Pharmaceutical statistics

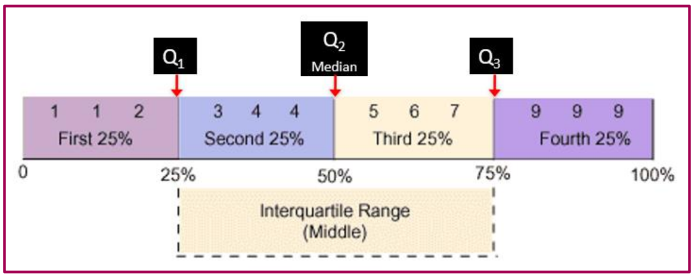
2025-2024

Dr. Rola Aqeel

ARKAN
ACADEMY

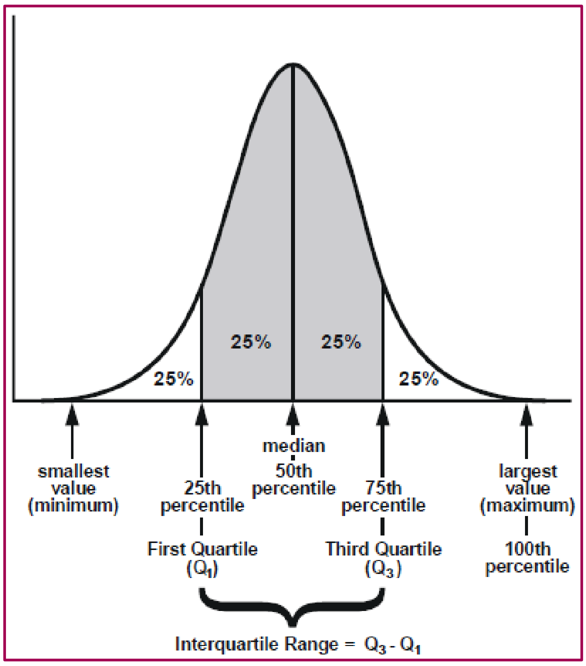
Quartiles

- Quartiles divide a data set into four equal parts. Each quartile represents 25% of the data.



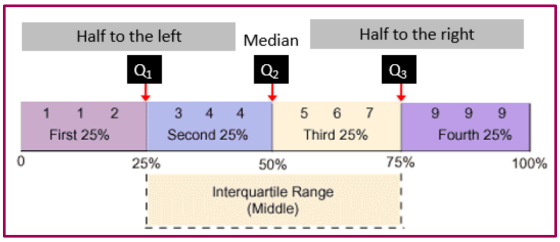
First Quartile	Q_1	cuts off lowest 25% of data	25th percentile
Second Quartile	Q_2	cuts data set in half	50th percentile
Third Quartile	Q_3	cuts off highest 25% of data, or lowest 75%	75th percentile

- ***There are three main quartiles:***
  - ***Q1 (First Quartile):*** This is the 25th percentile, meaning that 25% of the data falls below Q1. It's also the median of the lower half of the data.
  - ***Q2 (Second Quartile):*** This is the 50th percentile, which is also the median of the entire data set. 50% of the data falls below Q2.
  - ***Q3 (Third Quartile):*** This is the 75th percentile, meaning that 75% of the data falls below Q3. It's the median of the upper half of the data.



- ***Interquartile Range (IQR):*** The range between Q1 and Q3 (i.e., Q3 − Q1). It shows the spread of the middle 50% of the data and is useful in identifying outliers and it's more robust for extreme values because it gives the range of the middle values.

- ***Finding Quartiles***
  - To find the quartiles for a set of data, do the following:
    - ✓ Arrange the data from smallest to highest (ordered array).
    - ✓ Locate the median (Q2).
    - ✓ The half to the left: locate their median (Q1).
    - ✓ The half to the right: Locate their median (Q3).



★ Example with ***odd (n)***:

- ✓ Times needed for 15 tablets to disintegrate in minutes:
  5, 10, 10, **10,** 10, 12, 15, **20,** 20, 25, 30, **30,** 40, 40, 60
- ✓ Data is already in an order from smallest to highest.
- ✓ *Median* is the $(n+1/2)^{th} = 8^{th} =$ **20**
- ✓ For the half to the left: n=7, median = 4^{th} =**10** minutes
- ✓ For the half to the right: n=7, median = 4^{th} =**30** minutes
  - Q_1=10 minutes; Q_2= 20 minutes; Q_3=30 minutes.
  - IQR=Q_3-Q_1=20 minutes
- ✓ This means that 25% of tablets need less than 10 minutes to disintegrate. Also 50% of tablets need 20 minutes to disintegrate. Before 30 minutes, 75% of all

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

tables were disintegrated. 25% only of these tablets need more than 30 minutes to disintegrate.

★ Example with ***even (n):***

- ✓ Times needed for 20 capsules to disintegrate in minutes:
  5, 10, 10, 15, **15, 15**, 15, 20, 20, **20, 25**, 30, 30 40, **40, 45**, 60, 60, 65, 85

Disintegration time (min)	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

- ✓ Data is already in an order from smallest to highest
- ✓ Median is the mean of the two middle values $(n/2)^{th}$ and $((n/2) + 1)^{th}$
  Median $=10^{th}$ and 11^{th} =(20+25)/2= **22.5**
- ✓ For the half to the left: n=10, median=mean of 5^{th} & 6^{th} =**15** minutes
- ✓ For the half to the right: n=10, median=mean of 5^{th} & 6^{th} =**42.5**
  - Q_1=15minutes; Q_2= 22.5 minutes; Q_3=42.5 minutes.
  - IRQ = 42.5 – 15 = 27.5 minutes.

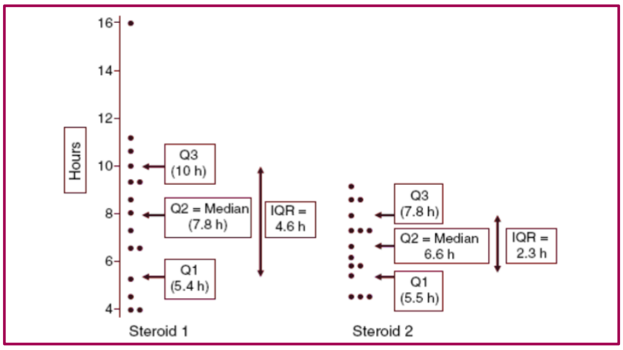
★ Example:

- ✓ Consider the elimination half-lives of two synthetic steroids have been determined using two groups, each containing 15 volunteers.
- ✓ The results are shown in the following table with the values ranked from lowest to highest for each steroid.

Table 2.5 Ranked half-lives for two steroids

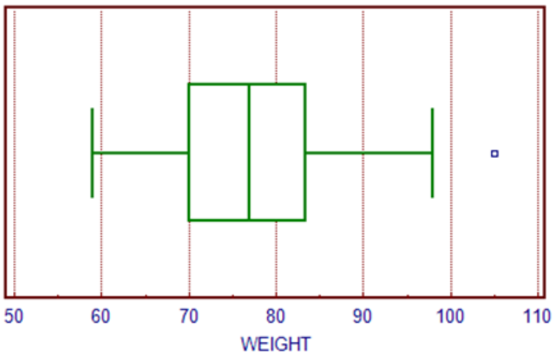
Steroid 1 Rank	Steroid 1 Half-life (h)	Steroid 2 Rank	Steroid 2 Half-life (h)	
1	3.9	1	4.4	
2	4.0	2	4.5	
4	**5.4**	4	**5.5**	4th ranked values = Q1
5	6.4	5	5.8	
6	6.5	6	5.9	
7	7.2	7	6.1	
8	**7.8**	8	**6.6**	8th ranked values = Q2 (Median)
9	8.6	9	7.2	
10	9.2	10	7.2	
11	9.3	11	7.3	
12	**10.0**	12	**7.8**	12th ranked values = Q3
13	10.6	13	8.5	
14	11.1	14	8.6	
15	15.8	15	9.1	

- ✓ The *median* (second quartile) indicates generally longer elimination half-lives for steroid 1 relative to steroid 2.



- ✓ The ***IQR*** for the half-life of steroid 2 is only half that for steroid 1, duly reflecting its less variable nature.
- ✓ The interquartile range indicates greater variability for the first steroid
- ✓ Just as the median is a robust indicator of central tendency, ***the interquartile range is a robust indicator of dispersion***.
- ✓ The interquartile range is a more useful measure of spread than range as it describes the middle 50% of the data values and thus *less affected by outliers*

- ***Box and whisker plot***
  - ➢ A box-and-whisker plot can be useful for handling many data values.
  - ➢ It shows only certain statistics rather than all the data.
  - ➢ ***Five-number summary*** is another name for the visual representations of the box-and-whisker plot.
  - ➢ The five-number summary consists of the *median*, the *quartiles*, and the *smallest* and *greatest values* in the distribution (***not including outliers***).
  - ➢ Immediate visuals of a box-and-whisker plot are the center, the spread, and the overall range of distribution.

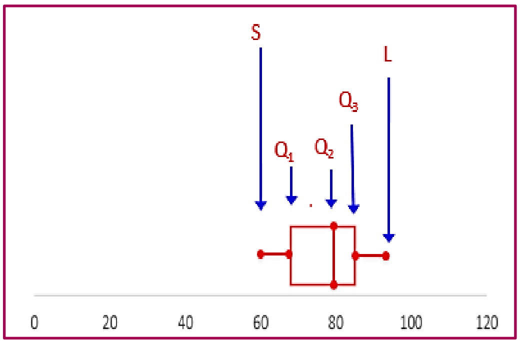


  - ➢ The ***first step*** in constructing a box-and-whisker plot is to first find the median (***Q_2***), the lower quartile (***Q_1***) and the upper quartile (***Q_3***) of a given set of data.

★ Example:

✓ The following set of numbers are weights of 10 patients in hospital (kgs):

75	1	62	Smallest (S)
62	2	67	
78	3	73	Q_1
96	4	75	
73	5	78	Median=78.5 (Q_2)
93	6	79	
85	7	81	
81	8	85	Q_3
67	9	93	
79	10	96	Largest (L)



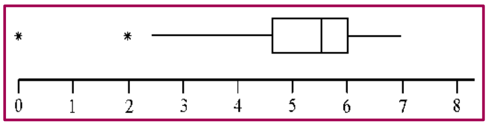
➢ *Outlier calculations*

✓ Outliers (extreme values) are values that are much bigger or smaller (distant) than the rest of the data.

✓ In order to be an outlier, the data value must be:

- Larger than Q3 by at least ***1.5*** times the interquartile range (IQR), ***or***
- Smaller than Q1 by at least ***1.5*** times the IQR.

✓ *Represented by a dot* on the box and whisker plot.



★ Example:

✓ Data set: 1, 55, 60, 70, 80, 85, 135

✓ Q1= 55, Q2= 70, Q3= 85, IQR = 85 – 55 = 30

✓ The lowest suggested value = Q1 – 1.5 IQR = 55 - 1.5(30) = 10

✓ The highest suggested value = Q3 + 1.5 IQR = 85 + 1.5(30) = 130

✓ 1<10 and 135>130 so they are outliers represented by a dot on the box and whisker plot.

- ***How to calculate Q1, Q2, Q3 for intervals?***

➢ Q2 is the median:

✓ Rank of Q2 = 169/2 = 84.5

✓ $\mathbf{Q2 = 29.5 + \left(\frac{84.5-70}{47}\right) * 10 = 32.6}$

Intervals	True limits	frequency	Cumulative frequency
10--19	9.5-19.5	4	4
20-29	19.5-29.5	66	70
30-39	29.5-39.5	47	117
40-49	39.5-49.5	36	153
50-59	49.5-59.5	12	165
60-69	59.5-69.5	4	169
Total		169	

➢ Q1 is the 25th percentile:

✓ Rank of Q1 = 169/4 = 42.25

✓ $\mathbf{Q1 = 19.5 + \left(\frac{42.25-4}{66}\right) * 10 = 25.29}$

➢ Q3 is the 75th percentile

✓ Rank of Q1 = 169*3/4 = 126.75

✓ $\mathbf{Q3 = 29.5 + \left(\frac{126.75-117}{36}\right) * 10 = 32.2}$

- ***Geometric standard deviation***

➢ The geometric standard deviation (GSD) measures the spread of a set of numbers that are best represented by their geometric mean.

➢ It is often used in data sets that involve multiplicative processes, such as growth rates or logarithmic distributions, where the geometric mean is more appropriate than the arithmetic mean.

- **Log GSD** = $\sqrt{\frac{\sum_{i=1}^{n}(\log xi - \log GM)^2}{n-1}}$ **Then anti-log for that value we obtained**

★ Example:

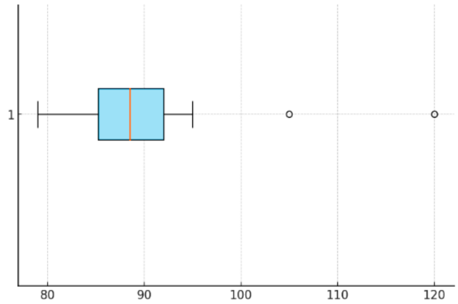
Data set: 89, 90, 87, 95, 86, 81, 120, 105, 83, 88, 91, 79

- ✓ Compute: mean, variance, standard deviation.
- ✓ Compute: median, the quartiles, and IQR.
- ✓ Construct a Box and Whisker plot for the data above.
- ✓ If the data point 79 changed to 197, calculate the new mean and the median and compare them to the old one? What do you conclude from this comparison?

- ***Let's start:***

The order array: 79, 81, 83, 86, 87, 88, 89, 90, 91, 95, 105, 120

- Mean (Average) = (89+90+87+95+86+81+120+105+83+88+91+79)/12 ≈ **91.17**
- Standard deviation: **SD** = $\sqrt{\frac{\sum_{i=1}^{n}(xi - \bar{x})^2}{n-1}}$ = **10.86**
- Variance: $S^2 = \frac{\sum(x-\bar{x})^2}{n-1}$ = **117.97**
- Median (Q2): the median is the average of the 6th and 7th values = (88+89)/2 = **88.5**
- First Quartile (Q1): the median of the lower half of the data (the first 6 numbers) = (83+86)/2 = **84.5**
- Third Quartile (Q3): the median of the upper half of the data (the last 6 numbers) = (91+95)/2 = **93**
- Construct a ***Box and Whisker plot***
  - ✓ Minimum: **79**
  - ✓ Q1: **84.5**
  - ✓ Median (Q2): **88.5**
  - ✓ Q3: **93**
  - ✓ Maximum: **120**



- If the data point 79 changed to 197, calculate the new mean and the median and compare them to the old one? What do you conclude from this comparison?
  - ✓ New order array: 81, 83, 86, 87, 88, 89, 90, 91, 95, 105, 120, 197
  - ✓ Mean= 101
  - ✓ Median = (89+90)/2 = 89.5
  - ✓ This shows that the ***mean*** is more sensitive to extreme values or outliers, while the ***median*** is a more robust measure of central tendency that is less affected by outliers.

ARKAN

ACADEMY

علم في كل مكان

Arkan academy

Arkanacademy

www.arkan-academy.com

+962 790408805