



Research Methods I

Quantitative Data: Parametric and
Non-Parametric Analyses





Lecture Outline



- Levels of measurement
- Parametric and non-parametric statistics and tests
- Why is this differentiation important?



Types of Data and Levels of Measurement

Types of Data

- We need to consider the TYPE of data we are using in any study design before we can decide what to do with that data
- But... how do you decide?

Types of Data

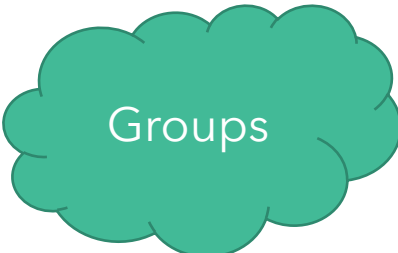
- There are two broad types of data:

CATEGORICAL

- E.g., gender, personality type
- Limited number of groups to which people can belong or can only belong to one group

CONTINUOUS

- E.G., Age, scores on a scale, seconds taken to complete a task
- Can take a vast range of values often from zero to an infinite number



Groups



Numbers

You can turn continuous data
into categorical but not vice
versa



Why does it matter?

- It will influence the quality of your results and the conclusions you are able to make
- It will also have an effect on the type of statistical analysis you are able to conduct
 - As you will learn!

Example

Categorical

- The results showed that a higher proportion of people in group B (60%) were anxious compared to those in group A (10%)

Continuous

Can use mean scores:

- The results showed that the mean anxiety score of participants in Group A was 2.3 and in Group B was 24.7 (measured on a scale of 0-30)
- Enables a more sophisticated option for statistical analysis

Types of data and levels of measurement

Scale type	Properties of scale
Nominal	Records differences A is red, B is blue
Ordinal	Records differences and magnitudes A is larger than B
Interval	Records differences and magnitudes, and has equal intervals A is two units larger than B
Ratio	Records differences and magnitudes, has equal intervals and a true zero A is twice as large as B

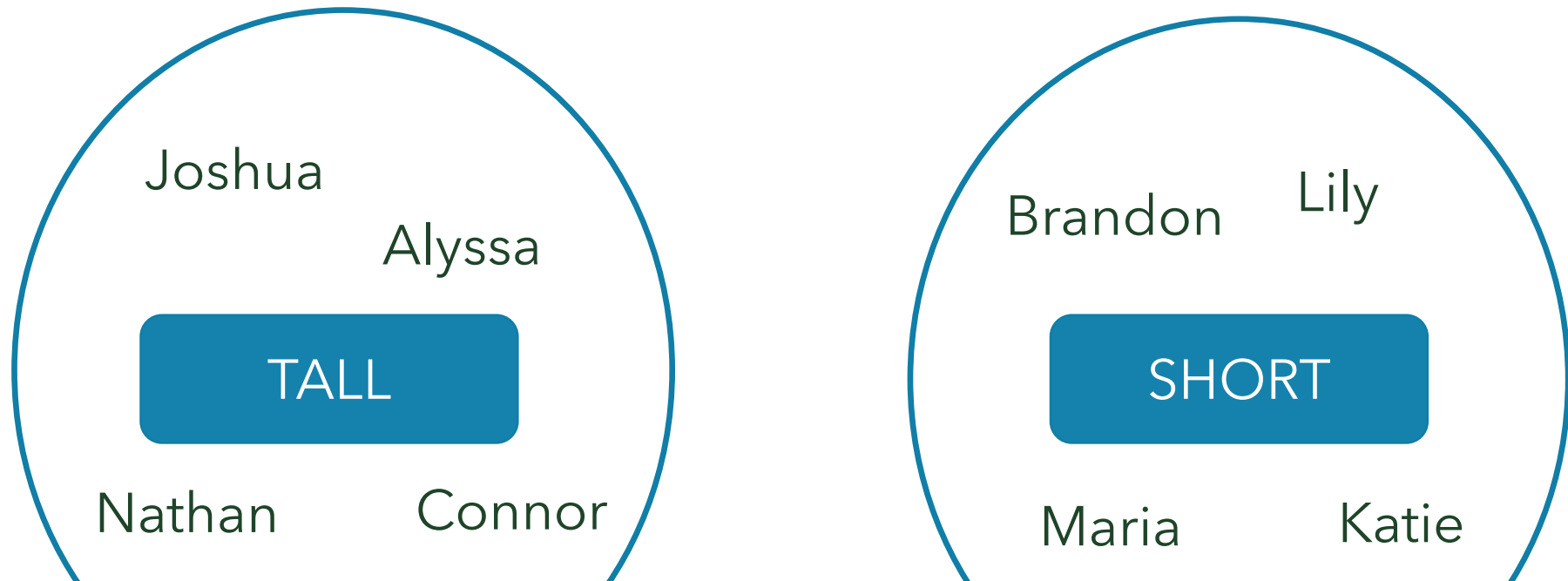
Nominal Data

- Only classifies people into groups;
- Refers to quality more than quantity
 - No 'numerical' significance - more 'labels or groups'

Example: organise a group of students into
'SHORT' and 'TALL'

Nominal Data

CATEGORICAL DATA



No distinction *at all* between the people in these two groups; i.e., no way of telling who is the tallest in the tall group, or who is the shortest in the short group

Types of data and levels of measurement

Scale type	Properties of scale
Nominal	Records differences A is red, B is blue
Ordinal	Records differences and magnitudes A is larger than B
Interval	Records differences and magnitudes, and has equal intervals A is two units larger than B
Ratio	Records differences and magnitudes, has equal intervals and a true zero A is twice as large as B

Ordinal Data

- Presents each piece of data in *order* of position
- Doesn't tell you what the relative *difference* is between these pieces of data
- Present *rank positions* in a group (i.e., 1st, 2nd, 3rd...)

Example: arrange students in height order from the tallest to the shortest

Ordinal Data

1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th
Connor	Nathan	Alyssa	Joshua	Brandon	Lily	Maria	Katie

Ordinal
classification
allows us to
see that
Nathan is taller
than Brandon

But it
doesn't tell
us the
difference in
sizes of
these
people

And that
Katie is the
shortest in
the group

Types of data and levels of measurement

Scale type	Properties of scale
Nominal	Records differences A is red, B is blue
Ordinal	Records differences and magnitudes A is larger than B
Interval	Records differences and magnitudes, and has equal intervals A is two units larger than B
Ratio	Records differences and magnitudes, has equal intervals and a true zero A is twice as large as B

Interval Data

- Numeric scales in which we know both the order and the exact differences between values
- Will give measured, real differences in height between people
- Interval data has a fixed size of interval between data points (e.g., degrees Celsius)
- But – no true zero

Interval Data

Connor	Nathan	Alyssa	Joshua	Brandon	Lily	Maria	Katie
180cm	176cm	170cm	169cm	166cm	158cm	54cm	152cm

Interval = fixed intervals between data
e.g., the difference between Connor's height (180cm) and Alyssa's height (170cm) is the same as the difference between Nathan's height (176cm) and Brandon's height (166cm)

Types of data and levels of measurement

Scale type	Properties of scale
Nominal	Records differences A is red, B is blue
Ordinal	Records differences and magnitudes A is larger than B
Interval	Records differences and magnitudes, and has equal intervals A is two units larger than B
Ratio	Records differences and magnitudes, has equal intervals and a true zero A is twice as large as B

Ratio Data

- A ratio scale has an *absolute zero* – a point where none of the quality being measured exists
 - E.g., mass; length; time; height
- Using a ratio scale permits comparisons such as twice as high, or one half as much

Ratio Data

Why is temperature (C/F) not ratio data?

'There is no such thing as 'no temperature'

20 degrees C, for example, is not twice as hot as 10 degrees C

If we convert C to F, it becomes clear:
 $10\text{C} = 50\text{F}$ and $20\text{C} = 68\text{F}$ - which is clearly NOT twice as hot

Ratio Data

- Ratio data scales tell us about the order, the exact value between units, and they have an absolute zero
- This allows for a wide range of both descriptive and inferential statistics to be applied

Ratio Data

E.G., WEIGHT

'Zero' means that something does not exist

A weight of zero does not exist, so weight is considered ratio data

Also permits comparisons such as 'twice as heavy', etc.

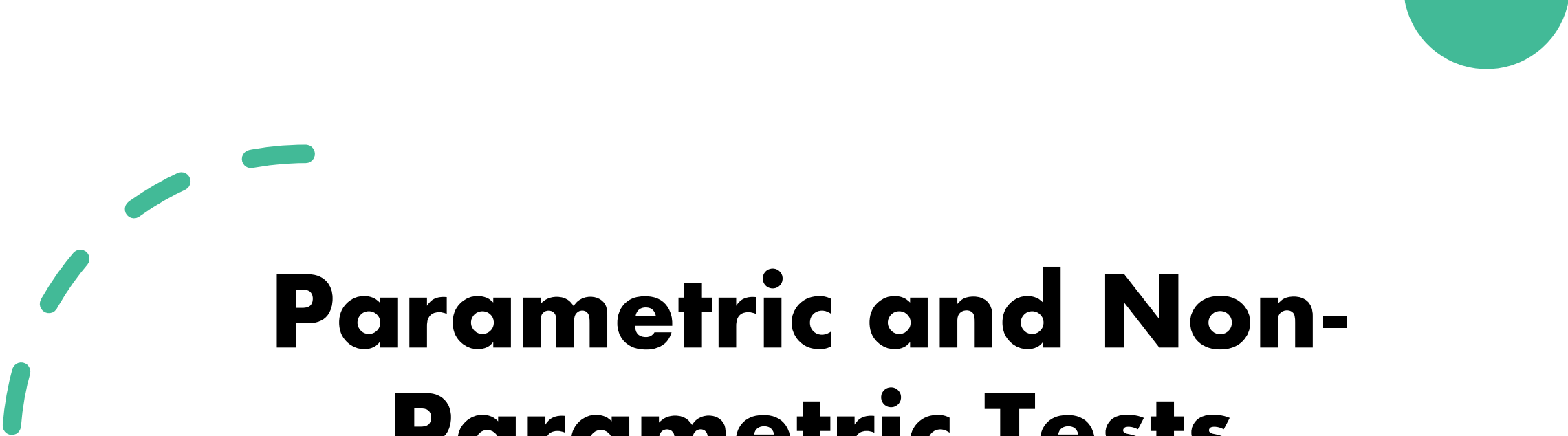




Using Data

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

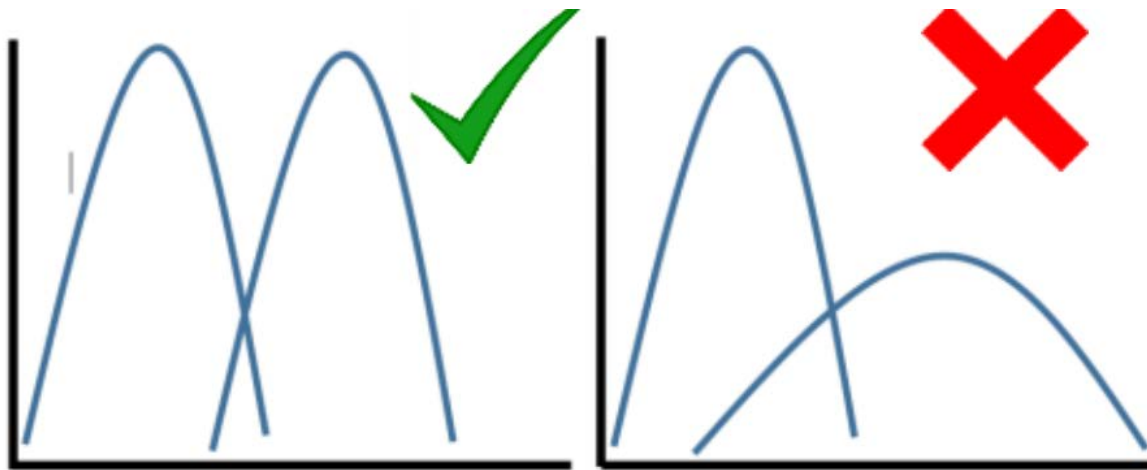
3. Large populations samples

- 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 basis

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 basis

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 basis

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 basis

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

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**