Pharmaceutical Statistics

Lecture 6 Part 1

Descriptive statistics

Measures of Dispersion

Prepared and Presented by

Dr. Muna Oqal

Measures of Dispersion for Grouped Frequency Table

A. Range:

The range of grouped frequency table can be calculated as follow:

R=(Upper limit of last class – Lower limit for the first Class)+1

Measures of Dispersion for Grouped Frequency Table

- **Example**

The following frequency table represents the time in minutes of 50 pharmacists to their work selected from the records of a given pharmaceutical industries company in Jordan:

Time to travel to work	Frequency
1-10	8
11-20	14
21-30	12
31-40	9
41-50	7

A. Calculate the range for the time travelled to the work for pharmacists?

Solution:

R=(Upper limit of last class – Lower limit for the first Class)+1

R=(50-1)+1=50 minutes

Measures of Dispersion for Grouped Frequency Table

B. Calculate the value of the sample standard deviation (S) for the time travelled to the work for the pharmacists?

B. Variance and Standard deviation

1. Find the midpoint for each class interval
2. For each class interval multiply the frequency with each midpoint (fx).
3. Find the sum $\sum(fx)$.
4. Find the square value for $\sum(fx)$ in step 3 .
5. Then divide the sum in step 4 by the sum of frequencies ($n=\sum f_i$) as follow $\frac{(\sum fx)^2}{n}$.
6. Find the square value of the midpoint for each interval (x^2)
7. For each class multiply the frequency (f) with each squared midpoint (X^2)=(fx^2)
8. Find the sum $\sum fx^2$
9. Calculate the sample variance (S^2) as follow:

$$S^2 = \frac{\sum fx^2 - \frac{(\sum fx)^2}{n}}{n-1}$$

Measures of Dispersion for Grouped Frequency Table

- **Solution:**

Class	Midpoint (x)	f	x^2	fx	fx^2
1 to 10	5.5	8	30.25	44	242
11 to 20	15.5	14	240.25	217	3363.5
21 to 30	25.5	12	650.25	306	7803
31 to 40	35.5	9	1260.25	319.5	11342.25
41 to 50	45.5	7	2070.25	318.5	14491.75
Total		50		1205	37242.5

- To get the sample mean as follow = $\bar{x} = \sum \frac{fx}{n} = \frac{1205}{50} = 24.1$
- The sum of fx is 1205
- The square value of **∑ (fx)** is 1452025
- The sum of fx^2 is 37242.5.
- To get sample variance as follow $\mathbf{S^2} = \dfrac{\sum fx^2 - \frac{(\sum fx)^2}{n}}{n-1}$
- $\mathbf{S^2} = \dfrac{37242.5 - \frac{1452025}{50}}{50-1} = 167.38$, **then S.D =** $\sqrt{167.38}$ **= 12.93**

Pharmaceutical Statistics

Lecture 6 Part 2

Descriptive statistics

Measures of Position

Prepared and Presented by

Dr. Muna Oqal

Measures of Position

Definitions

- **Measures of position are used to describe the relative location of an observation.**

- **Quantiles system: are the positions of values in the data set. Data set are divided to quartiles, percentiles, percent (25% each), etc., after ordering them from smallest to the largest.**

Measures of Position

- The median and quartiles are specific examples of quantiles.
- Quantile systems that cut data into more than four ranges are really only useful where there are quite large numbers of observations. Such as quintiles, deciles and centiles (percentiles).
- There are four quintiles, which divide data into five ranges, nine deciles for ten ranges and 99 centiles that produce 100 ranges.
- The ninth decile is thus equivalent to the 90th centile and both indicate a point that ranks 10% from the top of a set of values.

Measures of Position

Quantile systems

Quantile systems divide ranked data sets into groups with equal numbers of observations in each group. Specifically:

- 3 *Quartiles* divide data into four equal-sized groups.
- 4 *Quintiles* divide it five ways.
- 9 *Deciles* divide it ten ways.
- 99 *Centiles* divide it 100 ways.

Measures of Position

A. Percentiles
B. Quartiles
C. Five Number Summary
D. Boxplot

Measures of Position

- **Percentiles or (centiles): are measures which divide the data set into 100 equal parts (P_1, P_2,…,P_{100})**
- **Percentiles: are measures indicating the value below a given percentage of observations in group of observation falls.**
- **For example, the 40th percentile (P_{40}) is the value (or score) below which 40% of the observations may be found.**

Measures of Position

Finding the Score Given a Percentile

$$L = \frac{K}{100} \bullet n$$

n total number of values in the data set

k percentile being used

L locator that gives the *position* of a value

P_k k th percentile

Then, you add 0.5 to find the exact position

Measures of Position

Example: we want to know the 75th percentile of the ordered grades from smallest to largest, n=24

58,59,68,69,74,75,78,78,78,79,79,81,83,84,85,
85,86,86,88,88,89,90,90,92

Solution:

Use

$$L = \frac{K}{100} \bullet n$$

Then L=75 / 100 *24= 18 (position) + 0.5 = 18.5 which corresponds to (86+88)/2 = 87

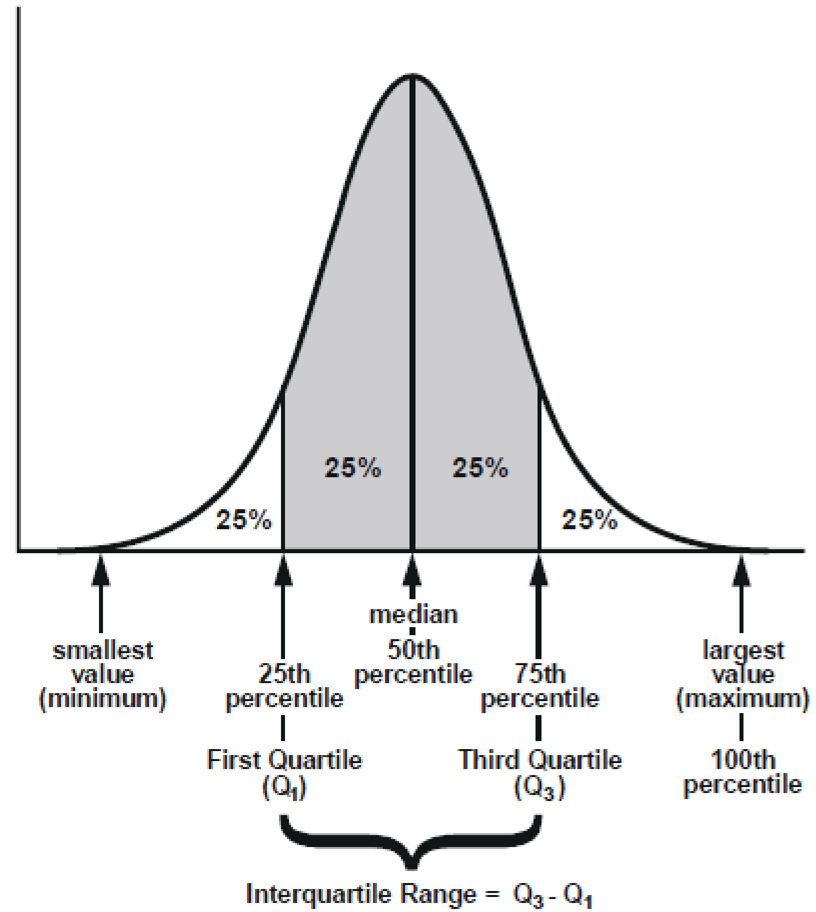
Measures of Position

- **The Quartiles**

The quartiles are position measures used in the educational and health-related fields to indicate the position of a individual in group.

Quartiles are the values of observations in a data set, when arranged in an ordered sequence, that can divided the data set into four equal parts, or quarters, using three quartiles namely Q_1, Q_2 and Q_3 each representing a quarter (fourth) of the population being sampled.

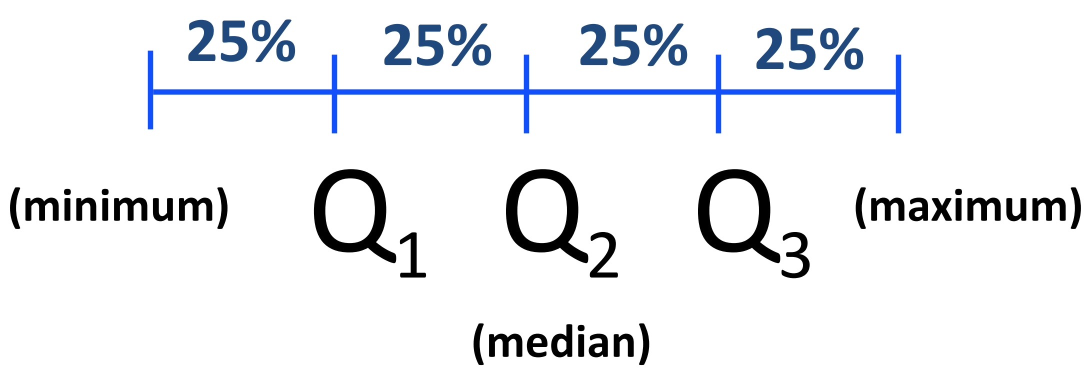
Figure 3.8
The middle half of the observations in a frequency distribution lie within the interquartile range



Quartiles

Q_1, Q_2, Q_3

divides ranked scores into four equal parts



$Q_1 = P_{25}$ $Q_2 = P_{50}$ $Q_3 = P_{75}$

Q=quartile, P= Percentile

Measures of Position

- **Definition**

(a) First (lower) Quartile (Q1)

The first (lower) quartile (Q1) is the median of the bottom half of the ordered observation for a data set (to the left of the median), or it is the median of the data set (lies at or below the median (MD)).

(b) Second Quartile (Q2)

The second quartile (Q2) is the median of the data set, that is, Q2= MD. It separates the lowest 50% of the data from highest 50%.

(c) Third (Upper) Quartile (Q3)

The third (upper) quartile (Q3) is the median of the top half of the ordered observations for a data set (to the right of the median), or it is the median of the data set (lies at or above the median (MD)).

Measures of Position

A. The Quartiles for Raw Data (ungrouped data)

To find the quartiles for a set of raw data, do the following:

1. Arrange the data set from the smallest to highest (ordered array).
2. Calculate the median (MD=Q2) for the data set.
3. For the half of the data set to the left of the MD calculate their median to get the first quartile (Q1).
4. For the half of the data set to the right of the MD, calculate their median to get the third quartile (Q3).

Measures of Position

- **Example (n is odd)**

The times in minutes needed for a random of 15 tablets to disintegrate are as follow:

(Q1) Median (Q2) (Q3)

5 10 10 10 10 12 15 20 20 25 30 30 40 40 60

Find the quartiles Q1, Q2 and Q3?

1. Data is already in an order from smallest to highest
2. The **MD** is $X_{((n+1)/2)}$ =$X_{(8)}$= **20** minutes.
3. For the half to the left of the median, n=8, then $X_{((n+1)/2)}$ =$X_{(4.5)}$ = **Q1**=(10+10)/2=**10** minutes.
4. The second quartiles is **Q2**=**MD**=**20** minutes.
5. For the half to the right with the MD, n=8, then then $X_{((n+1)/2)}$ =$X_{(4.5)}$ = the third quartile which is **Q3**=(30+30)/2=**30** minutes.
6. The quartiles are: **Q1=10 minutes**, **Q2=20 minutes**, **Q3=30 minutes**.

When n is odd, you can include the median in both quartiles **(Q_1 and Q_3)**

Measures of Position

Since $Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$, so we can find Q1, Q2, and Q3 as following:

Q1=25%=25/100*15=3.75 + 0.5= 4.25th (**position**),

which corresponds to **10** ((10+10)/2).

Q2=50/100*15=7.5 +0.5= 8th (**position**),

which corresponds to **20**.

Q3=75/100*15=11.25+0.5=11.75th (**position**),

which corresponds to **30** ((30+30)/2).

5	10	10	10	10	12	15	20	20	25	30	30	40	40	60

Measures of Position

Conclusion

- This means that 25% of the tablets need less than 10 minutes to disintegrate.
- 50% of the tablets need 20 minutes to disintegrate.
- Before 30 minutes 75% of all tables were disintegrated.
- 25% only of these tablets need more than 30 minutes to disintegrate.

Disintegration time (min)	Frequency
5	1
10	4
12	1
15	1
20	1
25	2
30	2
40	2
60	1
Total	15

Measures of Position

- **Example (n is even)**

The time in minutes needed a random sample of 20 capsule to disintegrate are as follow:

(Q1) **Median (Q2)** **(Q3)**

5 10 10 15 15 15 15 20 20 20 **25 30 30 40 40 45 60 60 65 85**

Find the quartiles Q1, Q2, and Q3?

1. Data is already in an order from smallest to highest
2. The **MD** is $X_{((n+1)/2)} = X_{(10.5)} = (20+25)/2 =$ **22.5** minutes.
3. For the half to the left of the median, n=10, then the first quartiles is **Q1**=(15+15)/2=**15** minutes
4. The second quartiles is Q2=MD=22.5 minutes.
5. For the half to the right with the MD, n=10, then the third quartile is **Q3**=(40+45)/2=**42.5** minutes.
6. The quartiles are: **Q1=15 minutes**, **Q2=22.5 minutes**, **Q3=42.5 minutes**.

Measures of Position

Since $Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$, so we can find Q1, Q2, and Q3 as following:

Q1=25%=25/100*20 = 5 + 0.5 = 5.5th (**position**),

which corresponds to **((15+15)/2)= 15**.

Q2=50/100*20= 10 +0.5 = 10.5th (**position**),

which corresponds to (**(20+25)/2)= 22.5**.

Q3=75/100*20= 15+0.5 = 15.5th (**position**),

which corresponds to **((40+45)/2)= 42.5**.

(Q1) **Median (Q2)** **(Q3)**

5 10 10 15 15 15 15 20 20 20 25 30 30 40 40 45 60 60 65 85

Measures of Position

Conclusion

- This means that 25% of the capsules need less than 15 minutes to disintegrate.
- 50% of the capsules need 22.5 minutes to disintegrate.
- Before 42.5 minutes 75% of all capsules were disintegrated.
- 25% only of these capsules need more than 42.5 minutes to disintegrate.

Disintegration Time (minutes)	Frequency
5	1
10	2
15	4
20	3
25	1
30	2
40	2
45	1
60	2
65	1
85	1
Total	20

Measures of Position

B. The quartiles for simple (ungrouped) frequency table

The quartiles for simple frequency table can be calculated as follows:

1. Find the cumulative frequency (cf).
2. Find the values of (n/4), (n/2), (3n/4), where n is the total number of observation (the sum of frequencies).
3. The first quartile (Q1) is the first value having cumulative frequency (cf) greater than or equal to (n/4).
4. The second quartile (Q2) is the first value having cumulative frequency (cf) greater than or equal to (n/2), that is the median (MD).
5. The third quartile (Q3), is the first value having cumulative frequency (cf) greater than or equal to (3n/4).

Measures of Position

- **Example**

The Gulf Pharmaceutical Industries Company in UAE planning to improve safety plan in its factory. For this, accident data for the last 50 weeks was compiled. These data were listed into simple frequency table as shown below:

Number of Accidents (x)	**2**	**7**	**12**	**17**	**22**
Number of weeks (f)	**5**	**19**	**13**	**8**	**5**

Find the quartiles Q1, Q2 and Q3 for number of accidents per week?

Measures of Position

- **Solution:**

Step-1: Calculate the value of cumulative frequency (cf) as follows:

	Number of Accidents (x)	Frequency (f)	cf
	2	5	5
Q1	7	19	24
Q2	12	13	37
Q3	17	8	45
	22	5	50
	Total	50	

Measures of Position

- **Step-2:** Find the following values:
- **n/4= 50/4 = 12.5**
- **n/2= 50/2 = 25**
- **3n/4 = (3)(50)/4=37.5**
- **Step-3:** The first quartile (Q1) is the first value having cumulative frequency (cf) greater than or equal to 12.5, then from the table the value of the first quartile (Q1) is **Q1=7** accident per week.
- **Step-4:** the second quartile (median) (Q2=MD) is the first value having cumulative frequency (cf) greater than or equal to 25, then from the table, the value of the second quartile (Q2) is **Q2=12** accident per week.
- **Step-5:** The third quartile (Q3) is the first value having cumulative frequency (fc) greater than or equal to 37.5, then from the table, the value of the first quartile (Q3) is **Q3=17** accident per weeks.

Measures of Position

- **Conclusion:**

The results mean that:

- ❑ 25% of weeks have less than 7 accidents and 75% of weeks have more than 7 accidents.
- ❑ 50% of weeks have less than 12 accidents and 50% of weeks have more than 12 accidents.
- ❑ 75% of weeks have less than 17 accidents and 25% of weeks have more than 17 accidents.

Measures of Position

C. The quartiles for grouped frequency table

The quartiles for grouped frequency table (frequency distribution) can be calculated as follows:

For the first quartile (Q1)

Step-1: Construct the cumulative frequency distribution.

Step-2: Determine the first quartile class interval (**First Quartile Class**),

that is, the first class having cumulative frequency (cf) greater than or equal to (n/4).

Step-3: Find the first quartile (Q1) by using the following formula:

$$Q_1 = L_1 + \left(\frac{\left(\frac{n}{4}\right) - cf_1}{f_1} \right) * i$$

Where:

n= the total number of frequencies.

cf_1= cumulative frequency prior to the first quartile class interval.

i = the class interval width.

L_1 = the lower boundary (limit) of the first quartile class interval.

f_1= the frequency of the first quartile class interval.

Measures of Position

C. The quartiles for grouped frequency table

For the second quartile (Median) (Q2=MD)

Step-1: Construct the cumulative frequency distribution.

Step-2: Determine the second quartile class interval (**Second Quartile Class**),

that is, the first class having cumulative frequency (cf) greater than or equal to (n/2).

Step-3: Find the second quartile (Q2) by using the following formula:

$$Q_2 = L_2 + \left(\frac{\left(\frac{n}{2}\right) - cf_2}{f_2} \right) * i$$

Where:

n= the total number of frequencies.

cf_2= cumulative frequency prior to the second quartile class interval.

i = the class interval width.

L_2 = the lower boundary (limit) of the second quartile class interval.

f_2= the frequency of the second quartile class interval.

Measures of Position

C. The quartiles for grouped frequency table

For the third quartile (Q3)

Step-1: Construct the cumulative frequency distribution.

Step-2: Determine the third quartile class interval (**Third Quartile Class**),

that is, the first class having cumulative frequency (cf) greater than or equal to (3n/4).

Step-3: Find the third quartile (Q2) by using the following formula:

$$Q_3 = L_3 + \left(\frac{\left(\frac{3n}{4}\right) - cf_3}{f_3} \right) * i$$

Where:

n= the total number of frequencies.

cf_3= cumulative frequency prior to the third quartile class interval.

i = the class interval width.

L_3 = the lower boundary (limit) of the third quartile class interval.

f_3= the frequency of the third quartile class interval.

Measures of Position

- **Example:**

The following frequency table represents the daily sales volume in JD for a period of 30 days selected from the sales records of given pharmacy in Jordan:

Class (sales volume in JD)	53 - 56	57 - 60	61 - 64	65 - 68	69 - 72
Number of Days (f)	3	5	9	7	6

Find the quartiles Q_1, Q_2, and Q_3 for the sales volume in JD?

Measures of Position

Solution

Step-1: calculate the value of cumulative frequency (cf) as follows:

Class Interval (Daily Sales Voume in JD)	f_i	cf_i	
53 - 56	3	3	
57 - 60	5	8	Q1 Class
61 - 64	9	17	Q2 Class
65 - 68	7	24	Q3 Class
69 - 72	6	30	
Total	30		

Step-2: Calculate the interval width (i)

i=57-53=4 so i=4

Measures of Position

- **Step-3: calculate the value of the Q1 a follows:**

n/4= 30/4 =7.5, then (57-60) is the first quartile class, then

$$Q_1 = L_1 + \left(\frac{\left(\frac{n}{4}\right) - cf_1}{f_1}\right) * i = 56.5 + \left(\frac{7.5-3}{5}\right) * 4 = 60.1 \text{ JD.}$$

- **Step-4: calculate the value of the Q2 a follows:**

n/2= 30/2 = 15, then (61-64) is the second quartile class (median class), then

$$Q_2 = L_2 + \left(\frac{\left(\frac{n}{2}\right) - cf_2}{f_2}\right) * i = 60.5 + \left(\frac{15-8}{9}\right) * 4 = 63.61 \text{ JD.}$$

- **Step-5: calculate the value of the Q3 a follows:**

3n/4=(3*30)/4= 22.5, then (65-68) is the third quartile class, then

$$Q_3 = L_3 + \left(\frac{\left(\frac{3n}{4}\right) - cf_3}{f_3}\right) * i = 64.5 + \left(\frac{22.5-17}{7}\right) * 4 = 67.64 \text{ JD.}$$

Measures of Position

Conclusion

For this pharmacy, the results mean that:

- 25% of days have sales volume less than 61.1 JD and 75% of days have sales volume more than 61.1 JD.

- 50% of days have sales volume less than 63.61 and 50% of days have sales volume more than 63.61 JD.

- 75% of days have sales volume less than 67.64 JD and 25% of days have sales volume more than 67.64 JD.

Measures of Position

Quartiles

$Q_1 = P_{25}$

$Q_2 = P_{50}$

$Q_3 = P_{75}$

P= Percentile

Interquartile Range (or IQR): $Q_3 - Q_1$

Semi-interquartile Range: $\frac{Q_3 - Q_1}{2}$

Midquartile Range: $\frac{Q_1 + Q_3}{2}$

10 - 90 Percentile Range: $P_{90} - P_{10}$

Measures of Position

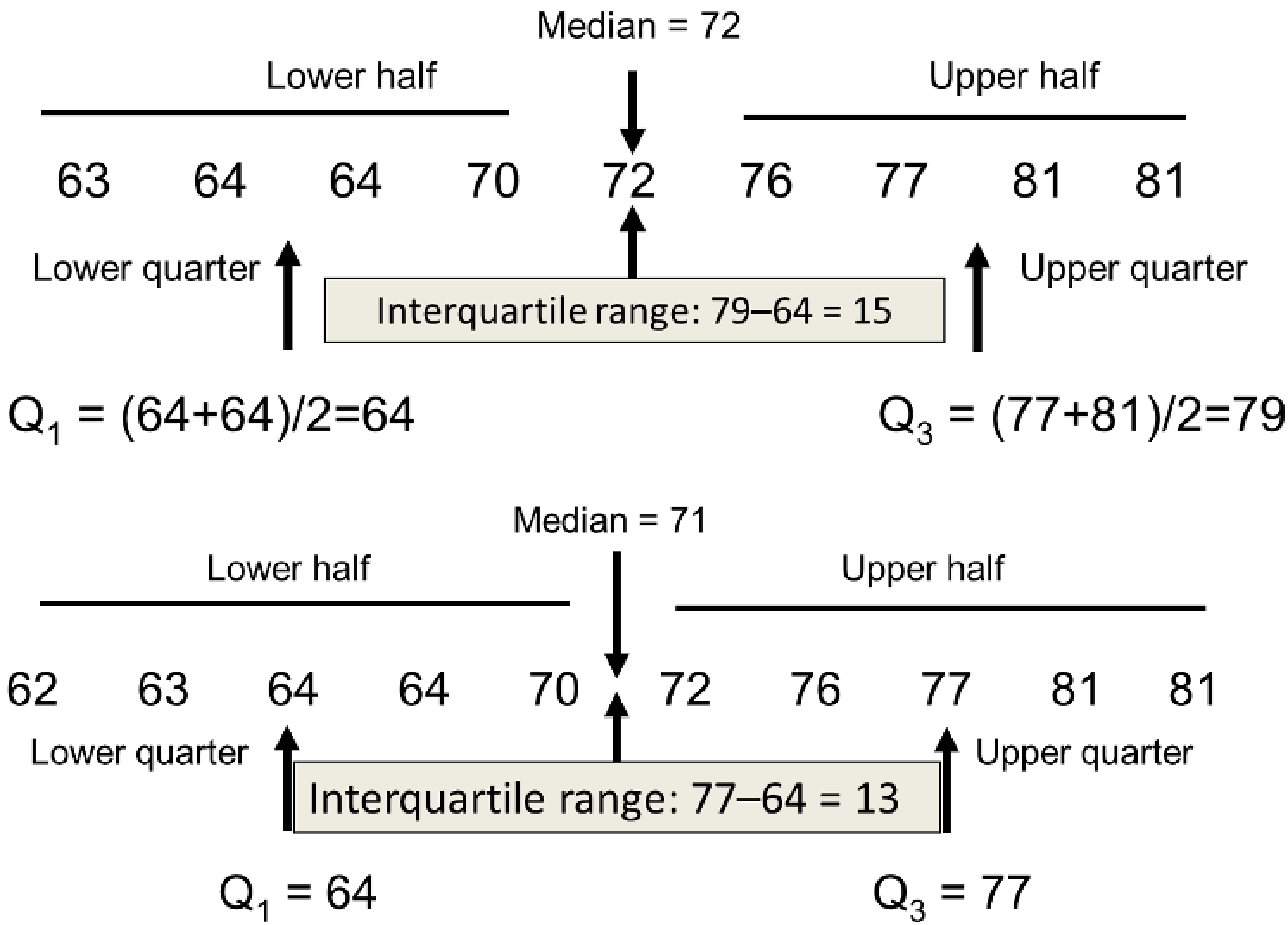
The Interquartile Range (IQR)

- The interquartile range (IQR) is a robust measure of variation that is based on the quartiles.

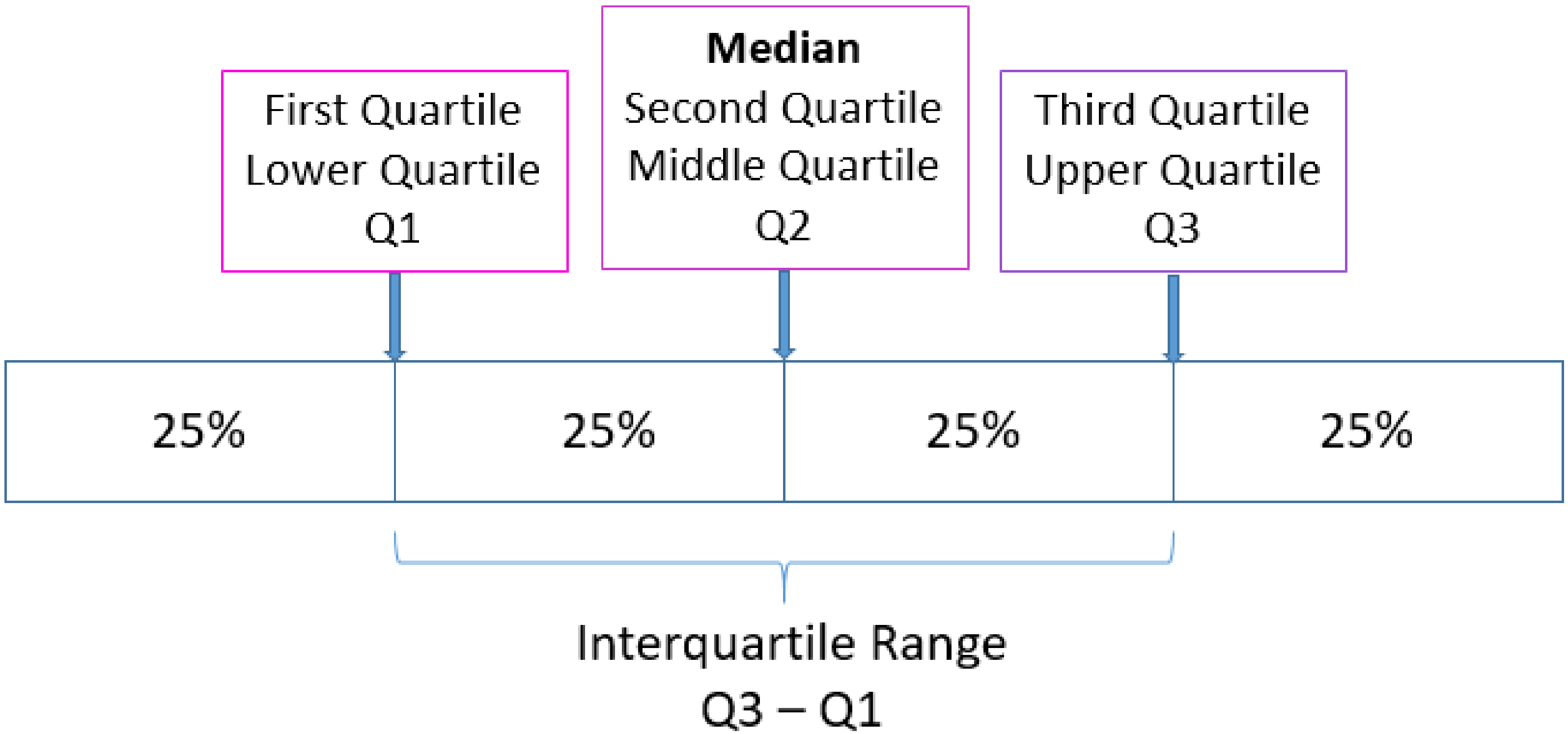
- The IQR is defined as the range of middle 50% of observations in the data set.

- It is the difference between the third quartile (Q3) and the first quartile (Q1), and it is found by using following formula:

$$IQR = Q3 - Q1$$


Median = 72
Lower half
Upper half
63 64 64 70 72 76 77 81 81
Lower quarter
Upper quarter
Interquartile range: 79–64 = 15
Q_1 = (64+64)/2=64
Q_3 = (77+81)/2=79
Median = 71
Lower half
Upper half
62 63 64 64 70 72 76 77 81 81
Lower quarter
Upper quarter
Interquartile range: 77–64 = 13
Q_1 = 64
Q_3 = 77

Median and Quartiles



Median and inter-quartile range are robust indicators of central tendency and dispersion

The median (second quartile) and inter-quartile range can be used as an alternative method for describing the central tendency and dispersion of a set of measured data. Both are robust and can be useful where there are occasional extreme values.

Measures of Position

- Example: we want to know the **IQR**, **Semi-interquartile Range**, **and the** **Midquartile** of the ordered grades from smallest to largest, **n=24**

58,59,68,69,74,75,78,78,78,79,79,81,83,84,85, 85,86,86,88,88,89,90,90,92

Measures of Position

Example: we want to know the Q1, Q2, Q3, IQR, Semi-interquartile Range , and the Midquartile, and the 10-90 percentile range of the ordered grades from smallest to largest, n=24

58,59,68,69,74,75,78,78,78,79,79,81,83,84,85,85,86,86,88,88,89,90,90,92

Q1=25%=25/100*24= 6 + 0.5 = 6.5 (position) which corresponds to 76.5.

Q2=50 / 100 *24= 12 +0.5 (position) which corresponds to 82.

Q3=75 / 100 *24= 18+0.5 (position) which corresponds to 87.

IQR= Q3–Q1= 87-76.5=10.5

Semi-interquartile Range= 10.5/2=5.25

Midquartile= Q1+Q3/2 = 81.75

10-90 percentile range=P90-P10.

P90= 0.9*24= 21.6+0.5=22=90, P10=0.1*24=2.4+0.5=2.9=68

10-90 percentile range=P90-P10= 90-68= 22