


MEASURES OF LOCATION— PERCENTILES and QUARTILES 2019-2020

Prof Dr Najlaa Fawzi



Another approach that addresses some of the shortcomings of the range is in quantifying the spread in the data set is the use of quartiles or percentiles.

Percentiles

They are values of a variable such as height, weight, age, etc. which divide the total observations by an imaginary line into two parts, expressed in percentages such as 10% and 90% or 25% and 75%, etc.

In all, there are 99 percentiles.

Centiles or percentiles are values in a series of observations arranged in ascending order of magnitude which divide the distribution into 100 equal parts.

Percentiles

A percentile is a measure of relative standing, meaning we get information about the position of a value relative to the rest of the data set.

Thus, the median is 50th centile. The 50th percentile will have 50% observations on either side.

Accordingly, 10th percentile should have 10% observations to the left and 90% to the right.

The 10th percentile , is designated $P 10$. The 50th percentile is the median , $P 50$.

PERCENTILE

- A value below which a certain percentage of observations lie.
- Data set:
 - 2,2,3,4,5,5,5,6,7,8,8,8,8,8,9,9,10,11,11,12
 - What is the **percentile** ranking of “10”?
 - Percentile rank of $x = \frac{\text{\# of values below } x}{n} * 100$

$$\text{Percentile rank of '10'} = \frac{16}{20} * 100 = 80\%$$

An Example of a Percentile

A class of 20 students had the following scores on their most recent test: 75, 77, 78, 78, 80, 81, 81, 82, 83, 84, 84, 84, 85, 87, 87, 88, 88, 88, 89, 90.

The score of 80 has four scores below it. Since $4/20 = 20\%$, 80 is the 20th percentile of the class.

The score of 90 has 19 scores below it. Since $19/20 = 95\%$, 90 corresponds to the 95th percentile of the class.

- Data set:
- 2,2,3,4,5,5,5,6,7,8,8,8,8,8,9,9,10,11,11,12
- What Value Exist at the **percentile** ranking of 25%?

$$\text{Value \#} = \frac{\text{percentile}}{100} (n+1)$$

$$\text{Value \#} = \frac{25}{100} (20 + 1) = 5.25 \quad \boxed{\text{Position}}$$

There is no "5.25th", so I take the average of the 5th & 6th values to find what value exist at the 25th percentile.

$$\frac{5+5}{2} = 5$$

Applications of Percentiles

Percentile scores have a variety of uses. Anytime that a set of data needs to be broken into portions, percentiles are helpful.

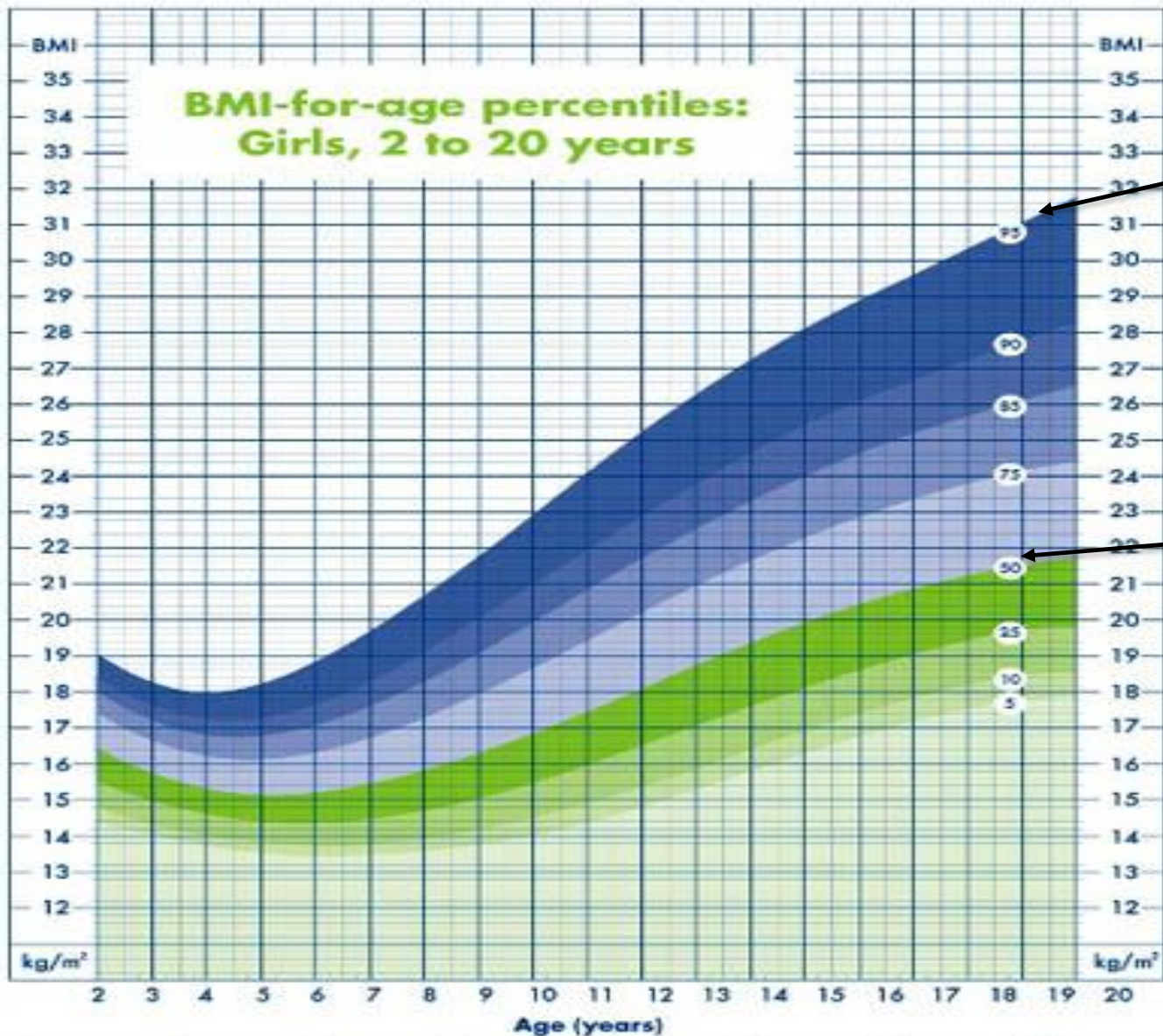
One common application of percentiles is being used in children's growth charts.

In addition to a physical height or weight measurement, pediatricians typically state this in terms of a percentile score.

A percentile is used in this situation in order to compare the height or weight of a given child to all children of that age.

This allows for an effective means of comparison.

BMI-for-age percentiles: Girls, 2 to 20 years



50

Weight status category	Percentile range
Underweight	Less than 5th percentile
Healthy weight	5th - 85th percentile
At risk of overweight	85th - 95th percentile
Overweight	95th percentile and greater

5th percentile 10th percentile 25th percentile 50th percentile 75th percentile 85th percentile 90th percentile 95th percentile

Published May 30, 2000 (revised 10/16/00).
SOURCE: Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).
<http://www.cdc.gov/growthcharts>

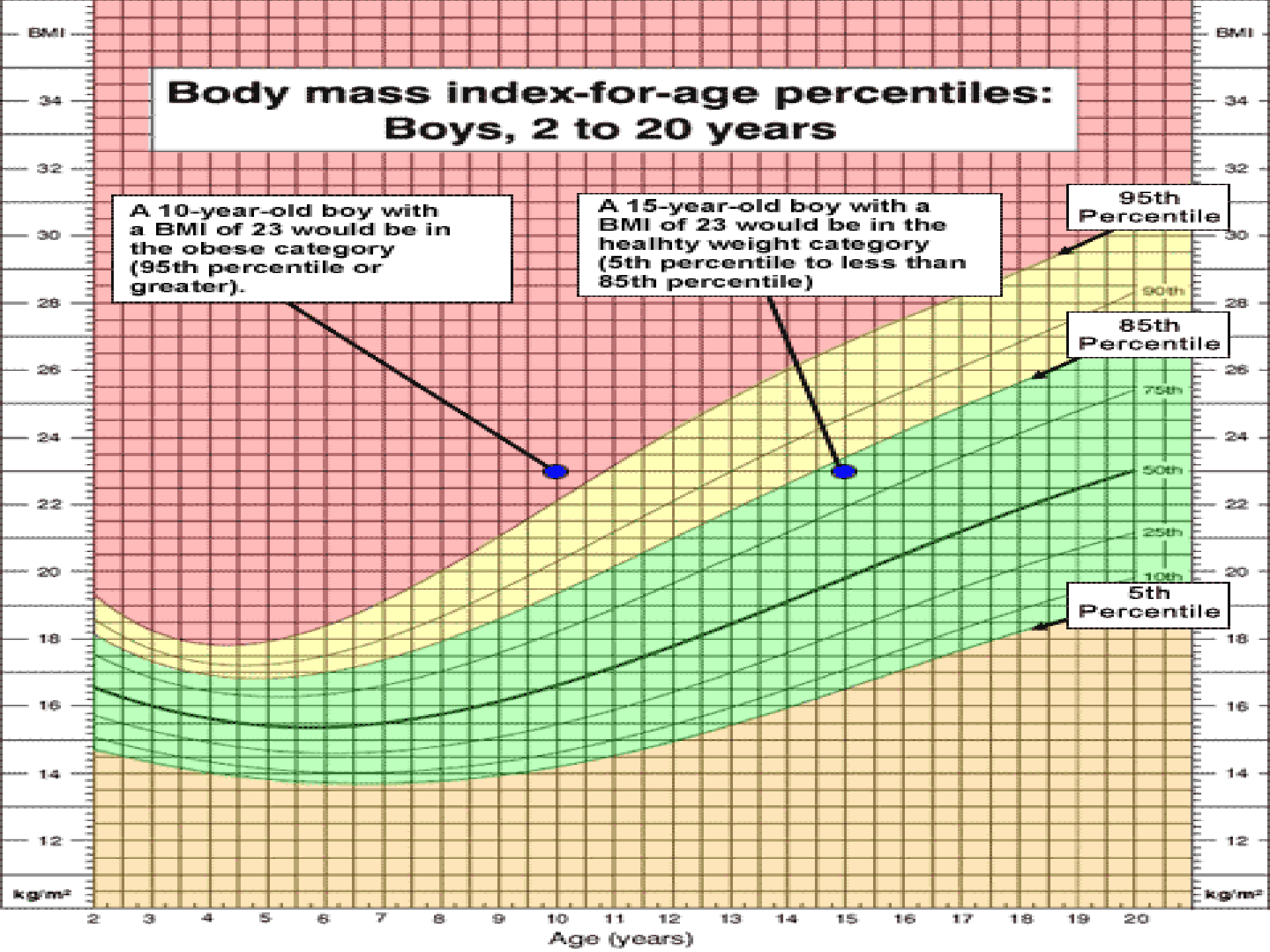


SAFER • HEALTHIER • PEOPLE™

Body mass index-for-age percentiles: Boys, 2 to 20 years

A 10-year-old boy with a BMI of 23 would be in the obese category (95th percentile or greater).

A 15-year-old boy with a BMI of 23 would be in the healthy weight category (5th percentile to less than 85th percentile).





Percentile vs. Percentage

Be careful with the word's percentile and percentage.

A percentage score indicates the proportion of a test that someone has completed correctly.

A percentile score tells us what percent of other scores are less than the data point we are investigating.

Example: Compute the 10th and 90th percentile for the birth weight data.

2069, 2581, 2759, 2834, 2838, 2841, 3031, 3101,
3200, 3245, 3248, 3260, 3265, 3314, 3323, 3484,
3541, 3609, 3649, 4146

Solution: Since 10th $10/100(20+1)=\underline{2.1}$
 $90/100(20+1)=\underline{18.9}$

10th percentile = the average of the 2nd and 3rd largest values = $(2581+2759)/2 = 2670$ g

90th percentile = the average of the 18th and 19th values = $(3609+3649)/2 = 3629$ grams.

Quartiles:

They are 3 different points located on the entire range of a variable such as height—Q1, Q2 and Q3.

Q1 or lower quartile will have 25% observations of heights falling on its left and 75% on its right.

Q2 or median will have 50% observations on either side


Q3 or upper will have 75% observations on its left and 25% on its right.

The 25th percentile is often referred to as


the **first quartile** and denoted **Q1** .


The 50th percentile is referred to as the second or **middle quartile** and written **Q2**.

The 75th percentile is referred to as the **third quartile** , **Q3**.




The first quartile is the point which gives us 25% of the area to the left of it and 75% to the right of it. This means that 25% of the observations are less than or equal to the first quartile and 75% of the observations greater than or equal to the first quartile. The first quartile is also called the 25th percentile.

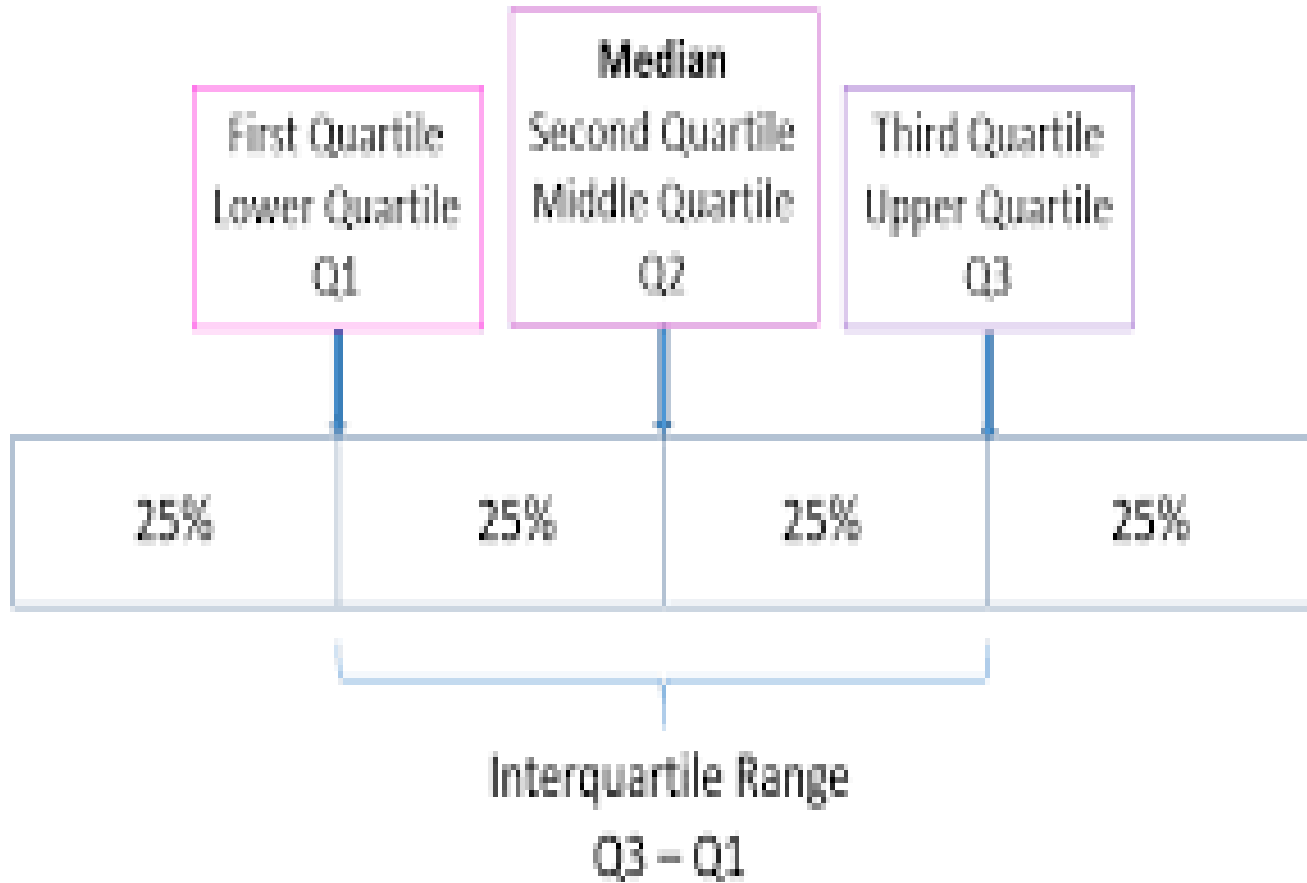




The third quartile is the point which gives us 75% of the area to the left of it and 25% of the area to the right of it. This means that 75% of the observations are less than or equal to the third quartile and 25% of the observation are greater than or equal to the third quartile. The third quartile is also called the 75th percentile.



Median and Quartiles



median(50th percentile)

lower
quartile(6th #)

upper 50%

1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 5, 6

Lower 50%
(11 numbers)

upper
quartile

**When we wish to find the quartiles for a set of data ,
the following formulas are used:**

$Q_1 = (n+1) / 4$ **th ordered observation**

**$Q_2 = 2(n+1) / 4 = (n+1) / 2$ **th ordered
observation****

$Q_3 = 3(n+1) / 4$ **th ordered observation**

The weight measurements (lb) of 20 children :

**14.6 24.3 24.9 27.0 (27.2 27.4) 28.2 28.8 29.9 (30.7
31.5) 31.6 32.3 32.8 (33.3 33.6) 34.3 36.9 38.3 44.0**

**The first quartile is the $Q_1 = (20+1)/4 = 5.25$ th measurements
which is $27.2 + (0.25) (27.4 - 27.2) = 27.25$**

**The median is the $Q_2 = (20+1)/2 = 10.5$ th measurement
 $30.7 + (0.5) (31.5 - 30.7) = 31.1$**

**Third quartile $Q_3 = 3(20+1)/4 = 15.75$ th measurement
 $33.3 + (0.75) (33.6 - 33.3) = 33.525$**

IQR = $33.525 - 27.25 = 6.275$

The R = 29.4

$100(6.275 / 29.4) = 21\%$ of the range

1, 11, 15, 19, 20, 24, 28, 34, 37, 47, 50, 57



Q_1



Lower quartile



17



Q_2



Median



26



Q_3



Upper quartile



42

IQR (Interquartile range)

The interquartile range is another useful measure of dispersion or spread. It is used when the median is used as the measure of central tendency. It gives the range in which the middle 50% of the distribution lies.

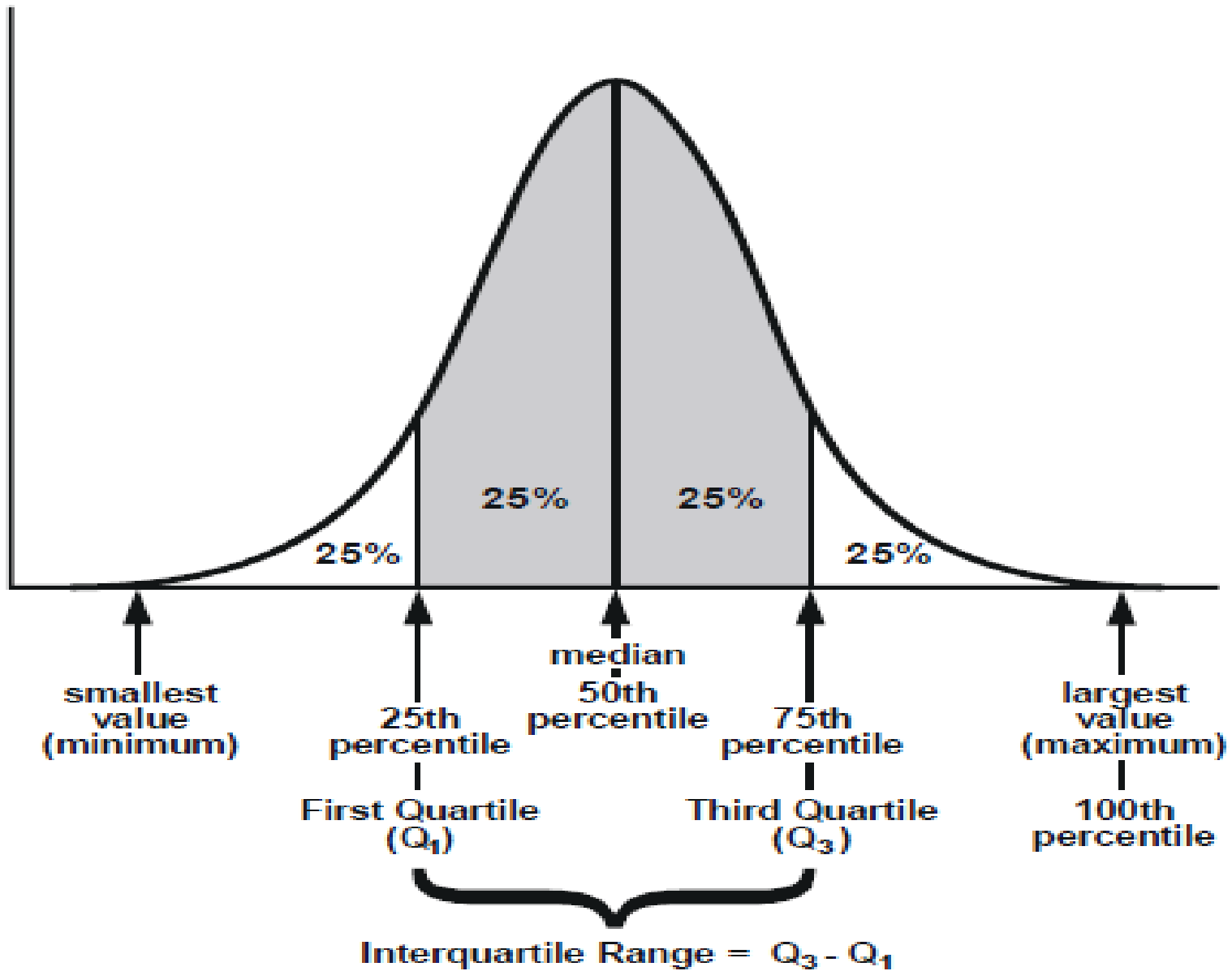
This summary statistic, like the median, is based on the ordered (ranked) scores.

The interquartile range gives the lower and upper values of where 50% of the ranked values are of which 25% are on either side of the median.

You will remember that the median represents the point where 50% of the ranked scores lie which is also called the 50th percentile so the range between 25% and 75% of the ranked scores lie is called the interquartile range.

This results in the sorted (ranked) data being divided up into what are known as four equal quartiles

Inter-quartile range is the middle (50%) of all observations i.e. observations limited between 25% & 75% of observations).



Interquartile Range: A measure that reflects the variability among the middle 50 percent of the observations in a data set .

IQR is the difference between the third and first quartile : that is

$$\text{IQR} = Q3 - Q1$$

A large IQR indicates a large amount of variability among the middle 50 percent of the relevant observations, and a small IQR indicates a small amount of variability among the relevant observations . It is more informative to compare the IQR with the range of the entire data set. A comparison may be made by forming the ratio of the IQR to the range (R) and multiply by 100. That is , $100(\text{IQR}/ R)$ tell as what percent the IQR is of the overall range .

Using Different Measures of Dispersion:

The following guidelines are useful in deciding which measure of dispersion is most appropriate for a given set of data.

1-The standard deviation is used when the mean is used (i.e., with symmetric numerical data).

2-Percentiles and the interquartile range are used in two situations:

a. When the median is used (i.e., with ordinal data or with skewed numerical data).

b. When the mean is used but the objective is to compare individual observations with a set of norms.

3-The interquartile range is used to describe the central 50% of a distribution, regardless of its shape.

4-The range is used with numerical data when the purpose is to emphasize extreme values.

5-The coefficient of variation is used when the intent is to compare distributions measured on different scales.

Choosing Appropriate Descriptive Statistics

Important in the selection of an appropriate descriptive measure is the level of measurement achieved by the data.

A- Description of nominal data:

1- Statistical methods : numerical operations are not performed on nominal scale variables even if the numerical codes are used to define categories . The appropriate descriptive statistics are the frequency of observation in the various categories (frequency table and the mode).

Example : A group of 100 patients with mental diseases are divided into groups on the basis of a psychiatric diagnosis [a nominal variable].

Distribution of psychiatric diagnosis in 100 patients

Diagnosis	frequency	Rf %
Organic brain syndrome	20	20
Psychosis	35	35
Mental retardation	20	20
Personality disorders	25	25
Total	100	100%

B- Description of ordinal data : The most appropriate statistics are frequency count ,median ,mode .

Example : A group of 20 geriatric patients is asked to rate the quality of **nursing home care on the following scale : 1= very poor ; 2= poor ; 3= fair ; 4= good ; 5= very good .**

Rating	frequency	RF%
Very poor	3	15
Poor	3	15
Fair	4	20
Good	7	35
V. good	3	15
Total	20	100%

The median response is a rating of 3.5 [usually with data of this type it is more meaningful to specify that the median response is between " fair " and "good" rather than to define it numerically].

The model response for the data is the rating " good".