

Applied Research Design and Analysis

Lecture 1b:
Applied Research Methods,
Design, and Analysis

1

Lecture outline

- Descriptive statistics
 - Measures of central tendency:
 - Mean, mode, median
 - Variance and standard deviation
 - Normal distribution
- Statistical inference and hypothesis testing
- Levels of measurement
- Parametric vs. non-parametric tests
 - Why this differentiation is important

2



- We measure a sample of a population on some factor
- Intelligence, attitudes towards refugees, levels of cancer distress, anything you can think of!

Usually impractical to measure *entire* population!

- Too big, too expensive, not really needed

3

Applied Research

- We want to INFER from our sample to the wider population
- E.g., a coping intervention aimed at decreasing the distress experienced by patients waiting to hear the outcome of a cancer diagnosis in our sample
- Would the intervention also decrease distress experienced in the wider population

4

Descriptive statistics – the sample group

Sample
↓
Population

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

- To infer from our sample to the population, we need two things;
 - Accurate data from our sample
 - A way to extrapolate from that group to the wider population

5

Quantitative methods

- Treatment of the Data: Used to explain how the data was prepared for analysis
- 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

6

Descriptive Statistics

Using the individual pieces of data from each participant

- Scores on individual scale for each participant
- Number of questions answered correctly
- Level of distress felt
- Reported narcissism scores
- Number of images correctly recognised

7

Descriptive Statistics

◦ Produces 'raw' data

- And in applied research usually A LOT of it!
- We need a way to reduce this to a manageable level

8

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

9

Descriptive Statistics: Presentation

Participants

Number of participants, age range, number of males/females, average age of groups

- This is standard for ANY study and typically described in the method section of a report!

However, with Applied Research, demographic information on participants is sometimes included in the results section.

- For further understanding sub-analyses
- Understanding info on drop-outs

10

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 the mean
 - Variance and standard deviation from the average score (if applicable)

11

Central Tendency

People will score differently on almost ANY measure

- Individual differences; error; fatigue

12

Central Tendency

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

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

13

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

14

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

15

The Mode

Issues with the mode?

Bimodal!

Non-modal!

Only useful for certain types of data!

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

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

16

The Median

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

- Put in rank/magnitude order

3 5 7 8 8 9 10 12 18

Central tendency is the middle number

Median = 8

17

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 two

$(8+9) / 2 = 8.5$

Median = 8.5

18

The Median

- Major problem with the median...

1 1 1 2 5 7 8 9 18

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

19

The Mean

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

- Takes into account each individual 'data point' within the sample
- To obtain the arithmetic mean you add together all the values of the variables from each case and then divide this total by the number of cases

20

The Mean

Equation for calculating the mean:

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

Thus, if you had the following numbers

4, 2, 5, 1, 7, 10, 6

$$\begin{aligned} \bar{x} &= \frac{\sum x}{n} \\ &= \frac{4+2+5+1+7+10+6}{7} \\ &= \frac{35}{7} \\ &= 5 \end{aligned}$$

21

BUT – can be a weakness if a very extreme score (outlier) is present

E.g., 10 12 1 9 8 500

(mean = 133)

- Can often occur in applied research
- Larger samples reduce this problem by weakening the effect of each data point individually.

22

Summed total divided by the number within the sample (N)

- For a single participant (over several trials)
- For a number of participants (sample)
- We most often look for the mean of all of participants scores in a group

23

ONE TRIAL

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

↓ ↓

$n = 10$ $n = 10$

- Imagine that we collect some data for **one participant** as follows:
 - Reaction times to shapes displayed on a screen
- We can generate our mean scores for each condition

24

- Calculate the sum for each condition
- Divide each sum by the number of trials (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

25

Research and Data

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

26

Research and Data

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
- The effect of extreme participants is also reduced by a larger sample

27

Research and Data

- Once we have the means for each participant, 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

28

ONE PARTICIPANT

- This box shows the mean scores for ten participants on the two tasks
- We can now sum the scores for each condition and find the means by dividing by the sample size (10)

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

29

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?

30

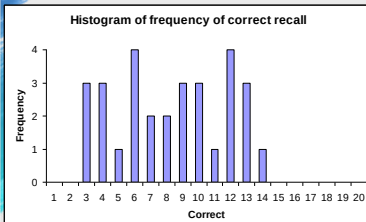
Calculating statistical inference from descriptive statistics

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'

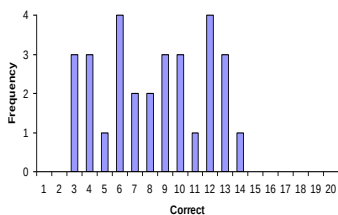
31

Histogram



32

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

33

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

34

Calculating statistical inference from descriptive statistics

Why is 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

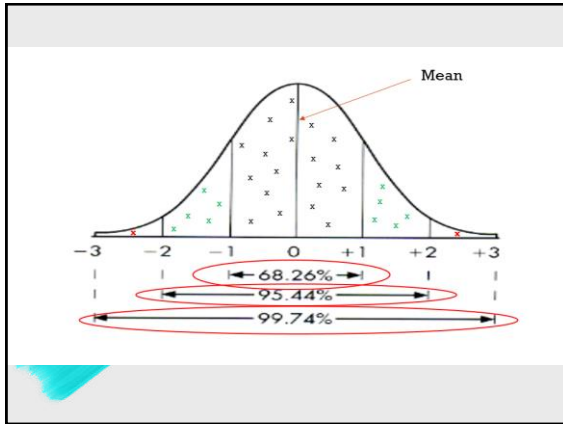
35

Skew

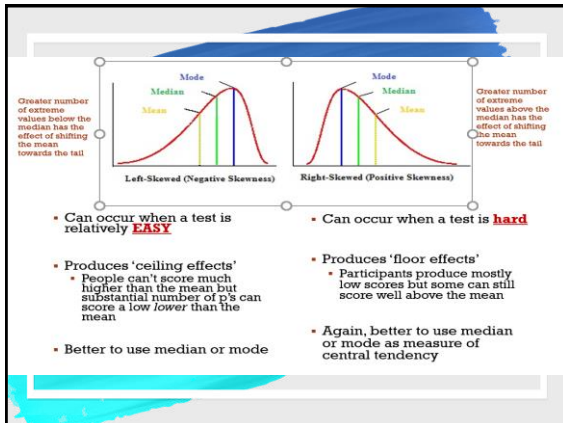
- Some distributions obtained from psychological measures that might be expected to be normal in fact turn out to be **SKEWED**.

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
- This occurs frequently in data from applied research

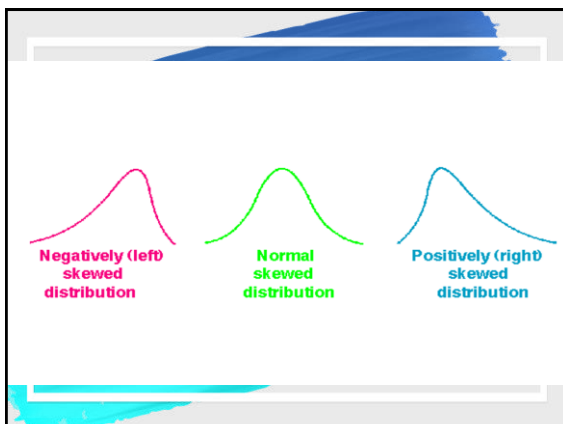
36



37



38



39

Measures of Variation and Dispersion

40

1. RANGE Difference between lowest and 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

41

1. RANGE Difference between lowest and highest score

- 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

42

**2.
VARIANCE**
Degree to
which all
scores vary
from the
mean

Compute how **each** score deviates from the mean by subtracting the score from the mean

e.g., Set C: 2, 15, 28, 14, 16
Mean = 15

$$\begin{array}{ll} 15-2 = 13, & 15-14 = 1, \\ 15-15 = 0 & 15-16 = -1 \\ 15-28 = -13 \end{array}$$

43

**2.
VARIANCE**
Degree to
which all
scores
vary from
the mean

Square each deviation score and add together

- Square to eliminate sign (negative scores)
E.g., $-13 \times -13 = 169$

$$13^2 + 0^2 + -13^2 + 1^2 + 1^2 =$$

$$169 + 0 + 169 + 1 + 1 = 340$$

44

**2.
VARIANCE**
Degree to
which all
scores vary
from the
mean

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

(there were five scores in total, so $5-1 = 4$)

$$340 / 4 = 85$$

45

3. STANDARD DEVIATION

Square
root of
variance

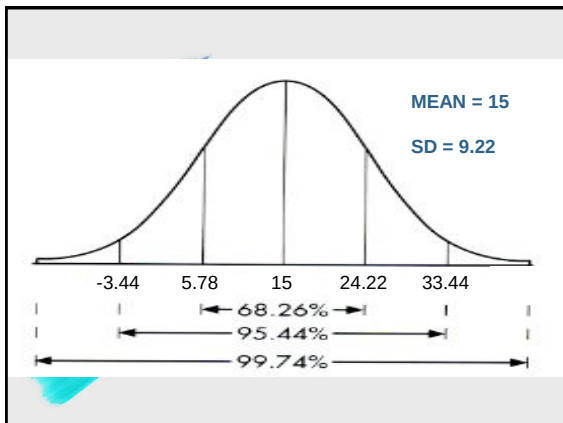
The standard Deviation (SD) produces a measure of dispersion which relates to the actual scores

Set C: 2, 14, 15, 16, 28

- MEAN = 15
- RANGE = 26
- VARIANCE = 85
- STANDARD DEVIATION = $\sqrt{85} = 9.22$

The SD from the mean is 9.22

46



47

Means

- Just knowing the mean of a group is not enough
- Group A: $M = 11$
- Group B: $M = 9$

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?

48

Statistical Inference and Hypothesis Testing

49

Statistical Inference

We have seen how we can start to test hypotheses by looking at differences in groups of data (e.g., means)

- Aim to generate a measure of probability that our results were purely 'chance' results
- A measure of 'significance'
- A measure of 'meaningfulness'

50

Hypothesis Testing

- Research Hypothesis vs. Null hypothesis
- We aim to calculate a probability value (or a p-value) that represents:
 - How likely it is that we would find out a set of results, if those results were purely down to 'chance' (i.e., if the null hypothesis were true)

51

Hypothesis Testing

The p-value represents this as a % chance

P-values are written as values between 0 and 1

- So, a 10% chance would be 0.10

The level we accept a result as significant is below a 5% chance (0.05)

$p < 0.05$ is a significant result

$p > 0.05$ is a non-significant result

52

Statistical Inference

- Essentially we are saying we are at least 95% confident that the outcomes from our data are due to our study manipulation

- But there is up to a 5% chance that the outcomes are due to error or are a fluke.

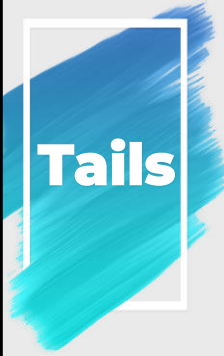
- After running statistical analyses on our data, if the p value is less than 0.05 ($p < 0.05$) we can accept our study hypothesis and reject the null

- If the p value is greater than 0.05 ($p > 0.05$), then we reject the study hypothesis and accept the null (i.e., no difference/relation)

53

Tails and Errors

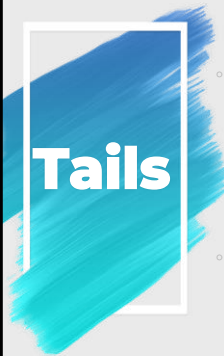
54



ONE-TAILED:

- Allocating all of your alpha (sig value) to testing the statistical significance in one direction of interest.
- Basically, you make a **directional** prediction about your results!

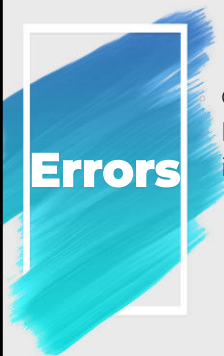
55



TWO-TAILED:

- Allocating half of your alpha (sig value) to test the statistical significance in one direction, and half of your alpha to test the statistical significant in the other direction.
- Basically, you are testing for the possibility of the relationship in both directions.

56

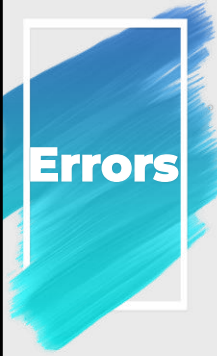


Type I Error

Concluding that the null hypothesis is false when it is actually true

- *(Thinking we have found something significant when it is not)*

57



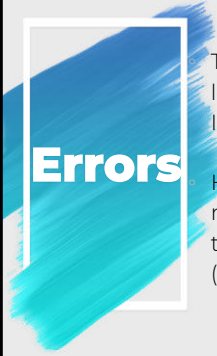
Errors

Type II Error

Concluding that the null hypothesis is true when it is actually false

- *(Not realising a significant difference is actually there)*

58



Errors

The more power in our data, the less likely we are to have a Type II error;

- Hence the desire to use the most powerful test possible for the data that we have (parametric tests where we can)

59

Types of data and Levels of measurement



60

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?



61

Types of data: Two broad types

- **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



62

**You can turn
continuous data
into categorical
but not vice
versa**



63

Why does it matter?

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

64

Example

- *Categorical*
 - 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:
 - Results showed the mean anxiety score of participants in Group A was 2.3 and in Group B it was 24.7 (measured on scale of 0-30)
 - Enables more sophisticated option for statistical analysis

65

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

66

Nominal Data

- Only classifies people into groups
- Refers to quality more than quantity

Example to Consider: organising a group of students into 'TALL' and 'SHORT'



67

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

68

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

69

Ordinal Data

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



70

Ordinal (tallest to shortest)

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

71

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

72

Interval Data

- Gives measured, real differences in height between people
- Interval data has a fixed size of interval between data points (e.g., degrees Centigrade)

73

Interval Data

Connor	Nathan	Alyssa	Joshua	Brandon	Lily	Maria	Katie
180cm	176cm	170cm	169cm	166cm	158cm	154cm	152cm

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

74

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

75

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

76

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

77

Parametric and Non- parametric tests

78

What is the difference between parametric and non-parametric tests?

And why does it matter?!



79

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 will detect them easily
- **HOWEVER:**
 - Parametric tests require a number of assumptions to be met if they are to be used in analyses.....

80

Parametric Tests

- Parametric tests make assumptions about the parameters (defining properties) of the population distribution
 - Normality / normal distribution of data
- The observations must be independent
 - E.g., Participants need to have completed measure of DV separately, not in groups
- Populations must have (roughly) same variances (Homogeneity of variance)

81

Parametric – the basics

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 P-Value is larger than 0.05, we assume normal distribution. If P-Value is smaller than 0.05, we do not assume normal distribution



82

Parametric – the basics

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**
 - They use the original numerical scores

83

Parametric – the basics

3. Large population samples

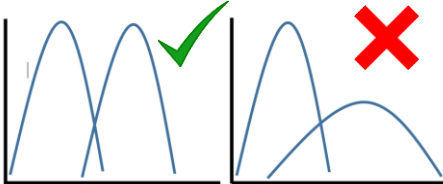
- When analysing data in SPSS, a number of assumptions will be checked as part of the process
- **But – you must be aware of minimum sample/data points when designing studies**
 - E.g., some complex analyses (multiple regression) require a minimum of 15 data points per variable.

84

Parametric – the basics

4. Homogeneity of Variance

- 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.

85

Parametric tests

- T-test (within and between)
 - Differences between two variables
- Pearson's r Correlation
 - Relationship between two variables
- Analysis of Variance (ANOVA)
 - Can investigate more than two levels of an IV and the effects of more than one IV
 - Within, Between, or Mixed
- Analysis of Covariance (ANCOVA)
 - Looks at effect of one or more IVs on DV, whilst *partialling out* effect of another variable – *Covariate*
- Multivariate Analysis of Variance (MANOVA)
 - Includes more than one DV

86

Non-parametric tests

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

87

Nonparametric Statistics

- Data ranked from lowest to highest
 - Regardless of group
- Statistics then conducted on ranks, not raw scores.

Group			Rank		
A	B	C	A	B	C
1	7	10	1	5	8
5	4	8	3	2	6
6	9	11	4	7	9

When scores are equal, they are both given a mean rank.
E.g., if 2 scores are tied and would have ranks '3' and '4', both are ranked at 3.5, but if there were three tied scores, that would have ranks '7', '8', and '9', then they would all be ranked at 8.

88

Non-parametric – the basics

1. Mostly use the Median and do not assume anything about the distribution of the data
 - Known as 'Distribution free' tests
 - Thus, are used when the assumptions for running a parametric test are broken

89

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 compared
 - Good for use with categorical data

90

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

91

Non-parametric – the basics

4. 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

92

Non-parametric tests

- Wilcoxon (2 conditions: related)
 - Assess if population mean ranks differ
- Mann-Whitney (2 conditions: unrelated)
 - Rank of one group compared to another
- Friedman (>2 conditions: related)
 - Analyses rank differences between multiple conditions
- Kruskal-Wallis (>2 conditions: unrelated)
 - Analyses rank differences between multiple groups
- Spearman rank (correlation)
- Chi-square (categories)
- Trends predicted:
 - Page's L (related)
 - Jonckheere (unrelated)

93

Non-parametric – Advantages

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

94

HOWEVER!
Some additional aspects to
consider

95

- 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.

96

Analysing Data – A decision chart

For Parametric Tests		
	Two Groups	> Two Groups
Between Groups	Independent t-test	Between-subjects ANOVA
Within Groups	Repeated-measures t-test	Within-subjects ANOVA
For Non-Parametric Tests		
	Two Groups	> Two Groups
Between Groups	Mann-Whitney	Kruskal-Wallis
Within Groups	Wilcoxon signed ranks	Friedman

97

Data Distribution in Applied Research

- Data obtained from Applied research in areas such as health, education, and social science in general typically yield values of...
 - Skewness (distorted curve)
 - Kurtosis (the sharpness of the peak of a frequency-distribution curve)
- I.e., they deviate from normal distribution.
- Think about why this might be the case.

98

Some key points.....

- Some studies include more than one type of analysis that include both simple and complex analysis
- May have multiple hypotheses
- May need to do different things with data at different stages
- E.g. Randomised controlled trial could have:
 - Main outcome analysis of ANCOVA design
 - Multiple regression to test validity of theoretical model used
 - T-tests to check differential drop-out/demographic differences between groups
 - Correlations to check assumptions for other analyses
- May need to run certain analyses as an alternative due to assumptions of desired analyses being broken.

99
