



Research Methods I

Quantitative Data

Stacey coleman

Lecture Outline

Descriptive statistics

- Mean, mode, median
- Range, variance, standard deviation

Parametric and non-parametric statistics

Why do we use numbers?



They add value
to what we are
saying...



Which of the
below sounds
better:

Many students at UWTSD
have a part time job

75% of students at UWTSD
have a part time job

Can we trust numbers?



Many people have strong opinions about numbers, especially statistics...



They don't "PROVE" anything...

Correct – they don't...
But as psychologists we don't want to PROVE anything, either



Numbers provide us with information, the value of which depends on what else you know and whether it suits your purpose



And numbers fit... where?

- When we examine numerical data, its relevance largely depends on what is being asked of your participants...
- Ask yourself:
 - Does this provide a useful insight?
 - Does it indicate outliers that may need further investigation?
 - Does it help me spot any trends?
 - Does it affect how you should think about a topic area?



Quantitative methods

In using quantitative measures, we then need to understand the techniques used to describe and analyse data:

- **Descriptive statistics:** used to summarise – for example, how many people took part, their gender, ages, measures of central tendency on data, etc.
- **Inferential statistics:** the use of statistical tests to analyse data



Experiments and studies

- **We measure a sample of a population on some factor**
 - Intelligence, cancer remission, face recognition; anything you can think of!
- **It is usually impractical to measure an entire population!**
 - Too big, too expensive, not really needed



Experiments and studies

We want to INFER from our sample to the wider population

- E.g., that measuring a tool which increases intelligence works on our sample
- That the tool will also increase intelligence in the wider population

Sample

To infer from our sample to the population, we need two things;

- 1) Accurate data from our sample group
- 2) A way to extrapolate from that group to the wider population

**Descriptive statistics
– the sample group**



**Inferential statistics –
statistical information about
the group and a link to the
wider population**



Descriptive Statistics



Descriptive statistics

- **Using the individual pieces of data from each participant**
 - Scores on individual tasks for each participant
 - *Number of questions answered correctly*
 - *Level of anxiety felt*
 - *Reported narcissism scores*
 - *Number of words remembered correctly*

Descriptive statistics

- **Produces 'raw' data**
 - And sometimes A LOT of it!
 - We need a way to reduce this to a manageable level





Conventions

✓Conventions in writing

✓Conventions in data presentation

- The principles are the same;
 - Set out in a standard format that everyone (with a little training) can clearly understand and interpret
 - Standardises information from the simplest to the most complex of datasets



Descriptive statistics: Presentation

Participants (in your method section)

- Number of, age range, number of males/females, average age of groups
 - This is standard for ANY study!

E.G., The sample used for this study consisted of 30 (10 male, 20 female) psychology students who ranged in age from 18 to 41 ($M = 21.50$, $SD = 5.60$)



Descriptive statistics: Presentation

Results

- Descriptive statistics of the dataset
 - Central tendency of the variable being measured (if/where applicable)
 - **Remember:** no measure of central tendency for categorical data!
 - Dispersion of scores around that mean
 - Variance and standard deviation from the average score (if applicable)



Measures of central tendency



Central tendency

People will score differently on almost ANY measure

- Individual differences; error; fatigue

If we give 100 people a memory test, some will score more or less than others. This could be because they have better or worse memory, are more or less tired, or have more or less things on their mind!

So, we need to find the ‘average’ of all the participants in the group

Central tendency

Three main ways to measure central tendency:

MODE

- The most commonly occurring score

MEDIAN

- The middle score when lined up in rank order (smallest to largest)

MEAN

- The sum of all scores divided by the number of participants

1. The mode

- Take all eligible responses (scores, reaction times, etc)

10, 5, 8, 12, 8, 3, 9, 18, 7

- The most commonly occurring score:

10, 5, 8, 12, 8, 3, 9, 18, 7

Mode = 8

1. The mode

- *Issues with the mode?*

10, 3, 8, 12, 8, 3, 9, 18, 7

Bi-modal!

Non-modal!

10, 3, 8, 12, 6, 2, 9, 18, 7

Only useful for certain types of data!

2. The median

- Take all eligible responses and put in rank/magnitude order

10, 5, 8, 12, 8, 3, 9, 18, 7

- Median is the middle number

3, 5, 7, 8, 8, 9, 10, 12, 18

Median = 8

2. The median

- *But what about even numbers of data?*

3, 5, 7, 8, 8, 9, 10, 12, 14, 18

- Add the numbers and divide by 2
- $(8+9) / 2 = 8.5$

Median = 8.5

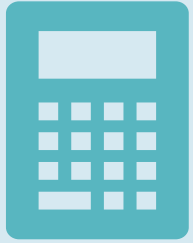
2. The median

- *Major problem with the median...*

1 1 1 2 5 6 8 9 18

- Takes no account of the other scores!
- Purely about the centre

3. The mean



The measure of arithmetic mean provides us with a powerful description of a large set of data



Considers every 'data point' within the sample

Calculated as the sum of all the values obtained from each case (participant) and then divide this number by the total number of cases

3. The mean

- Equation for calculating the mean

- Dataset: 12, 7, 8, 10, 2, 1, 9

- $$= \frac{12+7+8+10+2+1+9}{7}$$

- $$= \frac{49}{7}$$

$$= 7$$

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

3. The mean

- BUT – can be a weakness if there is a very extreme score (outlier) present
- E.g., 10 12 1 9 8 500
- (mean = 133)
- Larger samples reduce this problem by weakening the effect of each data point individually

3. The mean

- Summed total divided by the number within the sample (N)
- For a single participant (over several trials)
- For many participants (sample)
- We most often look for the mean of all participants scores in a group

- Imagine that we collect data for **one participant** across ten trials on two different shaped warning signs
 - We measure reaction times to the shapes displayed on a screen
- We can generate a mean score for each condition

Square	Circle
401	280
420	380
355	329
290	370
340	423
380	390
202	296
674	327
340	310
408	399

One row
=
one trial



$n = 10$



$n = 10$

- We can calculate the sum for each condition
- And then divide the summed total by the number of trails (*in this case, 10*)

Square	Circle
401	280
420	380
355	329
290	370
340	423
380	390
202	296
674	327
340	310
408	399

3810 / 10



Mean for
square = 381

3504 / 10



Mean for
circle = 350.4

Research and data

- These means relate to **one** of our participants
- We need to calculate the means for many participants...

How many??

- A representative sample...
- Just as we can have an extreme sample for a participant, we can also have an extreme participant in a sample

Research and data

- Once we have the means for each participants, we can put them together to find the 'grand means' for the conditions...
 - The mean of the means for each condition
- Aiming for the strongest measure possible
 - Measure taken many times across many participants

- This table now shows the mean scores for **ten participants** on the two tasks
 - Mean reaction times to square and circle warnings
- We can now sum the scores for each condition and find the means by dividing by the sample size (10)

M(square)	M(circle)
381	350
290	390
340	287
360	365
382	378
298	420
402	297
384	332
342	389
389	364

3568 / 10



Mean for
square = 356.8

3572 / 10



Mean for
circle = 357.2

HOW CAN WE TELL IF THE DIFFERENCE BETWEEN THESE MEANS IS STATISTICALLY SIGNIFICANT?

I.E., THESE RESULTS ARE NOT DUE TO RANDOM
FLUCTUATION, ERROR, OR CHANCE?



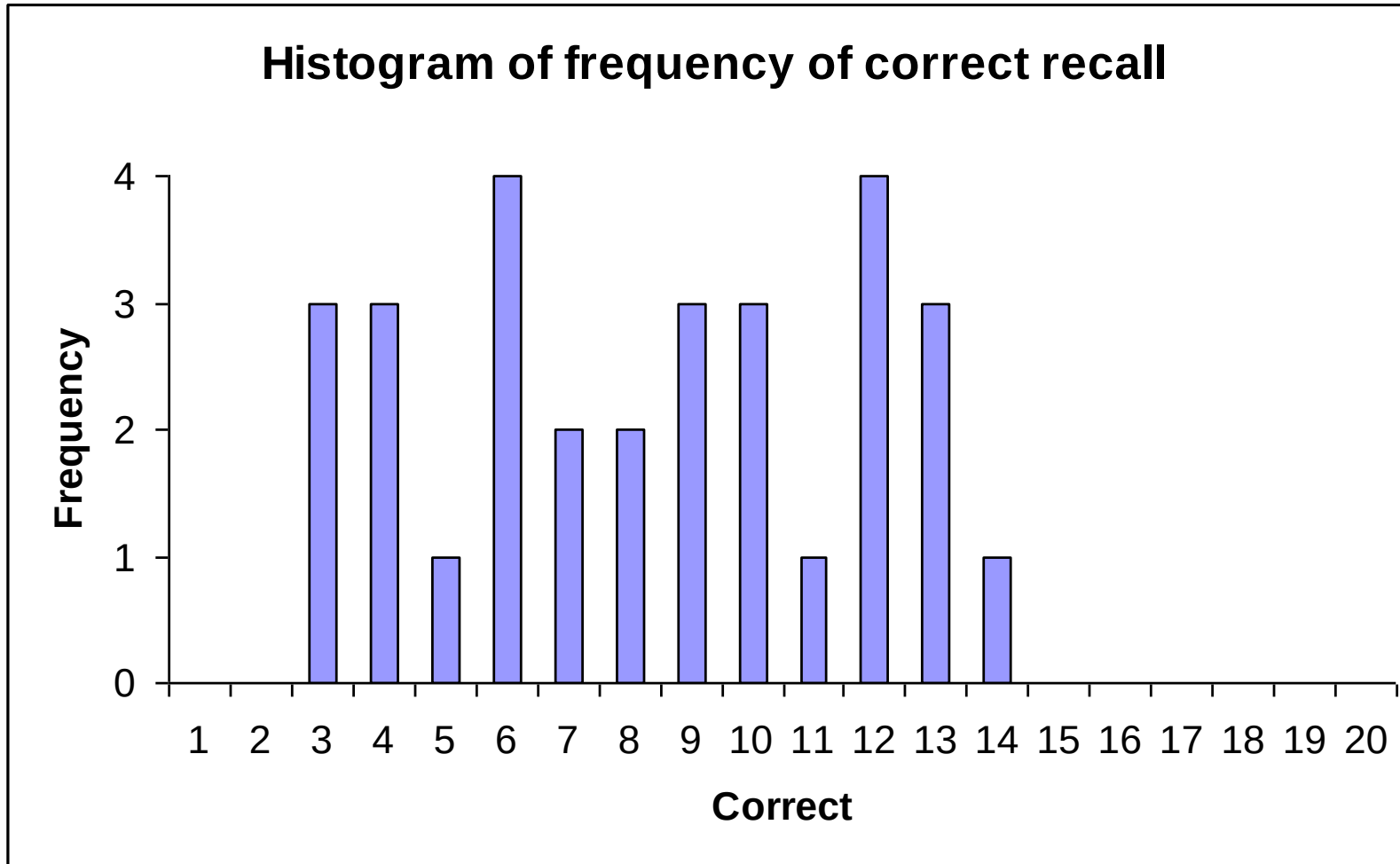
Calculating statistical inference from descriptive statistics

- Just knowing the mean of a group is not enough
 - Group A: $M = 11$
 - Group B: $M = 10$
- Is the difference between these two groups down to chance?
Or is it a reliable effect of something?
 - How many people have we tested?
 - What is the variation of the scores in the groups?

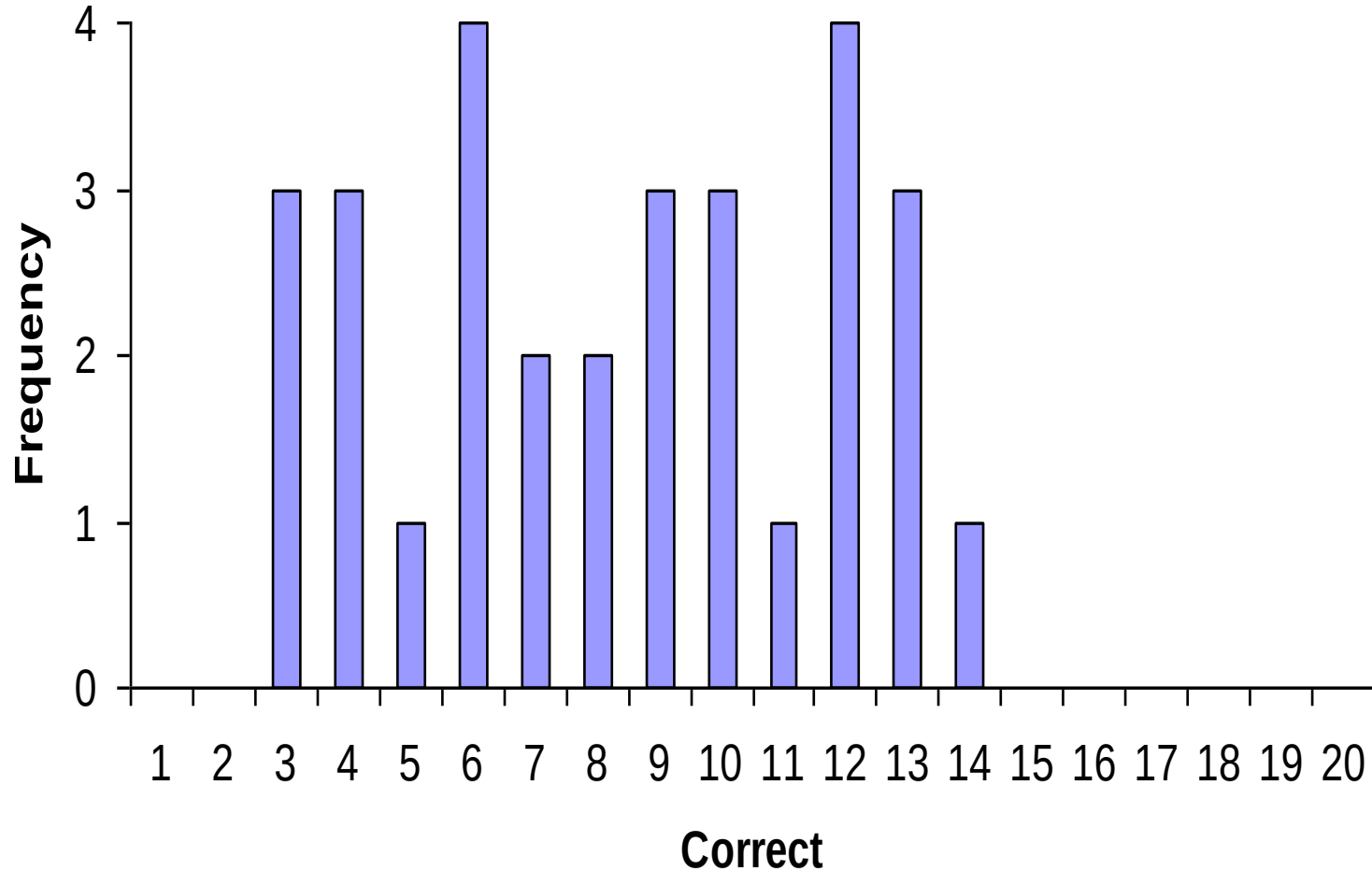
Calculating statistical inference from descriptive statistics

- First, descriptives can consider the 'frequencies' of certain scores
 - How many times each score occurs in a sample
 - Can represent this graphically by plotting out the 'frequency' for each score in a **histogram**

Histogram



Histogram of frequency of correct recall



What does this histogram show us?

1. The frequency for each individual score (how many)
2. How symmetrical the data is 'skew'
3. Any values markedly different / extreme scores

Calculating statistical inference from descriptive statistics

- By plotting histograms, we show one important quality of our sample
- **Variability or ‘dispersion’**
 - How much variability there is between the individual scores in our sample

Calculating statistical inference from descriptive statistics

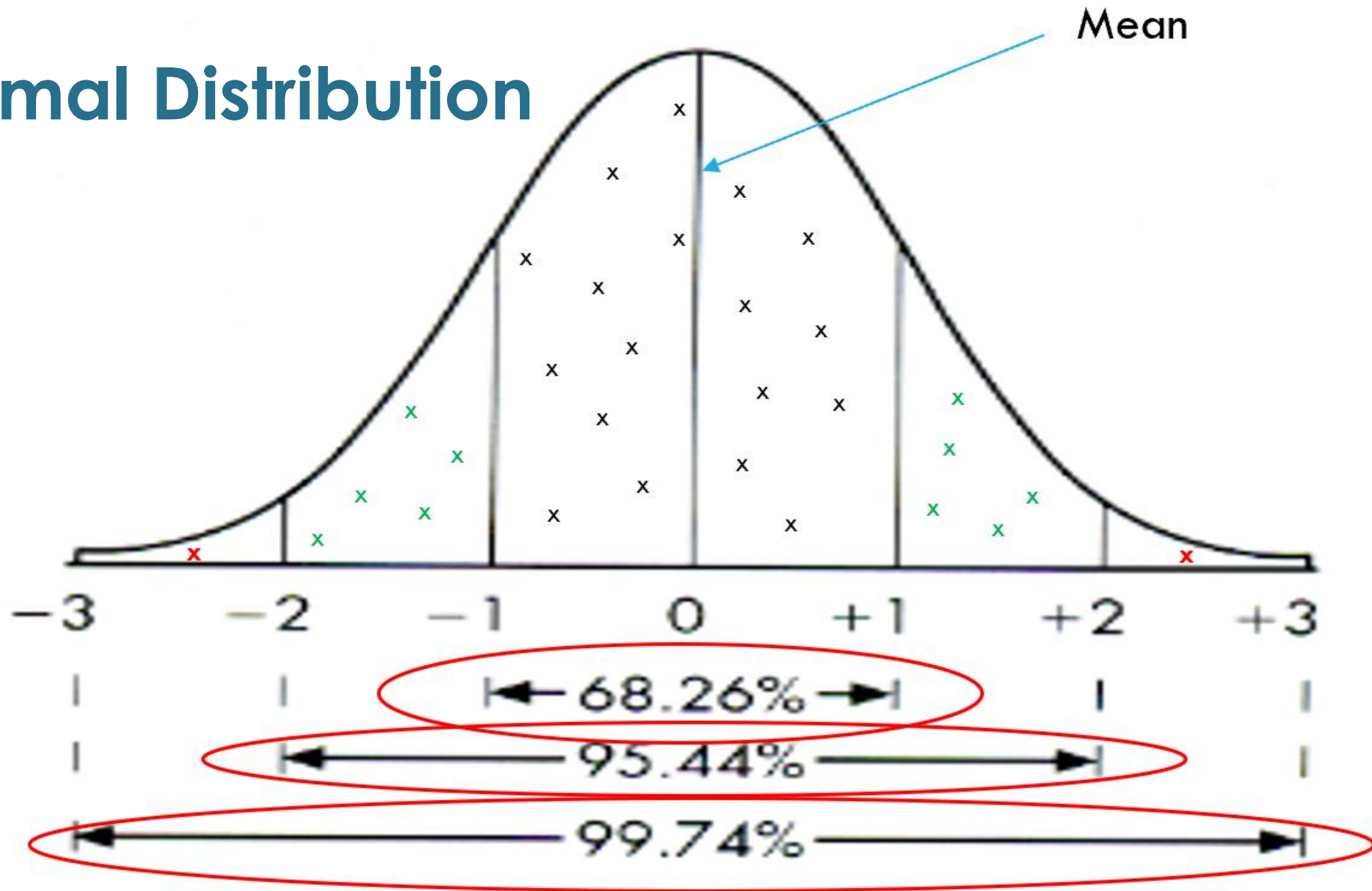
- **Why this this important?**
- Amount of dispersion tells us how reliable our mean scores are likely to be for a sample
- Will be used to judge any differences between samples in terms of significance

Skew

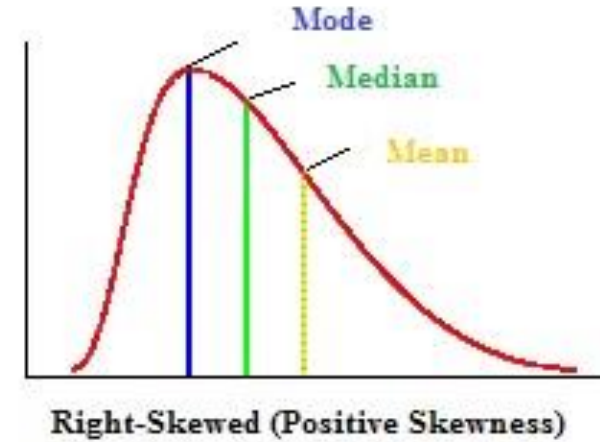
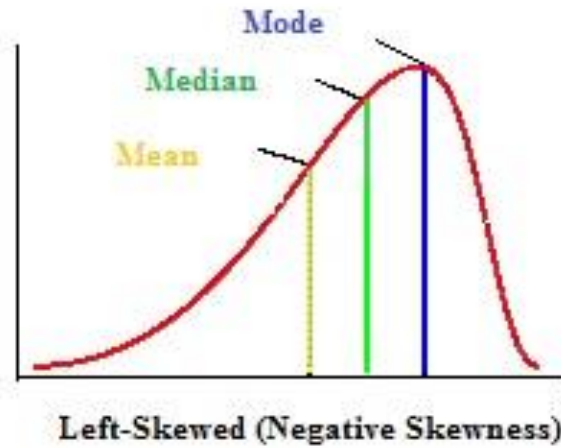


- Some distributions obtained from psychological measures that might be expected to be normal (and fit a **normal distribution**) in fact turn out to be **SKewed**.
- This means that they are 'lop-sided' – they have the peak (mode) to one side and a distinctive tail on the side where more extreme values occur

Normal Distribution

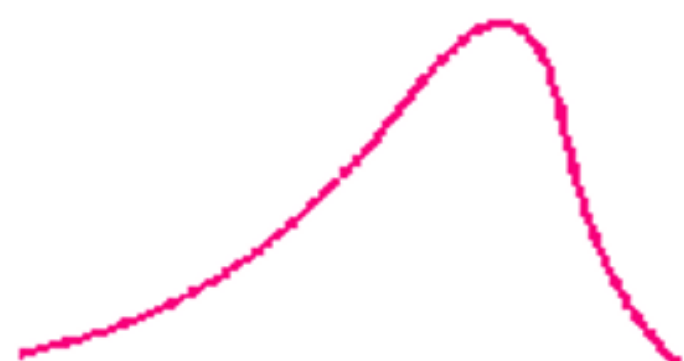


Greater number of extreme values below the median has the effect of shifting the mean towards the tail

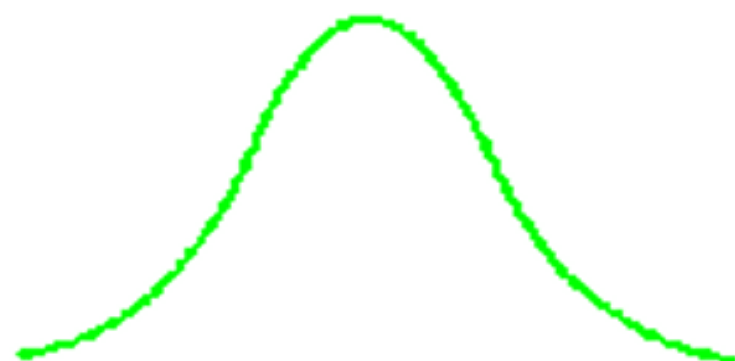


Greater number of extreme values above the median has the effect of shifting the mean towards the tail

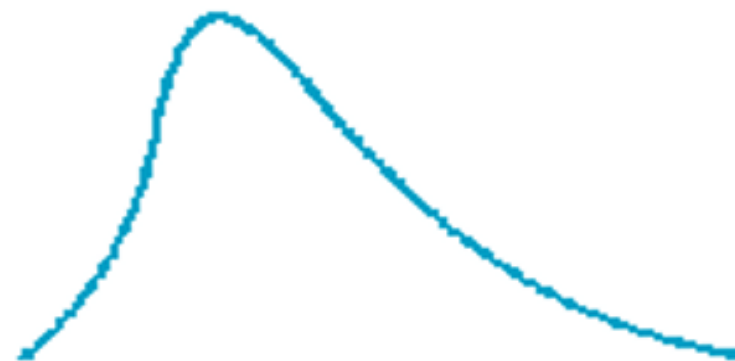
- Can occur when a test is relatively **EASY**
 - Produces 'ceiling effects'
 - *Can occur when a test isn't challenging enough – many Ps may have reached the highest score that can be achieved*
 - Better to use median or mode
- Can occur when a test is **hard**
 - Produces 'floor effects'
 - Participants produce mostly low scores, but some can still score well above the mean
 - Again, better to use median or mode as measure of central tendency



**Negatively (left)
skewed
distribution**



**Normal
skewed
distribution**



**Positively (right)
skewed
distribution**



Measures of variation and dispersion

Measures of variation and dispersion

1. RANGE

The difference between the lowest and the highest score

- Imagine two sets of data with the same mean
- Set A: 12, 14, 15, 16, 18 =
- **mean 15**
- Set B: 2, 3, 15, 27, 28 =
- **mean 15**

The **range** can distinguish between these two data sets

Set A: range is $18 - 12 = 6$

Set B: range is $28 - 2 = 26$

Set B contains more dispersion

Measures of variation and dispersion

2. VARIANCE

The degree to which all scores vary from the mean

Part 1: Compute how **each** score deviates from the mean by subtracting each score from the mean

E.g, Set C: 2, 14, 15, 16, 28

$M = 15$

$15 - 2 = \mathbf{13}$, $15 - 14 = \mathbf{1}$, $15 - 15 = \mathbf{0}$, $15 - 16 = \mathbf{-1}$, $15 - 28 = \mathbf{-13}$

Variance

- Part 2: Square the deviation scores and sum
 - Square them to get rid of the sign (negative scores)
 - E.g., $-13 \times -13 = 169$

$$169 + 1 + 0 + 1 + 169 = \mathbf{340}$$

- Divide the total by the number of scores MINUS ONE to get the variance

- $340 / 4 = \mathbf{85}$

$$13 \times 13 = 169$$

$$1 \times 1 = 1$$

$$0 \times 0 = 0$$

$$-1 \times -1 = 1$$

$$-13 \times -13 = 169$$

Number of data points
(N) = 5

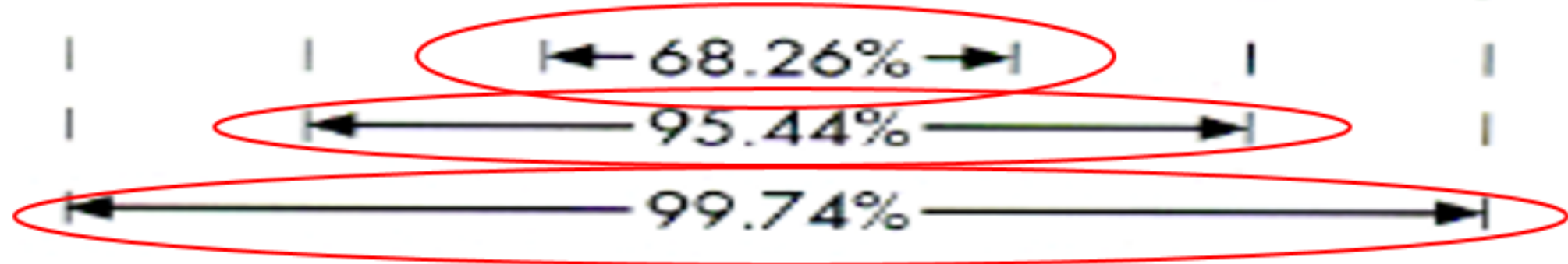
$$\text{So } N - 1 = 4$$

Measures of variation and dispersion

3. STANDARD DEVIATION

Square root of variance – produces a measure of dispersion which relates to the actual scores

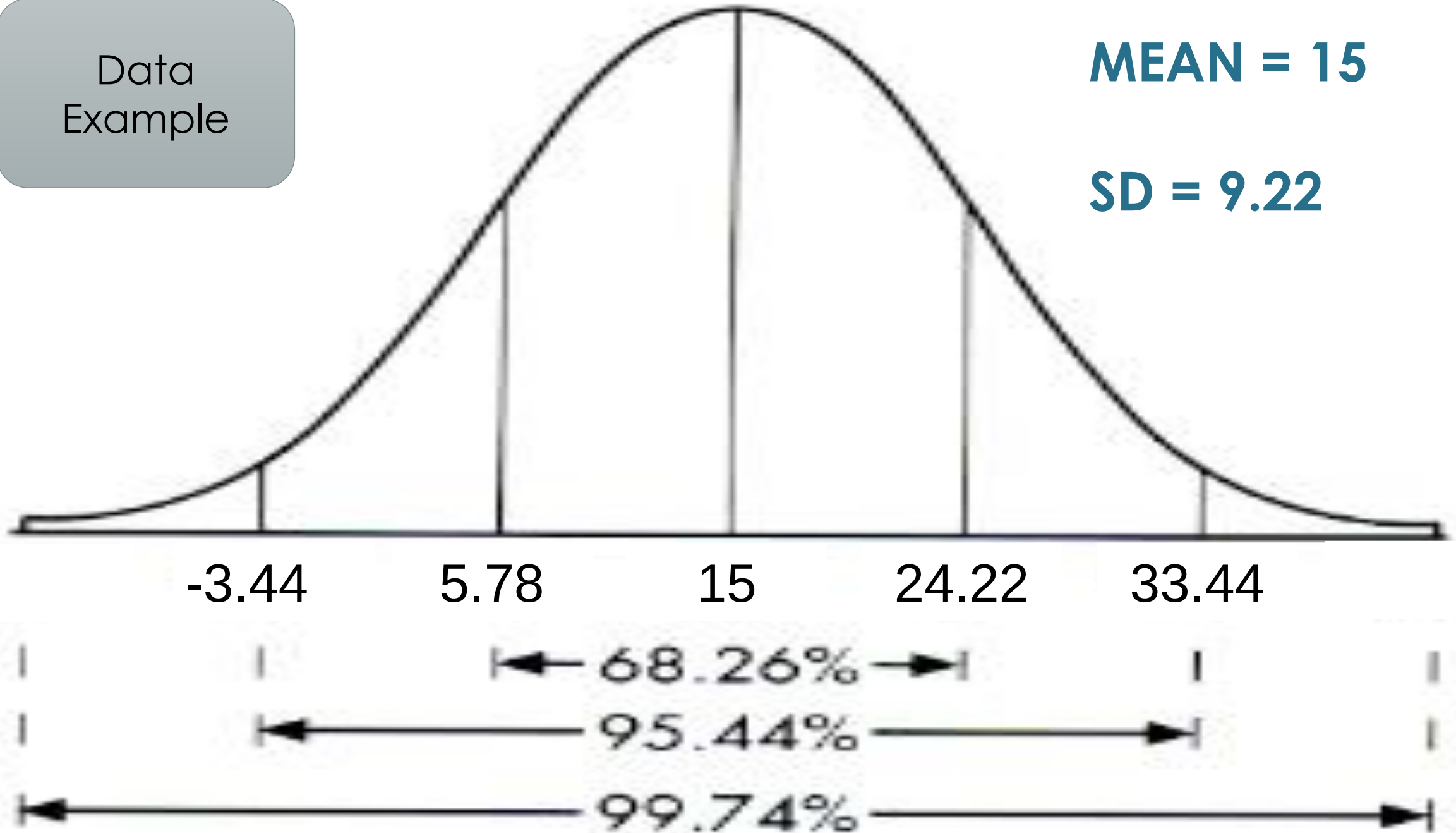
- Set C: 2, 14, 15, 16, 28
 - MEAN = 15
 - RANGE = 26
 - VARIANCE = 85
 - STANDARD DEVIATION (SD) = $\sqrt{85} = 9.22$
- **The standard deviation from the mean was 9.22**



Data
Example

MEAN = 15

SD = 9.22



Data
Example

MEAN = 15

SD = 9.22

More on normal distribution
and SD next week!!

-3.44

5.78

15

24.22

33.44

68.26%

95.44%

99.74%



Parametric and Non-Parametric Statistics

RECAP

Levels of measurement

Nominal

Ordinal

Interval

Ratio

Using Data

- Knowing the **type** of data you have can help you decide which statistical test you need to carry out to analyse any data
- It will also help you with *setting up* your data set prior to the analysis



Parametric and Non-Parametric Tests

**What's the difference
between parametric
and non-parametric
tests?**

**And why does it
matter?!**



Parametric vs. Non-Parametric

- Most research tends to use parametric tests to analyse data
 - Why?
- They are a much stronger and robust type of test
- If there **is** a difference/association, these tests can detect them easily
- But: parametric tests required a number of assumptions to be met if they are to be used in analyses...



Parametric Tests

Parametric vs. Non-Parametric

Parametric tests make assumptions about the parameters (defining properties) of the population distribution

- Normality / normal distribution

The observations must be independent

- E.g., P's need to have completed measure of DV separately, not in groups

Populations must have (roughly) same variances

- Homogeneity of variance

Parametric – the basis

1. Parametric tests are known as 'distribution tests'

- Need to assume data is normally distributed
- If not sure – can check for normality using histogram or check skewness/kurtosis
- Data analysis packages also test for parametric assumptions

Shapiro Wilk is a powerful test for normal distribution. It returns a test statistic called W and a p-value. If the p-value is larger than .05, we assume normal distribution. If it is smaller than .05, we do not assume normal distribution

Parametric – the basis

2. Parametric tests use numerical data

- Sounds obvious – all quantitative tests involve numerical data...
- This means data collected on a continuous scale - **interval or ratio level of measurement**
 - Original numerical scores

Parametric – the basis

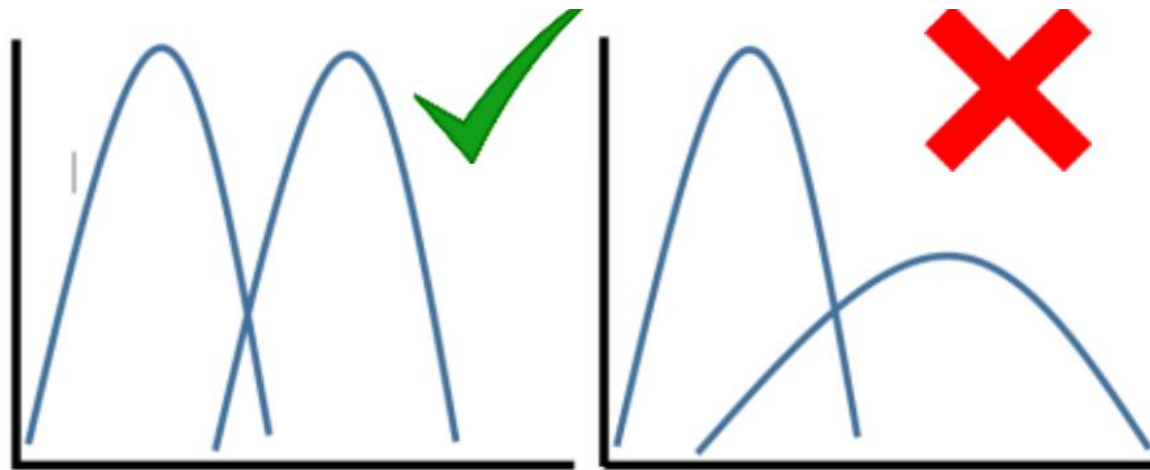
3. Large populations samples

- When analysing data using SPSS, a number of the assumptions will be checked as part of the process
- **But – you must aware of minimum sample/data points when designing studies**
 - E.g., some complex analyses (multiple regression) require minimum of 15 data points per variable.

Parametric – the basis

3. Homogeneity of Variance

- This refers to the expectation that population variances are all equal



Levene's test uses an F-test to test the null hypothesis that the variance is equal across groups. A p value less than .05 indicates a violation of the assumption. If this occurs, conducting a non-parametric equivalent of the analysis is more appropriate.

Parametric Tests



T-test
(independent and paired samples)

Exploration of differences between two groups/conditions



Pearson's correlation

Relationship between two variables



Analysis of Variance
(ANOVA)

Investigation of more than two levels of an IV and the effects of more than one IV
Can be within, between or mixed



Analysis of Covariance
(ANCOVA)

Looks at the effects of one or more IV's on a DV, whilst *partialling out* the effect of another variable (a covariate)



Multivariate Analysis of Covariance
(MANCOVA)

Includes more than one DV



Non-Parametric Tests

Non-Parametric Tests

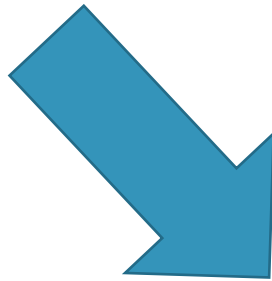
- Non-parametric tests **do not** rely on assumptions such as normality, and make no other assumptions about the data
- Non-parametric tests protect against some violations of assumptions and not others
- Employ ordinal data
 - (NB not chi-square)
- Obtained by ranking original scores

Non-Parametric Statistics

- Data are ranked from lowest to highest regardless of group
- Statistics are then conducted on ranks, not raw scores
- E.g.,

Non-Parametric Statistics

GROUP		
A	B	C
1	7	10
5	4	8
6	9	11



Rank values from
smallest to
highest

RANK		
A	B	C
1	5	8
3	2	6
4	7	9



Replace values
with new 'ranks'

1	2	3	4	5	6	7	8	9
1	4	5	6	7	8	9	10	11

Non-parametric – the basics

1. Mostly use the Median and do not assume anything about distribution

- Known as 'Distribution free' tests
- Are therefore used when the assumptions for conducting a parametric test are violated

Non-parametric – the basics

2. Mathematically, value is ignored

- The value of the numbers we get are more or less ignored
- Use of the median means the magnitude of differences are not compares
- Good for use with categorical data

Non-parametric – the basics

3. Data can be analysed according to rank

- NP methods widely used for studying populations that take on a ranked order
- NP methods may also be necessary when data have a ranking but no clear numerical interpretation
 - E.g., assessing preferences of colour

Non-parametric – the basics

3. Less powerful than parametric tests

- If you have the right data, always best to use a parametric technique where you can
- NP tests can be used more widely and are, in some cases, more applicable to the populations being studied
- Less fussy than parametric – which also makes them less sensitive or powerful

Non-parametric Tests



Wilcoxon
(2 conditions: paired samples)

Assess if population mean ranks differ



Mann-Whitney
(2 conditions: independent samples)

Rank of one group compared to another



Spearman rank

Correlation



Friedman
(>2 conditions: paired samples)

Analyses rank differences between multiple conditions



Kruskal-Wallis
(>2 conditions: independent samples)

Analyses rank differences between multiple groups)



Chi-square

Explores associations between categorical variables

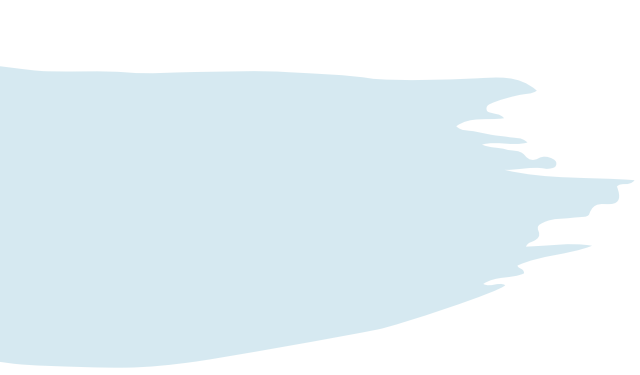
Non-parametric – Advantages

- Make no assumptions about data
- Can be used on small samples
- Generalise across samples drawn from populations with different distributions
- Less likely to lead to errors in outcomes



HOWEVER!

ADDITIONAL ASPECTS
TO CONSIDER

- 
- If sample size is too small for 'parametric stats' do not force the use of a parametric test
 - Do we really want to make decisions based on a sample that may be too small to detect any significance, anyway?
 - Sample size is important for stability!
 - Parametric stats are only more powerful when the assumptions upon which they are based are well met
 - If assumptions ARE violated, then NP stats are more powerful.

Which Test?

For Parametric Tests		
	Two Groups	> Two Groups
Independent samples	Independent samples t-test	Between-subjects ANOVA
Paired samples	Paired samples t-test	Repeated measures ANOVA

For Non-Parametric Tests		
	Two Groups	> Two Groups
Independent samples	Mann-Whitney	Kruskall-Wallis
Paired samples	Wilcoxon Signed Rank	Friedman Test



Statistical Analysis Decision Tree