" Chapter 1,2 "

" Q#1 "

لا يكتب في هذا الهامش

1) a) the sample

b) * it is a descriptive measurs computed from the data of a population

* the (mean - median - Variance = ...) variable of (all, every..) population

✓ d)

2) ✓ a)

b) Continuous Variable

c) qualitative - nominal

d) continuous variable

e) qualitative - ordinal

3) a) discrete variable

b) " "

✓ c)

d) discrete variable

e) " "

4) a) quantitative variable (continuous)

b) " " " (discrete)

✓ c)

d) quantitative variable (continuous)

e) " " "

5) ✓ d)

6) a) parameter

b) quantitative - discrete

c) parameter

✓ d)

e) qualitative - nominal

7) a) discrete variable

b) qualitative variable

✓ c)

8) ✓ b)

there is also an equivalent definitions which is ,

* is the distance from the true lower limit of an interval to the true upper limit of it

* the distance from the midpoint to the midpoint of the next one.

(1)

لا يكتب في هذا الهامش

9.) a) quantitative - discrete
✓b)
c) qualitative - nominal
d) quantitative - discrete
e) qualitative - ordinal

10) ✓c)

11) c) continuous Variable
✓d)
e) qualitative Variable

12) a) quantitative - continuous
b) " "
✓c)
d) quantitative - continuous
e) quantitative - discrete

13) a) discrete Variable
✓b)
c) continuous Variable

14) a) quantitative - discrete
b) " - continuous
c) qualitative - nominal
d) " "
✓e)

15) a) discrete Variable
b) continuous Variable
✓c)

16) ✓a)

17) a) Parametex
✓b)
c) Parameter
d) "

لا يكتب في هذا الهامش

"Q#2"

# of times that child has been to a doctor (variable)	frequency (# of child)	relative frequency	Percentage frequency
1	15	15/40 = .375	.375 X 100 = 37.5%
2	10	10/40 = .25	.25 X 100 = 25%
3	8	8/40 = .2	.2 X 100 = 20%
4	4	4/40 = .1	.1 X 100 = 10%
5	3	3/40 = .075	.075 X 100 = 7.5%
total or the sample size (n)	40	1	100%

A) the # of times that child has been to a doctor

B) quantitative = discrete

C) 15 + 10 + 8 + 4 + 3 = 40 = n

D) 8

E) 4 + 3 = 7

F) 10 + 8 + 4 = 22

G) 5

H) $\frac{10+15}{40} \times 100 = \frac{25}{40} \times 100 = 62.5\%$

I) we know that the variable has highest percentage who has the highest frequency, so it is 1

+ another paragraph

as the data from kind "simple frequency distribution" so the type of graph is bar chart

لا يكتب في هذا الهامش

"Q#3"

# of bruieses on body (Variable)	frequency	relative frequency	Percentage frequency
0	5	5/30 = .1667	.1667X100 = 16.67%
1	10	10/30 = .3333	.3333X100 = 33.33%
2	7	7/30 = .2333	.2333X100 = 23.33%
3	6	6/30 = .2	.2X100 = 20%
4	2	2/30 = .0667	.0667X100 = 6.67%
total	30	1	≈ 100%

A) bar chart

B) the # of bruieses on body

C) quantitative - discrete

D) 5+10+7+6+2 = 30 = n

E) 6+2 = 8

F) $\frac{7+10+5}{30} \times 100 = 73.3\%$

G) we know that the variable has largest percentage who has the largest frequency, so it is 1

H) 0

How to use calculator:

1- CASIO fx-991ES PLUS and fx-82ES (لونها رمادي)

الآله من نوع

أ. التاكد من ان الآلة الحاسبة لا تحتوي على اي بيانات مخزنة مسبقا وذلك بعمل مايلي

**مسح البيانات من الذاكرة

Push Shift ⟶ 9

⟶ to see Clear ⟶ Push 3 = ⟶ Push AC

**تحويل الآلة الحاسبة إلي الاستخدام الإحصائي

Push Mode ⟶ to see stat ⟶ Push 3 =

**ادخال البيانات

1. الحاله الاولى لو كانت البيانات بهذه الصورة

[Data: 3 , 5 ,8, 9]

Push Shift ⟶ 1

⟶ to see Data ⟶ Push 2

سوف يظهر عمود اسمه x والصفوف مرقمة

نكتب باول صف عند رقم 1 أول قيمة 3 وهكذا

3 = 5 = 8 = 9 = ⟶ Push AC

**ايجاد الوسط الحسابي والانحراف المعياري والتباين

Push Shift 1

⟶ to see Var ⟶ Push 4

سوف تظهر شاشة بهذه الصورة

1:n	2: $\bar{x}$
3: σx (xσn)	4:sx (xσn-1)

Then,

Push 2 = ⟶ get $\bar{X}$

Push 3 = ⟶ get σ

Push 4 = ⟶ get S

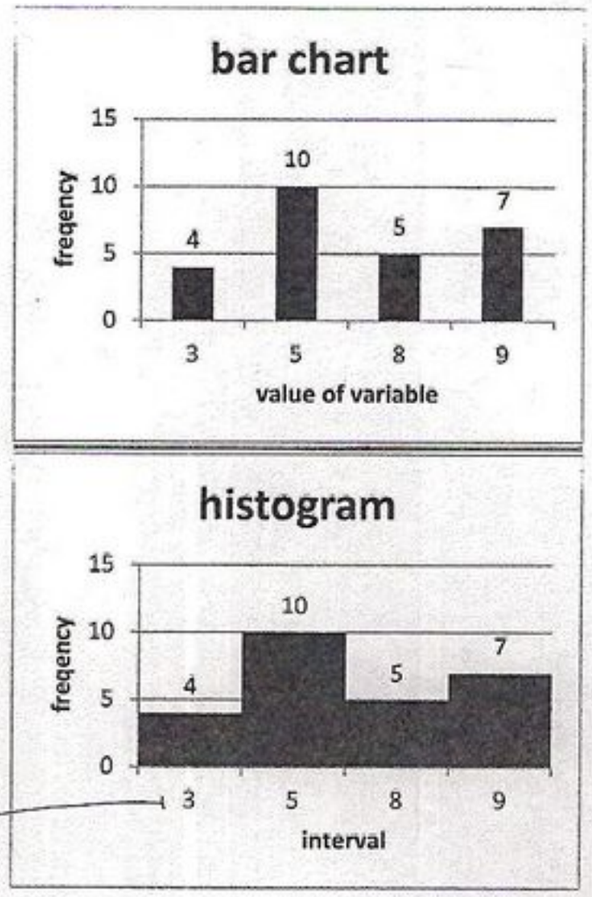
Where S: Sample standard deviation

To get S^2 : Sample variance , just square S.

2. الحاله الاولى لو كانت البيانات اما في Tables or Chart

سوف نأخذ التكرار (frequency) في الاعتبار

Value of variable (discrete variable) or mid interval (continues)	frequency
3	4
5	10
8	5
9	7



they are as midpoint

Push Shift ⟶ 1

⟶ to see type ⟶ Push 1 ⟶ Push 2

سوف يظهر عمودين x and y والصفوف مرقمة

نضع باول عمود (x) قيم المتغير ونضع قيم التكرار بالعمود y وبنفس الترتيب

يعني القيمة 3 بالصف الأول بالعمود x وقيمة التكرار له 4 نضعها بنفس الصف بالعمود y وهكذا

بالعمود x نكتب

3 = 5 = 8 = 9 =

وبالأسهم ننتقل الى العمود y ونكتب من اول صف

4 = 10 = 5 = 7 = ⟶ Push AC

**ايجاد الوسط الحسابى

Push Shift 1

⟶ to see Sum ⟶ Push 3

⟶ to see الشاشة

1: Σx^2	2: Σx
3: Σy^2	4: Σy
5: Σxy	6: Σx^3
7: $\Sigma x^2 y$	8: Σx^4

Then,

Push 5 = ⟶ get Σxf

Push 4 = ⟶ get Σf

ولايجاد الوسط الحسابي نوجد

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{\Sigma xy}{\Sigma y}$$

ولإيجاد تباين العينة نوجد:

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

وبالتالي الإنحراف المعياري للعينة هو: $s = \sqrt{s^2}$

How to use calculator:

1- الآله من نوع CASIO fx-991MS (لونها اسود)

أ. التاكد من ان الآلة الحاسبة لا تحتوى على اي بيانات مخزنة مسبقا وذلك بعمل مايلي

**مسح البيانات من الذاكرة

Push Shift ⟶ Mode

⟶ to see Scl ⟶ Push 3 = ⟶ Push AC

**تحويل الآلة الحاسبة إلى الاستخدام الإحصائي

Push Mode ⟶ Push Mode

⟶ to see SD ⟶ Push 1 =

**ادخال البيانات

1. الحاله الاولى لو كانت البيانات بهذه الصورة

[Data: 3 , 5 ,8, 9]

3 M^+ 5 M^+ 8 M^+ 9 M^+ ⟶ Push AC

**ايجاد الوسط الحسابي والانحراف المعيارى والتباين

Push Shift 2

⟶ to see الشاشة

	$\bar{x}$	xσn	xσn-1
	1	2	3

Then,

Push 1 = ⟶ get $\overline{X}$

Push 2 = ⟶ get σ

Push 3 = ⟶ get S

Where S: Sample standard deviation

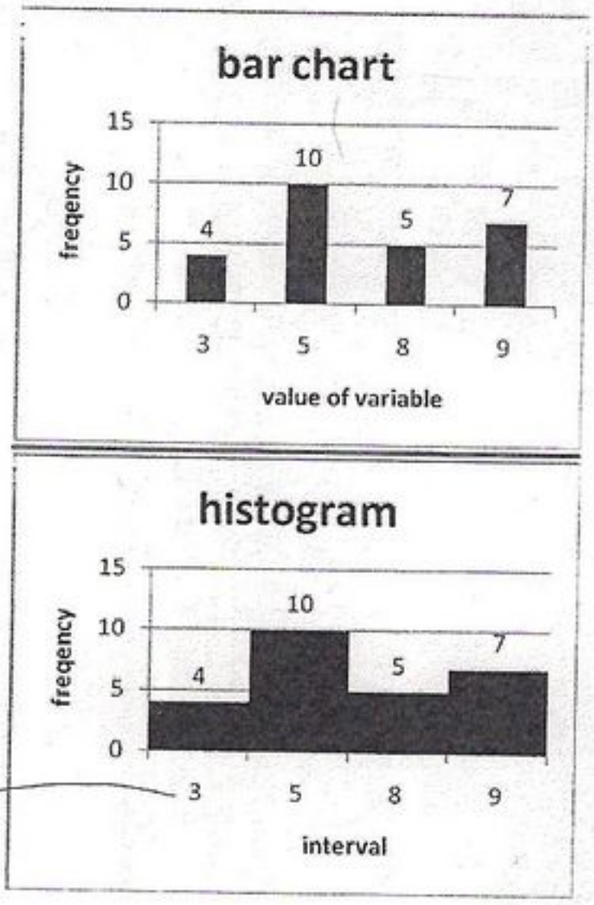
To get S^2 : Sample variance , just square S.

.2 الحاله الاولى لو كانت البيانات

اما في Tables or Chart

سوف ناخذ التكرار (frequency) في الاعتبار

Value of variable (discrete variable) or mid interval (continues)	frequency
3	4
5	10
8	5
9	7



They are as midpoint

سوف يكون ادخال البيانات في هذه الحالة كما يلي :

3 Push Shift , 4 M+ 5 Push Shift , 10 M+ 8 Push Shift , 5 M+ 9 Push Shift , 7 M+

Push AC

**ايجاد الوسط الحسابي والانحراف المعياري والتباين

Push Shift 2

⟶ to see الشاشة

	$\bar{x}$	xσn	xσn-1
	1	2	3

Then,

Push 1 = ⟶ get $\overline{X}$

Push 2 = ⟶ get σ

Push 3 = ⟶ get S

Where S: Sample standard deviation

To get S^2 : Sample variance , just square S.

لا يكتب في هذا الهامش

"Q#4"

*: means that it is given , X : value unknown

classes interval age (variable)	true classes interval	midpoints
*15-20	(15-.5) -< (20+.5) = 14.5 -< 20.5	(15+20)/2 = 17.5
* 21-26	(20.5) -< (26+.5) = 20.5 -< 26.5	(21+26)/2 = 23.5
27-32	(26.5) -< (32+.5) = 26.5 -< 32.5	(27+32)/2 = 29.5
33-38	(32.5) -< (38+.5) = 32.5 -< 88.5	(33+38)/2 = 35.5
39-44	(38.5) -< (44+.5) = 38.5 -< 44.5	(39+44)/2 = 41.5

frequency	relative frequency	cumulative frequency
*3	3/60 = .05	3
$\frac{X}{60} = .2 \Rightarrow X = 60 \times .2 = 12$	* .2	3+12 = 15
*18	18/60 = .3	3+12+18 = 33
*17	17/60 = .283	3+12+18+17 = 50
60 - (3+12+18+17) = 10	10/60 = .16667	3+12+18+17+10 = 60
*60 = total	≃ 1	

we can see that from the given :

The width (w) = 21 - 15 = 6 (from the classes interval)

The smallest unit (S.U.) = 1 $\Rightarrow \frac{S.U.}{2} = \frac{1}{2} = .5$ (for find the true classes inteval).

and as we have classes (from kind "Groupe frequency distribution"),

so the type of the graph is histogram.

The variable is the age of 60 women having children, so the type of it is quantitative - continuous.

1) d

2) c

3) b

4) a

5) repetition of the paragraph (4)

6) f

7) a

8) the age class has largest percentage who has the largest frequency, so it is 27-32

9) 12+18+17+10 = 57

(13)

لا يكتب في هذا الهامش

10) quantitative - continuous

11) $\frac{18+17+10}{60} \times 100 = 75\%$

12) and 13) :

the sample mean $\bar{x}$ and the sample standard deviation s ;

we will find them by using the calculator for :

midpoint	frequency
17.5	3
23.5	12
29.5	18
35.5	17
41.5	10

we will get that $\bar{x} = 31.4$, $s = 6.76 \Rightarrow s^2 = (6.76)^2$

14) the sample mode is the value of midpoint who has largest frequency , so it is 29.5

لا يكتب في هذا الهامش

"Q#5"

*: means that it is given , X: value unknown

Classes interval serum level (Mg/l) (variable)	true classes interval	mid points
* 5-24	(5-.5) -< (24+.5) = 4.5 -< 24.5	* 14.5
* 25-44	(24.5) -< (44+.5) = 24.5 -< 44.5	(25+44)/2 = 34.5
45-64	(44.5) -< (64+.5) = 44.5 -< 64.5	(45+64)/2 = 54.5
65-84	(64.5) -< (84+.5) = 64.5 -< 84.5	(65+84)/2 = 74.5
85-104	(84.5) -< (104+.5) = 84.5 -< 104.5	(85+104)/2 = 94.5
total		

	Frequency	relative frequency	cumulative frequency
	* 50	50/200 = .25	50
	* 63	63/200 = .315	50+63 = 113
	$\frac{X}{200} = .305 \Rightarrow X = 200 \times .305 = 61$	* .305	50+63+61 = 174
	*24	24/200 = .12	50+63+61+24 = 198
	200 - (50+63+61+24) = 2	2/200 = .01	50+63+61+24+2 = 200
total	n = 200	1	

we can see that from the given:

The width(w.) = 25-5 = 20 (from the classes interval)

The smallest unit (s.u.) = 1 $\Rightarrow \frac{s.u.}{2} = \frac{1}{2} = .5$ (for find the true classes interval)

and as we have classes (from kind grouped frequency distribution"),
so the type of the graph is histogram.

The variable is the serum level (mg/l), so the type of it is quantitative - continuous.

1) b

2) c

3) e

4) e

5) c

6) a

7) b

(15)

لا يكتب في هذا الهامش

"Q#6"

as we have the graph from the type histogram, so we have 6 classes.

and we can find the following:

i	midpoint	frequency	true classes interval
1	1	10	$TL_1 \leftarrow .5 - < 1.5 \rightarrow <TU_1$
2	2	15	$1.5 - < 2.5 \rightarrow <TU_2$
3	3	25	$2.5 - < 3.5 \rightarrow <TU_3$
4	4	15	$3.5 - < 4.5 \rightarrow <TU_4$
5	5	10	$4.5 - < 5.5 \rightarrow <TU_5$
6	6	5	$5.5 - < 6.5 \rightarrow <TU_6$
	total	$n = 80$	

to find the true classes interval, we will use the following steps:

a.) take the difference between the first and the second midpoints to find the width (w.)

$midpoint^{st} = 1$, $midpoint^{nd} = 2$

$\therefore w = midpoint^{nd} - midpoint^{st} = 2 - 1 = 1$

b.) $H.w. = \frac{w}{2} = \frac{1}{2} = .5$

c.) we will put the following codes: i: such as the counter

TL_i: the i^{th} true lower limit

$<TU_i$: the i^{th} true upper limit of at least, where it is at the same time the $(i+1)^{th}$ true lower limit

so,

$TL_1 = midpoint^{st} - H.w. = 1 - .5 = .5$

$<TU_1 = midpoint^{st} + H.w. = 1 + .5 = 1.5$

d) $<TU_2 = 1.5 + w = 1.5 + 1 = 2.5$

$<TU_3 = 2.5 + w = 2.5 + 1 = 3.5$

$<TU_4 = 3.5 + w = 3.5 + 1 = 4.5$

$<TU_5 = 4.5 + w = 4.5 + 1 = 5.5$

$<TU_6 = 5.5 + w = 5.5 + 1 = 6.6$

Note: the number of true classes such as last number in the counter i.

لا يكتب في هذا الهامش

The variable is the hours spent without pain after a surgery, so the type of it is quantitative - continuous.

1.) c

2) d

3) $\frac{15+10+5}{80} \times 100 = \frac{30}{80} \times 100 = 37.5\%$ ∴ a

4) b

5.) the sample mean $\bar{x}$ we will find it by using the calculator for:

midpoints	frequency
1	10
2	15
3	25
4	15
5	10
6	5
total	n = 80

extra part:

sample standard deviation

$s = 1.388 \Rightarrow s^2 = 1.926$

sample variance

we will get that $\bar{x} = 3.1875$ ∴ d

6.) the sample mode is the value of midpoint who has largest frequency, so it is 3 ∴ b

لا يكتب في هذا الهامش

"Q#7"

as we have the graph from the type histogram, so we have a classes.

and we can find the following:

i	midpoint	frequency	true classes interval
1	.5	13	$TL_1 \leftarrow .25 - < .75 \rightarrow < TU_1$
2	1	36	$.75 - < 1.25 \rightarrow < TU_2$
3	1.5	17	$1.25 - < 1.75 \rightarrow < TU_3$
4	2	18	$1.75 - < 2.25 \rightarrow < TU_4$
5	2.5	15	$2.25 - < 2.75 \rightarrow < TU_5$
6	3	11	$2.75 - < 3.25 \rightarrow < TU_6$
	total	$n = 110$	

to find the true classes interval, we will use the same steps in the last Question "Q#6":

a) $w = 1 - .5 = .5$

b) $H.W. = \frac{w}{2} = \frac{.5}{2} = .25$

c) $TL_1 = .5 - .25 = .25$

$< TU_1 = .5 + .25 = .75$

d) $< TU_2 = .75 + .5 = 1.25$

$< TU_3 = 1.25 + .5 = 1.75$

$< TU_4 = 1.75 + .5 = 2.25$

$< TU_5 = 2.25 + .5 = 2.75$

$< TU_6 = 2.75 + .5 = 3.25$

The variable is tumor size (in cm), so the type of it is quantitative - continuous.

1) $\frac{18}{110} \times 100 = 16.36\%$ $\therefore$ c

2) d

3) the size of tumor has the largest (or highest) percentage of patients who has the largest (or highest) frequency of patients, so it is 1 $\therefore$ b

4) to find the sample mean $\bar{x}$, we will do it as the same way from Question "Q#6" part "5" and we get it $\bar{x} = 1.586$ $\therefore$ e

extra part: $s = .774 \Rightarrow s^2 = .6$

5) the sample mode is the value of midpoint who has largest frequency, so it is 1 $\therefore$ a

لا يكتب في هذا الهامش

"Q#8"

as we have the graph from the type histagram, so we have ai classes and we can find the following:

i	midpoint	frequency	true classes interval
1	3.5	26	$TL_1 \leftarrow (3) - < (4) \rightarrow < TU_1$
2	4.5	29	$4 - < (5) \rightarrow < TU_2$
3	5.5	27	$5 - < (6) \rightarrow < TU_3$
4	6.5	15	$6 - < (7) \rightarrow < TU_4$
5	7.5	5	$7 - < (8) \rightarrow < TU_5$
	total	n=102	

to find the true classes interval, we will use the same steps in Question "Q#6":

a) $W = 4.5 - 3.5 = 1$

b) $H.W. = \frac{W}{2} = \frac{1}{2} = .5$

c) $TL_1 = 3.5 - .5 = 3$

$< TU_1 = 3.5 + .5 = 4$

d) $< TU_2 = 4 + 1 = 5$

$< TU_3 = 5 + 1 = 6$

$< TU_4 = 6 + 1 = 7$

$< TU_5 = 7 + 1 = 8$

The variable is Cholesteral level (mmol/l.), so the type of it is quantitative-Continuous.

1) C

2) C

3) $26 + 29 + 27 + 15 + 5 = 102$ $\therefore$ d

4) the Cholesteral level has the lowest percentage of women who has the lowest frequency of women, so it is 7.5 $\therefore$ b

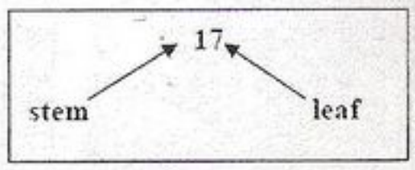
5) to find the sample mean $\bar{x}$, we will do it as the same way from "Q#6" part "5" and we get it $\bar{x} = 4.95$ $\therefore$ c

extra part: $s = 1.166 \Rightarrow s^2 = 1.359$

6) the sample mode is the value of midpoint who has the largest frequency, so it is 4.5 $\therefore$ d

Stem and Leaf Plots Examples

1. A stem and leaf plot is a method used to organize statistical data. The greatest common place value of the data is used to form the stem. The next greatest common place value is used to form the leaves.



2. **EXAMPLE**: Make a stem and leaf plot of the algebra test scores given below: Then complete each question.

56, 65, 98, 82, 64, 71, 78, 77, 86, 95, 91, 59, 69, 70, 80, 92, 76, 82, 85, 91, 92, 99, 73

1st put the scores in numerical order.

56, 59, 64, 65, 69, 70, 71, 73, 76, 77, 78, 80, 82, 82, 85, 86, 91, 91, 92, 92, 95, 98, 99

Since the data range from 56 to 99, the stems range from 5 to 9. To plot the data, make a vertical list of the stems. Each number is assigned to the graph by pairing the units digit, or leaf, with the correct stem. The score 56 is plotted by placing the units digit, 6, to the right of stem 5.

Stem	Leaf
5	6 9
6	4 5 9
7	0 1 3 6 7 8
8	0 2 2 5 6
9	1 1 2 2 5 8 9

a. What type of graph does a stem and leaf plot represent when turned vertically? **Histogram**

b. What was the lowest score on the algebra test? **56**

c. What was the highest score on the algebra test? **99**

d. In which interval did most students score? **90 to 99**

3. **EXAMPLE**: Make a stem and leaf plot of the history test scores given below. Then complete each question.

65, 82, 73, 91, 95, 86, 78, 69, 80, 88

1st put the scores in numerical order.

65, 69, 73, 78, 80, 82, 86, 88, 91, 95

Since the data range from 65 to 95, the stems range from 6 to 9. To plot the data, make a vertical list of the stems. Each number is assigned to the graph by pairing the units digit, or leaf, with the correct stem. The score 65 is plotted by placing the units digit, 5, to the right of stem 6.

Stem	Leaf
6	5 9
7	3 8
8	0 2 6 8
9	1 5

a. What was the lowest score on the history test? **65**

b. What was the highest score on the history test? **95**

c. In which interval did most students score? **80 to 89**

4)

	A	B
1	DATE	MINUTES
2	JULY 9	55
3	JULY 9	3
4	JULY 9	6
5	JULY 10	14
6	JULY 10	18
7	JULY 10	5
8	JULY 10	23
9	JULY 11	30
10	JULY 11	23
11	JULY 11	10
12	JULY 11	2
13	JULY 11	36

Make a stem-and-leaf plot of the length of the 12 cell phone calls.

Step 1: Order the data.

2, 3, 5, 6, 10, 14, 18, 23, 23, 30, 36, 55

Step 2: Choose the stems and leaves. Because the data values range from 2 to 55, use the *tens* digits for the stems and the *ones* digits for the leaves.

Step 3: Write the stems to the *left* of the vertical line.

Step 4: Write the leaves for each stem to the *right* of the vertical line.

Phone call lengths

Order the stems vertically. The stem for data values less than 10 is 0.

Include stems without leaves.

Write the leaves horizontally.

Stem	Leaf
0	2 3 5 6
1	0 4 8
2	3 3
3	0 6
4	
5	5

لا يكتب في
هذا الهامش

"Q#9"

1) the sample size = n = 25 students

2) 4 students which they [27, 35, 42, 49]

3) 10 students which they [70, 74, 75, 78, 78, 84, 85, 88, 93, 92]

4) 16 students which they [27, 35, 42, 49, 50, 51, 53, 57, 59, 62, 63, 64, 67, 68, 69, 70]

5) the student they have from 50 to less than 84 are

[50, 51, 53, 57, 59, 62, 63, 64, 67, 68, 69, 70, 74, 75, 78, 78]

so we have 16 students and the percentage of they are :

$$\frac{16}{25} \times 100 = 64\%$$

لا يكتب في هذا الهامش

"Q#10"

We have a group "A" which is a level of haemoglobin for 10 children:

A : 6.7 , 9.1 , 10 , 11.4 , 12.4 , 9.8 , 8.3 , 9.9 , 9.1 , 7.5

a) to find the sample mean $\bar{X}_A$ and the sample variance S_A^2, we will put the data in "A" in the following table:

haemoglobin level (in g/100ml)	frequency
6.7	1
7.5	1
8.3	1
9.1	2
9.8	1
9.9	1
10	1
11.4	1
12.4	1
total	n=10

and by using the calculator we will get that

$\bar{X}_A = 9.43$, $S_A^2 = 2.8751 \Rightarrow S_A = 1.70$

so we can see that $\bar{X}_A > S_A^2$

b) here we have another group "B", but we did not have the data of it which the size of it is 15=n.

but we have that $\sum_{i=1}^{15} X_i = 150$ and $\sum_{i=1}^{15} X_i^2 = 1521$

to find the $\bar{X}_B$ and S_B^2 we need to use some rules for the two which are:

[if we have $x_1, x_2, \dots, x_n$ the sample values of the variable where n is the sample size, then the sample mean is

$$\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

and the sample variance is $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2}{n-1}$

and the sample standard deviation is $S = \sqrt{S^2}$]

so $\bar{X}_B = \frac{\sum_{i=1}^{15} X_i}{15} = \frac{150}{15} = 10$ and $S_B^2 = \frac{\sum_{i=1}^{15} X_i^2 - n(\bar{X}_B)^2}{15-1} = \frac{1521-(15)(10)^2}{14} = 1.5 \Rightarrow S_B = \sqrt{1.5} = 1.22$

لا يكتب في هذا الهامش

Variation
or

Now, to answer which of two groups "A" and "B" is more variable we will find each one of them the coefficient of variation (C.V.) :

$$C.V_A = \frac{S_A}{\bar{x}_A} \times 100 = \frac{1.7}{9.43} \times 100 = 18.06\ \%$$

$$C.V_B = \frac{S_B}{\bar{x}_B} \times 100 = \frac{1.22}{10} \times 100 = 12.247\ \%$$

$$\therefore C.V_A > C.V_B$$

So the haemoglobin levels of group "A" are more variable then the haemaglobin levels of group "B"

or

the haemaglobin levels of group "B" are less variable than the haemaglobin levels of group "A"

لا يكتب في هذا الهامش

"Q# 11"

given: $\sum_{i=1}^{8} X_i = 495$, $\sum_{i=1}^{8} X_i^2 = 30659$, the sample size $= n = 8$

a) $\bar{X} = \frac{\sum_{i=1}^{8} X_i}{8} = \frac{495}{8} = 61.875$

b) $S^2 = \frac{\sum_{i=1}^{8} X_i^2 - (n)(\bar{X})^2}{8-1} = \frac{30659 - (8)(61.875)^2}{7} = 4.4107$

c) $S = \sqrt{S^2} = \sqrt{4.4107} = 2.1$

d) $CV = \frac{S}{\bar{X}} \times 100 = \frac{2.1}{61.875} \times 100 = 3.39\%$

لا يكتب في هذا الهامش

"Q#12"

we will calculate the coefficient of variation (C.V.) for each school :

$$C.V.(School\ A) = \frac{S(School\ A)}{\bar{x}(School\ A)} \times 100 = \frac{6}{30} \times 100 = 20\%$$

$$C.V.(school\ B) = \frac{S(school\ B)}{\bar{x}(School\ B)} \times 100 = \frac{10}{60} \times 100 = 16.67\%$$

$\therefore$ C.V.(school A) > C.V.(school B)

$\therefore$ the weights of students from school A are more variable than the weights of students from school B

or

the weights of students from school B are less variable than the weights of students from school A

لا يكتب في
هذا الهامش

" Q # 13 "

we will calculate the coefficient of variation (C.V.) for each one:

$$C.V.(\text{haemoglobin level}) = \frac{S(\text{haemoglobin level})}{\bar{x}(\text{haemoglobin level})} \times 100 = \frac{1.2}{13.7} \times 100 = 8.76\,\%$$

$$C.V.(RBC) = \frac{S(RBC)}{\bar{x}(RBC)} \times 100 = \frac{.5}{4.5} \times 100 = 11.1\,\%$$

$\therefore$ C.V.(haemoglobin level) < C.V.(RBC)

$\therefore$ the haemoglobin levels of women are less variable than the RBC of women

or

the RBC of women are more variable than the haemoglobin levels of women

//

لا يكتب في هذا الهامش

"Q#14"

we will calculate the coefficient of variation (C.V.) for each one:

$$C.V.(age) = \frac{S(age)}{\overline{X}(age)} \times 100 = \frac{6}{30} \times 100 = 20\%$$

$$C.V.(haemaglobin\ level) = \frac{S(haemoglobin\ level)}{\overline{X}(haemoglobin\ level)} \times 100 = \frac{10}{60} \times 100 = 16.67\%$$

$\therefore C.V.(age) > C.V.(haemoglobin\ level)$

∴ the ages of the patients are more variable than the haemaglobin levels of the patients.

or

the haemaglobin levels of the patients are less variable the the ages of the patients.

The sample mean and its properties

Suppose we have a sample of size n

$$X_1, X_2, \ldots, X_n$$

from a population that we are studying.

The sample mean ("average")

$$\bar{X} = (X_1 + \cdots + X_n)/n = \sum_i X_i/n$$

The sample variance

$$s_x^2 = \sum_i (X_i - \bar{X})^2/(n-1)$$

Theorem

If c is a constant, and $y_i = x_i + c$, then $\bar{y} = \bar{x} + c$.

If c is a constant, and $y_i = cx_i$, then $\bar{y} = c\bar{x}$

Theorem

If c is a constant, and $y_i = x_i + c$, then $s_y = s_x$, $s_y^2 = s_x^2$

If c is a constant, and $y_i = cx_i$, then $s_y = |c| s_x$

And

$$s_y^2 = c^2 s_x^2$$

Now we take the square root of both sides, recalling the basic algebraic fact that $\sqrt{c^2} = |c|$, to obtain $s_y = |c| s_x$, as claimed.

"Q # 15"

لا يكتب في هذا الهامش

1) The variable is the number of children for 15 women, so the type of it is quantitative - discrete ∴ e

for 2, 3, 4 and 5 we doing the following :

# of children	frequency	order of the sample data
0	2	$X_{(1)}, X_{(2)}$
1	2	$X_{(3)}, X_{(4)}$
2	3	$X_{(5)}, X_{(6)}, X_{(7)}$
3	2	$X_{(8)}, X_{(9)}$
4	3	$X_{(10)}, X_{(11)}, X_{(12)}$
5	2	$X_{(13)}, X_{(14)}$
6	1	$X_{(15)}$
total	15 women	

Note that the difference between ordered sample and unordered sample :

unordered sample ....

$X_1=3, X_2=5, X_3=2, X_4=1, X_5=4, X_6=3, X_7=5, X_8=4, X_9=0, X_{10}=1, X_{11}=2$

$X_{12}=6, X_{13}=4, X_{14}=0, X_{15}=2$ (as is given exactly in the question.)

ordered sample ....

$X_{(1)}=0, X_{(2)}=0, X_{(3)}=1, X_{(4)}=1, X_{(5)}=2, X_{(6)}=2, X_{(7)}=2, X_{(8)}=3, X_{(9)}=3, X_{(10)}=4, X_{(11)}=4$

$X_{(12)}=4, X_{(13)}=5, X_{(14)}=5, X_{(15)}=1$

2) by using the calculator we get $\bar{X}=2.8$ ∴ d

3) as the sample size is odd which is $n=15$ then

$\text{median} = X_{(\frac{n+1}{2})} = X_{(\frac{15+1}{2})} = X_{(\frac{16}{2})} = X_{(8)} = 3$ ∴ a

4) the sample mode is the value of the # of the children who has the largest frequency, so we have two sample mode which they 2 and 4

∴ b

5) by using the calculator we get $S_x = 1.859$ ∴ d

7) the coefficient of variation $= C.V._{\bar{X}} = \frac{S_x}{\bar{X}} \times 100 = \frac{1.859}{2.8} \times 100 = 66.39\%$ ∴ d

لا يكتب في هذا الهامش

* add 2 to each number of children i.e. $y_i = x_i + 2 \quad \forall \; 1 \le i \le (n = 15)$

8.) $\bar{y} = \bar{x} + 2 = 2.8 + 2$

So we can see that the sample mean $\bar{y}$ is increased (+) by 2 from $\bar{x}$

$\therefore d$

9.) $S_y^2 = S_x^2 = (1.859)^2$ So S_y^2 is the same as S_x^2 $\therefore a$

10.) $C.V_y = \frac{S_y}{\bar{y}} \times 100 = \frac{1.859}{2.8+2} \times 100 = 38.73\%$

So we can see that ($C.V_y = 38.73\%$) is smaller than ($C.V_x = 66.39\%$)

$\therefore c$

لا يكتب في هذا الهامش

"Q#16"

1) the variable is the # of asthma cases seen in past months in hospital, so the type of it is quantitative discrete i.e

for 2, 3, 4 and 5 we doing the following :

# of asthma cases	frequency	order of the sample data
6	1	$X_{(1)}$
14	1	$X_{(2)}$
15	1	$X_{(3)}$
16	1	$X_{(4)}$
20	3	$X_{(5)}, X_{(6)}, X_{(7)}$
25	1	$X_{(8)}$
29	1	$X_{(9)}$
30	1	$X_{(10)}$
35	1	$X_{(11)}$
49	1	$X_{(12)}$
total	n=12	

2) by using the calculator we get $\bar{X} = 23.25$ ∴ d

3) the sample mode is the value of the # of asthma cases in past months who has the largest frequency, so it is 20 ∴ e

4) as the sample size is even which is n=12 then

$$\text{median} = \left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\right)/2 = \left(X_{(\frac{12}{2})} + X_{(\frac{12}{2}+1)}\right)/2$$

$$= \left(X_{(6)} + X_{(7)}\right)/2 = (20+20)/2 = 40/2 = 20 \quad \therefore c$$

5) by using the calculator we get $S_x = 11.355$ ∴ f

7) the coefficient of variation $= CV_x = \frac{S_x}{\bar{X}} \times 100 = \frac{11.355}{23.25} \times 100 = 48.84\%$

لا يكتب في هذا الهامش

* multiply by 2 to each number asthma cases in past months

i.e. $y_i = 2x_i \quad \forall\ 1 \le i \le (n=12)$

8) $\bar{y} = 2\bar{x} = 2(23.25)$

So we can see that $\bar{y}$ is multiplied by 2 from $\bar{x}$ ∴ ✓

9) $S_y = |2| S_x = 2S_x = 2(11.355)$

So, S_y is multiplied by 2 from S_x ∴ ✓

10) $CV_y = \frac{S_y}{\bar{y}} \times 100 = \frac{(2)(11.355)}{(2)(23.25)} \times 100 = \frac{11.355}{23.25} \times 100 = 48.84\%$

So we can see that ($CV_y = 48.84\%$) is the same as ($CV_x = 48.84\%$)

∴ a