Department of Statistics
Faculty of Science
Yarmouk University

SATS 101

Introduction to Probability and Statistics

Yarmouk University

Second Semester

2009/2010

Done by: Osama Alkhoun
Mobile: 079648461З

Chapter 2

Describing Data with Numerical Measures

- Graphical methods may not always be sufficient for describing data.
- Numerical measures can be created for both populations and samples.
  - A **parameter** is a numerical descriptive measure calculated for a population.
  - A **statistic** is a numerical descriptive measure calculated for a sample.

مقاييس النزعة المركزية
تسمى مقاييس النزعة المركزية بمقاييس الموضع أو المتوسطات ، وهى القيم التي تتركز القيم حولها ، ومن هذه المقاييس ، الوسط الحسابي Mean ، والمنوال Mode، والوسيط Median ، والوسط الهندسي ، والوسط التوافقي ، والرباعيات Quartile ، والمئينات Percentage ، وفيما يلي عرض لأهم هذه المقاييس

Arithmetic Mean or Average

- The mean of a set of measurements is the sum of the measurements divided by the total number of measurements.
- where n = number of measurements $\bar{x} = \frac{\sum x_i}{n}$

$\sum x_i =$ sum of all the measuremen ts

الوسط الحسابي Arithmetic Mean
من أهم مقاييس النزعة المركزية ، وأكثرها استخداماً في النواحي التطبيقية ، ويمكن حسابه للبيانات المنفصلة Numbers وغير المنفصلة (على شكل فترات) Intervals ، كما يلي :

الوسط الحسابي للبيانات المنفصلة Numbers
يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها . فإذا كان لدينا n من القيم، ويرمز لها بالرمز : $x_1, x_2, \ldots, x_n$.
فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\bar{x}$ يحسب بالمعادلة التالية :

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

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

حيث يدل الرمز Σ على المجموع .

مثـال
فيما يلي درجات 8 طلاب في مادة الإحصاء 101.

40 36 40 35 37 42 32 34

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .
الحـل
لإيجاد الوسط الحسابي للدرجات تطبق المعادلة كما يلي:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$= \frac{34+32+42+37+35+40+36+40}{8} = \frac{296}{8} = 37$$

أي أن الوسط الحسابي لدرجة الطالب في اختبار م مادة الإحصاء 101 يساوي 37 درجة

خصائص الوسط الحسابي
يتصف الوسط الحسابي بعدد من الخصائص ، ومن هذه الخصائص ما يلي :

1. الوسط الحسابي للمقدار الثابت يساوى الثابت نفسه ، أي أنه إذا كانت قيم x هـ ، $x: a, a, \ldots, a$ ، فإن الوسط الحسابي هو:

$$\boxed{\bar{x} = \frac{a + a + \ldots + a}{n} = \frac{na}{n} = a}$$

ومثال على ذلك ، لو اخترنا مجموعة من 5 طلاب ، ووجدنا أن كل طالب وزنه 63 كيلوجرام، فإن متوسط وزن الطالب في هذه المجموعة هو :

$$\bar{x} = \frac{63+63+63+63+63}{5} = \frac{315}{5} = 63\ k.g$$

2. مجموع انحرافات القيم عن وسطها الحسابي يساوى صفرا ، ويعبر عن هذه الخاصية بالمعادلة $\boxed{\sum(x-\bar{x}) = 0}$

3. إذا أضيف مقدار ثابت إلى كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (بعد الإضافة) يساوى الوسط الحسابي للقيم الأصلية (قبل الإضافة) مضافا إليها هذا المقدار الثابت . فإذا كانت القيم هي : $x_1, x_2, \ldots, x_n$ ، وتم إضافة مقدار ثابت (a) إلى كل قيمة من القيم ، ونرمز للقيم الجديدة بالرمز y ، أي أن $y = x + a$ ، فإن : الوسط الحسابي لقيم الإضافة) هو: $\boxed{\bar{y} = \bar{x} + a}$

حيث أن $\bar{y}$ هو الوسط الحسابي للقيم الجديدة

4. إذا ضرب مقدار ثابت (a) في كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (القيم الناتجة بعد الضرب) يساوى الوسط الحسابي للقيم الأصلية (القيم بعد التعديل) مضروبا في هذا المقدار الثابت . أي أنه إذا كان : $y = a\ x$ ، ويكون الوسط الحسابي للقيم الجديدة

$$\boxed{\bar{y} = a\ \bar{x}}$$

5. مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن ، أي أن:

$$\boxed{\sum (x-\bar{x})^2 < \sum (x-a)^2 \ \ if \ \ a \neq \bar{x}}$$

مزايا وعيوب الوسط الحسابي

يتميز الوسط الحسابي بالمزايا التالية :

- أنه سهل الحساب .
- يأخذ في الاعتبار كل القيم .
- أنه أكثر المقاييس استخداما وفهما .

ومن عيوبه .

- أنه يتأثر بالقيم الشاذة والمتطرفة .
- يصعب حسابه في حالة البيانات الوصفية .
- يصعب حسابه في حالة الجداول التكرارية المفتوحة .

Median

- The median of a set of measurements is the middle measurement when the measurements are ranked from smallest to largest.
- The position of the median is

$$.5(n+1)$$

- once the measurements have been ordered.

الوسيط Median

هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم ، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم $(n/2)$ ، ويزيد عنها النصف الآخر $(n/2)$، أي أن 50% من القيم أقل منه، 50% من القيم أعلى منه. وفيما يلي كيفية حساب الوسيط في حالة البيانات غير مبوبة ، والبيانات المبوبة.

الوسيط للبيانات المنفصلة Numbers

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

- ترتب القيم تصاعديا .
- تحديد رتبة الوسيط، وهي : رتبة الوسيط = $\left(\frac{n+1}{2}\right)$
- إذا كان عدد القيم (n) فردي فإن الوسيط هو:

$$\boxed{\text{الوسيط = القيمة رقم } \left(\frac{n+1}{2}\right)}$$

- إذا كان عدد القيم (n) زوجي، فإن الوسيط يقع بين القيمة رقم $(n/2)$، والقيمة رقم $((n/2)+1)$، ومن ثم يحسب الوسيط بتطبيق المعادلة التالي:

$$\boxed{\text{الوسيط} = \frac{\text{القيمة رقم } \left(\frac{n}{2}\right) + \text{القيمة رقم } \left(\frac{n}{2}+1\right)}{2}}$$

مثـال:تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبية متشابهة، وتم زراعتها بمحصول القمح، وتم استخدام نوعين من التسميد هما : النوع (a) وجرب على 7 وحدات تجريبية، والنوع (b) وجرب على 10 وحدات تجريبية ، وبعد انتهاء الموسم الزراعي ، تم تسجيل إنتاجية الوحدة بالطن / هكتار ، وكانت على النحو التالي :

الوحدة	1	2	3	4	5	6	7	8	9	10
النوع (a)	1.2	2.75	3.25	2	3	2.3	1.5			
النوع (b)	4.5	1.8	3.5	3.75	2	2.5	1.5	4	2.5	3

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينها.

الحـل

أولا : حساب وسيط الإنتاج للنوع الأول (a)

- ترتيب القيم تصاعديا :

قيمة الوسيط

الإنتاج	1.2	1.5	2	2.3	2.75	3	3.25
الرتبة	1	2	2	4	5	6	7

رتبة الوسيط

- عدد القيم فردى $(n = 7)$
- إذا رتبة الوسيط هي: $((n+1)/2 = (7+1)/2 = 4)$.
- ويكون الوسيط هو القيمة رقم 4 ، أي أن وسيط الإنتاج للنوع a هو:

$$Med_a = 2.3$$

ثانيا : حساب وسيط الإنتاج للنوع الثاني (b) :

- ترتيب القيم تصاعديا .

قيمة الوسيط = $\frac{2.5 + 3}{2}$

2.75

الإنتاج	1.5	1.8	2	2.5	2.5	3	3.5	3.75	4	4.5
الرتبة	1	2	3	4	5	6	7	8	9	10

5.5

رتبة الوسيط

- عدد القيم زوجي $(n = 10)$ إذا
- رتبة الوسيط هي : $((n+1)/2 = (10+1)/2 = 5.5)$.
- الوسيط = الوسط الحسابي للقيمتين الواقعتين في المنتصف (رقم 5 ،6) .

$$Med_b = \frac{2.5+3}{2} = 2.75$$

وبمقارنة النوعين من السماد ، نجد أن وسيط إنتاجية النوع (a) أقل من وسيط إنتاجية النوع (b)

أي أن : $Med_b > Med_a$.

مزايا وعيوب الوسيط

من مزايا الوسيط

- لا يتأثر بالقيم الشاذة أو المتطرفة .
- كما أنه سهل في الحساب .
- مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيم أخرى . أي أن : $\sum|x - Med| \le \sum|x - a|\ , \quad a \ne Med$

ومن عيوب الوسيط

- أنه لا يأخذ عند حسابه كل القيم في الاعتبار، فهو يعتمد على قيمة أو قيمتين فقط.
- يصعب حسابه في حالة البيانات الوصفية التي تقاس بمعيار اسمي nominal

Mode

- The mode is the measurement which occurs most frequently.
- The set: 2, 4, 9, 8, 8, 5, 3
  - The mode is 8, which occurs twice
- The set: 2, 2, 9, 8, 8, 5, 3
  - There are two modes—8 and 2 (bimodal)
- The set: 2, 4, 9, 8, 5, 3
  - There is no mode (each value is unique).

المنوال Mode

يعرف المنوال بأنه القيمة الأكثر شيوعا أو تكرارا ، ويكثر استخدامه في حالة البيانات الوصفية، لمعرفة النمط (المستوى) الشائع، ويمكن حساب للبيانات المبوبة وغير المبوبة كما يلي:

حساب المنوال في حالة البيانات غير المبوبة

المنوال (Mod) = القيمة (المستوى) الأكثر تكرارا (٣-١٢)

مثـال

اختيرت عينات عشوائية من طلاب بعض أقسام كلية علوم الأغذية والزراعة ، وتم رصد درجات هؤلاء الطلاب في مقرر 122 إحصاء التطبيقي ، وكانت النتائج كالتالي:

قسم وقاية النباتات	80	77	75	77	77	77	65	70	58	67
قسم علوم الأغذية	88	68	60	75	93	65	77	85	95	90
قسم الاقتصاد	80	65	69	80	65	88	76	65	86	80

قسم الإنتاج الحيواني	85	73	69	85	73	69	69	73	72	85

والمطلوب حساب منوال الدرجات لكل قسم من الأقسام :

الحـل

هذه البيانات غير مبوبة ، لذا فإن :

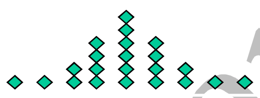
المنوال = القيمة الأكثر تكرارا

والجدول التالي يبين منوال الدرجة لكل قسم من الأقسام .

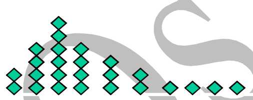
القسم	القيمة الأكثر تكرار	القيمة المنوالية
قسم وقاية النباتات	الدرجة 77 تكررت 4 مرات	المنوال = 77 درجة
قسم علوم الأغذية	جميع القيم ليس لها تكرار	لا يوجد منوال
قسم الاقتصاد	الدرجة 65 تكررت 3 مرات الدرجة 80 تكررت 3 مرات	يوجد منوالان هما : المنوال الأول = 65 المنوال الثاني = 80
قسم الإنتاج الحيواني	الدرجة 69 تكررت 3 مرات الدرجة 73 تكررت 3 مرات الدرجة 85 تكررت 3 مرات	يوجد ثلاثة منوال هي: المنوال الأول = 69 المنوال الثاني = 73 المنوال الثالث = 85

استخدام مقاييس النزعة المركزية في تحديد شكل توزيع البيانات

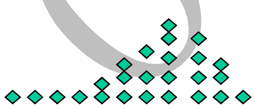
يمكن استخدام الوسط الحسابي والوسيط والمنوال في وصف المنحنى التكراري، والذي يعبر عن شكل توزيع البيانات ، كما يلي :



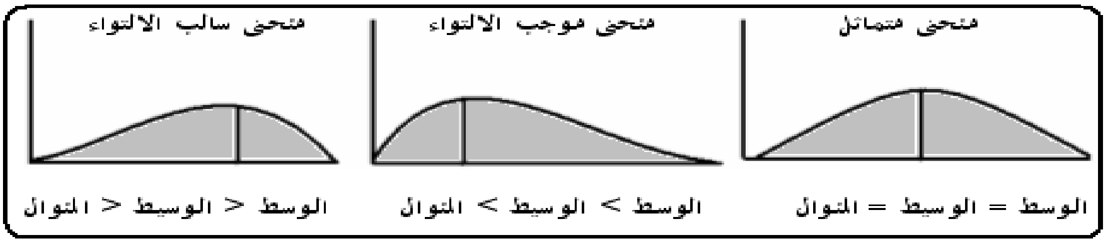
Symmetric: Mean = Median



Skewed right: Mean > Median



Skewed left: Mean < Median



- يكون المنحنى متماثل إذا كان : الوسط = الوسيط = المنوال .
- يكون المنحنى موجب الالتواء (ملتوي جهة اليمين) إذا كان: الوسط > الوسيط > المنوال
- يكون المنحنى سالب الالتواء (ملتوي جهة اليسار) إذا كان : الوسط < الوسيط < المنوال

مثال عام

قام مدير مراقبة الإنتاج بسحب عينة من 10 عبوات من المياه المعبأة للشرب ، ذات الحجم 5 لتر ، والمنتجة بواسطة إحدى شركات تعبئة المياه لفحص كمية الأملاح الذائبة، وكانت كالتالي:

115 123 119 123 124 119 123 121 123 121

والمطلوب : حساب الوسط الحسابي، والوسيط، والمنوال، ثم حدد شكل الالتواء لهذه البيانات .الحل:

- حساب الوسط الحسابي :

$$\overline{x} = \frac{\sum x}{n} = \frac{1211}{10} = 121.1$$

- حساب الوسيط :

رتبة الوسيط : $(n+1)/2 = (10+1)/2 = 5.5$

ترتيب القيم تصاعديا

قيمة الوسيط 122

الطاقة	115	119	119	121	121	123	123	123	123	124
الرتبة	1	2	3	4	5	6	7	8	9	10

رتبة الوسيط 5.5

عدد القيم = 10 ، وهو عدد زوجي. الوسيط = الوسط الحسابي للقيمتين رقم (6 , 5)

$$Med = \frac{121+123}{2} = \frac{244}{2} = 122$$

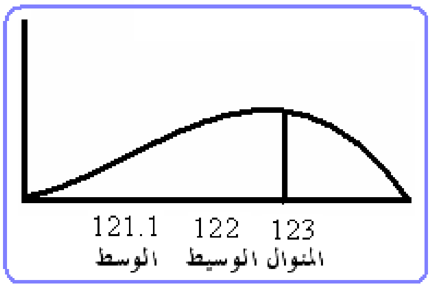
- حساب المنوال :

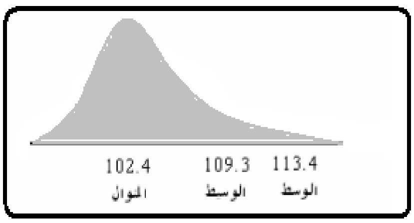
المنوال يساوى القيمة الأكثر تكرارا: القيمة 123 تكررت أكثر من غيرها ، إذا

$$Mod = 123$$

وبمقارنة الوسط والوسيط و المنوال نجد أن :

نجد أن : الوسط < الوسيط < المنوال ، إذا توزيع بيانات كمية الأملاح سالبة الالتواء.





The Range

- The range, R, of a set of *n* measurements is the difference between the largest and smallest measurements.
- Example: A botanist records the number of petals on 5 flowers:

5, 12, 6, 8, 14

- The range is **R = 14 – 5 = 9.**

Measures of Dispersion مقاييس التشتت

Range المدى

- هو أبسط مقاييس التشتت ,وأيسرها حسابا (وحدة القياس؟)
- في حالة البيانات غير المبوبة:

المدى في حالة البيانات غير المبوبة = أكبر قراءة – أقل قراءة

$$Rang = Max - Min \qquad (٤-١)$$

من مزايا المدى:

- أنه بسيط وسهل الحساب
- يكثر استخدامه عند الإعلان عن حالات الطقس، مثل درجات الحرارة، والرطوبة، والضغط الجوي.

ومن عيوبه:

- أنه يعتمد على قيمتين فقط ، ولا يأخذ جميع القيم في الحسبان
- يتأثر بالقيم الشاذة

The Variance

- The variance is measure of variability that uses all the measurements. It measures the average deviation of the measurements about their mean.
- Flower petals: 5, 12, 6, 8, 14
- The variance of a population of N measurements is the average of the squared deviations of the measurements about their mean (μ). $\bar{x} = \frac{45}{5} = 9$ $\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N}$
- The **variance of a sample** of *n* measurements is the sum of the squared deviations of the measurements about their mean, divided by *(n* – 1). $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$

التباين Variance

- أكثر مقاييس التشتت استخداما في النواحي التطبيقية.
- يعبر عن متوسط مربعات انحرافات القيم عن وسطها الحسابي
- يمكن حساب:
  - التباين في المجتمع (σ^2)
  - التباين في العينة (s^2)
- افرض لدينا كل مفردات المجتمع: ($x_1, x_2, x_3, \ldots\ldots\ldots, x_n$)
- التباين في هذا المجتمع، ويرمز له بالرمز (σ^2) يحسب باستخدام المعادلة التالية:

$$\boxed{\sigma^2 = \frac{\sum (x-u)^2}{N}} \qquad (٤-٦)$$

- حيث (μ) هو الوسط الحسابي في المجتمع ويحسب من:

$$\mu = \sum x / N$$

التباين في العينة - (S^2)

- غالبا يكون تباين المجتمع (σ^2) غير معلوم
- وعندئذ يتم سحب عينة من هذا المجتمع، ويحسب التباين من:

$$\boxed{s^2 = \frac{\sum (x-\bar{x})^2}{n-1}} \qquad (٤-٨)$$

The Standard Deviation

- In calculating the variance, we squared all of the deviations, and in doing so changed the scale of the measurements.
- To return this measure of variability to the original units of measure, we calculate the **standard deviation**, the positive square root of the variance.

Population standard deviation $: \sigma = \sqrt{\sigma^2}$

Sample standard deviation $: s = \sqrt{s^2}$

Two Ways to Calculate the Sample Variance

	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	5	-4	16
	12	3	9
	6	-3	9
	8	-1	1
	14	5	25
Sum	45	0	60

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

$$= \frac{60}{4} = 15$$

$$s = \sqrt{s^2} = \sqrt{15} = 3.87$$

	x_i	x_i^2
	5	25
	12	144
	6	36
	8	64
	14	196
Sum	45	465

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

$$= \frac{465 - \frac{45^2}{5}}{4} = 15$$

$$s = \sqrt{s^2} = \sqrt{15} = 3.87$$

Standard Deviation الانحراف المعياري

- من مقاييس التشتت يقاس بوحدات قياس المتغير محل الدراسة
- **هو الجذر التربيعي الموجب للتباين أي:**

$$\boxed{\text{الانحراف المعياري} = \sqrt{\text{التباين}}}$$ (١١-٤)

- الانحراف المعياري للمقدار الثابت يساوي صفرا
- إذا أضيف مقدار ثابت إلى كل قيمة لا يتأثر الانحراف الانحراف المعياري بذلك
- إذا ضرب كل قيمة من قيم المفردات في مقدار ثابت فإن الانحراف المعياري للقيم الجديدة ، يساوي الانحراف المعياري للقيم الأصلية مضروبا في الثابت
- إذا كان لدينا الاقتران الخطي : $y = ax + b$

فإن الانحراف المعياري للمتغير (y) يكون: $s_y = as_x$

مزايا وعيوب الانحراف المعياري

المزايا:

- أنه أكثر مقاييس التشتت استخداما .
- يسهل التعامل معه رياضيا .
- يأخذ كل القيم في الاعتبار .

العيوب:

- يتأثر بالقيم الشاذة

Using Measures of Center and Spread: Tchebysheff's Theorem
Given a number *k* greater than or equal to 1 and a set of *n* measurements, at least $1-(1/k^2)$ of the measurement will lie within *k* standard deviations of the mean.

- Can be used for either samples ($\overline{x}$ and *s*) or for a population (μ and σ).
- **Important results:**
  - If $k = 2$, at least $1 - 1/2^2 = 3/4$ of the measurements are within 2 standard deviations of the mean.
  - If $k = 3$, at least $1 - 1/3^2 = 8/9$ of the measurements are within 3 standard deviations of the mean.

Using Measures of Center and Spread: The Empirical Rule
Given a distribution of measurements that is approximately mound-shaped:

- The interval m ± s contains approximately 68% of the measurements.
- The interval m ± 2s contains approximately 95% of the measurements.

- The interval m ± 3s contains approximately 99.7% of the measurements.

قاعدة تشيبشيف النظرية Chebyshev's Theorem

- وفكرة هذه القاعدة: في أي توزيع من التوزيعات النظرية "تقريبا" كل القيم تكون "قريبة" من الوسط الحسابي وطبقا لهذه القاعدة، فإنه على الأقل 75% من قيم المشاهدات تقع في المدى، $\bar{x} \pm 2s$ على الأقل 89% من قيم المشاهدات تقع في المدى. $\bar{x} \pm 3s$

Approximating *s*

- From Tchebysheff's Theorem and the Empirical Rule, we know that

$$R \approx 4\text{-}6\, s$$

- To approximate the standard deviation of a set of measurements, we can use:

$$s \approx R/4$$
$$\text{or } s \approx R/6 \text{ for a large data set.}$$

Measures of Relative Standing

- Where does one particular measurement stand in relation to the other measurements in the data set?
- How many standard deviations away from the mean does the measurement lie? This is measured by the *z*-score.

$$z\text{-score} = \frac{x - \bar{x}}{s}$$

- From Tchebysheff's Theorem and the Empirical Rule
  - At least 3/4 and more likely 95% of measurements lie within 2 standard deviations of the mean.
  - At least 8/9 and more likely 99.7% of measurements lie within 3 standard deviations of the mean.
- z-scores between –2 and 2 are not unusual. z-scores should not be more than 3 in absolute value. z-scores larger than 3 in absolute value would indicate a possible outlier.

(Standardized degree(z-score الدرجة المعيارية

- تقيس الدرجة المعيارية لقيمة معينة عدد وحدات الانحراف المعياري التي تزيد أو تقل بها هذه القيمة عن الوسط الحسابي
- الدرجة المعيارية للقيمة (xi) ويرمز لها بالرمز(z) تحسب باستخدام

المعادلة التالية: $z = \frac{x_i - \bar{x}}{s}$

مثال

اختير أحد الأغنام من المجموعة الأولى بعد تطبيق البرنامج ، ووجد أن وزنه 178 كيلوجرام، وبالمثل أحد الأغنام من المجموعة الثانية، ووجد أن وزنه 180 كيلوجرام ، قارن بين هذين القيمتين من حيث أهمية كل منها في المجموعة التي تنتمي إليها:

- للمقارنة بين الوحدتين يتم حساب الدرجة المعيارية لوزن كل منها، بتطبيق المعادلة:
    - الدرجة المعيارية لوزن الوحدة المسحوبة من المجموعة الأولى (178 Kg.) هي: $z = \frac{x - \bar{x}}{s} = \frac{178 - 173}{23} = 0.22$
    - الدرجة المعيارية لوزن الوحدة المسحوبة من المجموعة الثانية (180 Kg.) هي $z = \frac{x - \bar{x}}{s} = \frac{180 - 198}{25} = -0.75$

من النتائج نجد أن الوزن 178 كيلوجرام يزيد عن الوسط الحسابي بـ 0.22 انحراف معياري ، بينما نجد أن الوزن 180 كيلوجرام يقل عن الوسط الحسابي بـ 0.75 انحراف معياري . ومن ثم في هذه الحالة الوزن الأول أهميته النسبية أعلى من الوزن الثاني.

القاعدة العملية-التوزيع الطبيعي

- للعينة $(x_1, x_2, x_3, \ldots\ldots\ldots, x_n)$، ذات الوسط الحسابي ($\bar{x}$)، والانحراف المعياري (s) ، يكون منحنى توزيع هذه العينة متماثل، إذا تحقق الآتي:
    - 68% تقريبا من قيم هذه المشاهدات تتراوح بين : $\bar{x} \pm s$
    - 95% تقريبا من قيم هذه المشاهدات تتراوح بين: $\bar{x} \pm 2s$
    - 99% تقريبا من قيم هذه المشاهدات تتراوح بين: $\bar{x} \pm 3s$

Measures of Relative Standing

- How many measurements lie below the measurement of interest? This is measured by the p^{th} percentile.

50^{th} Percentile	≡ Median
25^{th} Percentile	≡ Lower Quartile (Q_1)
75^{th} Percentile	≡ Upper Quartile (Q_3)

طريقة " المئين " لقياس الالتواء

- المئين ينتج من ترتيب البيانات تصاعديا و تقسيمها إلى 100 جزء متساوٍ يفصل بينها قيم تسمى المئين (p^i)
- مثلا: المئين 15 ويرمز له بالرمز (p^{15}) هو القيمة التي يقل عنها 15% من القيم.

- كيف نحسب مثلا المئين (p^i)؟؟
- يتبع نفس الفكرة المستخدمة في حساب الربيع .

حساب المئين

- نرتب القيم تصاعديا ونحدد رتبة المئين من: $R = (n+1)\left(\frac{i}{100}\right)$
- إذا كانت الرتبة (R) عدد صحيح فإن: $p^i = x(R)$

Quartiles and the IQR

- The lower quartile (Q_1) is the value of x which is larger than 25% and less than 75% of the ordered measurements.
- The upper quartile (Q_3) is the value of x which is larger than 75% and less than 25% of the ordered measurements.
- The range of the "middle 50%" of the measurements is the inter quartile range,

  IQR = $Q_3 - Q_1$

Calculating Sample Quartiles

- The lower and upper quartiles (Q_1 and Q_3), can be calculated as follows:
- The position of Q_1 is $.25(n+1)$
- The position of Q_3 is $.75(n+1)$

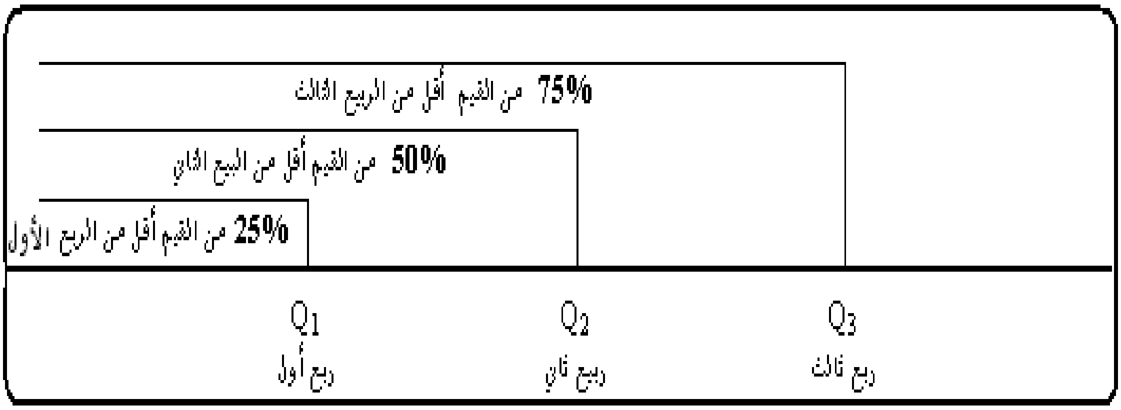
once the measurements have been ordered. If the positions are not integers, find the quartiles by interpolation.

الرباعيات Quartiles

عند تقسيم القيم إلى أربع أجزاء متساوية، يوجد ثلاث إحصاءات ترتيبي تسمى بالرباعيات، وهي:

- الربيع الأول: وهو القيمة التي يقل عنها ربع عدد القيم، أي يقل عنها 25% من القيم، ويرمز له بالرمز Q_1 .
- الربيع الثاني: وهو القيمة التي يقل عنها نصف عدد القيم، أي يقل عنها 50% من القيم، ويرمز له بالرمز Q_2 ، ومن ثم يعبر هذا الربيع عن الوسيط.
- الربيع الثالث: وهو القيمة التي يقل عنها ثلاث أرباع عدد القيم، أي يقل عنها 75% من القيم، ويرمز له بالرمز Q_3 .

الرباعيات



ولحساب أي من الرباعيات الثلاث، يتم إتباع الآتي:

- بفرض أن عدد القيم عددها n، وأنها مرتبة كالتالي:

القيم مرتبة: $X_{(1)} > X_{(2)} > X_{(3)} \quad > X_{(n)}$

: الرتبة $1 \quad 2 \quad 3 \quad n$

- نحدد الرتبة (Ri) للرباعي رقم (i=1,2,3) من القانون:

$$R = (n+1)\times\left(\frac{i}{4}\right)$$

- إذا كانت R عددا صحيحا فإن قيمة الربيع هو: $Q_i = X_{(R)}$.
- إذا كانت R عدد كسري، فإن الرباعي (Q_i) يقع في المدى : $X_{(l)} < Q_i < X_{(u)}$ ، ومن ثم يحسب (Q_i) بالمعادلة التالية:

$$\boxed{Q_i = x_{(l)} + (R-l)(x_{(u)} - x_{(l)})} \qquad (١٤-٣)$$

- حيث أن (i) هي رتبة القيمة السابقة للرباعي

مثال

فيما يلي كمية الإنتاج اليومي من الحليب باللتر للبقرة الواحدة لعينة حجمها 10 أبقار اختيرت من مزرعة معينة:

30 27 18 20 29 34 32 29 23 25

احسب الرباعيات الثلاث لكمية الإنتاج، وما هو تعليقك؟

الحل:

لحساب الرباعيات الثلاث، يتم إتباع الآتي:

- ترتيب القيم تصاعديا:

قمة الربيع		22.25			28				30.5	
القيم	**18**	**20**	**23**	**25**	**27**	**29**	**29**	**30**	**32**	**34**
الرتبة	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
رتبة الربيع		**2.75**			**5.5**			**8.25**		

- حساب الربيع الأول (Q_1):

رتبة الربيع الأول هي: $R=(n+1)\times\left(\frac{i}{4}\right)=(10+1)\times\left(\frac{1}{4}\right)=2.75$

يقع الربيع الأول بين القيمتين: $(20<Q_1<23)$، وبتطبيق المعادلة (3-14) نجد أن:

$l=2,\ R=2.75\ ,x_{(l)}=20\ .x_{(u)}=23$

إذا : $$Q_1=x_{(l)}+(R-l)\times(x_{(u)}-x_{(l)})=20+0.75(23-20)=22.25$$

- حساب الربيع الثاني (الوسيط) Q_2

رتبة الربيع الثاني هي: $R=(n+1)\times\left(\frac{i}{4}\right)=(10+1)\times\left(\frac{2}{4}\right)=5.5$

يقع الربيع الثاني بين القيمتين: $(27<Q_2<29)$، وبتطبيق المعادلة (3-14) نجد أن:

$l=5,\ R=5.5\ ,x_{(l)}=27\ .x_{(u)}=29$

إذا : $$Q_2=x_{(l)}+(R-l)\times(x_{(u)}-x_{(l)})=27+0.5(29-27)=28$$

- حساب الربيع الثالث Q_3

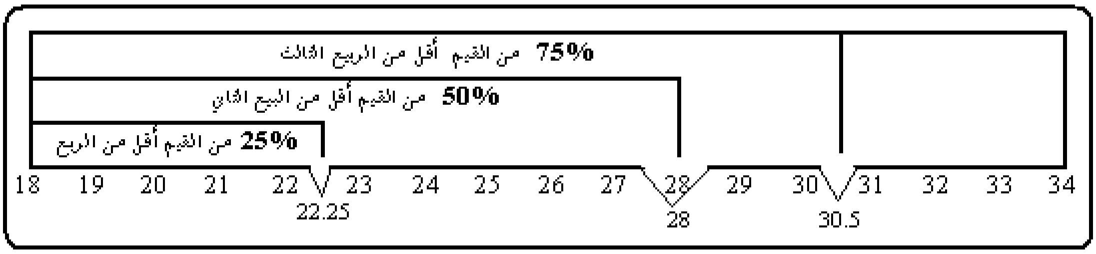
رتبة الربيع الثالث هي: $R=(n+1)\times\left(\frac{i}{4}\right)=(10+1)\times\left(\frac{3}{4}\right)=8.25$

يقع الربيع الثالث بين القيمتين: $(30<Q_3<32)$، وبتطبيق المعادلة (3-14) نجد أن:

$l=8,\ R=8.25\ ,x_{(l)}=30\ .x_{(u)}=32$

إذا : $$Q_3=x_{(l)}+(R-l)\times(x_{(u)}-x_{(l)})=30+0.25(32-30)=30.5$$

من النتائج السابقة نجد أن:



- 25% من الأبقار يقل إنتاجه عن 22.25 لتر يوميا.
- 50% من الأبقار يقل إنتاجه عن 28 لتر يوميا.
- 75% من الأبقار يقل إنتاجه عن 30.5 لتر يوميا.

Using Measures of Center and Spread: The Box Plot
The Five-Number Summary:

Min $\quad Q_1 \quad$ Median $\quad Q_3 \quad$ Max

- Divides the data into 4 sets containing an equal number of measurements.
- A quick summary of the data distribution.
- Use to form a box plot to describe the shape of the distribution and to detect outliers.

Constructing a Box Plot

- Calculate Q1, the median, Q3 and IQR.
- Draw a horizontal line to represent the scale of measurement.
- Draw a box using Q1, the median, Q3.
- Isolate outliers by calculating
  - Lower fence: Q_1-1.5 IQR
  - Upper fence: Q_3+1.5 IQR
- Measurements beyond the upper or lower fence is are outliers and are marked (*).
- Draw "whiskers" connecting the largest and smallest measurements that are NOT outliers to the box.

Example:
Amt of sodium in 8 brands of cheese:
260 290 300 320 330 340 340 520
Q_1 = 292.5
m = 325
Q_3 = 340

IQR = **$Q_3 - Q_1$**
= 340-292.5 = 47.5
Lower fence = **Q_1** – 1.5 (IRQ)
= 292.5-1.5(47.5) = 221.25
Upper fence = **Q_3** + 1.5 (IRQ)
= 340 + 1.5(47.5) = 411.25
Outlier: x = 520

Interpreting Box Plots

- ✓ Median line in center of box and whiskers of equal length—symmetric distribution
- ✓ Median line left of center and long right whisker—skewed right
- ✓ Median line right of center and long left whisker—skewed left