

Pharmaceutical Statistics

Lecture 7

Descriptive statistics

Measures of Position

Prepared and Presented by

Dr. Muna Oqal

Measures of Position

- **The Five-Number Summary (FNS)**

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

1. The **minimum** (smallest observation) (**Min**).
2. The **lower quartile** or first quartile (**Q_1**).
3. The **median** (the middle value) or second quartile (**Q_2**).
4. The **upper quartile** or third quartile (**Q_3**).
5. The **maximum** (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.
 - To determine whether or not any data points are outliers (extreme values).

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:

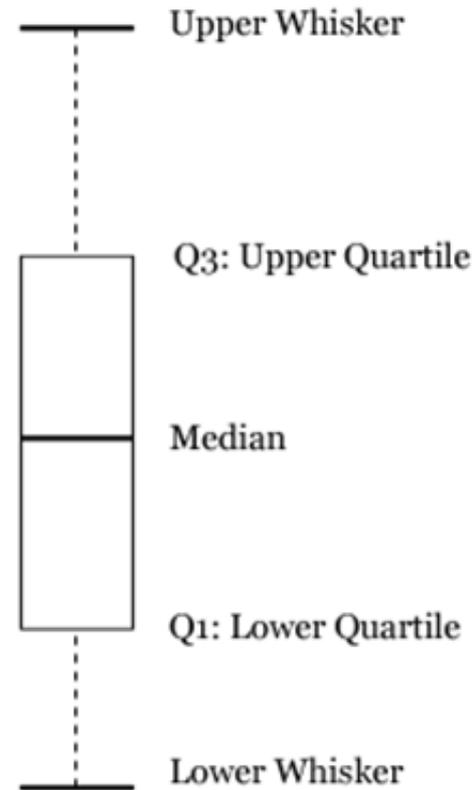
0.3 1.0 2.6 4.5 17.1

- 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

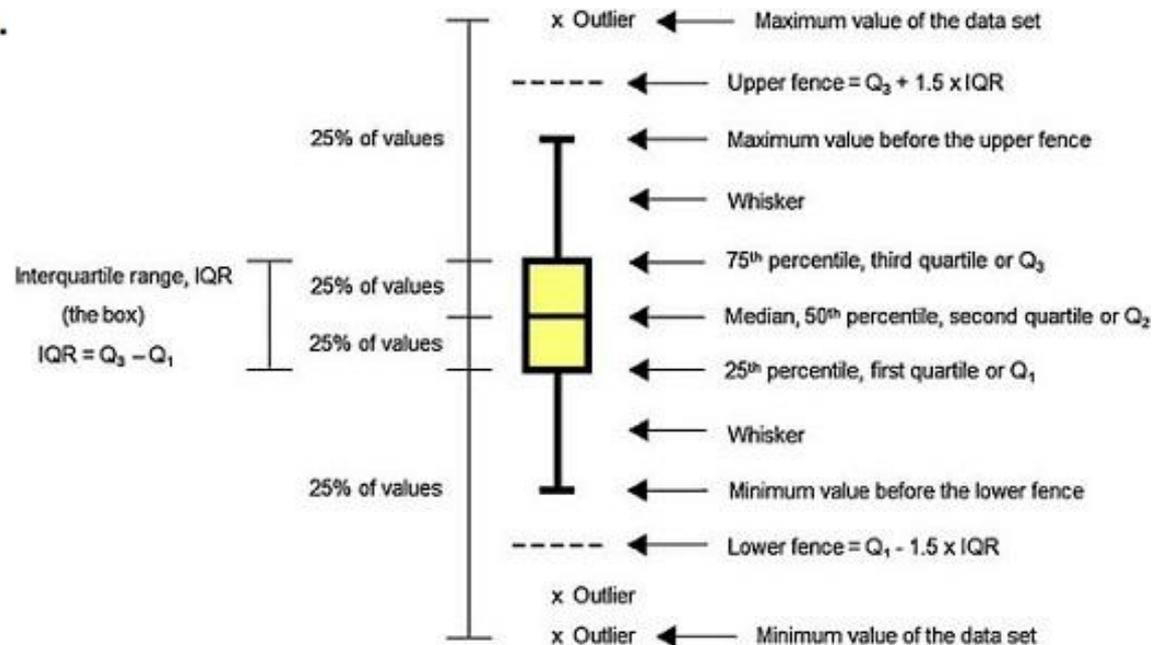
- **The Box and Whisker Plot (Boxplot):**

- 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.



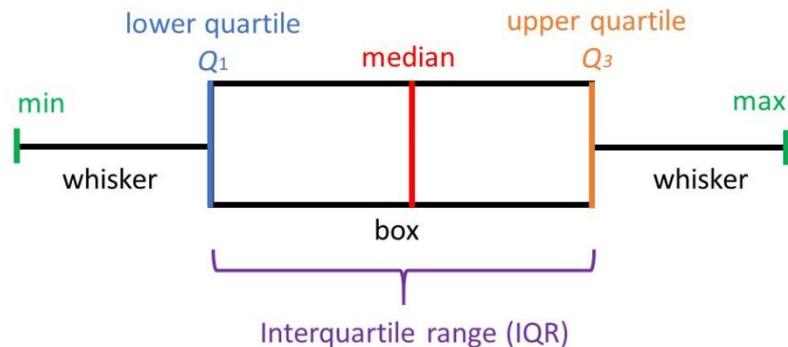
Boxplot

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



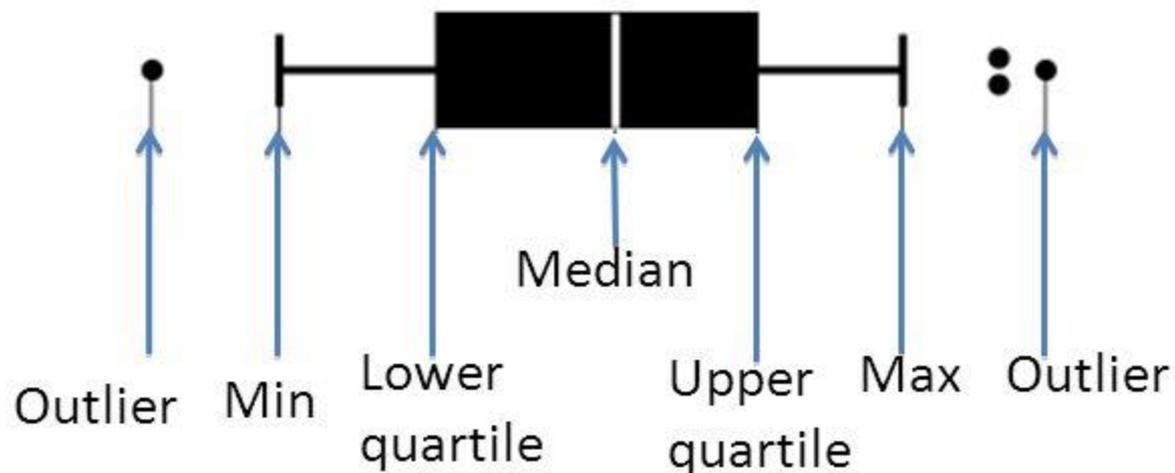
Construction of Boxplot

- To construct a Boxplot for a data set you have to do the following steps:
 1. Determine the FNS.
 2. Find the value of IQR.
 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_3 .
 4. Divide the box into two parts by drawing a vertical line through the box at the location of the median (Q_2).
 5. Draw a horizontal line called a whisker from the left end of the box to the minimum value and then draw another horizontal line or whisker from the right end of the box to the maximum value.

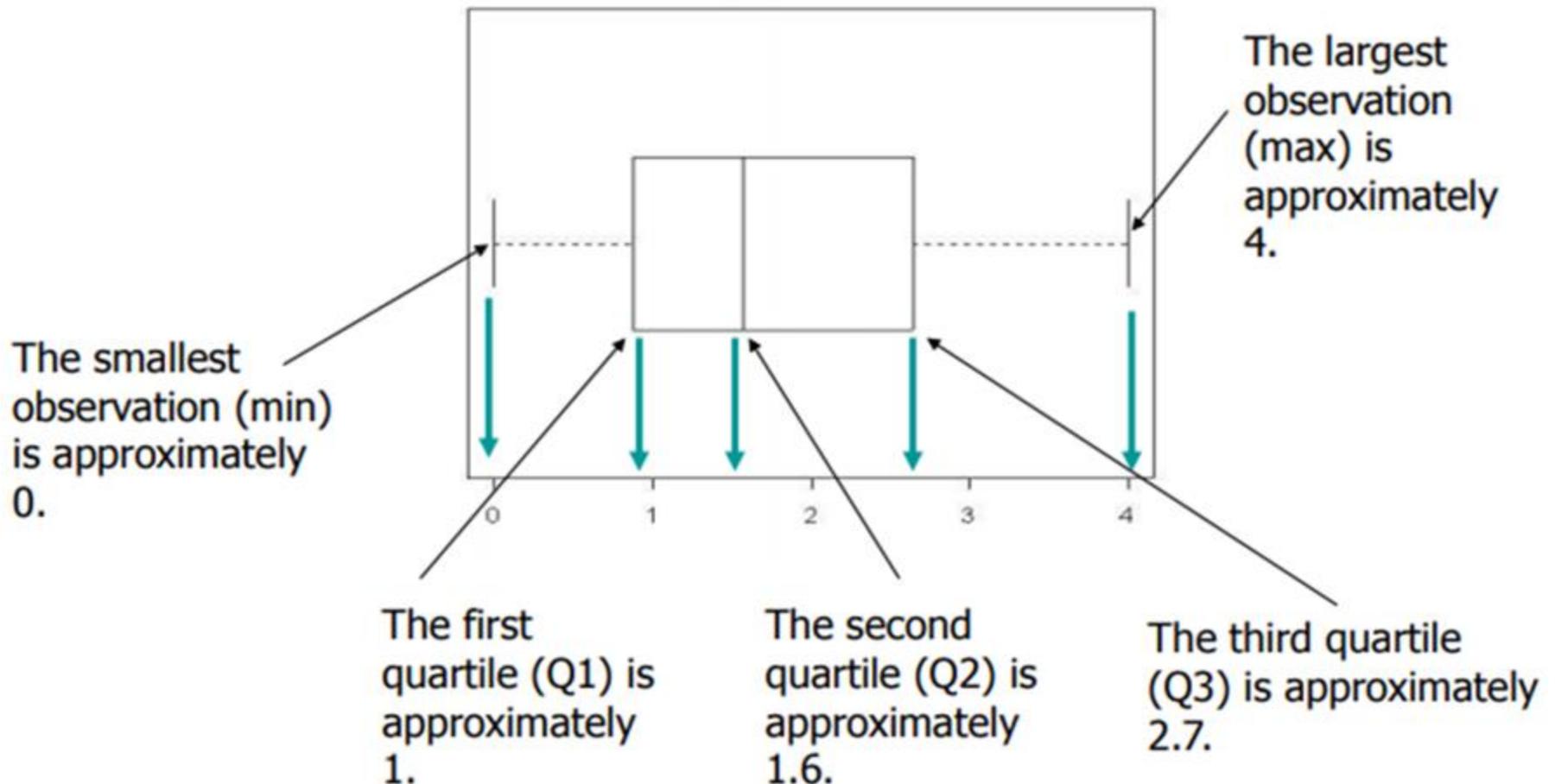


How to read a box-and-whisker plot.

- Dots outside of the box and whiskers are outliers.
- Maximum and minimum values are shown at the ends of each whisker.
- The upper and lower quartiles are the ends of the box.
- The median is the line in the middle of the box.
- The max, min, LQ, UQ, and Median make up the “Five Number Summary”.



Example Boxplot



Boxplot

- **Example**

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

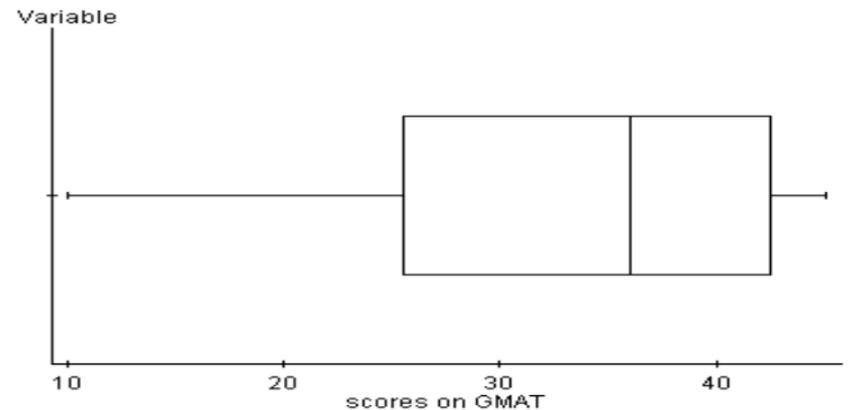
- The 5-number summary is:

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.
- Once they have been encountered, their origin should be investigated.
- 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 .

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

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

- 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.

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?

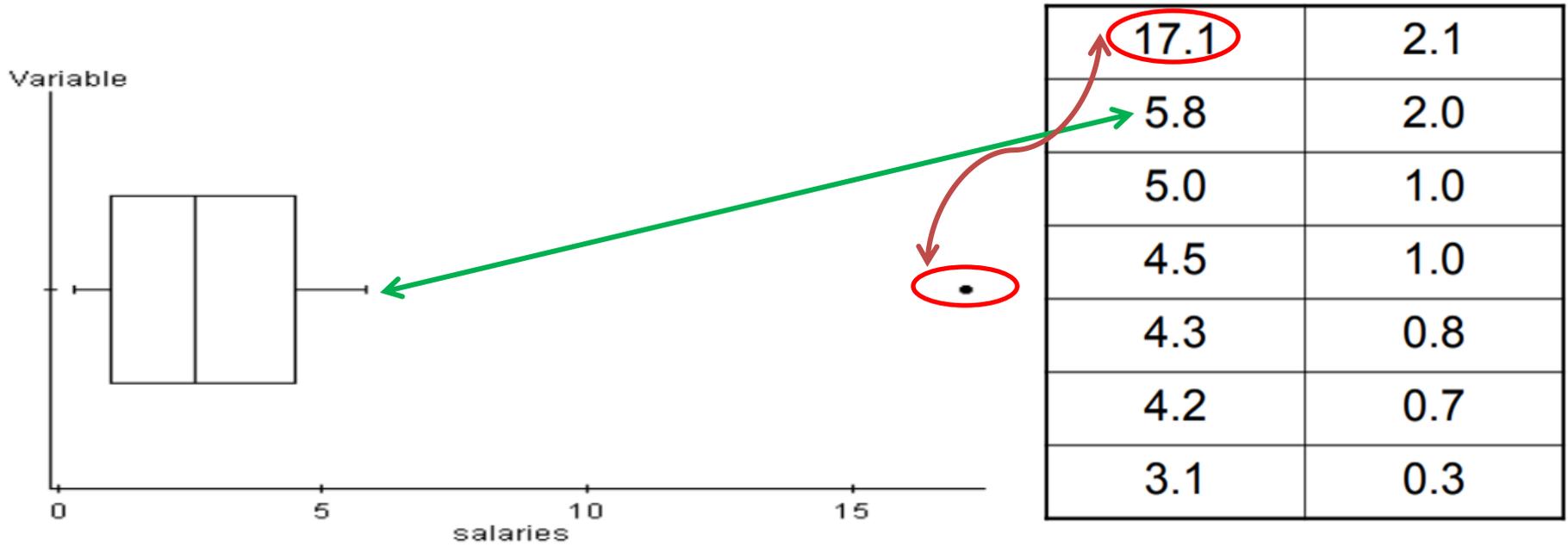
Testing Our Way to Outliers

- 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.

Modified 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.

Modified Boxplot



Boxplot & Skewness

- 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.
 - 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

Normal Distribution

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



Positive Skew

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



Negative Skew

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



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 Min=111 Max=439

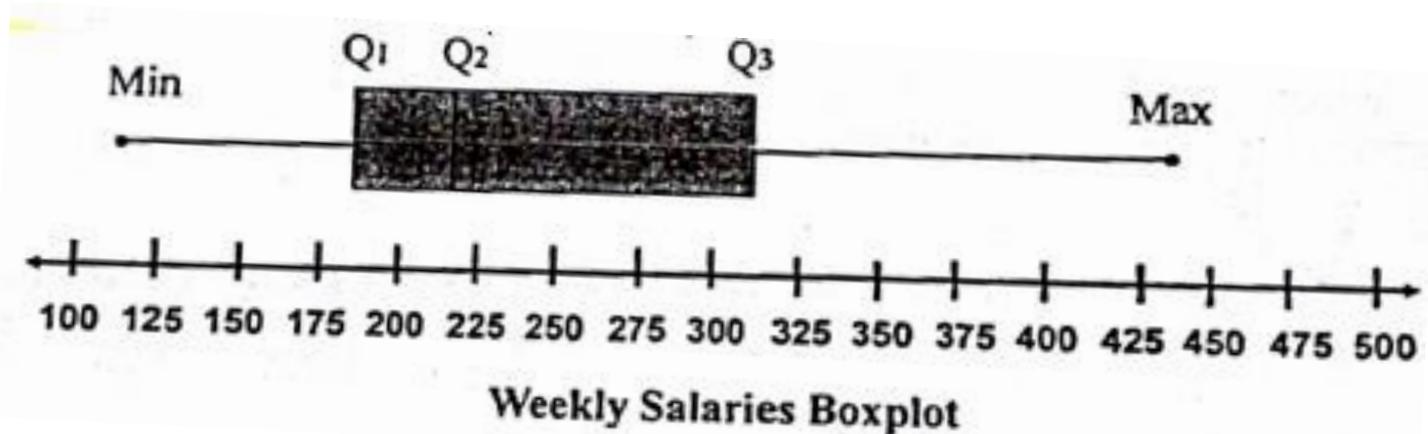
IQR= 319-182= 137

Range=439-111=328

LF=182-1.5(137)= -23.5

UF=319+1.5(137)=524.5

Boxplot



Conclusion

1. Approximate shape $Q_2 - Q_1 = 39.5 < Q_3 - Q_1 = 97.5$, 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.

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. Min = 110 Max=210 Range=210-110=100

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

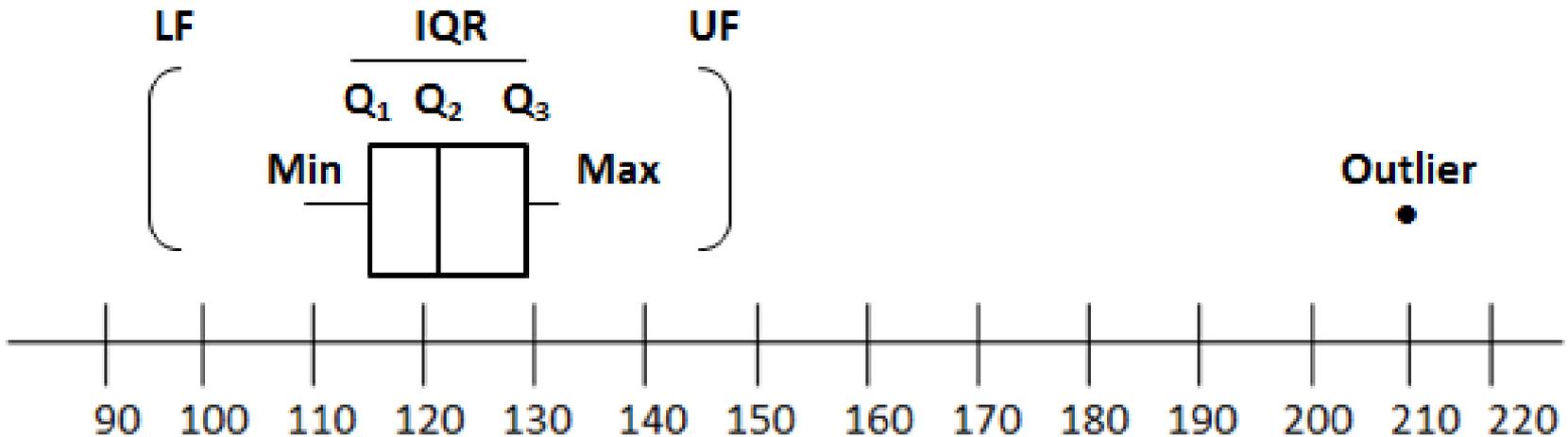
4. IQR=129-116=13

5. LF=96.5 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**

**Good
Luck**

