

RESEARCH METHODS

Lecture 4: Organising Data

Stacey Coleman



RECAP: IV'S AND DV'S

• Independent variables:

- Variable which the experimenter manipulates in order to determine the effect on a given behaviour (i.e., the effect on the chosen dependent variable)
- NOT influenced by any other variable in your study



RECAP: IV'S AND DV'S

- Simple IV's:
 - Groups of participants
 - E.g., age, gender
 - Psychological constructs
 - E.g., personality type
- Also!
 - Task, timing, environmental conditions, questions used...



RECAP: IV'S AND DV'S

• Dependent variable:

- The factor/variable that is assumed to be directly affected by changes in the IV in an experiment
- The dependent variable DEPENDS UPON the independent variable



STUDY EXAMPLES

- Physically punished children are more aggressive



- Memory deterioration can be the result of stress at work



RECAP: IV'S AND DV'S

- In these examples...
- Levels of aggression depends upon the level of physical punishment
- Memory deterioration depends upon the effect of stress at work



LECTURE OUTLINE

- Types of data
 - Continuous and categorical
- Descriptive statistics
 - Mean, mode, median
 - Variance & standard deviation
 - Normal distribution
- Intro to statistical inference and hypothesis testing



TYPES OF DATA

TYPES OF DATA

- We need to consider the **TYPE** of data we are using in any study design
- But... how do you decide

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

Groups



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

Numbers

Types of Data

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



WHY DOES IT MATTER??

▪ So 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 <i>A is red, B is blue</i>
Ordinal	Records differences and magnitudes <i>A is larger than B</i>
Interval	Records differences and magnitudes, and has equal intervals <i>A is two units larger than B</i>
Ratio	Records differences and magnitudes, has equal intervals and a true zero <i>A is twice as large as B</i>



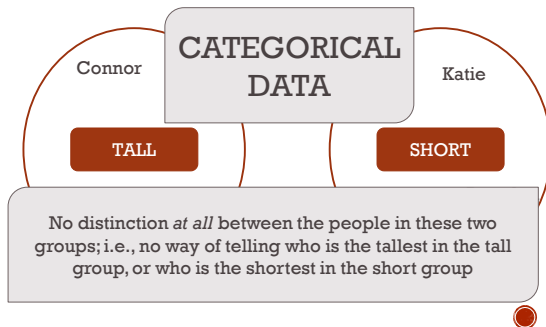
NOMINAL DATA

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

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



NOMINAL



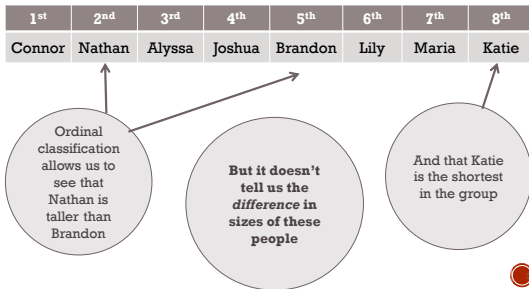
TYPES OF DATA AND LEVELS OF MEASUREMENT

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

ORDINAL

- 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...)

ORDINAL (TALLEST TO SHORTEST)



TYPES OF DATA AND LEVELS OF MEASUREMENT

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

INTERVAL

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

INTERVAL

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)

TYPES OF DATA AND LEVELS OF MEASUREMENT

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

RATIO

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

USING DATA

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



DESCRIPTIVE STATISTICS



EXPERIMENTS AND STUDIES

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

- To do this, 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

- Using the individual pieces of data from each participant

- Scores on individual tasks for each participant
 - Number of questions answered correctly on IQ test
 - Level of anxiety felt
 - Reported narcissism scores
- 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, 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 ($M = 21.50, SD = 8.6$).



DESCRIPTIVE STATISTICS - PRESENTATION

- Results
 - Descriptive statistics of the dataset
 - Central tendency of the variable being measured (if/where applicable)
 - Remember: you don't have a measure of central tendency for categorical data!
 - Dispersion of scores around that mean
 - Variance and standard deviation from the average score (if applicable)



CENTRAL TENDENCY??

- People will score differently on almost ANY measure
 - Individual differences; error
- If I test a group of 100 people for their IQ, some will score more or less than others. This could be because they are more or less intelligent, more or less tired, or have more or less things playing on their mind!
- So, we need to find the 'average' of all the participants in the group



CENTRAL TENDENCY

- There are three main ways to measure central tendency:

1) **MODE**

- The most commonly occurring score

2) **MEDIAN**

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

3) **MEAN**

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



THE MODE

- Take all eligible responses (scores on a test, 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



ISSUES WITH THE MODE?

10 3 8 12 8 3 9 18 7

Bimodal

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

Only useful for
certain types of data!

Non-
modal



THE MEDIAN

▪ 10 5 8 12 8 3 9 18 7

▪ Put in rank order;

▪ 3 5 7 8 8 9 10 12 18



Median=
8

▪ Central tendency is the middle number



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



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



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
 - However, can be a weakness if a very extreme score is 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

THE MEAN

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

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

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

- 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



RESEARCH AND DATA

- These means relate to **one** of our participants
- We need to calculate the means for a number of different participants
- How many??
 - A representative sample
 - Just as we can have extreme sample for a participant, we can also have extreme participants in a sample
 - A participant who was bored/distracted/drunk!
 - The effect of an extreme participant is also reduced by a larger sample



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



ONE PARTICIPANT

- This box shows the mean scores for **ten participants** on the two tasks

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., that these results are not due to random fluctuation, error, or chance?!

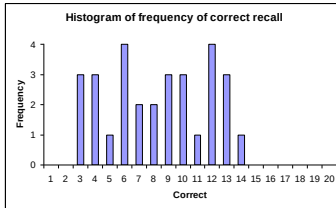


CALCULATING STATISTICAL INFERENCE FROM DESCRIPTIVE STATISTICS

- 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



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 these histograms we show one important quality of our sample
- The variability or 'dispersion' of our sample
 - How much variability is there between the individual scores in our sample
- 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



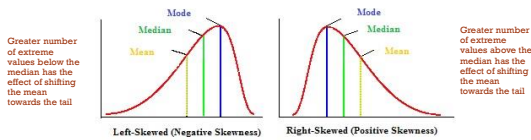
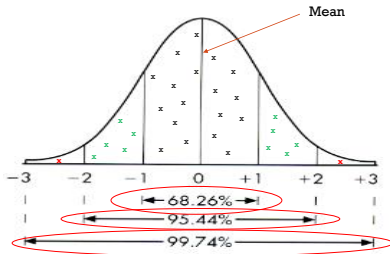
SKEW??



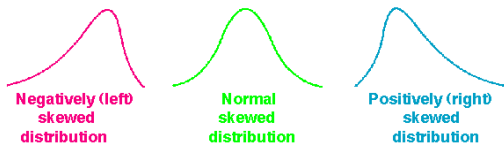
- Some distributions obtained from psychological measures that might be expected to be normal in fact turn out to be **SKEWED**.
- This means that they are 'lop-sided' – they have the peak (mode) to one side and a distinctive tail on the side where more extreme values occur



SO WHAT IS NORMAL DISTRIBUTION?



- Can occur when a test is relatively **EASY**
- Produces 'ceiling effects'
 - People can't score much higher than the mean but substantial number of p's can score a low lower than the mean
- 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



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

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

E.g, Set C

$15 - 2 = 13$, $15 - 14 = 1$, $15 - 15 = 0$, $15 - 16 = -1$, $15 - 28 = -13$



VARIANCE

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

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

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

$$340 / 4 = 85$$



MEASURES OF VARIATION AND DISPERSION

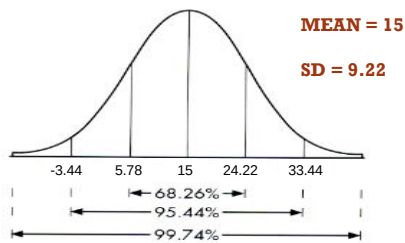
3. STANDARD DEVIATION

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 = $\sqrt{85} = 9.22$
- The standard deviation from the mean was 9.22



NORMAL DISTRIBUTION



- This principle allows us to start looking at different things

- Test 100 people on intelligence (IQ test) and reaction times to answer questions (seconds)
- Could look at correlation but no causation
- What if I have a theory that intelligence causes faster reactions?
- Have two different measurement types
 - But can turn them into z-scores and it doesn't matter



- 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?
- Depends on a number of issues
 - How many people have we tested?
 - What is the variation of the scores in the group?



STATISTICAL INFERENCE AND HYPOTHESIS TESTING



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' for our results
 - A measure of 'meaningfulness' for our results



HYPOTHESIS TESTING

- Experimental Hypothesis (H_1) – the way I expect my data to differ between the experimental and control group
 - or the relationship I expect to see between my data if correlational
 - My prediction, based upon the literature/prior experience
- Null hypothesis (H_0) – recognition that there may be no differences/relationships between my data
 - We tend not to report this explicitly in our hypothesis
 - Because it is (implicitly) always there



HYPOTHESIS TESTING

- We aim to calculate a probability value (or a p-value) that represents:
 - How likely it is that we would find out set of results, if those results were purely down to 'chance' (i.e., if the null hypothesis were true)
- The p-value represents this as a % chance
- P-values are written as values between 0 and 1



HYPOTHESIS TESTING

- So, a 50% chance would be written as 0.50
- The level we accept a result as significant is below a 5% chance
- i.e., **0.05**

$p < 0.05$ is a significant result
 $p > .05$ is a non-significant result



STATISTICAL INFERENCE

- What is the probability value representing as 'significant' in a test of difference between sets of data?
 - That the scores are significantly different
 - But also, that the groups tested are not from the same global 'population'
 - To do this we need to use certain 'descriptive' statistics for each groups data
 - This can depend upon the test that we are using



FINALLY – TAILS AND ERRORS



TAILS

- We can have one-tailed and two-tailed tests
- **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!



TAILS

• 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.



ERRORS

• 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)

• TYPE II ERROR

- Concluding that the null hypothesis is true when it is actually false (not realising that a significant difference is there)
- 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)



WORKSHOP THIS WEEK

- Introduction to SPSS

