

Measures of central tendency, dispersion, and frequency table

Prepared by Dr. Hesham Mousli

MHA 613
Statistics in Healthcare

Lecture 3



Chapter

03

Numerical Descriptive Statistics (Numerical data) (Measures of central tendency)

- It is very difficult to have any **‘feeling’** for a set of numerical measurements unless we can summarize the data in a meaningful way.
- We can also condense the information by providing measures that describe the important characteristics of the data.
- The arithmetic mean (**average**) is a general term for a measure of **location**; it describes a typical measurement for data **centeredness**.

1. The arithmetic mean

- The **arithmetic mean**, often simply called the mean, of a set of values is calculated by adding up all the values and dividing this sum by the number of values in the set.
- It is useful to be able to summarize this verbal description by an **algebraic formula**.

1. The arithmetic mean

- Using mathematical notation, we write our set of n observations of a variable, x , as $x_1, x_2, x_3, \dots, x_n$.
- **For example**, x might represent an individual's height (cm), so that x_1 represents the height of the first individual, and x_i represents the height of the i^{th} individual, etc.
- We can write the formula for the arithmetic mean of the observations, which is written $\bar{x}$ and pronounced 'x bar', as:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

2. The median

- If we **arrange** our data in order of magnitude, starting with the **smallest** value and ending with the **largest** value, then the **median** is the **middle value** of this ordered set.
- The **median** divides the ordered values into **two halves**, with an equal number of values both **above** and **below** it.

2. The median

- It is easy to calculate the median if the number of observations, n , is odd.
- It is the $(n + 1)/2^{\text{th}}$ observation in the ordered set ***after ordering the data set in either an ascending or descending way.***
- So, for example, if $n = 11$, then the median is the $(11 + 1)/2 = 12/2 = 6^{\text{th}}$ observation in the ordered set.

2. The median

- If n is even then, we usually calculate it as the arithmetic mean of the two middle observations in the ordered set (i.e. the $n/2^{\text{th}}$ and the $(n/2 + 1)^{\text{th}}$).
- So, for example, if $n = 20$, the median is the arithmetic mean of the $20/2 = 10^{\text{th}}$ and the $(20/2) + 1 = (10 + 1) = 11^{\text{th}}$ observations in the ordered set.

2. The median

- If n is even then, we usually calculate it as the arithmetic mean of the two middle observations in the ordered set (i.e. the $n/2^{\text{th}}$ and the $(n/2 + 1)^{\text{th}}$).
- So, for example, if $n = 20$, the median is the arithmetic mean of the $20/2 = 10^{\text{th}}$ and the $(20/2) + 1 = (10 + 1) = 11^{\text{th}}$ observations in the **ordered set**.

2. The median

1, 3, 3, **6**, 7, 8, 9

Median = **6**

1, 2, 3, **4**, **5**, 6, 8, 9

Median = $(4 + 5) \div 2$
= **4.5**

3. The mode

- The **mode** is the value that occurs most frequently in a data set.
- Some data sets do not have a mode because each value only occurs once.
- When two or more values occur with the same highest frequency, they are referred to as modes (**Bimodal or Multimodal**).
- We **rarely** use the mode as a **summary measure**.

3. The mode

Definition

**Mode of
Categorical
Data**

**For non-numerical data, the mode
is the measure of central tendency
to use.**



Example:
A bar graph of the
categorical
variables of ice
cream
preferences.

Pros. & Cons.

Type of measure	Advantages	Disadvantages
Mean	• Use all the data values • Algebraically defined so, mathematically manageable	• Distorted by outliers
Median	• Not distorted by outliers	• Ignores most of the data
Mode	• Easily determined for the categorical data	• Ignores most of the data

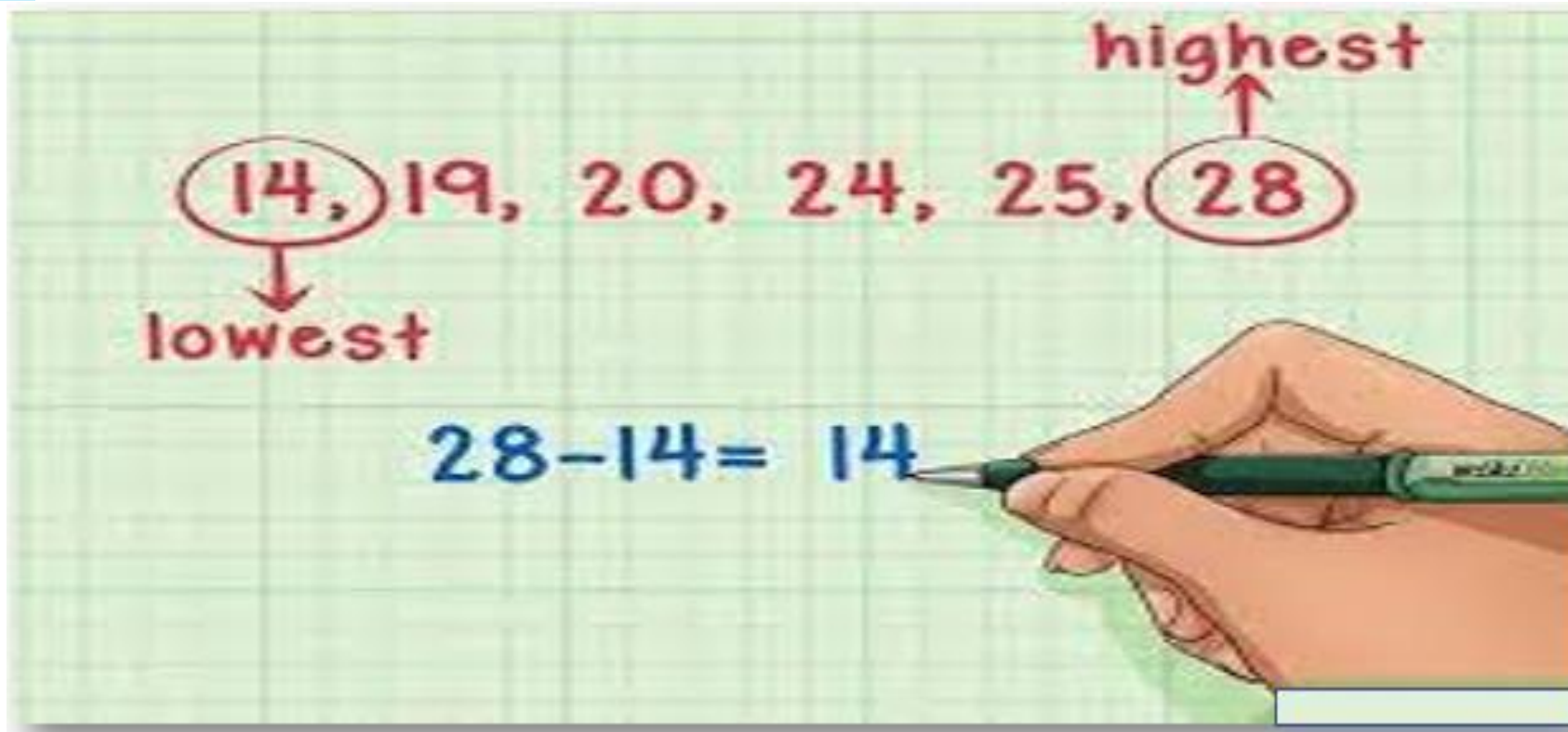
Chapter

03

Numerical Descriptive Statistics (Numerical data) (Measures of dispersion)

- We can provide two summary measures of a continuous variable: one indicating the 'average' and the other describing the 'spread' of observations.
- Then we have condensed the data in a meaningful way. **(summarized data)**.
- Main measures of **spread (dispersion)** are: **(Range, Interquartile Range (IQR), Variance, Standard deviation)**.

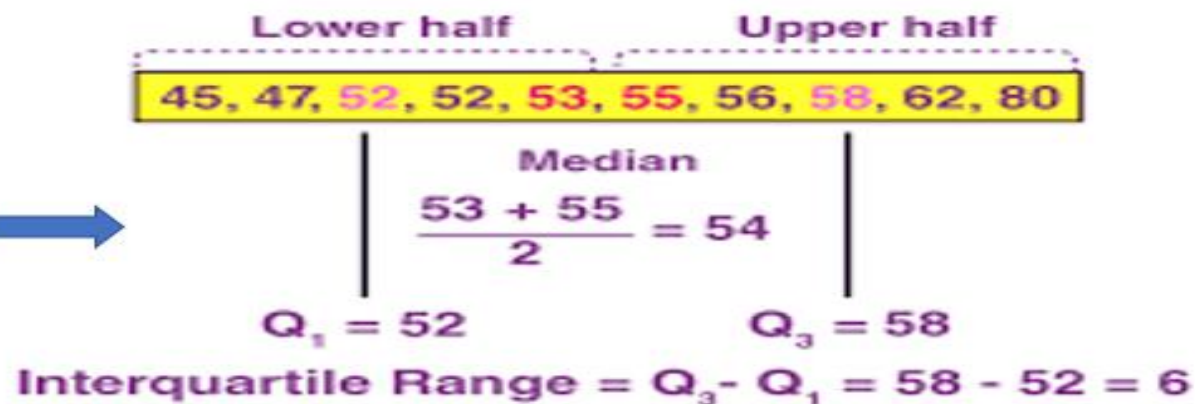
1. The Range



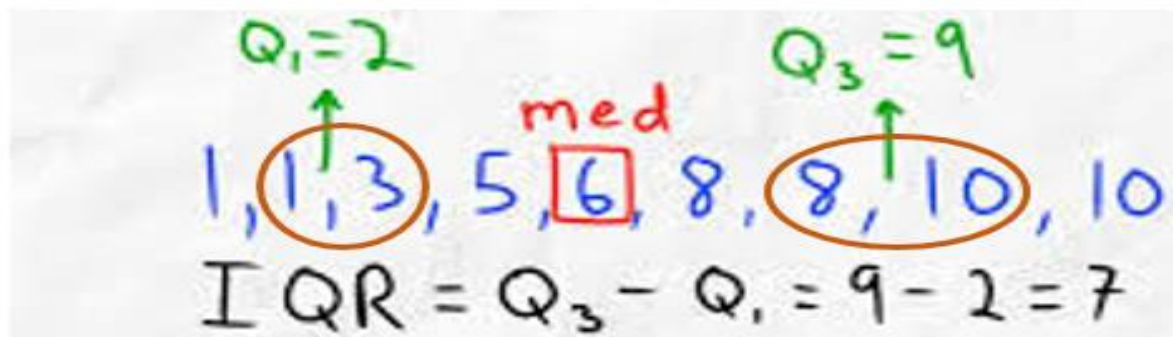
2. The Interquartile Range (IQR)

- To calculate the IQR, the data set is divided into quartiles, or four rank-ordered even parts.
- These quartiles are denoted by **Q1** (also called the lower quartile), **Q2** (the median), and **Q3** (also called the upper quartile).
- The lower quartile corresponds with the **25th percentile** and the upper quartile corresponds with the **75th percentile**, so $IQR = Q3 - Q1$.

1. Even number of observations



1. Odd number of observations



3. The Variance & Standard Deviation

Sample Variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Sample Standard Deviation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Calculating Variance & Standard Deviation

2. Subtract the mean from every data value

3. Square the previous column

1. Add your data and divide to find the mean value (μ)

Survey 1

x	$x - \mu$	$(x - \mu)^2$
3	-4	16
7	0	0
6	-1	1
8	1	1
11	4	16
35	Total	34

4. Add up this column

$$\therefore \mu = \frac{35}{5} = 7$$

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{n}} = \sqrt{\frac{34}{5}} \approx 2.61$$

5. Calculate Variance and Standard Deviation

Pros. & Cons.

Type of measure	Advantages	Disadvantages
Range	• Easy determined	• Uses only two observations • Distorted by outliers
Variance	• Uses every observation	• Sensitive to outliers
Standard Deviation	• Uses every observation • Easily interpreted	• Sensitive to outliers

Chapter

03

Numerical Descriptive Statistics (Categorical data) (Frequencies)

Categories	Tally Marks	Frequency
Red	 	5
Green	 	6
Blue		2
Black		1
Yellow		3

Score (x)	Tally Marks	Frequency (f)
13		2
14		4
15	 	5
16	 	8
17	 	6
Total		25

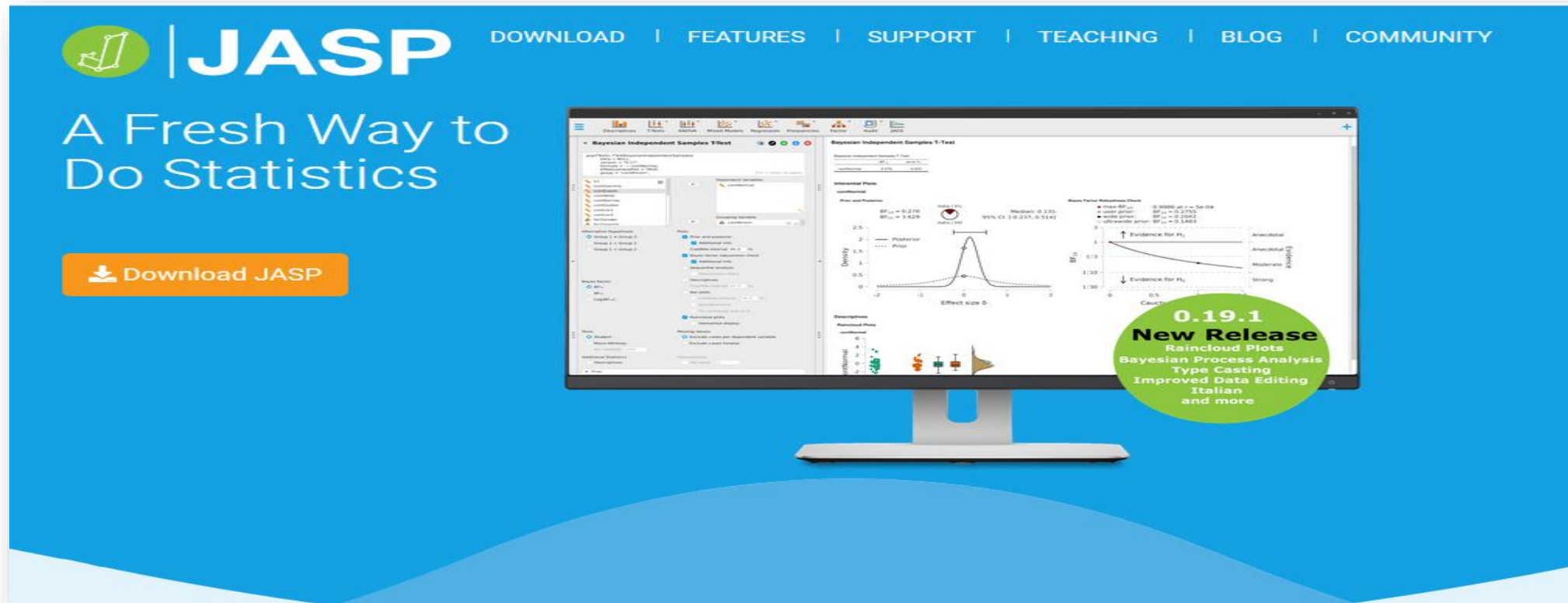
Case Scenario (1)

➤ **Medication Error Reduction Program:**

- A healthcare system wants to evaluate the effectiveness of a program to reduce medication errors in its pharmacy department over six months.

Case Scenario (1)

Month	Total Prescriptions	Medication Errors	Error Rate (%)
Jan.	10,000	30	0.30%
Feb.	11,500	25	0.22%
Mar.	12,000	28	0.23%
Apr.	11,000	18	0.16%
May	12,500	15	0.12%
June	13,000	10	0.08%



The image shows a promotional banner for JASP (JStatSoft) and a screenshot of the software interface. The banner has a blue background with the JASP logo and the text "A Fresh Way to Do Statistics". Below this is an orange "Download JASP" button. The software interface screenshot shows the "Bayesian Independent Samples T-Test" window. It includes a left sidebar with a tree view of the analysis, a central plot area showing a "Posterior" distribution curve, and a right sidebar with various statistical results and a "Bayesian Factor Reliability Check" plot. A green circular badge in the bottom right corner of the screenshot announces the "0.19.1 New Release" with features like "Raincloud Plots", "Bayesian Process Analysis", "Type Casting", "Improved Data Editing", and "Italian and more".

JASP DOWNLOAD | FEATURES | SUPPORT | TEACHING | BLOG | COMMUNITY

A Fresh Way to Do Statistics

Download JASP

0.19.1 New Release
Raincloud Plots
Bayesian Process Analysis
Type Casting
Improved Data Editing
Italian and more

JASP*

Analyses Synchronisation Resize Data Insert Remove

	Month	Total Prescriptions	Number of Medication Errors	Error Rate (%)
1	Jan.	10000	30	0.3
2	Feb.	11500	25	0.22
3	Mar.	12000	28	0.23
4	Apr.	11000	18	0.16
5	May	12500	15	0.12
6	June	13000	10	0.08

JASP*

	Month	Total Pres		Error Rate (%)
1	Jan.	10000		0.3
2	Feb.	11500		0.22
3	Mar.	12000	28	0.23
4	Apr.	11000	18	0.16
5	May	12500	15	0.12
6	June	13000	10	0.08

▼ Descriptive Statistics

Month
Error Rate (%)



Variables

Total Prescriptions
Number of Medication Errors

Split



☐ Transpose descriptives table

▼ Statistics

Sample size

- ☒ Valid
☒ Missing

Central tendency

- ☒ Mode
☒ Median
☒ Mean

Dispersion

- | | |
|--|---|
| ☒ Std. deviation | ☐ Coefficient of variation |
| ☐ MAD | ☐ MAD robust |
| ☐ IQR | ☐ Variance |
| ☐ Range | ☒ Minimum |
| ☒ Maximum | |

Quantiles

- ☐ Quartiles
☐ Cut points for: equal groups
☐ Percentiles:

Distribution

- ☐ Skewness
☐ Kurtosis
☐ Shapiro-Wilk test
☐ Sum

Results

Descriptive Statistics

Descriptive Statistics

	Valid	Missing	Mode	Median	Mean	Std. Deviation	Minimum	Maximum
Total Prescriptions	6	0	10000.000 ^a	11750.000	11666.667	1080.123	10000.000	13000.000
Number of Medication Errors	6	0	10.000 ^a	21.500	21.000	7.899	10.000	30.000

^a The mode is computed assuming that variables are discreet.

➤ **Descriptive Statistics Applied:**

- **Mean and Median Error Rate:** To track progress over time.
- **Error Rate Trend:** Monthly percentage reduction in errors.
- **Proportion of errors to prescriptions:** Relative to the workload.

➤ **Request Additional frequency distribution** for 126 medication error incidents categorized by error types:

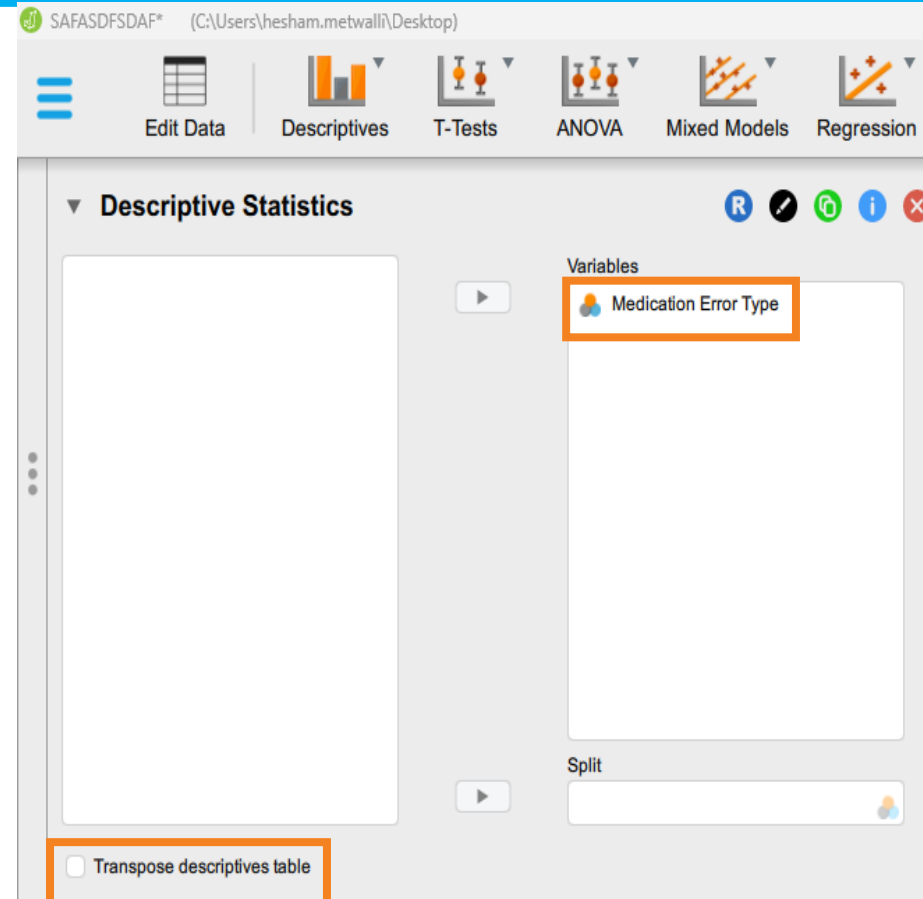
JASP*

Analyses Synchronisation Resize Data Insert Remove

	Month	Total Prescriptions	Number of Medication Errors	Error Rate (%)
1	Jan.	10000	30	0.3
2	Feb.	11500	25	0.22
3	Mar.	12000	28	0.23
4	Apr.	11000	18	0.16
5	May	12500	15	0.12
6	June	13000	10	0.08

SAFASDFSDF* (C:\Users\hesham.metwalli\Desktop)

	Edit Data	Descriptives	T-Tests
	 Medication Error Type		
1	Wrong Dose Administered.		
2	Incorrect Medication Given.		
3	Wrong Time of Administration.		
4	Wrong Time of Administration.		
5	Incorrect Route of Administration.		
6	Wrong Patient.		
7	Wrong Time of Administration.		



Results

Descriptive Statistics

Descriptive Statistics

Medication Error Type	
Valid	126
Missing	0
Mean	
Std. Deviation	
Minimum	
Maximum	

Frequency Tables

Frequencies for Medication Error Type

Medication Error Type	Frequency	Percent	Valid Percent	Cumulative Percent
Wrong Dose Administered.	32	25.397	25.397	25.397
Incorrect Medication Given.	25	19.841	19.841	45.238
Wrong Time of Administration.	18	14.286	14.286	59.524
Incorrect Route of Administration.	14	11.111	11.111	70.635
Wrong Patient.	9	7.143	7.143	77.778
Omitted Medication.	18	14.286	14.286	92.063
Medication Given Without Prescription.	10	7.937	7.937	100.000
Missing	0	0.000		
Total	126	100.000		

➤ **Decision Making:**

- **The descriptive statistics show a significant reduction in the medication error rate from 0.30% to 0.08%. The management concludes that the error reduction program is effective. They decide to:**
 1. Scale the program to other departments like surgery and inpatient wards.
 2. Provide additional training and improve communication between pharmacists and prescribers, since the results are promising but could still be improved.

Case Scenario (2)

➤ **Staffing Optimization in an Outpatient Clinic:**

- An outpatient clinic is experiencing long patient waiting times, which impacts patient satisfaction. The clinic's management aims to optimize staffing levels by analyzing patient flow and waiting time data.

JASP*

Analyses Synchronisation Resize Data Insert

	Patient No.	Waiting time (minutes)	Satisfaction Score (%)
1	1	15	80
2	2	63	75
3	3	14	91
4	4	32	75
5	5	22	81
6	6	24	69
7	7	15	85
8	8	36	90
9	9	35	77
10	10	24	91



Descriptive Statistics

Descriptive Statistics

Waiting time (minutes)	
Valid	100
Missing	0
Mode	24.000 ^a
Median	28.000
Mean	32.100
Std. Deviation	18.238
Minimum	14.000
Maximum	100.000

^a The mode is computed assuming that variables are discreet.

➤ **Descriptive Statistics Applied:**

- **Mean Waiting Time:** Identifies the busiest hours with the longest waits.
- **Mean & Median Patient Satisfaction Scores:** Used to gauge patient perception based on wait times.

➤ **Request Additional frequency distribution** for patients' reported complaints:

JASP*

	Analyses	Synchronis
	Service Complaints	
1	Inadequate seating/waiting	
2	Rushed Consultations	
3	Rushed Consultations	
4	Inadequate seating/waiting	
5	Poor Staff	
6	Slow processing by staff	
7	Difficulty Scheduling	
8	Inadequate seating/waiting	
9	Rushed Consultations	

Results

Descriptive Statistics

Descriptive Statistics

Service Complaints	
Valid	130
Missing	0
Mean	
Std. Deviation	
Minimum	
Maximum	

Frequency Tables

Frequencies for Service Complaints

Service Complaints	Frequency	Percent	Valid Percent	Cumulative Percent
Inadequate seating/waiting	44	33.846	33.846	33.846
Rushed Consultations	32	24.615	24.615	58.462
Poor Staff	20	15.385	15.385	73.846
Slow processing by staff	17	13.077	13.077	86.923
Difficulty Scheduling	17	13.077	13.077	100.000
Missing	0	0.000		
Total	130	100.000		

➤ **Decision Making:**

- **The clinic notices a peak in patient arrivals between 9 AM and 11 AM, with waiting times significantly increasing during those hours. The analysis suggests a clear relationship between low staff numbers and longer wait times. Management decides to:**
 1. Increase staffing during peak hours to reduce waiting times.
 2. Introduce an appointment scheduling system to smooth outpatient arrivals and prevent overcrowding.

TO WRAP UP

- Key statistical tools for measuring central tendency are the mean, median, mode.
- Key statistical tools for measuring dispersion are the range, interquartile range (IQR), variance, and standard deviation.
- Key statistical tools for analyzing frequencies are the frequency distribution and frequency table.
- In healthcare decision-making, these statistical tools are crucial for data analysis, trend identification, and informed decisions to improve patient care and efficiency.



Thank You