POLS0008
INTRODUCTION TO QUANTITATIVE RESEARCH METHODS

EXAMINING DATA (PART I)

Dr Anwar Musah (a.musah@ucl.ac.uk)
Lecturer in Social and Geographic Data Science
UCL Geography

RECAP OF LAST WEEK

Definition of Statistics?

- Statistics is the science of collecting, organising, analysing and interpreting numerical data to assist in making a more informed decision

- In short – statistics is the science of crunching data and getting some meaningful information from it

- It is typically use to summarising data points aka <u>samples</u>, and testing a hypothesis using such <u>sample</u> to make predictions about a <u>population</u> OR determine causal relationships

Definition of Statistics?

- Composed of three main facets: Description, Inference, and Design
- The descriptive and inferential elements are what we refer to as statistical analysis

Descriptive (explorative) statistics

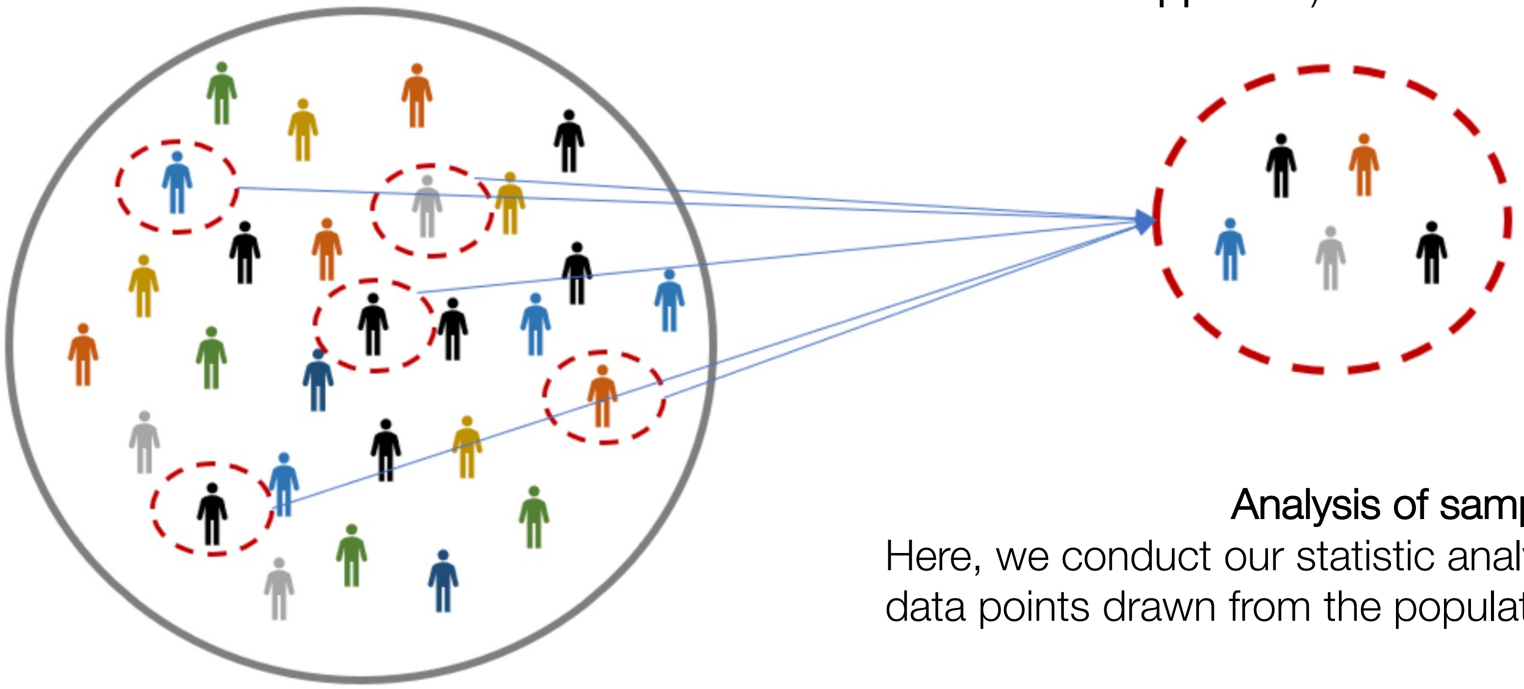
- Describing ways the data looks
- Summarizing the data that has been collected
- Design: Cross-sectional and ecological framework
- Hypothesis generating exercise

Inferential (evidence-based) statistics

- Making prediction (or future forecasts) about the wider population
- Evidence-based research for testing hypothesis and making conclusions
- Causal inference
- Design: Case-control, cohort and RCTs framework

Population
This is the entire set of data points we wish to study (e.g., individuals, businesses, geographical units etc.)

Sample
This is a subset of data points chosen for study through data collection (preferably, **probabilistic data collection approach**)



Analysis of sample
Here, we conduct our statistic analysis on the subset of data points drawn from the population.

Whatever descriptive summaries or inferences made on this sample – its **representative & generalisable** of the population its from.

More (very basic) Statistical Notation

sample statistic	population parameter	description
n	N	number of members of sample or population
$\bar{x}$ "x-bar"	μ "mu" or μ_x	mean
s (TIs say Sx)	σ "sigma" or σ_x	standard deviation For variance, apply a squared symbol (s^2 or σ^2).
$\hat{p}$ "p-hat"	p	proportion

Type of Variables

- A variable is anything that we can measure about the subjects in our sample
- Variables vary, that is they take on a range of values
- There are two classifications of variables

Continuous Variables	Categorical Variables
1. Interval 2. Ratio	1. Nominal 2. Ordinal

- Levels of measurement: (Lowest) Nominal << Ordinal << Interval << Ratio (Highest)

- We defined what independent and dependent variables are.

What are Descriptive Statistics?

Univariable analysis

- Analysis of only one variable on some characteristic
    - ❖Frequency Distributions – essentially a count or distribution of values on some single variable
    - ❖Other descriptive statistics – some summary measure that describes the data in a way not obvious by looking at the frequency distribution

Bivariable analysis

- Analysis of two variables – can be simple scatter plots or cross-tabulations

Multivariable analysis

- Analysis of three or more variables

We created a table that contains group categories for height (of 5cm) measurement, and compute the frequency and proportions. In addition, we compute the cumulative frequency and its cumulative proportion as well.

94, 95, 97, 97, 100, 100, 101, 102, 103, 105, 105, 108, 108, 109, 109, 112, 113, 113, 118, 119, 121

Height groups	Frequency	Relative Frequency or percentage (%)	Cumulative Frequency	Cumulative Relative Frequency
90-95	2	0.09523810 (9%)	2	0.09523810 (9%)
96-100	4	0.19047619 (19%)	6	0.2857143 (28%)
101-105	5	0.23809524 (24%)	11	0.5238095 (52%)
106-110	4	0.19047619 (19%)	15	0.7142857 (71%)
111-115	3	0.14285714 (14%)	18	0.8571429 (85%)
116-120	2	0.09523810 (9%)	20	0.9523810 (95%)
120+	1	0.04761905 (4%)	21	1.0000000 (100%)

We group the data points accordingly

94, 95

97, 97, 100, 100

101, 102, 103, 105, 105

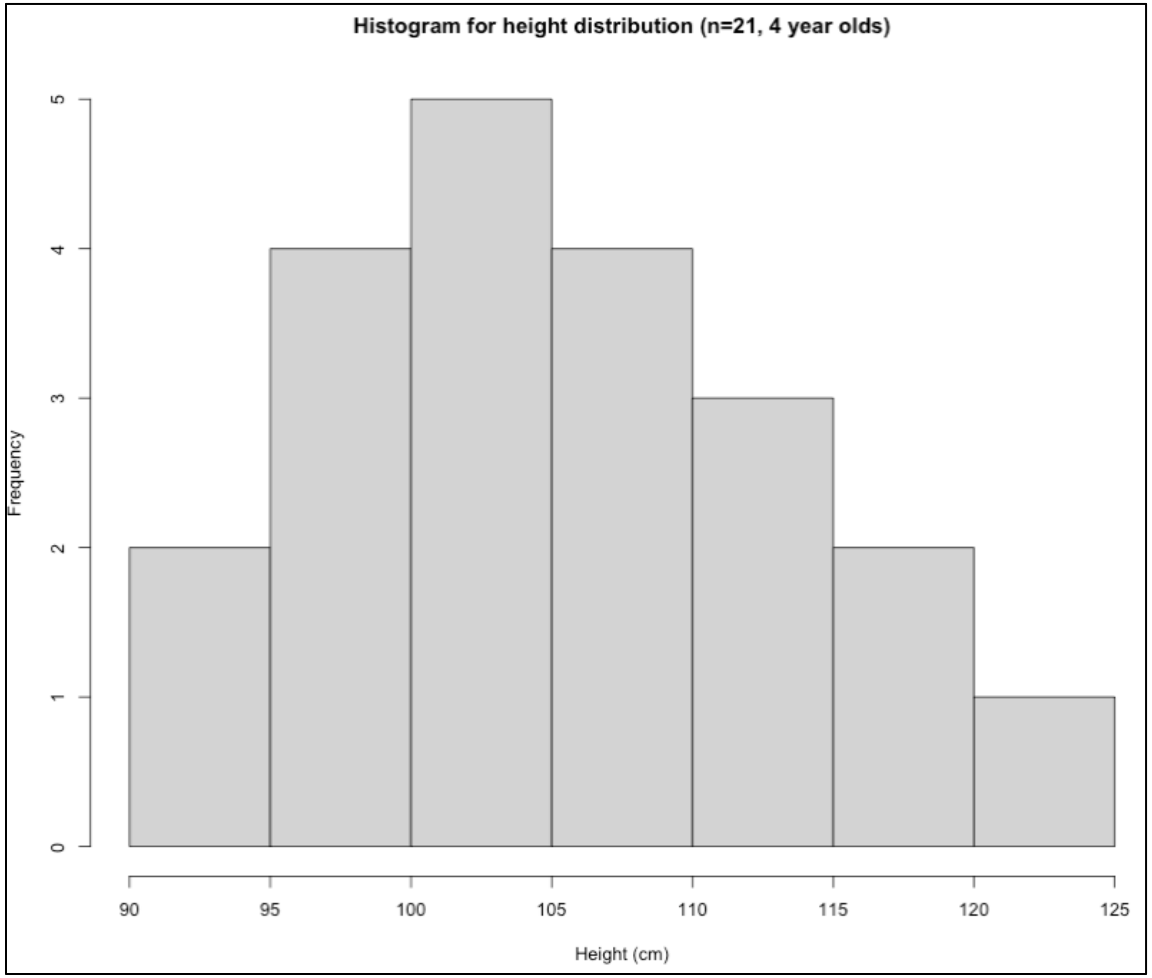
108, 108, 109, 109

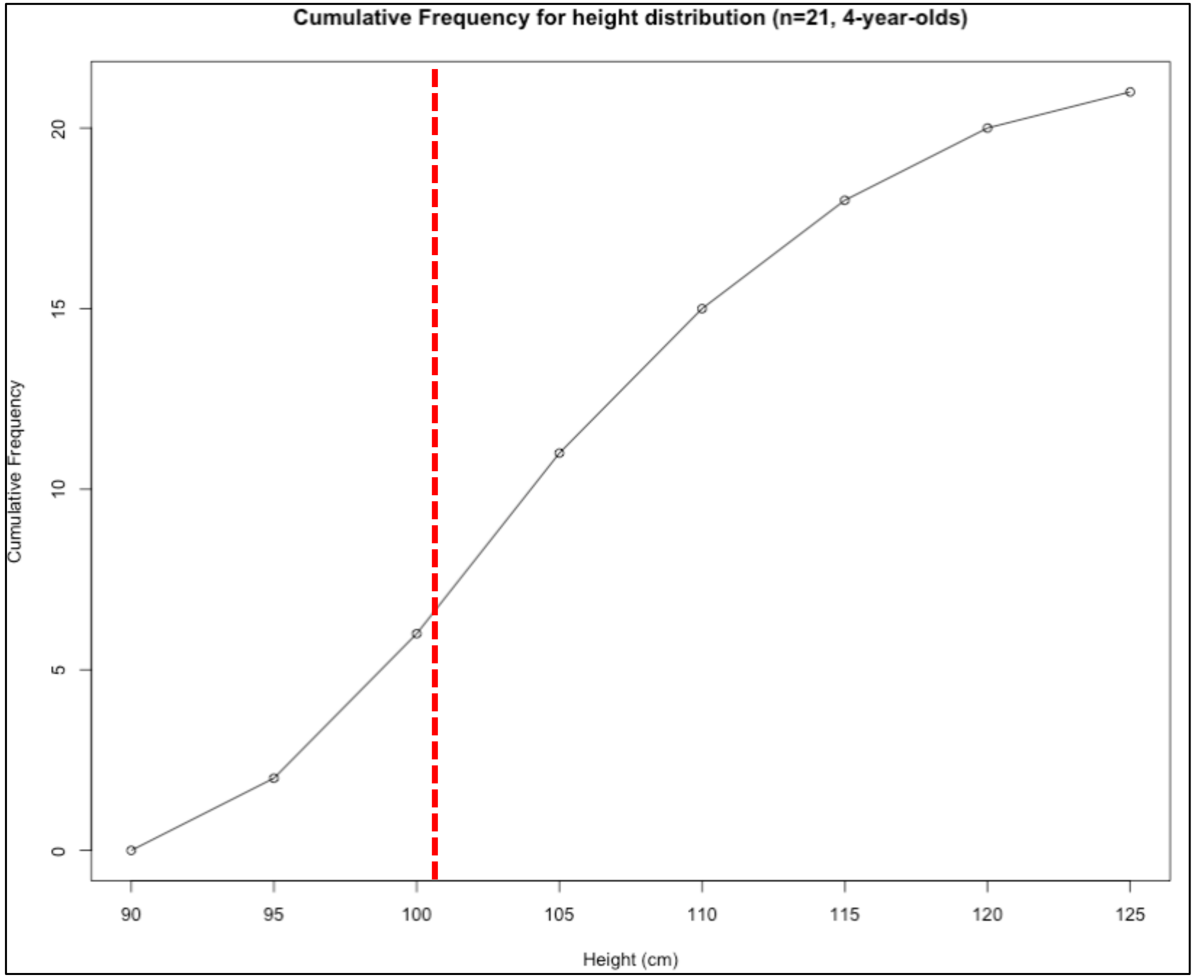
112, 113, 113

118, 119

121

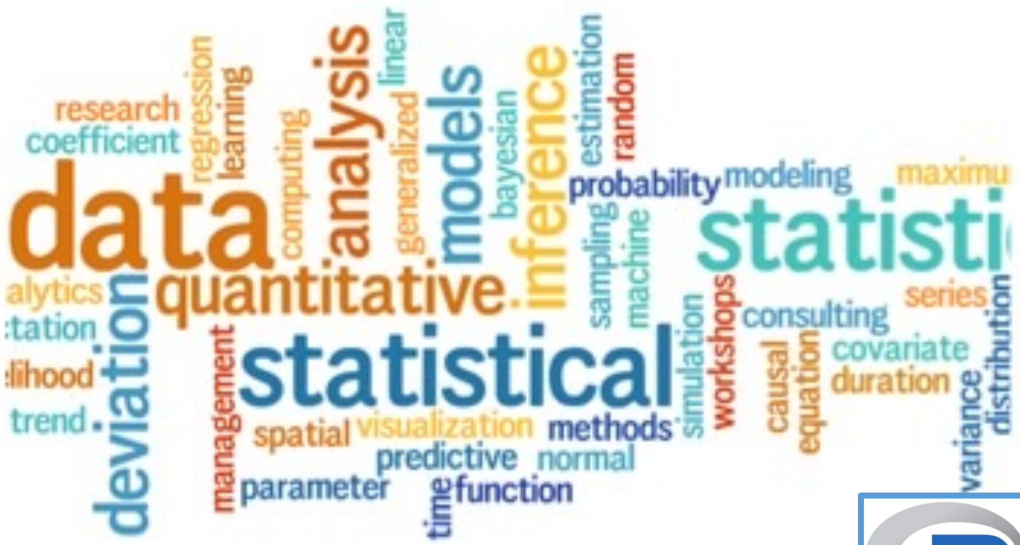
This output is called a "**Frequency Distribution table**", it's visual representation is a **histogram** for the data's **relative frequency** and **cumulative frequency plot** for the **cumulative frequency**.

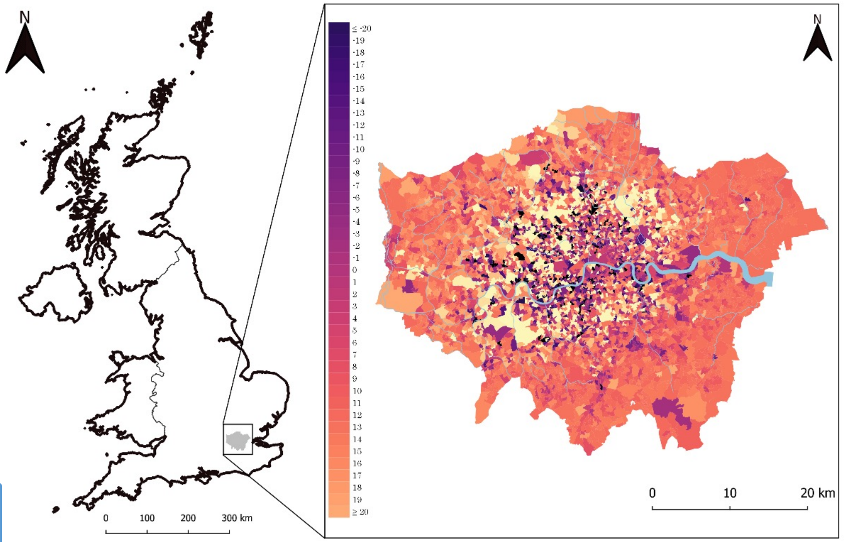


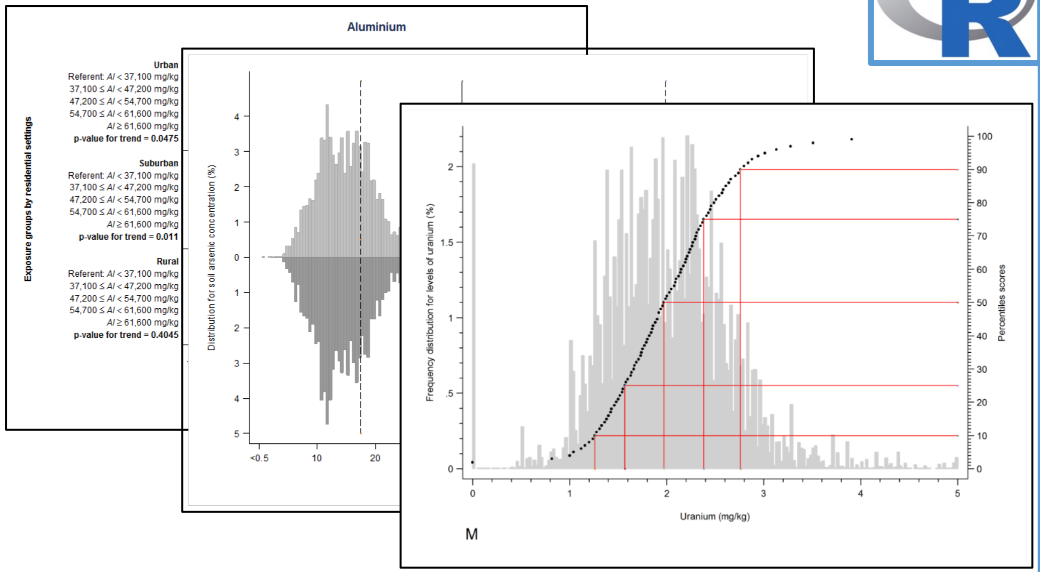


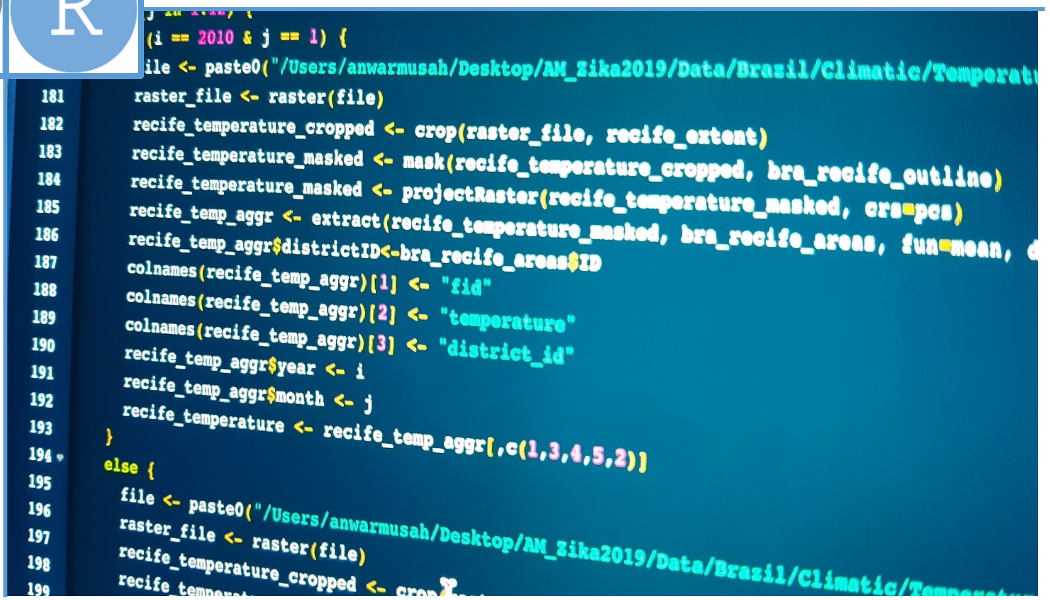
Interpretation: The above table output show the frequency distribution of heights (in cm) in kids who are 4 years of age entering in reception. The group with the highest frequency was 101-105 cm which accounts for 24% of the data. Health-wise, we can see from the cumulative frequency results that there are 6 kids with height values that are less than 101 cm. This corresponds to 0.2857 (29%) of the data – descriptively, these 6 kids growth is a cause for concern.

Statistical Inference









Summary Statistics

Summary measures

What kinds of analysis and summary statistics can you perform on a particular type of dataset?

1. Categorical Data

You can group the data by according to categories and perform the following:

- Compute the Frequencies (counts)
- Compute the Percentages (or Relative Frequency)
- Calculate the Cumulative Frequencies or Cumulative Percentages
- Graphical approaches also include bar plots and pie charts
- The Mode (category with that occurs most)

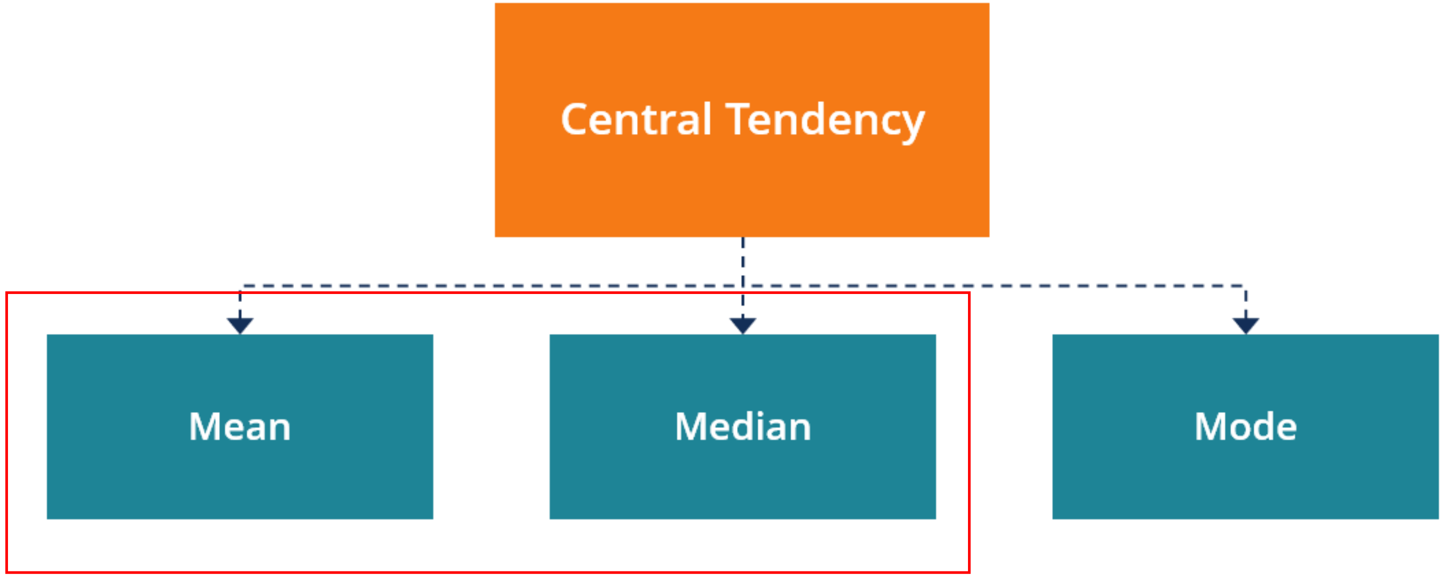
2. Numerical Data

You can perform the following analysis:

- Compute the mode (value that occur most)
- Compute the median
- Compute the mean
- Lowest (Minimum) & Highest (Maximum)
- Percentiles
- Variance
- Standard deviation
- Range
- Quartiles and Interquartile ranges

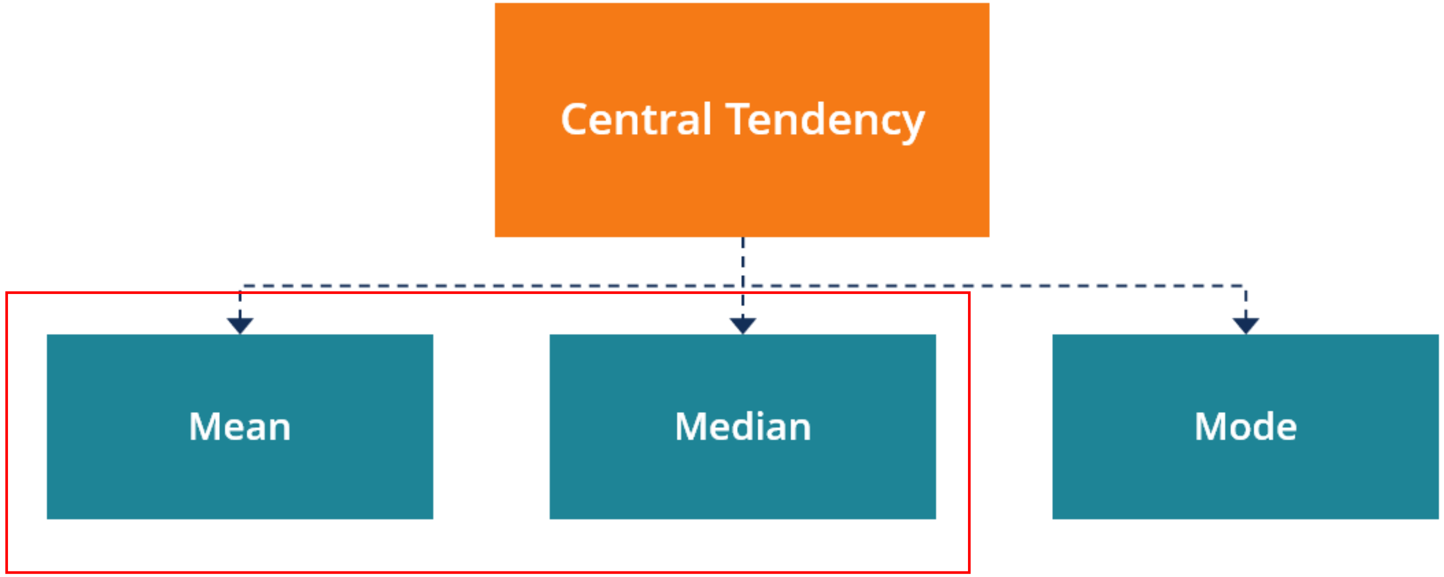
Central Tendency

- Indicate the location of the middle or the centre of a distribution
- Central tendency is the point at which the distribution is in balance



Measures of Central Tendency

- Indicate the location of the middle or the centre of a distribution
- Central tendency is the point at which the distribution is in balance



Mean

- Arithmetic Mean (also referred as the word Average) is a central estimate
- Takes into account all values
- Mean is the preferred measure of central tendency, except when there are extreme values
- Easily distorted by extreme values

$$\bar{x} = \frac{\sum x_i}{n}$$

where $\sum \mathbf{x_i}$ represents the sum of all observations and 'n' is the number of observations in the sample. The $\bar{\mathbf{x}}$ represent the mean

Example of Summarising Numerical Data using Mean

Finding the mean from these 9 data points: 13, 18, 13, 14, 13, 16, 14, 21, 13

$n = 9$
x_i represents each of these observations. $x_1 = 13,\ x_2 = 18,\ x_3 = 13, \ldots$ and $x_9 = 13$
$\bar{x}$ represents the mean which is to be calculated when summing all x_i and dividing by n

Solution:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{(13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13)}{9} = 15$$

Average or central value from that dataset is 15

In RStudio, the code for this would be:

```
datapoints <- c(13, 18, 13, 14, 13, 16, 14, 21,13)
mean(datapoints)
```

Median

- Median is the exact middle value
- Half of the values are smaller than the median and half are larger
- The median divides the distribution in two equal parts

$$\text{Median} = \frac{(\text{n}+1)^{\text{th}}}{2}$$ the value of ordered observations

Example of Summarising Numerical Data using Median

Finding the median from these 9 data points: 13, 18, 13, 14, 13, 16, 14, 21, 13

n = 9

Arrange the values in ascending order:
Before: 13, 18, 13, 14, 13, 16, 14, 21, 13
After: 13, 13, 13, 13, 14, 14, 16, 18, 21

In RStudio, the code for this would be:

```
datapoints <- c(13, 18, 13, 14, 13, 16, 14, 21,13)
median(datapoints)
```

Solution:

$$\text{Median} = \frac{(n+1)\text{th}}{2} = \frac{(9+1)}{2} = 5^{\text{th}} \text{ position in that arranged dataset}$$

13, 13, 13, 13, $\boxed{14,}$ 14, 16, 18, 21
So, the median is 14

Things to know about the Mean & Median [1]

1. The Mean is influenced by extreme values
2. The Median is not influenced by extreme values

Distribution A

2 3 4 5 **20**

- Mean

  2+ 3+ 4+ 5+ 20 / 5

  34 / 5 = 6.8

- Median

  5 + 1 / 2 = 3rd obs.

  = 4

Distribution B

2 3 4 5 **6**

- Mean

  2+ 3+ 4+ 5+ 6 / 5

  20 / 5 = 4

- Median

  5 + 1 / 2 = 3rd obs.

  = 4

Things to know about the Mean & Median [2]

Median values are different when sample size contains an odd or even number of data points

Rank data from the lowest to the highest value median is rank calculated from (n+1)/2

Odd (n = 5)

14, 17, 18, 20, 21

Median = (n +1)/ 2

Median = (5 + 1)/ 2 = 3rd obs

14, 17, $\boxed{18,}$ 20, 21

Median = 3rd obs = 18

Even (n = 4)

14, 17, 18, 20

Median = (n +1)/ 2

Median = (4+1/2 = 2.5th obs

14, 17 || 18, 20

Median = 2.5th obs = 17.5

Measure for Dispersion & Variation [1]

- NOTE: Measures of central tendency give us measures of where the middle of a set of data occurs - it is not enough to report only measures of central tendencies
- This is where measure for variation and dispersion steps in…
- These measure describe how the data is clustered or dispersed around the mean.
- Measures of variation determine the range of the distribution relative to the measures of central tendency

Measure for Dispersion & Variation [2]

- These estimates basically determine the spread and range of values
- It determines how the distribution are relative to the measures of central tendency:
  - Range (Maximum & Minimum value)
  - Upper and lower quartiles (i.e., 75th & 25th Percentiles), and Interquartile range
  - Variance to estimate the Standard Deviation

Range (minimum & maximum)

Range gives the extreme values

The difference between the min and max values

= highest value – lowest value

Lowest value 13, 13, 13, 13, 14, 14, 16, 18, 21 Highest value

Range = 21 – 13 = 8

In RStudio, the code for this would be:

```
datapoints <- c(13, 18, 13, 14, 13, 16, 14, 21,13)

max(datapoints)
Min(datapoints)

max(datapoints) – min(datapoints)
```

UCL

Quartiles and Interquartile range

Divide a range of data into four equal parts 100/4 parts = 25%

Lower quartile (1st quartile, Q1) (aka 25th percentile)

(n+1)/4 value of ordered obs.

It's the number below which lies the 25% of the bottom data

2nd quartile (the median, Q2) (aka 50th percentile)

Divides the range in the middle & has 50% of the data below it

Upper quartile (3rd quartile, Q3) (aka 75th percentile)

3*(n+1)/4 value of ordered obs.

It has 75% of the data below it & the top 25% of the data above it

Example of Summarising data using quartiles and IQRs

Find the median, the lower and upper quartile from these data points: 1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27

Solution:
1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27
n = 11

$\textbf{Median} = \frac{(n+1)}{2} = \frac{(11+1)}{2} = \frac{(12)}{2} = 6^{th}$ position; So, the median is 9

$\textbf{Lower quartile} = \frac{(n+1)}{4} = \frac{(11+1)}{4} = \frac{(12)}{4} = 3^{rd}$ position; So, the lower quartile is 5

$\textbf{Upper quartile} = \frac{3(n+1)}{4} = \frac{3(11+1)}{4} = \frac{3(12)}{4} = \frac{36}{4} =$ 9th position; upper quartile is 18

To find the middle 50% of the data; IQR = 18 – 5 = 13

In RStudio, the code for this would be:

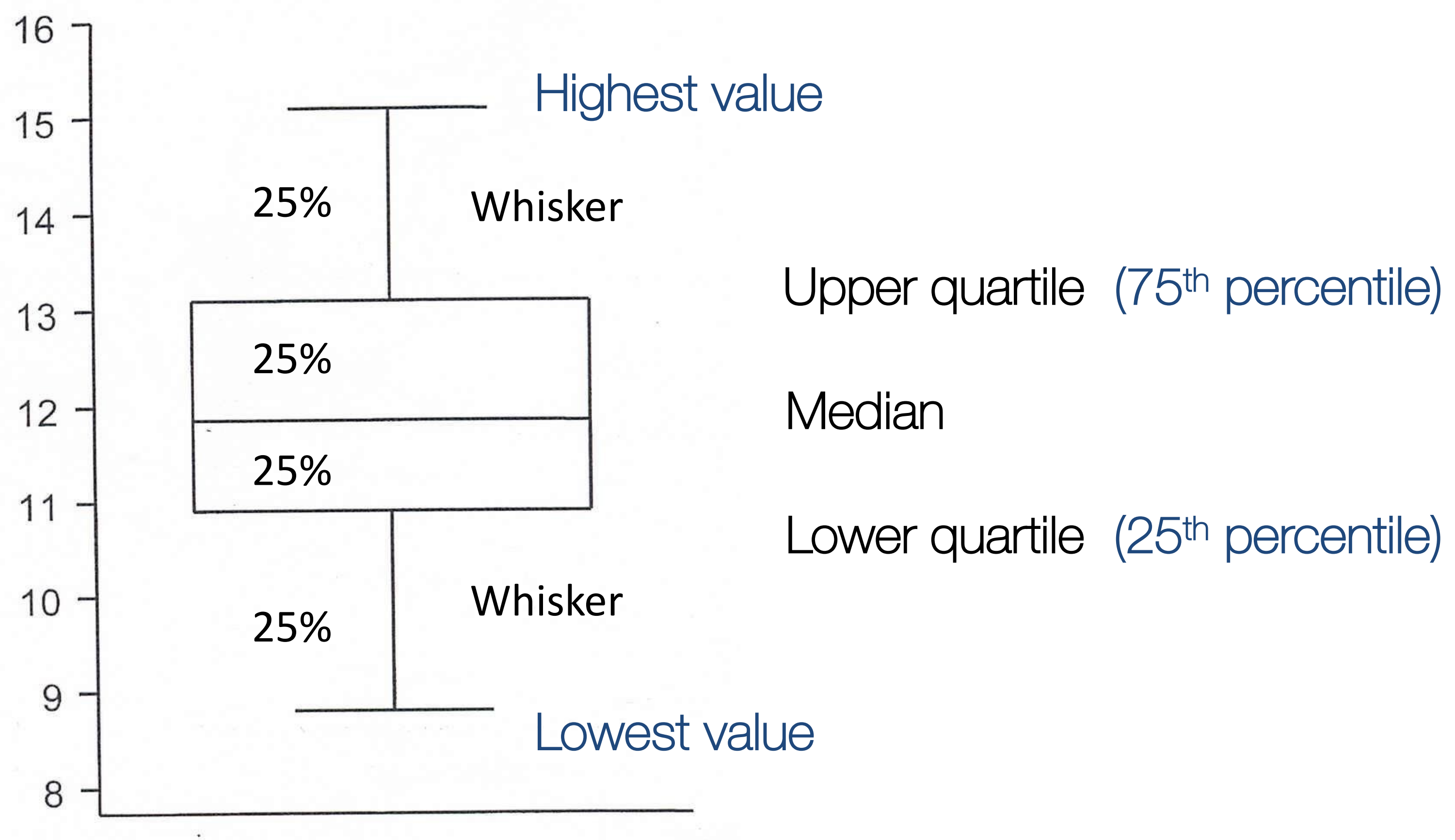
```
datapoints <- c(1, 2, 5, 6, 7, 8, 9, 12, 15, 18)

summary(datapoints)
```

1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27

Lower 25%	Middle 50% of the data	Upper 25%

Box (whiskers) plot [1]



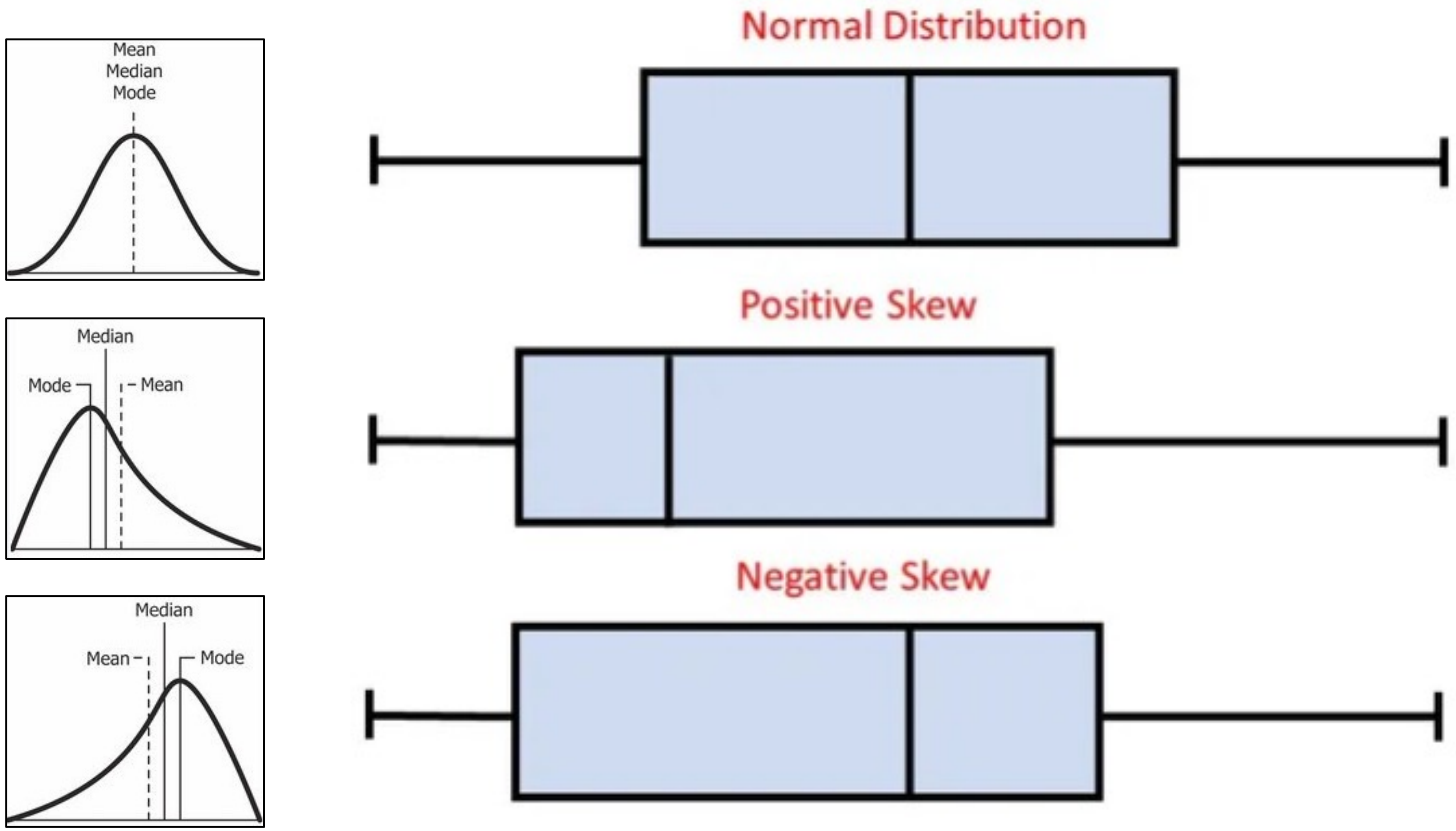
Why are box plots useful?

Box plots divides the data into section that each contain 25% of the data in the section

They provide a visual summary to quickly identify the pattern i.e., central and dispersion.

If its perfectly symmetrical (which most box plots are not) it is safe to assume the data is normally distributed, otherwise its skewed

Box (whiskers) plot [2]



The box plot looks really centred on the median – so the distribution is considered normal. Traits: Mean, Median and Mode are the same

The box plot shows the median centred on the left – so the distribution is considered positively skewed. Traits: Mode > Median > Mean

The box plot shows the median centred on the right – so the distribution is considered negatively skewed. Traits: Mode < Median < Mean

Variance

- It is defined in terms of the deviations of the observations from the mean
- Measures the spread about the mean
- Measures ‘somewhat’ the average distance between to observation and mean value
- We need to calculate this in order to get the standard deviation

Low value for (↓) S^2 = data are clustered about the mean

High value (↑) S^2 = data are spread Or disperse

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

Standard Deviation

It is more convenient to express the variance in the original unit by taking the square root of the variance

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

Example: Variance and Standard Deviation

Recall finding the mean from these 9 data points: 13, 18, 13, 14, 13, 16, 14, 21, 13
We estimated it as 15

Let measure how this data spreads around this value by calculating the variance and standard deviation

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \quad \blacktriangleright \quad \boldsymbol{S} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Step 1: List the observations accordingly in a table
i.e., 13, 18, 13, 14, 13, 16, 14, 21, 13,

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

i	x_i			
1	13			
2	18			
3	13			
4	14			
5	13			
6	16			
7	14			
8	21			
9	13			

Step 2: Calculate the mean from the list of observed values

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

i	x_i			
1	13			
2	18			
3	13			
4	14			
5	13			
6	16			
7	14			
8	21			
9	13			

Mean: $\bar{x} = 15$

Step 3: Calculate the difference between the observed value and the mean

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

i	x_i	$(x_i - \bar{x})$		
1	13	13 – 15 = -2		
2	18	18 – 15 = 3		
3	13	13 – 15 = -2		
4	14	14 – 15 = -1		
5	13	13 – 15 = -2		
6	16	16 – 15 = 1		
7	14	14 – 15 = -1		
8	21	21 – 15 = 6		
9	13	13 – 15 = -2		

Mean: $\bar{x} = 15$

Step 4: Calculate the squared difference between the observed value and the mean

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

i	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	
1	13	13 – 15 = -2	$(-2)^2 = 4$	
2	18	18 – 15 = 3	$(3)^2 = 9$	
3	13	13 – 15 = -2	$(-2)^2 = 4$	
4	14	14 – 15 = -1	$(-1)^2 = 1$	
5	13	13 – 15 = -2	$(-2)^2 = 4$	
6	16	16 – 15 = 1	$(1)^2 = 1$	
7	14	14 – 15 = -1	$(-1)^2 = 1$	
8	21	21 – 15 = 6	$(6)^2 = 36$	
9	13	13 – 15 = -2	$(-2)^2 = 4$	

Mean: $\bar{x} = 15$

Step 5: Calculate the SUM of the squared differences between the observed value and the mean

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

i	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	13	13 – 15 = -2	$(-2)^2 = 4$
2	18	18 – 15 = 3	$(3)^2 = 9$
3	13	13 – 15 = -2	$(-2)^2 = 4$
4	14	14 – 15 = -1	$(-1)^2 = 1$
5	13	13 – 15 = -2	$(-2)^2 = 4$
6	16	16 – 15 = 1	$(1)^2 = 1$
7	14	14 – 15 = -1	$(-1)^2 = 1$
8	21	21 – 15 = 6	$(6)^2 = 36$
9	13	13 – 15 = -2	$(-2)^2 = 4$

Mean: $\bar{x} = 15$

$\sum (x_i - \bar{x})^2 = 64$

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

i	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	13	13 – 15 = -2	$(-2)^2 = 4$
2	18	18 – 15 = 3	$(3)^2 = 9$
3	13	13 – 15 = -2	$(-2)^2 = 4$
4	14	14 – 15 = -1	$(-1)^2 = 1$
5	13	13 – 15 = -2	$(-2)^2 = 4$
6	16	16 – 15 = 1	$(1)^2 = 1$
7	14	14 – 15 = -1	$(-1)^2 = 1$
8	21	21 – 15 = 6	$(6)^2 = 36$
9	13	13 – 15 = -2	$(-2)^2 = 4$

Mean: $\bar{x} = 15$

$\sum(x_i - \bar{x})^2 = 64$

$S^2 = 64/(9-1) = 64/8 = 8$

S^2 which is the variance is estimated to be ± 8.

We added a plus/minus sign because we are measuring the spread around the mean.

$S = \sqrt{8} = \pm 2.828$. This is the result we are seeking to show the were the mean lies in the data (or how the data deviates away from the mean)

Interpretation: the average is 15 with a ± 2.828 error or deviation

原作　矢立肇

Coffee break

Presentation of Descriptive Results

Here, we present some examples of how you should present a result for descriptive analysis

Example: Suppose you conducted a survey among 654 respondents in living in the East Midlands. You want to determine the overall distribution of Lung Capacity Volume among these individuals.

Variables	Type	Codebook
Lung Capacity Volume (LCV) (in litres)	Continuous (Dependent)	Ratio
Gender	Categorical (Independent)	Male Female
Age Groups (in years)	Categorical (Independent)	< 55 years 55-59 years 60+ years
Altitude	Categorical (Independent)	High land Low land

An example show the overall distribution of a dependent variable that is continuous

Data for LCV, 654 response:

3.124, 3.172, 3.160, 2.674, 3.685, 5.008, 3.757, 2.245, 3.961, 3.826, 2.806, 3.205, 4.579, 4.354, 4.774, 3.796, 2.416, 3.634, 5.056, 5.812, 2.200, 1.768, 0.517, 5.734, …

Full interpretation: The overall mean LCV among the 654 respondents from East Midlands was 5.91 litres (with SD of ±2.60) with the following quantiles being: Median = 5.64; IQRs 3.94 to 7.35. The lower observed LCV was 0.37 and the highest was 15.37.

No need to worry about presenting any of this in a table etc. Just provide a direct interpretation like above.

Use a table when you have to provide several summary estimates on the fly, or summaries estimates broken down by several other categorical variables

An example of a table showing the overall breakdown of a dependent variable that is continuous by various categorical attributes that are treated as independent variables

Table 1: Shows a descriptive breakdown (by characteristics)of the lung capacity function (i.e., volume) among 654 respondents in the East Midlands.

Variables	n	Mean (±SD)	IQR [Median (Q1-Q4)]	Ranges (Min-Max)
Gender				
Female	318	5.35 (±1.937)	5.46 (3.85 to 6.98)	0.37 to 9.51
Male	336	6.44 (±3.011)	5.82 (4.03 to 8.6)	0.39 to 15.38
Age Groups				
<55 years	130	3.07 (±1.043)	3.06 (2.42 to 3.63)	0.37 to 5.73
55-59 years	407	5.99 (±1.979)	5.76 (4.57 to 7.15)	1.88 to 13.67
60+ years	117	8.8 (±2.387)	8.56 (7.17 to 10.68)	4.59 to 15.38
Altitude Type				
High land	65	7.83 (±2.25)	7.51 (6.38 to 9.25)	3.08 to 12.62
Low land	589	5.7 (±2.552)	5.39 (3.76 to 7.14)	0.37 to 15.38

Total sample size (N) = 654

Interpretation

- In terms of gender - overall, men on average have a higher LCV than women [mean: 6.44 (±3.011) vs. 5.35 (±1.937)]
- In terms of age group - those in the lowest age group on average have higher LCV compared to the other age groups
- Those who live in an area that’s high land have higher LCV than Low landers [mean: 7.83 (±2.25) vs. 5.7 (±2.55)]

An example of a table showing the overall distribution of a dependent variable that is continuous across other characteristics grouped/stratified by a categorical attribute

Table 2: Shows a descriptive analysis of the lung capacity function (i.e., volume) stratified by gender.

Variables	Men				Women			
	n	Mean (±SD)	Median (IQR)	Ranges (Min-Max)	n	Mean (±SD)	Median (IQR)	Ranges (Min-Max)
Age Groups								
<55 years	65	3.25 ±1.015)	3.15 (2.72 to 3.8)	0.39 to 5.73	65	2.89 (±1.047)	2.83 (2.24 to 3.39)	0.37 to 5.69
55-59 years	209	6.26 (±2.293)	5.94 (4.49 to 7.67)	2.29 to 13.67	198	5.7 (±1.535)	5.69 (4.67 to 6.96)	1.88 to 9.51
60+ years	62	10.37 (±2.028)	10.5 (9.2 to 11.46)	4.83 to 15.38	55	7.02 (±1.238)	7.18 (6.07 to 7.89)	4.59 to 9.45
Altitude								
High land	26	9.23 (±2.668)	9.63 (8.08 to 11.14)	3.08 to 12.62	39	6.9 (±1.269)	7.22 (6.03 to 7.59)	4.59 to 9.51
Low land	310	6.2 (±2.922)	5.64 (3.89 to 8.06)	0.39 to 15.38	279	5.14 (±1.918)	5.25 (3.63 to 6.6)	0.37 to 9.45

Total sample size = 654 (Men = 336 and Women = 318)

Some note on interpretation (should be comparative)

- You can provide a descriptive interpretation that is gender-specific
- You can also do a descriptive cross-comparisons to compare the magnitude in differences
- You can also do checks for dose-response relationships – e.g., altitude from low --> high shows LCV increase; age groups, from lowest to highest, shows (weirdly) LCV increasing.

Presentation of tables [1]

Table 2: Shows a descriptive analysis of the lung capacity function (i.e., volume) stratified by gender. Data is from among 654 respondents from the East Midlands.

Variables	n	Mean (±SD)	IQR [Median (Q1-Q4)]	Ranges (Min-Max)
Gender				
Female	318	5.35 (±1.937)	5.46 (3.85 to 6.98)	0.37 to 9.51
Male	336	6.44 (±3.011)	5.82 (4.03 to 8.6)	0.39 to 15.38
Age Groups				
<55 years	130	3.07 (±1.043)	3.06 (2.42 to 3.63)	0.37 to 5.73
55-59 years	407	5.99 (±1.979)	5.76 (4.57 to 7.15)	1.88 to 13.67
60+ years	117	8.8 (±2.387)	8.56 (7.17 to 10.68)	4.59 to 15.38
Altitude Type				
High land	65	7.83 (±2.25)	7.51 (6.38 to 9.25)	3.08 to 12.62
Low land	589	5.7 (±2.552)	5.39 (3.76 to 7.14)	0.37 to 15.38

Total sample size (N) = 654

Variables	n	Mean (SD)	IQR [Median (Q1-Q4)]	Ranges (Min-Max)
Female	318	5.35 (1.937)	5.46 (3.85 to 6.98)	0.37 to 9.51
Male	336	6.44 (3.011)	5.82 (4.03 to 8.6)	0.39 to 15.38
<55 years	130	3.07 (1.043)	3.06 (2.42 to 3.63)	0.37 to 5.73
55-59 years	407	5.99 (1.979)	5.76 (4.57 to 7.15)	1.88 to 13.67
60+ years	117	8.8 (2.387)	8.56 (7.17 to 10.68)	4.59 to 15.38
High land	65	7.83 (±2.25)	7.51 (6.38 to 9.25)	3.08 to 12.62
Low land	589	5.7 (±2.552)	5.39 (3.76 to 7.14)	0.37 to 15.38

Best Standards:

Fully formatted table, with table legends which looks great, and its of the standards that is considered publication-worthy. This type will yield you full marks i.e., correct results and show an eye details.

Normal Standards:

Partially formatted table, with no table legends which looks rudimentary. No way near considered publication-worthy (i.e., in a report, research article, thesis etc.). The marker will be like “meh”… but if s/he (i.e., marker) is pissed-off or woke up on the wrong side of the bed, they may deduct marks.

Presentation of tables [2]

```
# A tibble: 2 × 9
  gender      n  mean    sd median    q1    q3   min   max
  <chr>   <int> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 female    318  5.35  1.94   5.46  3.85  6.98 0.373  9.51
2 male      336  6.44  3.01   5.82  4.03  8.60 0.388 15.4

# A tibble: 3 × 9
  agegroup        n  mean    sd median    q1    q3   min   max
  <chr>       <int> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 <55 years     130  3.07  1.04   3.06  2.42  3.63 0.373  5.73
2 55-59 years   407  5.99  1.98   5.76  4.57  7.15 1.88  13.7
3 60+ years     117  8.80  2.39   8.56  7.17 10.7  4.59  15.4

# A tibble: 2 × 9
  altitude      n  mean    sd median    q1    q3   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 high land    65  7.83  2.25   7.51  6.38  9.25 3.08   12.6
2 low land    589  5.70  2.55   5.40  3.76  7.14 0.373  15.4
```

Crap standards:

Not even worth the marker's time.

If you want to alienate the markers, we dare you to submit an assignment with tables not formatted…

We double dare you to submit an output that is copied and pasted from R…

Presentation of tables [3]

A tibble: 7 × 8

	n	mean	sd	median	q1	q3	min	max
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	130	3.07	1.04	3.06	2.42	3.63	0.373	5.73
2	407	5.99	1.98	5.76	4.57	7.15	1.88	13.7
3	117	8.80	2.39	8.56	7.17	10.7	4.59	15.4
4	318	5.35	1.94	5.46	3.85	6.98	0.373	9.51
5	336	6.44	3.01	5.82	4.03	8.60	0.388	15.4
6	65	7.83	2.25	7.51	6.38	9.25	3.08	12.6
7	589	5.70	2.55	5.40	3.76	7.14	0.373	15.4

Really crap standards:

Even worse… we triple dare you to take a screenshot and paste it into your assignment…

“We will shut that sh*t down! No exceptions”
[Quote: Negan (The Walking Dead, Season 6, Episode 16 [Last Day on Earth])

Any questions?