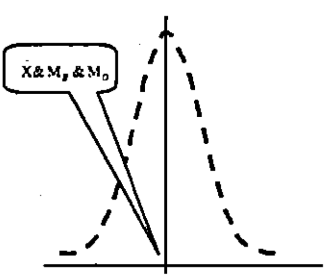
$$T_1 = 50 + 10 * \frac{X - \bar{X}}{S} = 50 + 10 * \frac{82 - 78.5}{2.17} \cong 66$$

$$T_2 = 50 + 10 * \frac{X - \bar{X}}{S} = 50 + 10 * \frac{82 - 78.5}{2.17} \cong 62$$

وحيث إن قيمة T_1 أكبر من قيمة T_2 فإن اللاعب الأول أفضل وأكثر جدارة بالتكريم من اللاعب الثانى.

11.2.4.مقاييس تحديد الشكل Measure of Shape

في كثير من الأحيان لا تكفى معرفة المتوسط والانحراف المعياري لوصف البيانات المتاحة وصفا تاما، ويكون من الضروري تحديد ما إذا كانت البيانات متماثلة Symmetrical أم لا ؟ وذلك من خلال تحديد شكل المنحنى الممثل لهذه البيانات حيث يظهر شكل المنحنى ما إذا كانت البيانات مفرطحة أم مدببة؟ وللإجابة علي هذا يوجد مقياسين لتحديد شكل المنحنى الممثل لهذه البيانات هما الالتواء Skewness، والتفرطح Kurtosis.



شكل رقم 11- 11 منحنى التوزيع الطبيعى المتماثل

أ- الالتواء Skewness

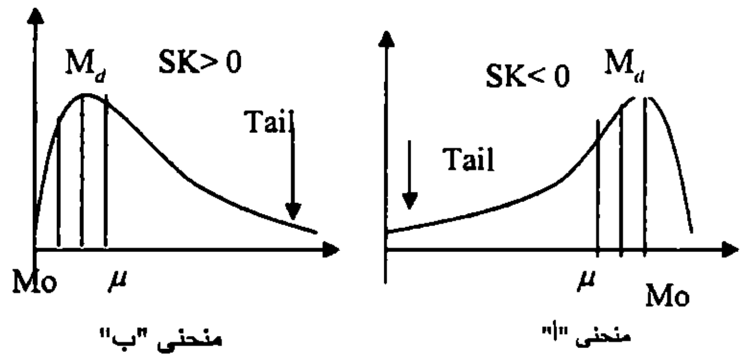
يقال أن البيانات الممثلة في الشكل 11-11 في حالة تماثل تام Perfectly Symmetric إذا كان كل من $\bar{X}\& M_B \& M_O$ ، المتوسط Mean والوسيط Median والمنوال Mode منطبقة، ويكون نصف البيانات علي يمين المتوسط $\bar{X}$ مساويا ومشابها للنصف علي يسار المتوسط، وهذه حالة مثالية لا توجد كثيرا في الواقع، وإنما يحدث أن يكون هناك ترحيل أو تشوه في هذا الشكل ويعبر عن هذا التشوه في الشكل بقيمة الالتواء Skewness والذي

يعرف على أنه بعد واختلاف المنحنى عن منحنى التماثل كما فى شكل 11-12 ويمكن حسابه بعدة طرق نذكر منها ما يلي:

$$\& \ Skewness = \frac{(\mu - M_o)}{\sigma} \ \& \ Skewness = \frac{3(\mu - M_d)}{\sigma}$$

$$Skewness = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

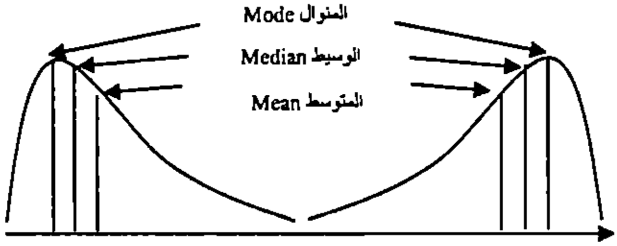
حيث μ هي المتوسط Mean & Md هي الوسيط Median & و Mo هو المنوال Mode & و σ هو الانحراف المعياري، و Q_1 & و Q_2 & و Q_3 هي على الربيع الأول والثاني والثالث على التوالي.



شكل رقم 11- 12 منحنى طبيعي ملتوى

ومقدار الاختلاف عند حساب قيمة الالتواء من المعادلات السابقة يكون بسيطا وفى الحدود المقبولة، وقيمة هذا المعامل تتراوح من 3 حتي 3-، وإذا كانت قيمته صفرا فإن المنحنى ينطبق على المنحنى الطبيعي Normal Distribution فى هذه الحالة بمعنى أن البيانات متجانسة.

وحيث إن المتوسط يمثل مركز ثقل الشكل فإنه يتحرك في اتجاه الذيل Tail، واتجاه الذيل Tail يحدد اسم المنحني، ففي المنحنى "ب" من الشكل 11-12 نجد أن الذيل متجه جهة اليمين أي جهة + أي أن قيمة الالتواء SK تكون موجبة، ويقال أن

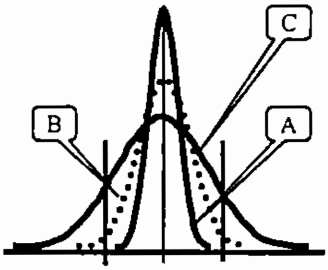


شكل رقم 11- 13 موقع المتوسط الوسط والوسيط لمنحنى ملتوى

هذا الشكل ملتو لليمين Right Skewed أو يمينى الذيل Right Tail أو موجب الالتواء Positive Skew، وفيه نجد أن القيم الكبيرة من البيانات تقع جهة الذيل، أما المنحنى "أ" فيقال أنه ملتوى لليسار Left Skewed أو يسارى الذيل Left Tail أو سالب الالتواء Negative Skew، وفيه نجد أن القيم الصغيرة من البيانات تقع جهة الذيل.

والقيم المتطرفة الموجودة بالبيانات تعمل على تحريك وسحب المتوسط الحسابي جهتها، فمثلاً إذا كانت القيم المتطرفة جهة اليمين فإنها تسحب المتوسط الحسابي إلى اليمين، وحينئذ يكون المتوسط الحسابي أكبر من الوسيط ويكون التوزيع ملتويا نحو اليمين أى موجب الالتواء والعكس صحيح كما فى شكل 11-13 (مجدى عبد الكريم حبيب 1996 صفحة 106).

ب- التفرطح Kurtosis



شكل رقم 11- 14 منحنى ذى درجات تفلطح مختلفة

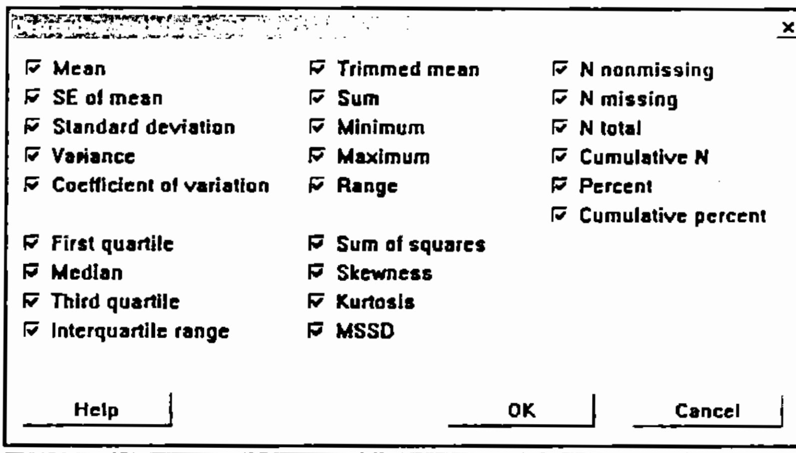
من الخصائص الأخرى التي ينبغي تحديدها هي درجة تفرطح الشكل وهذا مقياس آخر لمدي اقتراب الشكل من منحنى التوزيع الطبيعي، أى هل يميل الشكل إلي التفرطح والانتشار؟ أم إلى التدبب و الانكماش ؟.

وإذا كانت قيمة هذا المعامل صفرا يكون شكل المنحنى منطبقا على منحنى التوزيع الطبيعى كما فى الشكل B، وحينئذ يقال أن المنحنى ذو تفرطح معتدل أو Mesokurtic، وكلما قلت هذه القيمة عن الصفر وزادت سالبيتها، فإن ذلك يعنى أن القيم القريبة من المتوسط تقل والقيم البعيدة عن المتوسط تزيد (أى أن التوزيع تكون قيمه أقل تركزا فى المنتصف بدرجة أكثر من التوزيع الطبيعي المساوي له فى الانحراف المعياري)، أي أن الشكل يكون مفرطحا و منخفضا، وحينئذ يقال أن المنحنى ذو تفرطح كبير أو Platykurtic كما فى المنحنى C من الشكل 11-14، أما إذا زادت قيمة هذا المعامل عن الصفر وزادت فى الاتجاه الموجب، فإن ذلك يعنى أن القيم القريبة من المتوسط تكثر، والقيم البعيدة عن المتوسط تقل (أى تتركز قيمه التوزيع فى المنتصف بدرجة أكثر من التوزيع الطبيعي المساوي له فى الانحراف المعياري)، أي أن الشكل يكون مدببا وعاليا، وحينئذ يقال أن المنحنى ذو تفرطح قليل Leptokurtic كما فى المنحنى A. (مصطفى زايد 1989 صفحة 189).

وبتطبيق ذلك على شكل 11-14 نجد أن العملية ذات التوزيع العريض (المفلطحة) أو Flat تكون سيئة، أى أن العملية "C " هي الأكثر عرضا أي الأكثر تفلطحا أي الأسوأ، أما العملية "B " فهي أقل عرضا وتفلطحا فهي أفضل من "C " وكذلك "A " أفضل من كل من "B " و "C ".

وفى نهاية هذا الجزء أحب أن أذكرك عزيزي القارئ بما أوردناه فى مقدمة هذا الكتاب بأن هناك العديد والعديد من البرامج الإحصائية التي ستقوم بتحمل عبء الحسابات عنك بالإضافة إلى دقتها، فلا تنزعج بما تم عرضه من كثرة الأمثلة، و إسهابنا فى شرحها وتعمدنا حلها يدويا، فقد كان ذلك لتوضيح المعنى ولبيان الفكرة، والدليل على ذلك أنك بأمر بسيط للكمبيوتر – لأى مجموعة بيانات مهما كان حجمها – ستتمكن من حساب كل المقاييس والمعايير التي تحدثنا عنها فى هذا الجزء، فمثلاً يمكن استخدام برنامج الميني‌تاب Minitab فى ذلك كما فى شكل 11-15:

Minitab > Stat > Basic Statistics > Display Descriptive Statistics



شكل رقم 11- 15 إستخدام برنامج مينيتاب لحساب المقاييس المختلفة

كذلك يمكن الاعتماد على برنامج الإكسل Excel فى ذلك أيضا كما فى شكل 11-16 بعد إدخال الأوامر التالية:

Excel > Tools > Data Analysis> Descriptive Statistics

Mean	6.21375	Range	0.05
Standard Error	0.006529466	Minimum	6.19
Median	6.21	Maximum	6.24
Mode	6.2	Sum	49.71
Standard Deviation	0.018468119	Count	8
Sample Variance	0.000341071	Largest(1)	6.24
Kurtosis	-0.94712316	Smallest(1)	6.19
Skewness	0.535803772		

شكل رقم 11- 16 إستخدام برنامج الاكسل فى حساب المقاييس

11.2.5. الأرقام القياسية Index Numbers

الأرقام القياسية وسيلة فعالة وطريق مختصر لوصف المتغيرات الاقتصادية، مثل التباين فى الأسعار أو الكميات أو القيمة مع الزمن، والرقم القياسي لتكاليف المعيشة يعطينا نسبة التباين فى تكاليف الحياة بين سنة وأخرى مقاسة بتغير أسعار سلة من المواد من وقت لآخر، فمثلا لو قلنا إن الرقم القياسي لتكاليف المعيشة فى بلد ما بلغ 180 دولار عام 2005 مقارنة بعام 2000، فإن هذا يعنى أن ما تحتاجه عائلة من تكاليف عام 2005 لشراء عدد من المواد (مجموعة هذه المواد يطلق عليها سلة وتحدد عند تعريف الرقم القياسي) يزيد بمقدار 80% عما كانت تحتاجه هذه العائلة عام 2000.

ومن أهم وجوه استخدام الأرقام القياسية تقدير الدخل الحقيقي للفرد ويحسب من العلاقة:

$$\text{الرقم القياسى الحقيقى للدخل} = \frac{\text{الرقم القياسى للدخل}}{\text{الرقم القياسى لتكاليف المعيشة}}$$

فمثلا إذا كان الرقم القياسي للدخل عام 2005 بالنسبة لعام 2000 يساوى 160، و كان الرقم القياسي لتكاليف المعيشة عام 2005 بالنسبة لعام 2000 يساوى 195، فإن الرقم القياسي الحقيقي للدخل يساوى $= \frac{160}{195} * 100 = 82.05$ أى أن القوة الشرائية لدخل للفرد أصبحت 82% مما كانت عليه عام 2000.

كما تستخدم الأرقام القياسية كذلك لقياس التباين فى الإنتاجية، والبطالة، ونمو الموارد، والأجور وغيرها، وسنكتفي بتناول وعرض بعض طرق حساب الأرقام القياسية للأسعار والكميات والنقود فقط كما يلي:

أ- الرقم القياسي للأسعار Price Index Number

الرقم القياسي للأسعار Price Index Number هو عدد واحد يوضح لنا كيف تغيرت مجموعة كاملة من الأسعار خلال فترة زمنية محددة، ويتم حسابه بأكثر من طريقة نتناول منها طريقتين فقط:

1. الرقم القياسي التجميعي Simple Aggregate Index وهو نسبة مجموع أسعار عدة سلع فى سنة ما (تسمى سنة المقارنة) إلى مجموع أسعار هذه السلع فى سنة أخرى (تسمى سنة الأساس)، ويحسب من العلاقة:

$$I_{Agregate} = \frac{\sum P_n}{\sum P_o} * 100$$

2. الرقم القياسي النسبي البسيط Simple Relative Index وهو الوسط الحسابي للأرقام القياسية للسلع، أى أننا نوجد الرقم القياسي لكل سلعة ثم نوجد الوسط الحسابي لهذه الأرقام القياسية ويحسب من العلاقة :

$$I_{Relative} = \frac{1}{N} * \frac{\sum P_n}{\sum P_o} * 100$$

حيث p_n هى الأسعار فى سنة المقارنة، وحيث p_o هى الأسعار فى سنة الأساس، وحيث N هو عدد السلع.

ب- الرقم القياسي للكميات Quantity Index Number

لما كانت كميات الإنتاج أو الاستهلاك تتغير مع الزمن، فقد أصبح ضروريا أن نقارن التباين فى الكميات من فترة لأخرى، ويتم حساب الأرقام القياسية للكميات بنفس طريقة حساب الأرقام القياسية للأسعار كما يلي:

1. الرقم القياسي التجميعي Simple Aggregate Index، ويحسب من العلاقة

$$I_{Aggregate} = \frac{\Sigma Q_n}{\Sigma Q_o} * 100$$

2. انرقم القياسي النسبي البسيط Simple Relative Index، ويحسب من العلاقة

$$I_{Relative} = \frac{1}{N} * \frac{\Sigma Q_n}{\Sigma Q_o} * 100$$

حيث Q_n هى الكميات فى سنة المقارنة، وحيث Q_o هى الكميات فى سنة الأساس، وحيث N هو عدد السلع.

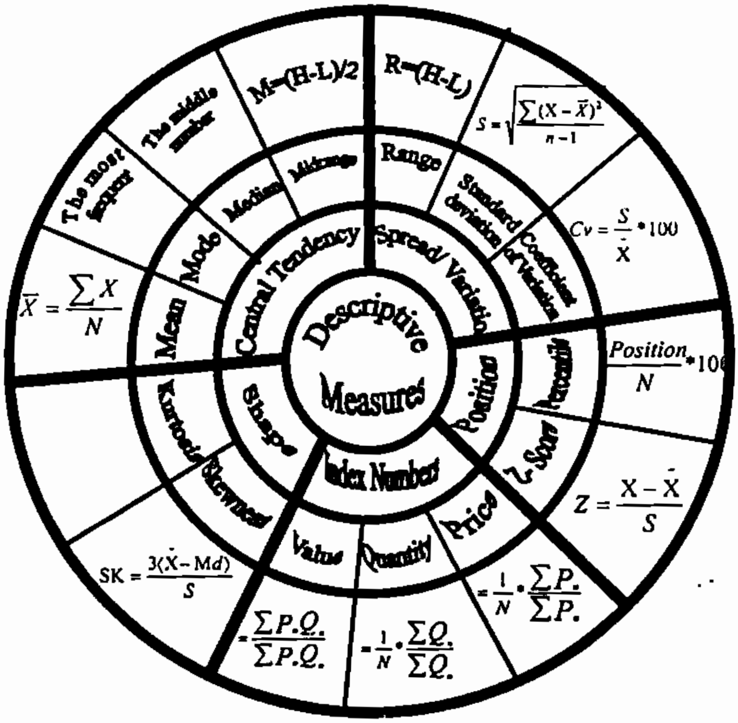
ج- الرقم القياسي للقيمة Value Index Numbers

إضافة إلى الأرقام القياسية للأسعار والكميات، فإننا نحتاج فى بعض الأحيان استعمال الأرقام القياسية للقيمة النقدية للسلع الاستهلاكية المستعملة والداخلة فى حساب الرقم القياسي فى فترة المقارنة T، وتحسب من العلاقة:

$$I_t = \frac{V_n}{V_o} = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_o}$$

والقيمة القياسية Value Index تقيس القيمة المالية للمتغير، أى تقيس القيمة النقدية بالجنيه مثلا لمتغير معين، و الرقم القياسي للقيمة هو أكثر هذه الأرقام دلالة، إذ أنه يأخذ الرقمين السابقين فى الاعتبار عند حسابه (محمد صبحي ابوصالح 2001).

والشكل 11-17 يعرض ملخصا للمقاييس الوصفية التى تم شرحها حتى الآن:



شكل رقم 11- 17 ملخص للمقاييس الوصفية المختلفة (للمؤلف)

11.3. عرض البيانات باستخدام الرسوم البيانية Data Presenting Using Graphs

تعرفنا من خلال الأجزاء السابقة علي أنواع مختلفة من البيانات، وتعرفنا كذلك على طريقة عرض هذه البيانات عن طريق اختزال وتحويل الأرقام الكثيرة إلي قيم ومعايير أقل عددا ومتفق عليها وتعطي انطباعا عن إجمالي الأرقام الأصلية مثل المتوسط Mean والوسيط Median والانحراف المعيارى Standard Deviation و......

والآن سنحاول معرفة كيفية اختزال وتحويل الأرقام الكثيرة إلي رسومات بيانية Graphs أسهل فهما وأكثر تعبيرا (راجع شكل 11-1).

وسوف نقدم فى هذا الفصل ملخصاً لأكثر طرق عرض البيانات ونتائج الدراسات شيوعاً ومنها:

أ- مخطط التوزيع التكراري Frequency Distribution

كما قلنا أن المشكلة تكمن فى وجود كم كبير من البيانات يصعب فهمه واستقراؤه، وأولي الطرق البيانية التي تختزل هذه البيانات الكثيرة والهائلة إلي رسومات يسهل عرضها وفهمها هي التوزيع التكرارى Frequency Distribution فلو تخيلنا أن لدينا مجموعة أرقام تتألف من 500 رقم وهي تعبر مثلا عن أعمار مجموعة من الأشخاص فبوضع الـ500 رقم أمامك لن تستطيع استخلاص أي معلومة، ولكن إذا قسمناهم إلي شرائح كما بالجدول 11-12:

جدول رقم 11- 12 أعمار مجموعة من الأشخاص

Class الفئة او الشريحة	Age العمر	Frequency التكرار
1	$\le 10\ years$	50
2	$10 \le age < 25$	73
3	$25 \le age < 40$	132
4	$40 \le age < 60$	207
5	$60 \le age$	38
		500

فإن ذلك يعطي صورة عن أن لدينا شريحة من الأشخاص الذين تتراوح أعمارهم بين 60 & 40 سنة بنسبة 41%= $\left(\frac{207}{500}\right)$ وأن الشريحة التي تليها هي الأشخاص التي يتراوح

أعمارها بين 25 & 40 سنة بنسبة =26% $\left(\frac{132}{500}\right)$ وأن نسبة المواليد والذين تقل أعمارهم عن 10 سنوات هي 10%= . $\left(\frac{50}{500}\right)$

وعلي ما سبق فإن التوزيع التكرارى Frequency Distributions قلص عدد البيانات الكثيرة إلي عدد أقل أسهل فهماً و استقراءً.

وعند تكوين التوزيع التكرارى يجب ملاحظة ما يلى:

1- عدد الشرائح Classes أو (K)اختياري وأنت الذي تحدده ومع ذلك يفضل أن يكون من 5 إلي 20 وبالطبع فإن عدد هذه الشرائح (K) تتناسب مع حجم البيانات المعالجة.

2- عرض الشريحة Class Width أو $CW = \frac{H-L}{K}$ حيث H هي أكبر قيمة في مجموعة الأرقام & L هي أصغر قيمة في مجموعة الأرقام، ويفضل أن تكون CW قيمة واحدة لكل الشرائح.

خطوات تجهيز جدول التوزيع التكرارى Frequency Distributions

1- نرتب مجموعة الأرقام المعطاة ترتيبا تصاعديا من الأصغر إلي الأكبر.

2- تختار K بقيمة تتناسب مع الأرقام.

3- نحسب عرض الشريحة Class Width من العلاقة $CW = \frac{H-L}{K}$ مع تقريب قيمة عرض الشريحة CW إلى قيمة يسهل التعامل معها فإذا كان عرض الشريحة مساويا 89.1 نختار 100، وإذا كان عرض الشريحة 1.37 نختار 1.5، وإذا كان عرض الشريحة 47.3 نختار 50، وإذا كان عرض الشريحة 12.2 نختار 10، وهكذا .

4- نعد أرقام كل شريحة من الأرقام المعطاة ثم نكون جدول التكرارات.

وسوف تتضح تلك الخطوات من المثال التالى:

مثال رقم 11-13

المطلوب رسم منحنى التوزيع التكرارى Frequency Distributions لمجموعة البيانات المذكورة فى جدول 11-13 و التى تعبر عن الأعمار السنية لمجموعة من الشباب

جدول رقم 11- 13 الاعمار السنية لمجموعة من الشباب

31.5	29.1	30.9	25.9	27.4	29.5	30.3	29.3	31.6	26.6
25.7	33.7	27.0	31.1	30.6	28.0	32.4	25.7	31.4	29.2
29.3	33.8	28.5	33.0	26.3	25.6	26.2	28.1	24.8	28.1
26.5	29.5	27.9	24.3	26.8	23.8	25.0	27.8	28.7	27.2
28.2	27.0	29.7	28.8	25.2	26.2	31.1	26.2	25.7	22.8

الحل: نرتب هذه الأرقام من الأصغر للأكبر (كما فى جدول 11-14) ثم نختار مثلا 6 شرائح ثم نحسب عرض الشريحة Class Width

جدول رقم 11- 14 التوزيع التكرارى للمثال11-13 بعدد ست شرائح

Class # رقم الشريحة	Class الشريحة	Frequency التكرار	Relative frequency
1	$22 \le C < 24$	2	4%
2	$24 \le C < 26$	9	18%
3	$26 \le C < 28$	13	26%
4	$28 \le C < 30$	13	28%
5	$30 \le C < 32$	8	16%
6	$32 \le C < 43$	4	4
		50	100%

ويمكن أيضا اختيار 12 شريحة كما فى جدول11-15:

جدول رقم 11- 15 التوزيع التكرارى للمثال11-13 بعدد 12 شريحة

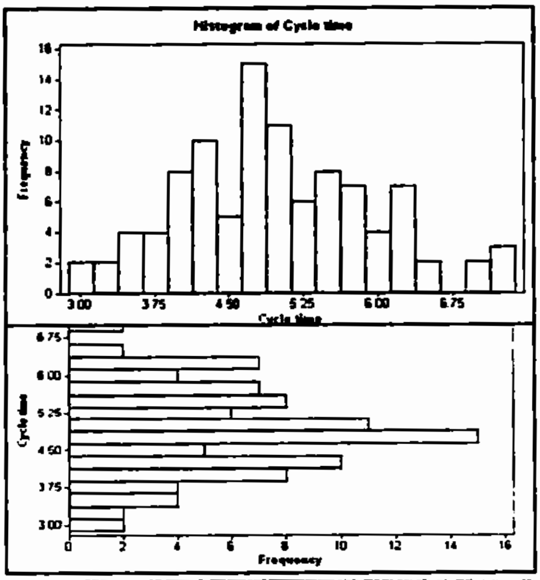
Class # رقم الشريحة	الشريحة Class	Frequency التكرار	Relative frequency
1	$22 \le C < 23$	1	2%
2	$23 \le C < 24$	1	2%
3	$24 \le C < 25$	2	4%
4	$25 \le C < 26$	7	14%
5	$26 \le C < 27$	7	14%
6	$27 \le C < 28$	6	12%
7	$28 \le C < 29$	7	14%
8	$29 \le C < 30$	7	14%
9	$30 \le C < 31$	3	6%
10	$31 \le C < 32$	5	10%
11	$32 \le C < 33$	1	2%
12	$33 \le C < 34$	3	6%
		50	100%

ب- الهستوجرام **Histogram**

بعد أن نكون قد فرغنا من تجهيز التوزيع التكرارى Frequency Distribution كجدول، يمكننا الآن تحويله إلى شكل بيانى يعبر عن هذه البيانات والتي بالطبع ستكون أكثر توضيحا لهذه البيانات.

والهستوجرام Histogram عبارة عن رسم بياني يوضح تكرار البيانات Frequency Distribution و يعكس شكلها، ومن خلاله يمكن الإجابة سريعا علي الأسئلة: هل البيانات متماثلة أم لا ؟ وأين تقع أكثر القيم تكرارا؟، وارتفاع أعمدة الهستوجرام تتناسب مع تكرار البيانات، ويجب أن تكون الأعمدة متجاورة بدون فواصل، وليس شرطا أن يبدأ ترقيم المحور الأفقي من الصفر.

و يظهر الشكل العام للهستوجرام كما فى شكل 11-18.



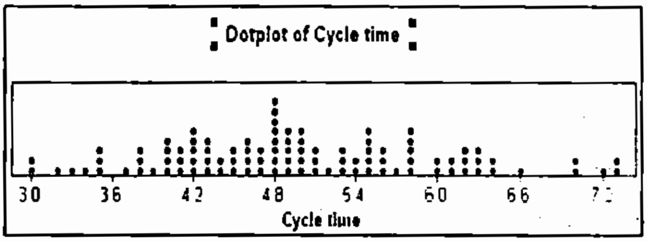
شكل رقم 11- 18 نماذج للهستوجرام

وفى بعض الحالات يمكن رسم الهستوجرام أفقيا بدلا من رأسيا وذلك لتسهيل الكتابة داخل الأعمدة.

وإذا تم قسمة قيمة كل عمود (التكرارات) من أعمدة الهستوجرام على إجمالى قيم الأعمدة (إجمالى التكرارات) فإننا نحصل على ما نطلق عليه دالة التوزيع الاحتمالية Probability Distribution Function (PDF) (يرجى مراجعة الفصل الثانى عشر تحت عنوان الاحتمال والتوزيعات الاحتمالية فى هذا الكتاب).

ج- رسم النقطة **Dot Plot**

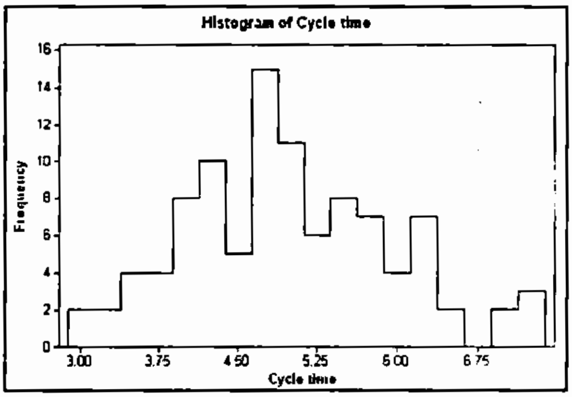
و فى بعض الحالات تستبدل الأعمدة بنقاط ● كما فى شكل 11-19 وفى هذه الحالة يسمى هذا الرسم برسم النقطة أو Dot Plot.



شكل رقم 11- 19 نموذج لرسم النقطة

د- رسم المساحة **Area Plot**

وهو وسيلة أخري لرسم البيانات وهو يأخذ منتصف قمة أعمدة الهستوجرام ثم يوصلها ببعضها و هنا نلاحظ أن هذا الشكل يبدأ بالصفر وينتهي بالصفر كما بالشكل 11-20، ولرسم هاتين النقطتين نتخيل وجود شريحة قيمتها صفر وعرضها كعرض باقي الشرائح.



شكل رقم 11- 20 نموذج لرسم المساحة

ه- مخطط الساق والورقة **Stem and Leaf Diagram**

ينسب هذا الشكل إلي السيد جون توكى John Tukey ، وهذا المخطط وسيلة مفيدة جدا في رسم البيانات التي يقل عددها عن 150 رقم، - ويجدر بالإشارة هنا إلي الهستوجرام يكون أكثر تعبيرا وأسهل تكوينا عن الساق والورقية Stem & Leaf في حالة الأرقام الكثيرة (أكثر من 150 رقم)- ويتميز عن الهستوجرام في أنه لا يتسبب في ضياع الأرقام الأصلية، فمنه يمكن استرجاع الأرقام مرة أخري بعكس الهستوجرام فإنني لا أستطيع استعادة الأرقام

الأصلية مرة أخري، كذلك نستطيع بعد رسم الساق و الورقة Stem And Leaf تحديد قيمة الوسيط Median وهو الرقم الذي يتوسط الأرقام المتاحة.

ويستخدم الساق الورقية لرسم صورة بيانية للبيانات المعطاة وفيه نقسم الأرقام إلي ساق أو جزء رأسي Stem وبجواره الأوراق Leafs فى خط أفقي، وإنشاء هذا المخطط أمر فى غاية السهولة، ولا يتطلب إلا فصل أرقام كل عدد فى البيانات المتاحة إلى جزأين أحدهما هو الساق Stem والأخر يسمى الورقة Leaf، والمعتاد أن تؤخذ الورقة على أنها الرقم الأخير فى العدد أى الرقم الذي فى أقصى اليمين، أما بقية الأرقام فهي الساق.

مثال رقم 11- 14:

فإذا كان لدي مجموعة من الأرقام مثل التي بالجدول رقم 11-16

جدول رقم 11- 16 مجموعة ارقام تشير الى القطر الخارجي لماسورة معدنية

3.4	4.1	3.6	4.7
4.5	3.8	4.1	3.4
2.3	3.9	3.4	4.1
2.7	3.9	5.1	4.1
3.8	4.1	3.4	2.7
5.9	4.5	2.4	2.3

فإن شكل الساق الورقية Stem & Leaf سيكون كما هو موضح بشكل 11-21

```
3     2   334
5     2   77
8     3   444
(5)   3   68899
9     4   11111
4     4   557
1     5
1     5   9
```

شكل رقم 11- 21 رسم الساق والورقة لبيانات الجدول 11-16

وبالنظر إلي هذا الشكل نجد أننا اعتبرنا الأرقام الصحيحة هي الساق Stem وتم كتابتها رأسيا تصاعديا، واعتبرنا الأرقام العشرية هي الورقة Leaf وكتبناها أفقيا كلما تكررت فى سطرين (الأول أقل من 0.5 والثاني أكبر من 0.5). أي أمام الرقم 2 نبحث في الأرقام عن

أي رقم يحتوي على 2 كعدد صحيح فنجد أن أول رقم يقابلنا هو 2.3، إذن أمام الرقم 2 الأول أكتب 3، ثم يقابلني رقم 2.3 أكتب 3، ثم يقابلني رقم 2.4 أكتب 4(كل ذلك فى السطر الأول)، ثم يقابلني الرقم 2.7 مرتين أكتب 7 مرتين، ونكرر ذلك مع الرقم 3 & 4 & 5.

أما العمود الأول الرأسي فيشير إلى عدد الأرقام العشرية التي تم تسجيلها (التكرارات)، فالرقم 3 يعنى أننا سجلنا ثلاثة أرقام، والرقم 5 يعنى أننا سجلنا خمسة أرقام، ويظل الحال هكذا حتى نصل إلى الرقم بين القوسين والذي يشير إلى أن هذا السطر يحتوى على قيمة الوسيط Median، وحينما نصل إلى هذا الرقم نبدأ تسجيل التكرارات من النهاية الأخرى للساق والورقة Stem And Leaf، وبالتالي يهمنا ثلاثة أرقام فى هذا العمود وهى:

- الرقم بين القوسين (إن وجد) يشير إلى أن الوسيط Median يقع فى هذا السطر، وقيمة هذا الرقم تشير إلى عدد الأرقام العشرية.
- الرقم الذي يسبق الرقم بين القوسين يشير إلى مجموع التكرارات التي تسبقه.
- الرقم الذي يلي الرقم بين القوسين يشير إلى مجموع التكرارات التي تليه.

ويلاحظ أن مجموع هذه الأرقام الثلاثة يساوى مجموع الأرقام المتاحة بالجدول. ولمعرفة عدد مرات تكرار قيمة ما يتم طرح قيمة الرقم الأيسر(فى نفس السطر الأفقى) من الرقم الذي يسبقه مباشرة.

مثال رقم 11- 15:

سجل أحد المراصد القراءات الموضحة بالجدول 11-17 والتي تعبر عن شدة الزلازل فى أحد المناطق بقياس ريختر Richter، والمطلوب رسم الساق والورقة Stem And Leaf والتعليق على الشكل الناتج.

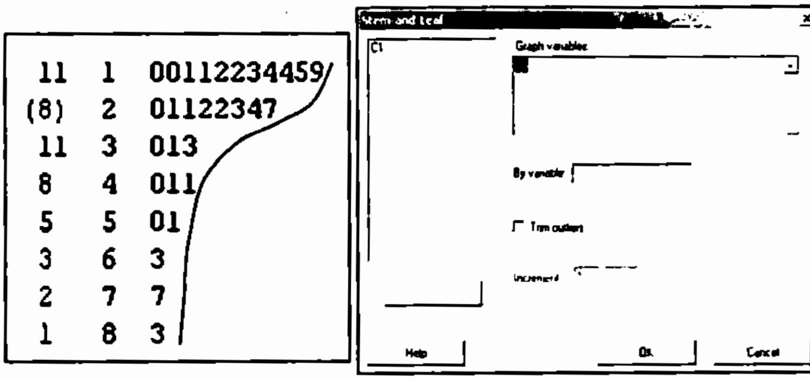
جدول رقم 11- 17 التوزيع التكرارى للمثال11-15 بعدد ست شرائح

1	1.2	2	3.3	1.4	5	1.1	1.1	1.4	2.1
8.3	1	1.9	2.2	2.7	2.2	5.1	4	1.3	2.1
3.1	4.1	6.3	2.3	2.4	1.2	3	7.7	4.1	1.5

الحل:

نعتبر أن الآحاد فى مجموعة الأرقام المتاحة هي الساق، وأن الجزء العشري هو الأوراق، ونكتب الأرقام من 1 حتّى 8 فى خط رأسى، ثم نسجل الأوراق كلّ أمام الساق المناسبة.

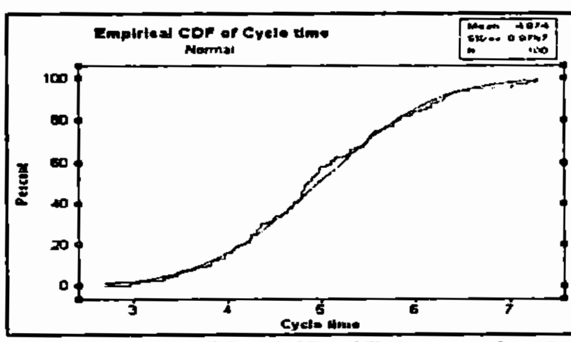
وباستخدام المينيتاب Minitab ندخل البيانات المتاحة، ثم من القوائم Graph > Stem and Leaf فيظهر الجدول الحواري ثم يليه الشكل النهائي كم فى شكل 11-22 موضحا به الساق والورقة Stem and Leaf .



شكل رقم 11- 22 رسم الساق والورقة للمثال 11-15 بالمنيتاب

ومنه نستخلص أن شدة الزلازل تتشتت وتتوزع بين القيمتين 1 و 8.3، غير أن البيانات تميل إلى التجمع حول القيم الصغرى وتقل تدريجيا فى اتجاه القيم الكبرى (هناك التواء إلى اليمين) وهذا يعنى أن معظم الزلازل فى هذه العينة كانت خفيفة، وإذا كانت هذه العينة تعبر عن المجتمع ككل، فإن وقوع زلازل شديدة فى هذه المنطقة يكون أمرا بعيد الاحتمال.

و- مخطط اجمالى التوزيعات Cumulative Frequencies



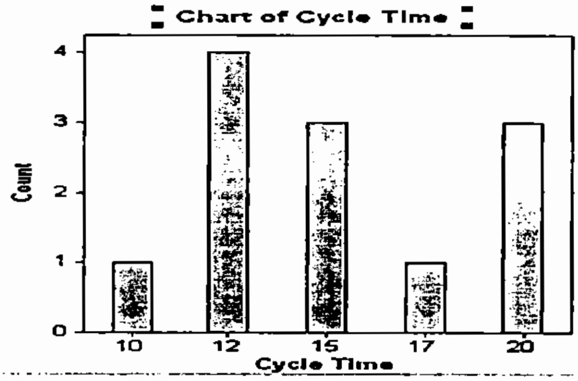
شكل رقم 11- 23 نموذج لمخطط اجمالى التوزيعات

وهو طريقة أخرى لعرض البيانات ومن اسمه يتضح أن كل نقطة تجمع ما قبلها فبعد ترتيب الأرقام وتصنيفهم تبعا للتوزيع التكرارى Frequency Distribution، نضيف عمودا يتم فيه جمع القيم السابقة كما فى شكل 11-23، وفائدة هذا النوع أنه يفيد في تحديد وحساب نسبة الأرقام الأقل من أو الأكبر من قيمة معينة وهو يبدأ بالصفر وينتهي ب 100%.

ملاحظة هامة:-يستخدم كل من Histogram & Cumulative Area Plot & Frequenciesلتمثيل البيانات من النوع الفترى Interval، والنسبى Ratio .

ز- مخطط الأعمدة Bar Chart

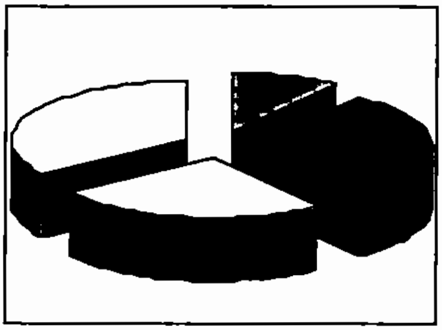
وهو نوع يشبه الهستوجرام غير أنه يستخدم لعرض البيانات مع المتغيرات الاسمية Nominal، و الترتيبية Ordinal وغالبا ما توجد فواصل وثغرات بين الأعمدة، ويمكن رسمه رأسيا أو أفقيا لسهولة كتابة تعليمات داخل الأعمدة والشكل 11-24 يعرض نموذجا لهذا النوع.



شكل رقم 11- 24 نموذج لمخطط الاعمدة

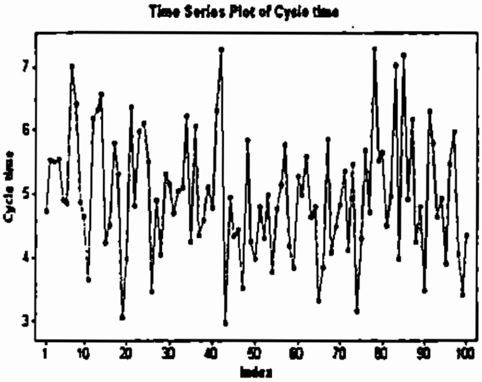
ح- مخطط الدائرة Pie Chart

ويستعمل هذا الأسلوب أيضا لعرض البيانات مع المتغيرات الاسمية Nominal، والترتيبية Ordinal كما فى شكل 11-25، وهو يوضح الأجزاء المكونة لعنصر ما، وغالبا ما تكون هذه الأجزاء عبارة عن نسب ويعبر عن الأجزاء الممثلة لهذه النسب بمساحة محصورة بخطين يخرجان من مركز الدائرة حتي محيطها.



شكل رقم 11- 25 نموذج لمخطط الدائرة

ط- السلاسل الزمنية Time Series



شكل رقم 11- 26 نموذج للسلاسل الزمنية

وفيها يتم توقيع القيم مباشرة على الرسم (شكل 11-26) بترتيبها الزمني وهو يعطى صورة مرتبطة بتوقيت حدوث القيم.

والسلاسل الزمنية هي مجموعة من البيانات المأخوذة خلال فترة زمنية محددة، وتختلف هذه الفترة فتكون ساعات في بعض الحالات كدراسة سلوك البكتريا في المعمل أو معدلات الارتفاع والانخفاض في سعر الأسهم، وتكون شهورا عند دراسة مبيعات الشركات و المؤسسات، وتكون سنين عند دراسة الفرص التسويقية، وتكون عقودا عند دراسة التغيرات الاجتماعية والمناخية.

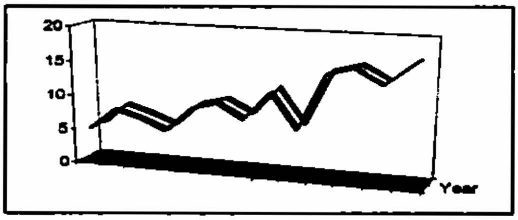
والهدف الرئيسي من تحليل هذه السلاسل الزمنية هو اكتشاف هذه التأثيرات وقياس نتائجها بشكل تقريبي إذ لا يمكن ذلك بشكل يقيني.

ودراسة هذا النوع من طرق عرض البيانات له أهميته لكونه يشكل أداة قوية تعين على فهم السلوك الماضي، و تساعد في تقويم الإنجازات في الحاضر، وتسهم في تخطيط العمليات في المستقبل.

و تتأثر بيانات السلاسل الزمنية عادة بأربعة بعوامل وهي:

1. **الاتجاه العام (T) Trend** : وهو مؤشر لبيان تزايد ظاهرة أو تراجعها على مدى الفترة الزمنية تحت الدراسة، وهناك طرق عديدة لتحديد الاتجاه سنتناول منها نوعين هما:

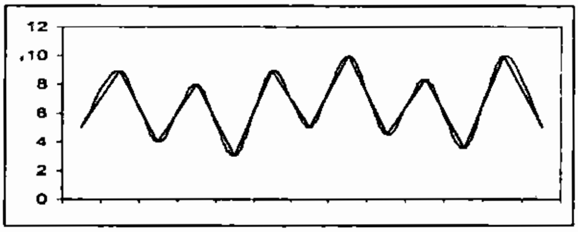
- **طريقة المعدل النصفي Semi-Average Method**: (شكل 11-27) وفيها نقوم بتقسيم البيانات المتاحة إلى قسمين متساويين ونحسب متوسط كل قسم ثم نصل بين هاتين النقطتين.(إذا كان عدد البيانات المتاحة عددا فرديا يتم حذف الرقم الأوسط).



شكل رقم 11- 27 نموذج لمنحنى المعدل النصفى

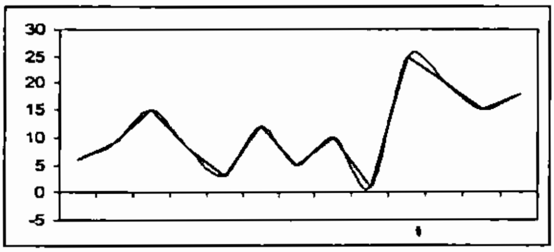
- طريقة مجموع المربعات الصغرى **Least Sum of Squares** : وهى الطريقة المستعملة فى تعيين خط الانحدار البسيط (أى رسم خط مستقيم يتوسط البيانات المتاحة) كما سيأتى تفصيله فى الفصل الخامس عشر تحت عنوان الانحدار الخطى فى هذا الكتاب إن شاء الله.

2. التغيرات الفصلية**Seasonal Variations (S)** : وهى تغيرات تحدث فى تتابع متسق على فترات زمنية محددة لا تتجاوز السنة الواحدة (شكل 11-28)، وكلمة "فصلية" لا تعنى بالضرورة فصول السنة، ولكنها تعنى أى تباين له طبيعة دورية، وتكون دوراته المتكررة ذات مدة زمنية قصيرة نسبيا، ومثال ذلك أسعار السلع الزراعية التي تكون عالية فى بداية الموسم ثم تنخفض بشدة عند ذروة الإنتاج ثم ترتفع ثانية عند انتهاء الموسم.



شكل رقم 11- 28 نموذج لمنحنى التغيرات الفصلية

3. التقلبات الدورية **Cyclical Fluctuation (C)** : وهى تغيرات تحدث أيضا فى تتابع متسق و على فترات زمنية محددة ولكنها تكون فترات طويلة نسبيا (شكل 11-29). مثل الدورات الاقتصادية والموضة.



شكل رقم 11- 29 نموذج لمنحنى التغيرات الدورية

التغيرات العشوائية **Random Or Erratic Variations (R)**: وهى تغيرات تحدث بشكل لا نمطي وعشوائي وتعزى لأسباب غير مرنية وغير متوقعة مثل الحرائق و الزلازل و الأحداث السياسية والحروب، وعموما هي تلك التغيرات التي لا تفسر عن طريق الاتجاه العام، أو التغيرات الفصلية، أو التقلبات الدورية.

وفى الواقع فإنه توجد طرق عديدة لوصف السلاسل الزمنية بدلالة هذه العوامل مشتركة مثل:

النموذج الضربى $Y = T*S*C*R$

النموذج الجمعى $Y = T+S+C+R$

النموذج المختلط $Y = (T*S*C)+R$ & $Y = (T*S)+(C*R)$

و أكثر هذه الطرق شهرة واستخداما هو الطريقة الأولى والذي يطلق عليه "النموذج الضربى"، كذلك توجد طرق كثيرة وعديدة لتحديد وفصل التأثيرات السابق الإشارة إليها عن بعضها البعض.

ي- مخطط الصندوق والشنب **Box & Whisker (Box Plot)**

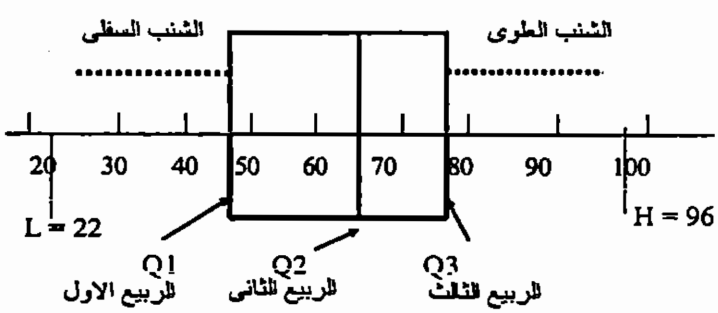
يعد مخطط الصندوق Box Plot من الأدوات الهامة لبيان توزيع البيانات و مدى تماثلها، وهو عبارة عن رسم بياني يشبه الصندوق، ومن خلاله يمكن الإجابة سريعا علي السؤال هل البيانات متماثلة أم لا ؟ وإلى أى جهة تميل هذه البيانات؟

وكما أشرنا في الجزء الأول من تقديم البيانات من خلال القياسات Measures تحت عنوان الربعيات Quartiles فإنه أمكننا حساب أقل قيمة في الأرقام وكذلك حساب القيمة التي يسبقها ربع الأرقام والقيمة التي يسبقها نصف الأرقام والقيمة التي يسبقها ثلاثة أرباع الأرقام وأكبر قيمة في الأرقام، وفي المثال 11-6 الذي تم شرحه في هذا الجزء كانت القيم كما يلي:

$$L = 22 \,\&\, Q_1 = 46 \,\&\, Q_2 = M_d = 62 \,\&\, Q_3 = 75$$

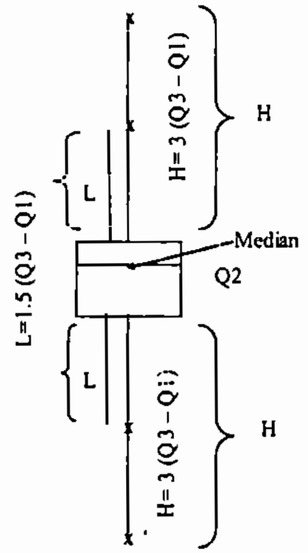
$$H = 96 \,\&\, IQR = Q_3 - Q_1 = 29$$

وبرسم هذه القيم وعمل صندوق Box حدوده هي $Q_1 \& Q_3$ وله شنب من كل جهة طوله L &H كما فى شكل 11-30



شكل رقم 11- 30 مخطط الصندوق والشنب للمثال 11-6

ملاحظات علي مخطط الصندوق Box Plot



شكل رقم 11- 31 منحنى الصندوق والشنب

- بداية ونهاية الصندوق Boxهي $Q_1 \& Q_3$.
- ارتفاع الصندوق Box هو Inter Quartile = $Q_3 - Q_1$.
- لا يشترط أن يكون الوسيط Median في منتصف الصندوق Box ولكن وجوده في المنتصف يدل علي تماثل البياناتSymmetrical Data.
- وجود خط الوسيط Median في النصف الأعلى من الصندوق Box يدل علي أن توزيع البيانات ملتو جهة اليمين Right Skewed، أو موجب الالتواء Positive Skewed، أو مذيل يمينا Right Tail.
- وجود الـMedian في النصف الأسفل من الصندوق Box يدل علي أن توزيع البيانات ملتو جهة اليسار Left Skewed، أو سالب الالتواء Negative Skewed، أو مذيل يسارا Left Tail.
- يوجد شنب وهمي قصير L طوله مرة ونصف من عرض الصندوق Box، أي يساوى $1.5*(Q_3 - Q_1)$.
- يوجد شنب وهمي طويل H طوله = ثلاثة أمثال عرض الصندوق، أى $3*(Q_3 - Q_1)$.
- القراءات التي تقع في مدي الشنب L القصير تسمي Mild Outlier، بينما القراءات التي تقع في مدي الشنبH الطويل تسمي Extreme Outlier.
- سيقع 25% القراءات أسفل الصندوق Box و25% أعلي الصندوق Box و50% داخله.
- طول الشنبين هما L & H ويكون الالتواء Skewness في اتجاه الشنب الأطول.
- إذا كانت البيانات المتاحة طبيعية فى توزيعها Normal، فإن عدد قراءات Mild Outlier لا يتجاوز 10% وعدد قراءات Extreme Outlier لا يتجاوز 1%.
- يستخدم الـBox Plot لتحديد النقاط الغريبة/الشاذة Outlier كما يكون مؤشرا لشكل البيانات Data Pattern وكذلك يوضح مدى التباين أو الاختلاف Variability.
- كلما كان حجم الصندوق Box صغيرا ، دل ذلك علي جودة العملية وقلة وانخفاض التباين أو الاختلاف Variability.

- عند مقارنة أكثر من عملية تكون العملية الممثلة بصندوقBox أصغر هي الأفضل.
- عند ظهور حرف (S) في مخطط الصندوق Box Plot مرسوم بالكمبيوتر يعني أنه يوجد Outlier.

خطوات رسم مخطط الصندوق Box Plot

- نرتب الأرقام ترتيبا تصاعديا.
- نحدد $Q_1 \& Q_3$ والوسيط $Q_2 =$ Median، وكذلك عرض الصندوق $= IRQ = Q_3 - Q_1$
- نحسب طول الشنب الصغير 1.5*IRQ، و طول الشنب الطويل 3*IRQ
- نحدد طول الشنب الأصلي (الحقيقي) من L & H
- نحدد النقط الغريبة/الشاذة Mild Outlier وExtreme Outlier ونحاول تحديد موقفها وتفسير وجودها.

ك- الرسوم المبهرة Deceptive Graphs

وهي أنواع تعتمد علي الخلق والإبداع كالرسوم المساحية و الثلاثية الأبعاد وكذلك توقيع منحنيين علي Graph واحد ويساعد كثيرا في هذا برنامج Excel

والشكل 11-1 الذى تم تقديمه فى مستهل هذا الفصل يعرض ملخصاً للطرق المختلفة لعرض البيانات الرقمية سواء كان بالمقاييس الرقمية Measures أم بالأشكال Graphs والتى تم شرحها جميعاً.