



لجان الرُّفَعَات

STATISTICS



MORPHINE ACADEMY

MORPHINE
ACADEMY

Pharmaceutical Statistics

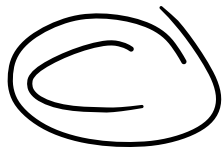
Lecture 7

Descriptive statistics

Measures of Position

Prepared and Presented by

Dr. Muna Oqal



Measures of Position

ملاحظات عن DATA
- يعطى $Q_1, Q_2, Q_3, \max$
- ما يقدر أحسن أو mean بشكل طبيعي
- خذ ما عندك كل الـ DATA

• The Five-Number Summary (FNS)

The five-number summary (FNS) is a set of descriptive measures (توفر مقاييس) that provide information (تزود معلومات) about a data set. It consists of the following five numbers:

1. The **minimum** (smallest observation) (أصغر مشاهدة) (**Min**).
2. The **lower quartile** (السفلي) or first quartile (أول ربع) (**Q_1**).
3. The **median** (الربط الحسابي) (the middle value) or second quartile (**Q_2**) (المتوسط).
4. The **upper quartile** or third quartile (**Q_3**).
5. The **maximum** (أكبر مشاهدة) (largest observation) (**Max**).

Measures of Position

FNS: Min, Q_1 , Q_2 , Q_3 , Max



- FNS helps:

- To describe the spread or variation (approximate distribution shape) of the data. *انتشار و تباين*
وصف لانتشار و تباين او DATA.
- To determine whether or not any data points are outliers (extreme values). *يا بغيره كثير يا قليلا كثير بالنسبة لـ DATA.*
و بنقدر نكتشف الـ outliers.

The 5-Number Summary

FNS: Min, Q_1 , Q_2 , Q_3 , Max

Example

if these are a scores on a 22 point quiz from a class.

5, 12, 14, 17, 21
min Q_1 Q_2 Q_3 max

Example

Find the FNS of the following data:

$$\text{min} = 0.3$$

$$Q_1 =$$

$$Q_2 =$$

$$Q_3 =$$

$$\text{max} = 17.1$$

17.1	2.1
5.8	2.0
5.0	1.0
4.5	1.0
4.3	0.8
4.2	0.7
3.1	0.3

min

$$\hat{Q}_2 \Rightarrow \frac{14+1}{2} = \frac{15}{2} = 7.5$$

$$\text{Median} = Q_2 = \frac{3.1+2.1}{2} = 2.6$$

$$\hat{Q}_1 \Rightarrow \frac{7+1}{2} = 4$$

$$Q_1 = 1$$

$$\hat{Q}_3 \Rightarrow \frac{7+1}{2} = 4$$

$$Q_3 = 4.5$$

FNS

0.3 1.0 2.6 4.5 17.1

The 5-number summary

Example

5 12 14 17 21

So, if these are a scores on a 22 point quiz from a class...

- The lowest score in the class was 5 points
- 25% of students earned 12 or fewer points (75% earned 12 or more)
- 50% of students earned 14 or fewer points (50% earned 14 or more) – the median
- 75% of students earned 17 or fewer points (25% earned 17 or more)
- The highest score in the class was 21 points

The 5-number summary

Example

- Find the 5-number summary of the following data (which are salaries (in millions) of an NBA team:
- Go to the next slide to check your work.

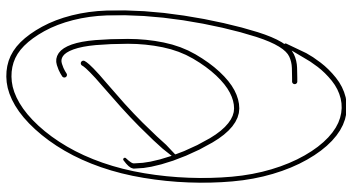
17.1	2.1
5.8	2.0
5.0	1.0
4.5	1.0
4.3	0.8
4.2	0.7
3.1	0.3

Measures of Position

- Your answer should be:

0.3 1.0 2.6 4.5 17.1

- Another measure of spread is found from the 5-number summary, the interquartile range (or IQR).
- The IQR is simply the 3rd quartile (Q3) minus the 1st quartile (Q1).
- So, $IQR = Q3 - Q1$
- This is simply a variation on the definition of range.



Boxplot

test
صن اُحد ار outliers

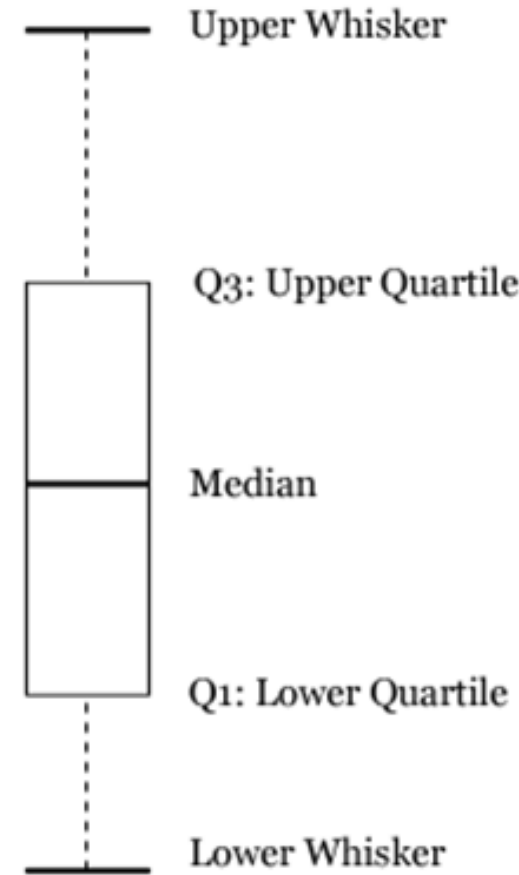
- **The Box and Whisker Plot (Boxplot):**

- Can be useful for handling many data. It shows only certain statistics rather than all the data set.

- The FNS can be used to construct a simple graph called Box-and Whisker plot or simply Boxplot.

- It can be helpful in:

1. Comparing between several groups of data sets.
2. Identifying outliers in the data set.
3. Identifying the approximate shape of the distribution of a data set.

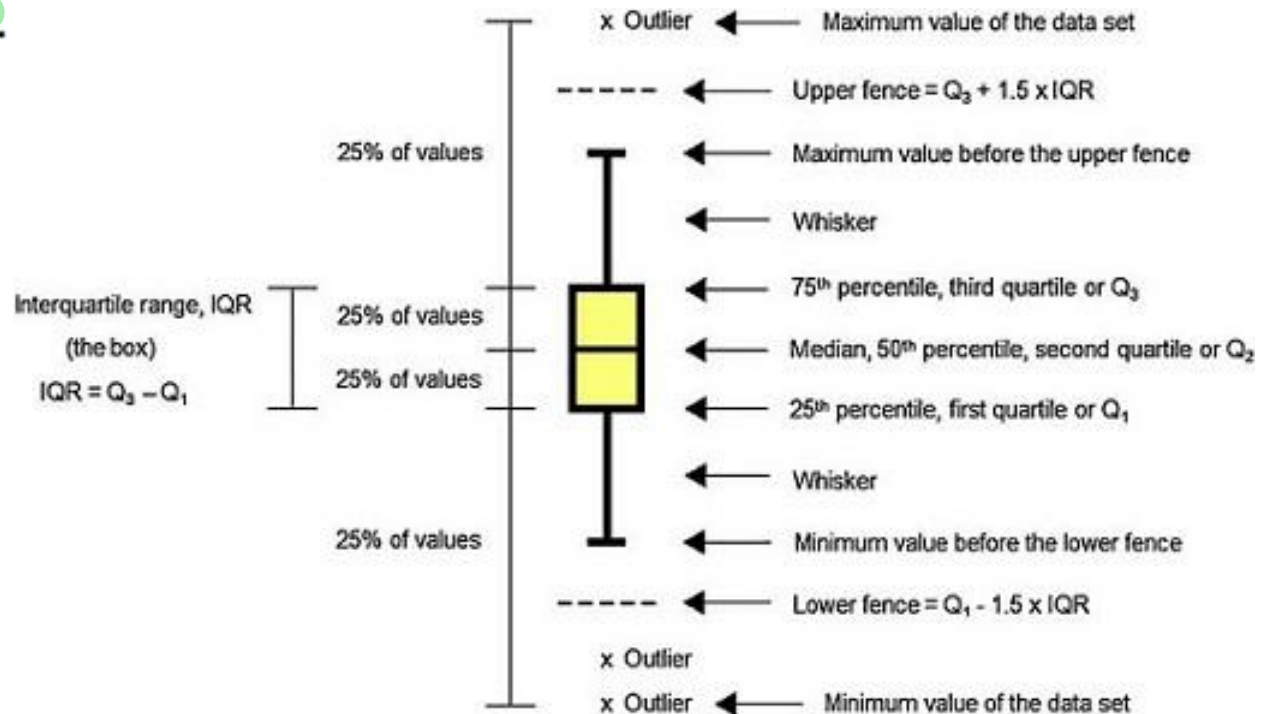


Boxplot

تعرض

FNs

- The boxplot displays the 5-number summary: minimum, lower quartile (Q1), median upper quartile (Q3), maximum.)
- It also shows the Inter-quartile range (IQR) ($Q_3 - Q_1$) and outliers.
- It also gives us information about the symmetry of the distribution.



BoxPlot

كيف أعمل BoxPlot ؟؟

① حدد ال FNS

(min Q_1 Q_2 Q_3 max)

② حدد IQR

$$= Q_3 - Q_1$$

③ نرسم خط أفقي ثم نرسم فوهة مثل

④ نقسم الصندوق إلى نصفين من طرف رسم خط عمودي داخل الصندوق عند Q_2 .

⑤ نرسم خط أفقي صغير (whisker) من طرف

الصندوق الأيسر إلى (min) ونرسم خط أفقي من الطرف الأيمن إلى (max)

كيف أقرأ ال box and whisker plot ؟؟

① النقاط التي تقع خارج ال box و whiskers تسمى **outliers**.

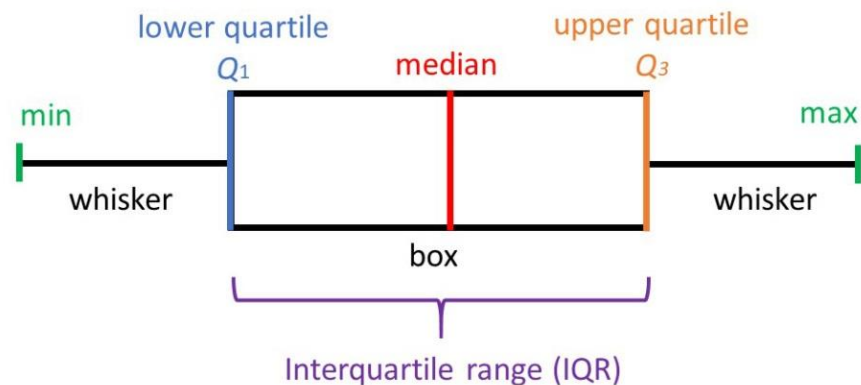
② أقصى وأدنى قيم تظهر في نهاية كل whisker.

③ ال Q_1 و Q_3 يقعان في نهاية ال box.

BoxPlot کيف اُصل Construction of Boxplot

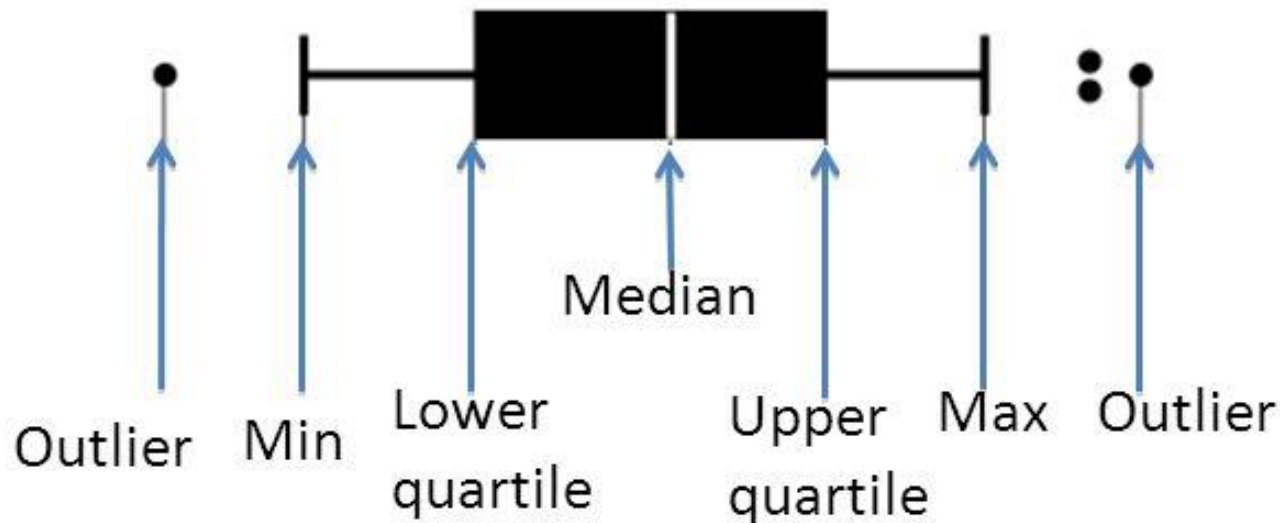
- To construct a Boxplot for a data set you have to do the following steps:

1. Determine the FNS. *فرد FNS* $\begin{matrix} \rightarrow \text{min} \\ \rightarrow Q_1, Q_2, Q_3 \\ \rightarrow \text{max} \end{matrix}$
2. Find the value of IQR. $IQR = Q_3 - Q_1$ *د برسم فوقيه مسمايل*
3. Draw a horizontal line and draw a box above it with the right and left ends of the box at the location of Q_1 and Q_2 .
4. *نقسم* Divide the box into two parts by drawing a vertical line *سمودي* through the box at the location of the median (Q_2). *کنند Q_2*
5. Draw a horizontal line *افقي* called a **whisker** *(min)* from the left end of the box to the minimum value *(max)* and then draw another horizontal line or whisker from the right end of the box to the maximum value.

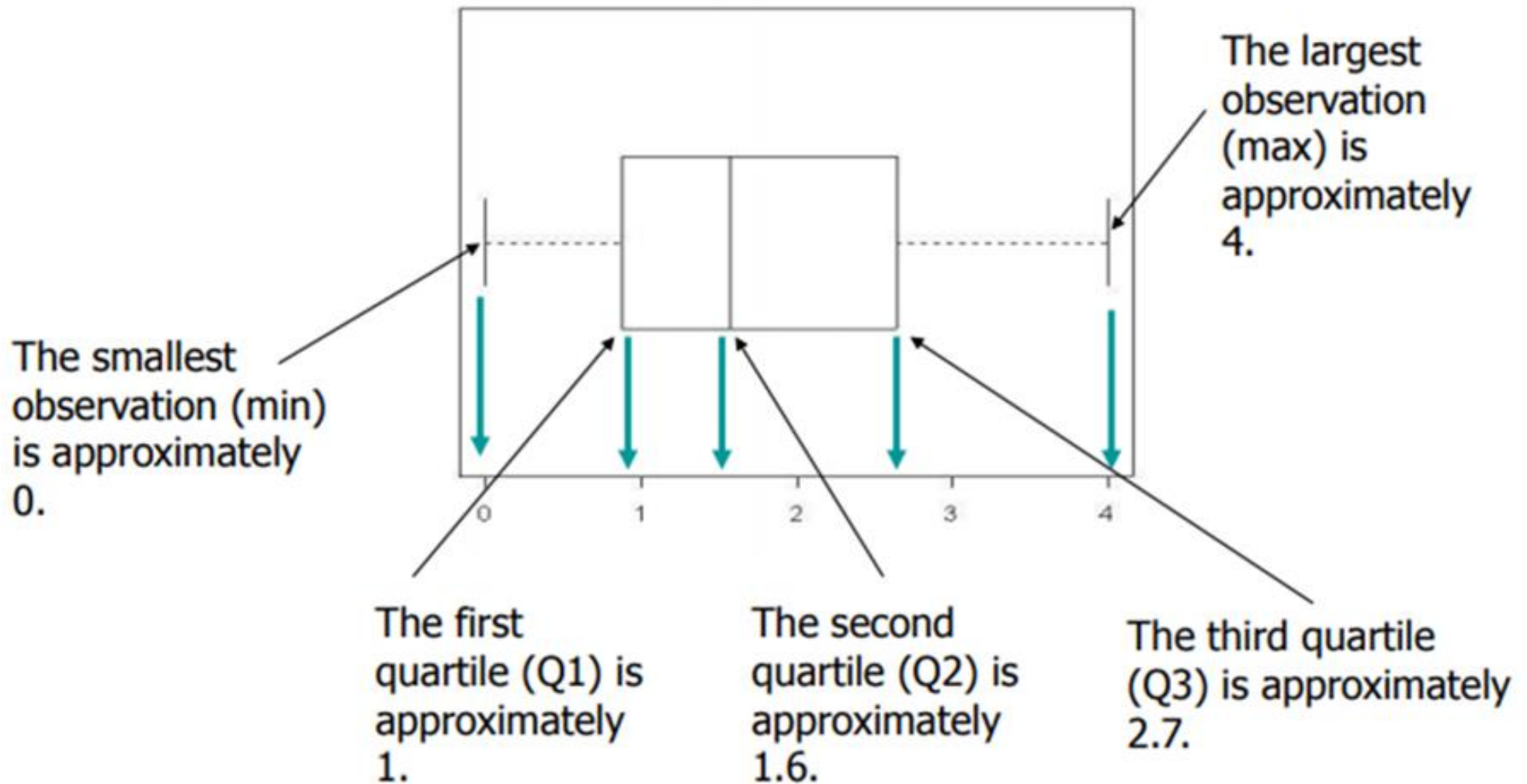


How to read a box-and-whisker plot.

- Dots outside of the box and whiskers are **outliers**.
خارجی نقاط *السياجات* *الفناری*
- Maximum and minimum values are shown at the ends of each whisker.
(max) *(min)* *موجود بنہايات* *whiskers*
- The upper and lower quartiles are the ends of the box.
Q3 *Q1* *موجودان بنہايات* *Box*
- The median is the line in the middle of the box.
Q2
- The max, min, LQ, UQ, and Median make up the “Five Number Summary”.
Q1 *Q3* *lower* *upper*



Example Boxplot



Example

Example

– Verbal GMAT scores of 12 students: 10, 22, 24, 27, 31, 33, 39, 40, 42, 43, 44, 45

5 6 6.5 7 8 9 10 11 12

1 2 3 4

(FNS)

ألف ختمو بي انظروا

min
= 10

Q₁

Q₂

Q₃

max

= 45

موقع

Q₁

$$= \frac{6+1}{2}$$

$$= 3.5$$

$$Q_1 = \frac{24+27}{2}$$

$$= 25.5$$

موقع

Q₂

$$= \frac{12+1}{2}$$

$$= 6.5$$

$$Q_2 = \frac{39+40}{2}$$

$$= 39.5$$

موقع

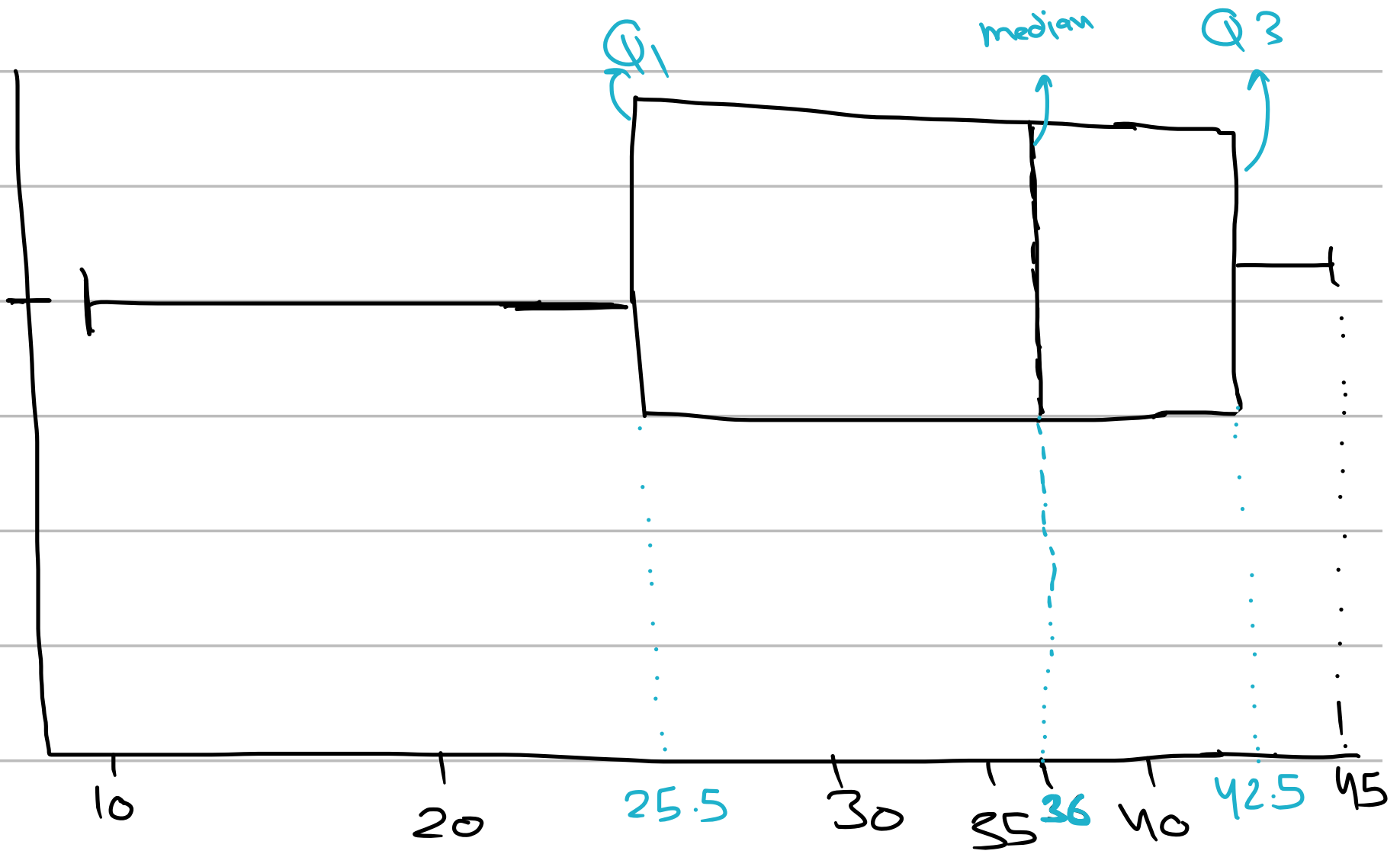
Q₃

$$= \frac{6+1}{2}$$

$$= 3.5$$

$$Q_3 = \frac{42+43}{2}$$

$$= 42.5$$

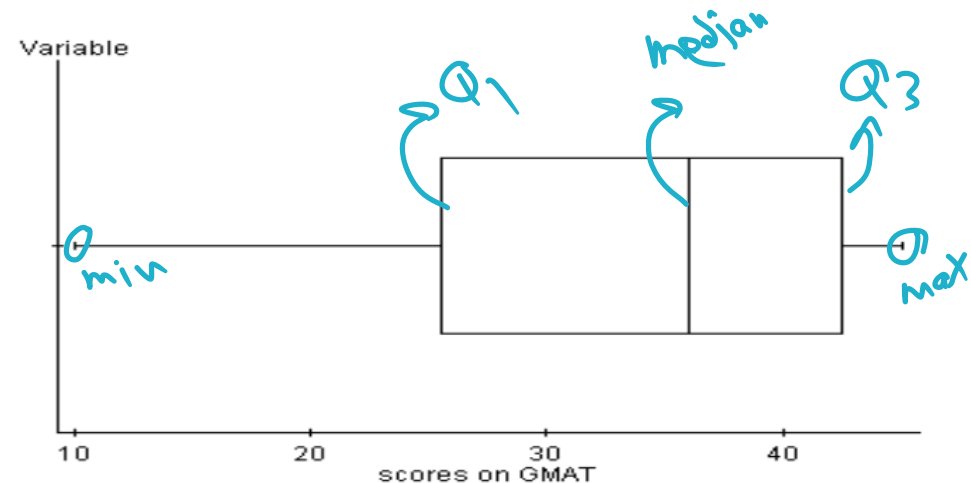


Boxplot

- **Example**

- Verbal GMAT scores of 12 students: 10, 22, 24, 27, 31, 33, 39, 40, 42, 43, 44, 45
- The 5-number summary is:
10 25.5 36 42.5 45
- Now the box-plot is constructed as follows:
 - The line inside the box indicates the median.
 - The left side of this box indicates the lower quartile (Q1).
 - The right side of this box indicates the upper quartile (Q3).
 - A straight line is then drawn from the lowest value of this distribution to the box (at Q1) and another straight line from the box (at Q3) to the highest value of this distribution.

Boxplot



Testing Our Way to Outliers

- Outliers can occur by chance, as a result of an error in during data entry, or as an error in measurement of a variable, or from error in sampling.

المشوار
بالصدفة
نتيجة
خطأ
ادخال
قياس
نظام الـ outliers: ← خطأ في ادخال البيانات
← خطأ في القياسات
← خطأ في أخذ العينة
← هدرها
- Once they have been encountered, their origin should be investigated.

التحري
قد يكون انتاج القيمة الشاذة
- The quartiles and IQR both are used to determine the outliers for the distribution of data set by calculating the Fence following Tukey's rule which considers a data point is an outlier if it falls more than 1.5 IQR below than Q_1 or 1.5 IQR above Q_3 .

قيمة (outlier)
أكثر من
أقل من
من
من
تعتبر قيمة شاذة
تعتبر قيمة شاذة
كبيرة
صغيرة

صياح (هدر لا يشافد outliers)

Lower Fence (LF) = $Q_1 - 1.5(IQR)$

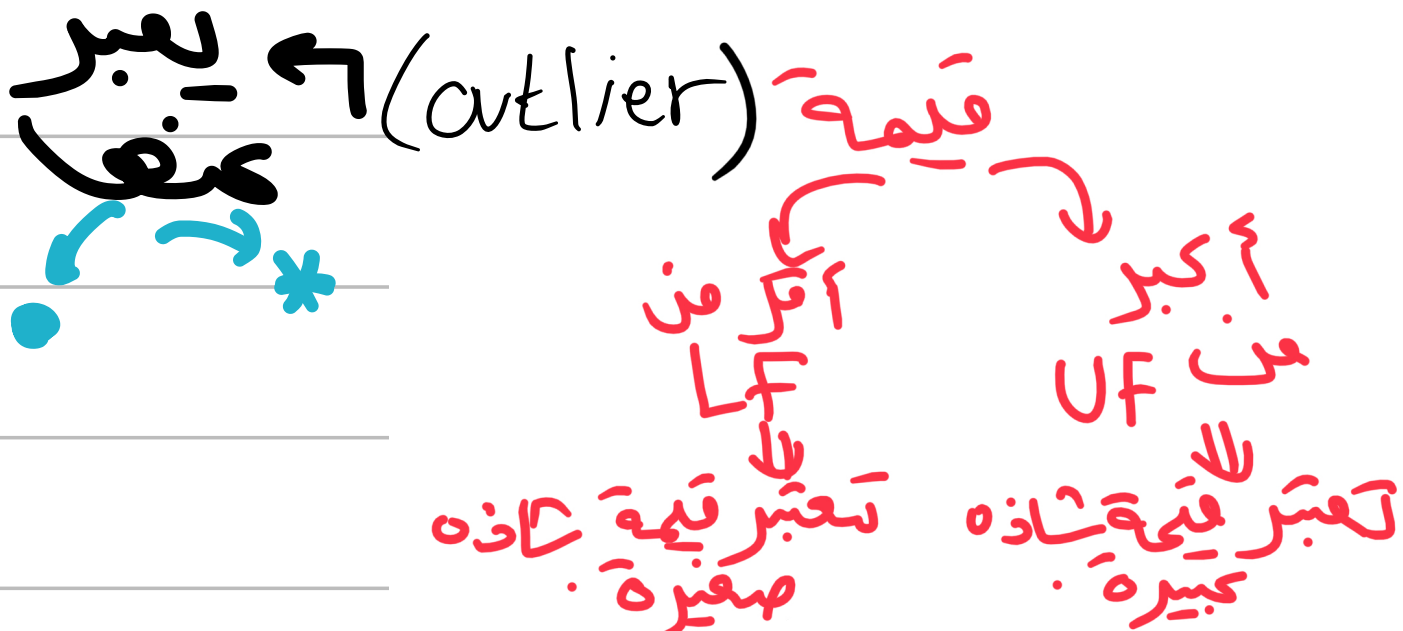
Upper Fence (UF) = $Q_3 + 1.5(IQR)$
- Then, any data value that is less than LF, or greater than the UF, will be considered as an outlier (unusually large or unusually small value), and presented on the graph as an asterisk (*) or solid point (●) outside the whiskers of the Boxplot for a set of data.

في نقاط Boxplot له whiskers

طريقة Tukey's Outliers Rule

$$\text{Lower Fence (LF)} = Q_1 - 1.5(\text{IQR})$$

$$\text{Upper Fence (UF)} = Q_3 + 1.5(\text{IQR})$$



Example

Back to the example which presents the salaries (in millions) of an NBA team and the FNS for the data set was found to be

min Q_1 Q_2 Q_3 max
0.3 1.0 2.6 4.5 17.1

- The 17.1 million dollar salary is quite high.
- Is it an outlier among the data?

$$\begin{aligned} \# \text{ IQR} &= Q_3 - Q_1 \\ &= 4.5 - 1 = 3.5 \end{aligned}$$

ندر
اضرب
UF

$$\begin{aligned} \text{Upper } F &= Q_3 + 1.5(IQR) \\ &= 4.5 + 1.5(3.5) \\ &= 9.75 \end{aligned}$$

$$IS \quad 17.1 > 9.75$$

outlier في 17.1 نعم، بالتالي
وكبير.

Testing Our Way to Outliers

Back to the example which presents the salaries (in millions) of an NBA team and the FNS for the data set was found to be

0.3 1.0 2.6 4.5 17.1

- The 17.1 million dollar salary is quite high.
- Is it an outlier among the data?

Testing Our Way to Outliers

- Recall the summary: 0.3 1.0 2.6 4.5 17.1
 - The IQR = $4.5 - 1.0 = 3.5$
 - Check for the high outlier:
 - **So the Upper Fence = $4.5 + 1.5(3.5) = 9.75$**
 - Is 17.1 more than 1.5IQR above the 3rd quartile (4.5)?
 - Is $17.1 > 1.5(3.5) + 4.5$?
 - Is $17.1 > 9.75$?
 - Yes, so the \$17.1 million salary is an outlier on the team's payroll.
-

(معدل)

Modified Boxplot

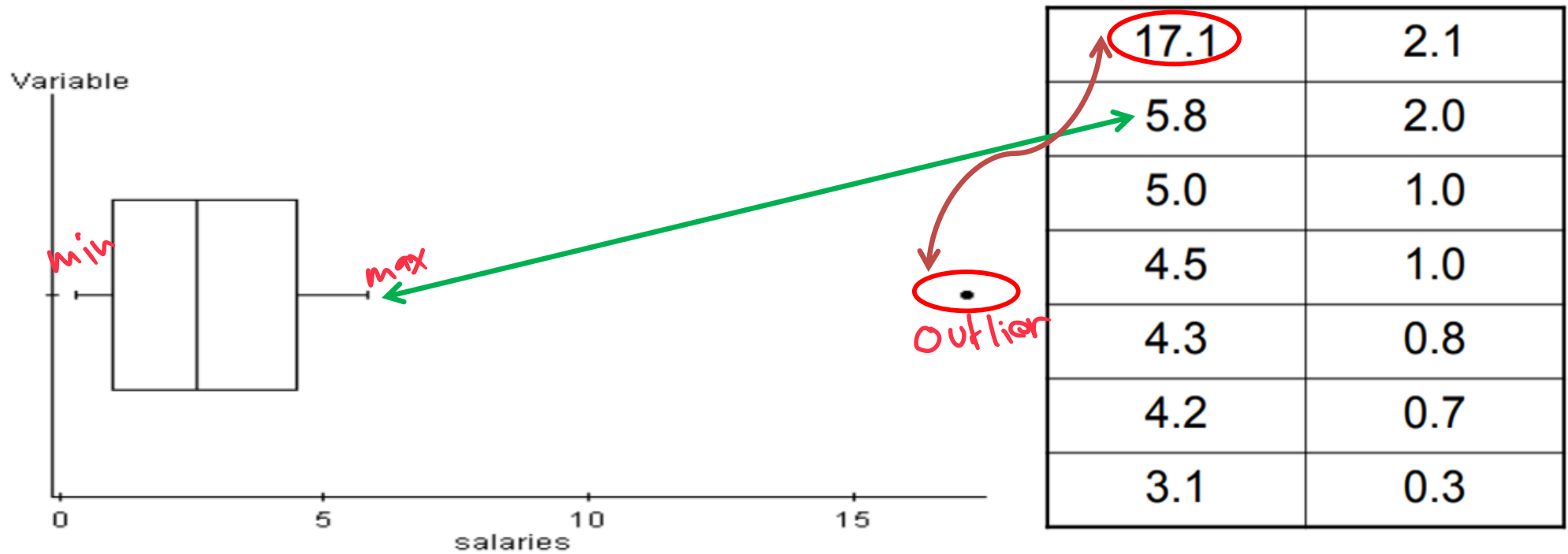
نسخه معدله
من Boxplot

- Consider the 5-number summary of NBA salaries: 0.3 1.0 2.6 4.5 17.1
- The modified boxplot shows outliers separately.
- On the next slide, note how the outlier 17.1 is plotted separately.
- The line only goes to the maximum (or minimum) point that is NOT an outlier.

الفرق بين Boxplot و Modified Boxplot ؟

لو في outliers فابرسها وكأها هين
الحد max و min نرسها كنقاط منفصلة (تكون واضحة بالرسم)

Modified Boxplot



Boxplot & Skewness

هل توزيع البيانات
مثالي لجهة معينة؟
كرد

- The quartiles and boxplot are used to identify the approximate shape for the distribution of the data set as following:
 - If only $Q_2 - Q_1 = Q_3 - Q_2$, then the approximate shape is symmetric or Bell-shaped.
الشكل التقريبي
زِي شكل الجرس المعكوب .
 - If $Q_2 - Q_1 > Q_3 - Q_2$, then the approximate shape is Left-skewed (- ve).
 - $Q_2 - Q_1 < Q_3 - Q_2$, then the approximate shape is Right-skewed (+ ve).

Boxplot & Skewness

طبيعي

Normal Distribution

$$(\text{Quartile 3} - \text{Quartile 2}) = (\text{Quartile 2} - \text{Quartile 1})$$

(Symmetric) (Bell-shaped)



رأى لليسار

Positive Skew

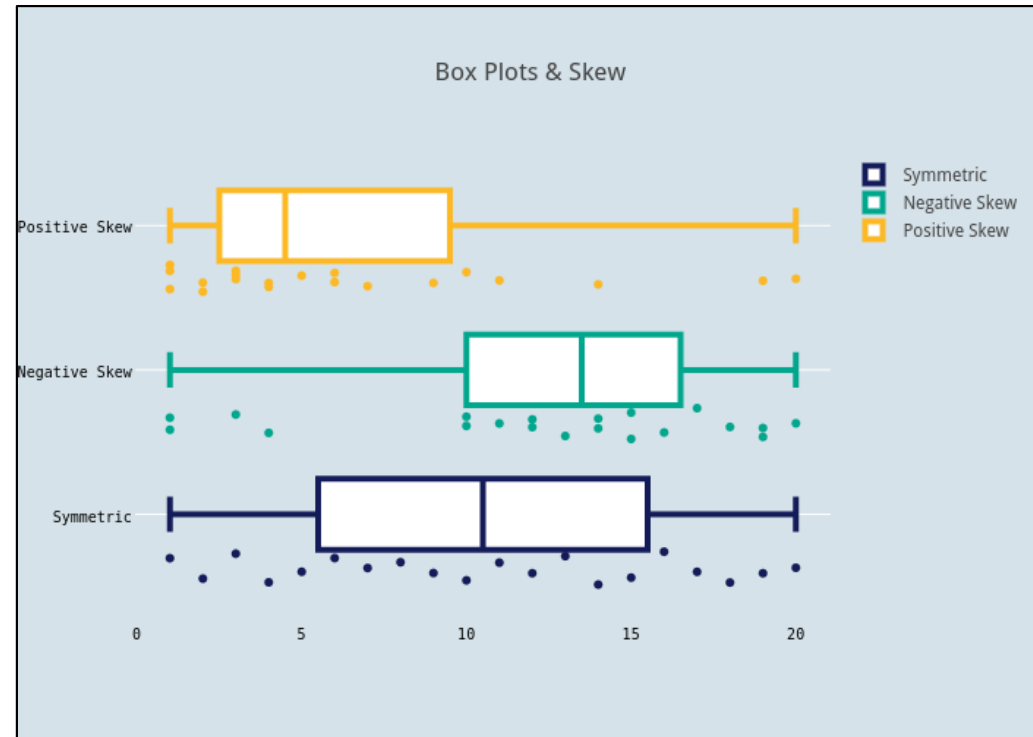
$$(\text{Quartile 3} - \text{Quartile 2}) > (\text{Quartile 2} - \text{Quartile 1})$$



رأى لليمين

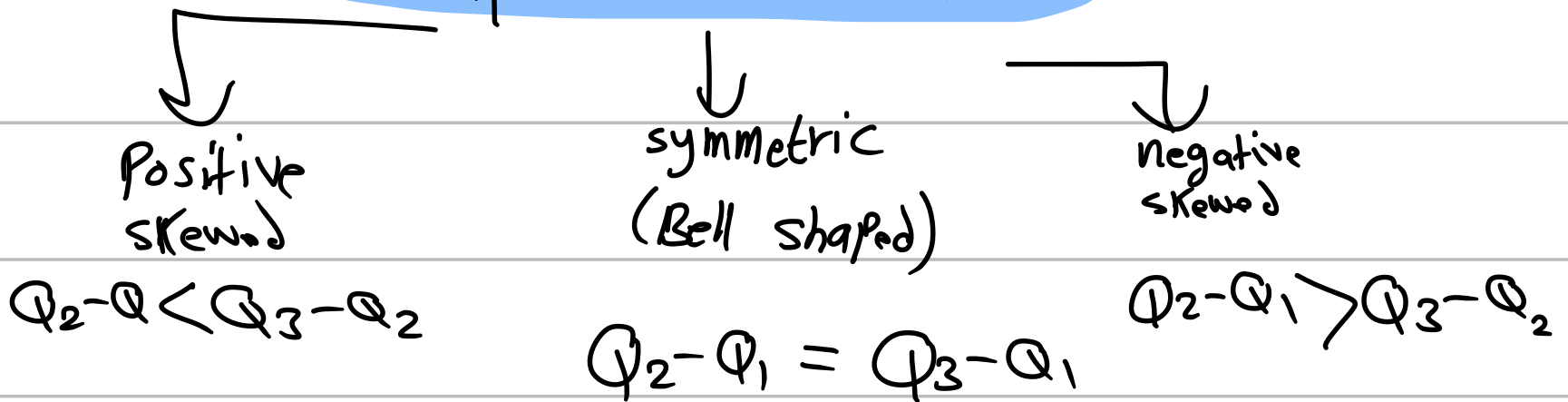
Negative Skew

$$(\text{Quartile 3} - \text{Quartile 2}) < (\text{Quartile 2} - \text{Quartile 1})$$



عكسه وابتكر
يعني لو شفته رأى لليمين
Positive يعني رأى على يالك انه
عكس تفكيرك اذا هو negative
لو شفته رأى لليسار
يعني رأى على يالك انه negative
عكس تفكيرك اذا هو positive

Boxplot and Skewness



Example

بين لوسفته زايح لليمين
 Positive انكس
 بين لوسفته زايح لليسار
 Negative انكس
 بين لوسفته زايح لليسار
 Negative انكس

• Example

The FNS for the weekly salaries in JD for a random sample of 22 pharmacists selected from the records of the ministry of health in Jordan are given as follows:

111 182 221.5 319 439
 min Q_1 Q_2 Q_3 max

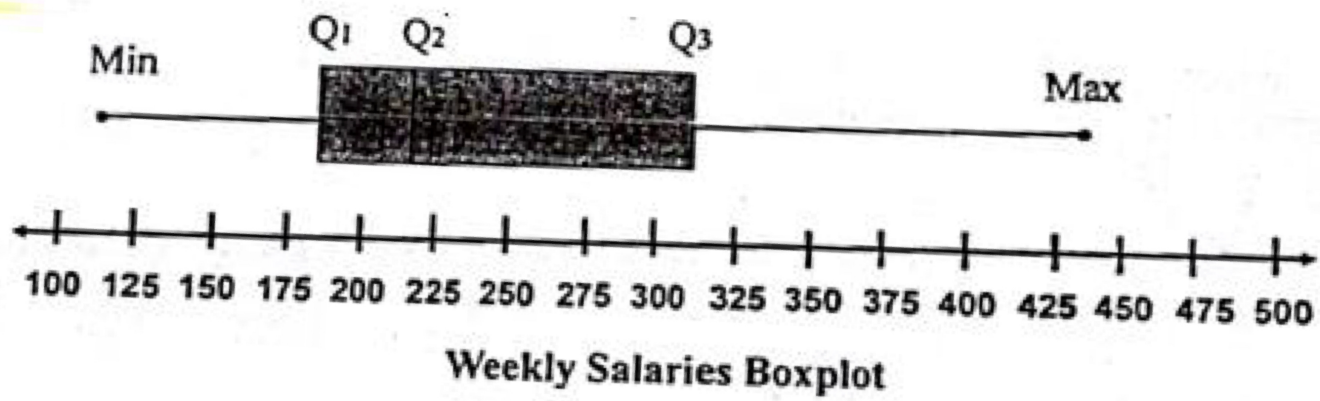
Construct a boxplot for the data above?

$$IQR = Q_3 - Q_1 = 319 - 182 = 137$$

$$Range = 439 - 111 = 328$$

$$LF = Q_1 - 1.5(IQR) = -23.5$$

$$UF = Q_3 + 1.5(IQR) = 521.5$$



1

Positive
Skewed

زائج الیسا از او

بیا کیفردت؟

$$Q_2 - Q_1 \quad \boxed{<} \quad Q_3 - Q_2$$

$$221.5 - 182 \quad \boxed{<} \quad 319 - 221.5$$

$$39.5 \quad \boxed{<} \quad 97.5$$

2) مانی ولا قیمة لال DATA اقل

من L.F ولا أكبر من U.F معناه لا يوجد outliers.

Boxplot

- **Example**

The FNS for the weekly salaries in JD for a random sample of 22 pharmacists selected from the records of the ministry of health in Jordan are given as follows:

111 182 221.5 319 439

Construct a boxplot for the data above?

$Q1=182$ $Q2=221.5$ $Q3=319$ $\text{Min}=111$ $\text{Max}=439$

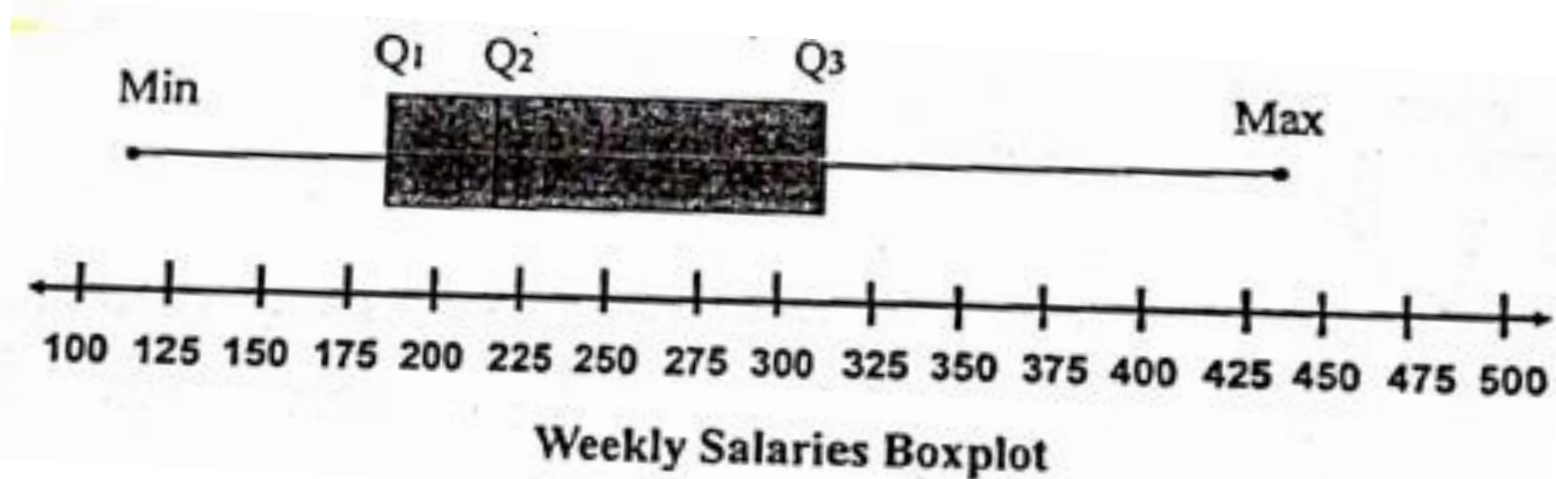
$\text{IQR}=319-182=137$

$\text{Range}=439-111=328$

$\text{LF}=182-1.5(137)=-23.5$

$\text{UF}=319+1.5(137)=524.5$

Boxplot



Conclusion

1. Approximate shape $Q_2 - Q_1 = 39.5 < Q_3 - Q_2 = 97.5$, then the approximate shape is right-skewed (+ve).
2. Outliers
3. Because there is no value in the data set that is less than the LF, or greater than the UF, then there is no outliers in the weekly salaries data considered.

Example

• Example

The systolic blood pressure in mmHg for a random sample of size 9 middle aged patients selected from king Abdullah University Hospital (KAUK) were as follows:

122, 129, 113, 119, 124, 132, 210, 110, 116

Construct a boxplot for the data above?

• Solution

نرتب / DATA

1 2 3 4 5 6 7 8 9
110, 113, 116, 119, 122, 124, 129, 132, 210

$$\textcircled{1} \max = 210$$
$$\min = 110$$

$$\text{Range} = 210 - 110$$
$$= 100$$

$$\textcircled{2} \hat{Q}_2 = \frac{9+1}{2} = 5$$

$$Q_2 = 122$$

$$\textcircled{3} \hat{Q}_1 = \frac{5+1}{2} = 3$$

$$Q_1 = 116$$

$$\textcircled{4} \hat{Q}_3 = \frac{5+1}{2} = 3$$
$$Q_3 = 210$$

درد Q_2
مع داده ها
از اکان
فردی
لا بار
 Q_1, Q_3

$$\textcircled{5} \text{ IQR} = Q_3 - Q_1 \\ = 13$$

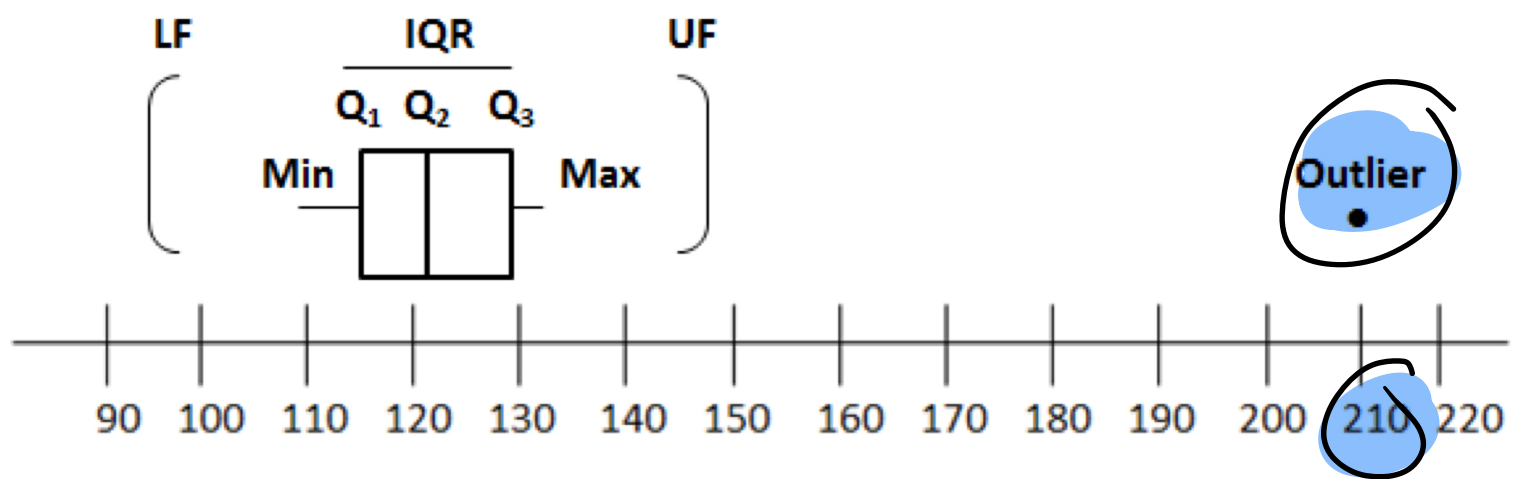
$$\textcircled{6} \text{ LF} = Q_1 - 1.5(\text{IQR}) = 96.5$$

$$\textcircled{7} \text{ UF} = Q_3 + 1.5(\text{IQR}) = 148.5$$

$$Q_2 - Q_1 \quad \boxed{<} \quad Q_3 - Q_2$$

$$6 \quad \boxed{<} \quad 7$$

(right skewed)



$$210 > UF$$

$$210 > 148.5$$

210 ← Outlier د کبیر

Boxplot

- Example

The systolic blood pressure in mmHg for a random sample of size 9 middle aged patients selected from king Abdullah University Hospital (KAUK) were as follows:

122, 129, 113, 119, 124, 132, 210, 110, 116

Construct a boxplot for the data above?

- Solution:

1. Ordered data from smallest to highest value

110, 113, 116, 119, 122, 124, 129, 132, 210

2. $\text{Min} = 110$ $\text{Max} = 210$ $\text{Range} = 210 - 110 = 100$

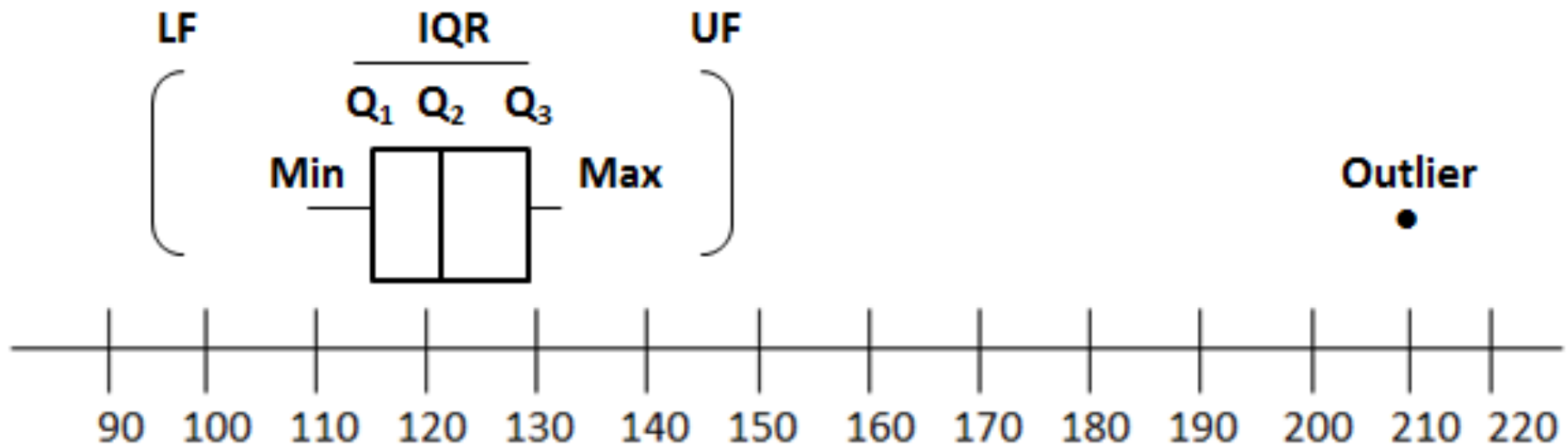
3. $Q_1 = 116$ $Q_2 = 122$ $Q_3 = 129$

4. $\text{IQR} = 129 - 116 = 13$

5. $\text{LF} = 96.5$ $\text{UF} = 148.5$

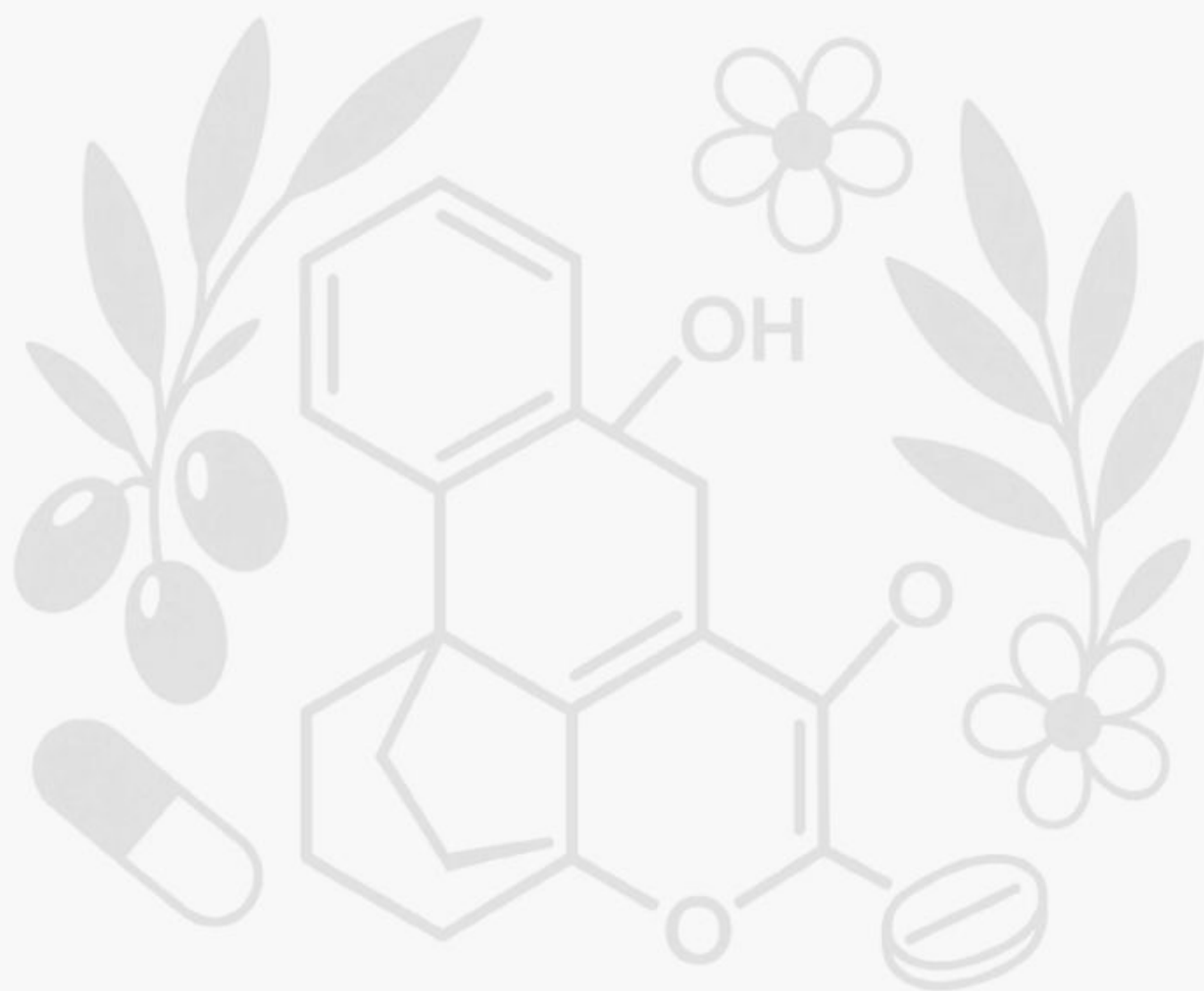
Boxplot

- Conclusion
 - Approximate shape is right-skewed (+ve), as $Q_2 - Q_1 = 6 < Q_3 - Q_2 = 7$.
 - The value 210 mmHg is greater than the UF (148.5), then this value is considered an outlier value.



**The rewards for hard-working
is success; never lose focus,
good luck on your coming exam**





MORPHINE
ACADEMY