Introduction to quantitative methods

Cara Featherstone

Overview

- Objective
  - By the end of this session, you should be able to carry out a t-test and a correlation, and know when to use which

- To do that, we will
  - Define classical test theory
  - Define different sources of variance
  - Have a go at a t-test
  - Have a go at a correlation

Before we start

- Any questions from last time?

Just to get you thinking...

- What type of question requires a quantitative (number-based) approach?
- OR how could a quantitative approach be used to tackle your research question?

Why use quantitative methods?

- Observed score = true score + error
  - Error = random variability

- So...
  - “Weight” = real weight + imprecision of scales
  - “Ability to play piano” = real ability to play + nerves induced by exam/experiment situation
  - “Attitudes towards a minority group” = real attitudes + willingness to disclose attitude

Why use quantitative methods?

- Aim of good research
  - Reliable and valid studies
- Aim of good design
  - Minimise “error” as much as possible
    - check the weighing scales are accurate
- Aim of statistical analysis
  - Determine how much variance is “unexplained” compared to how much is explained by our model

Types of data

- Nominal
  - Category, e.g. male or female, musician or non-musician...
- Ordinal
  - In order, but not separated by equal intervals (1st, 2nd and 3rd in a race)
- Interval
  - The intervals between them are equal but there is no absolute zero (can be negatives) e.g. Celsius, IQ (?), personality measures...
- Ratio
  - There is an absolute zero, which means zero. Time, distance, number of publications, years of musical training...

Generalisability

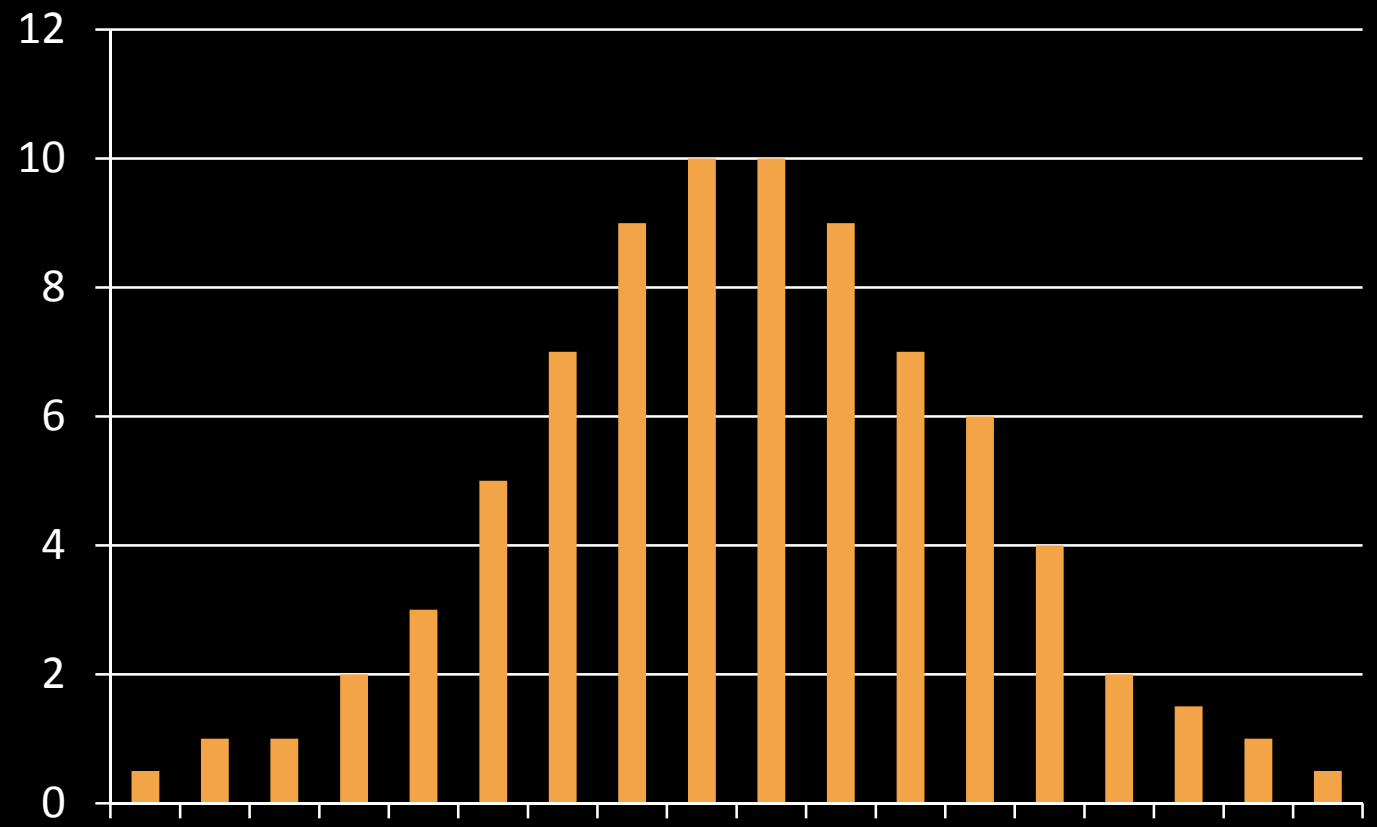
- How generalisable is the study's outcome?
  - People are different
  - Situations are different
  - Can't test all peoples in all situations!

  - Statistical analyses come in handy here

  - Test a representative sample of the population
  - Carry out some "inferential statistics"

Normal distribution

- An assumption in most statistical analyses



Descriptive statistics

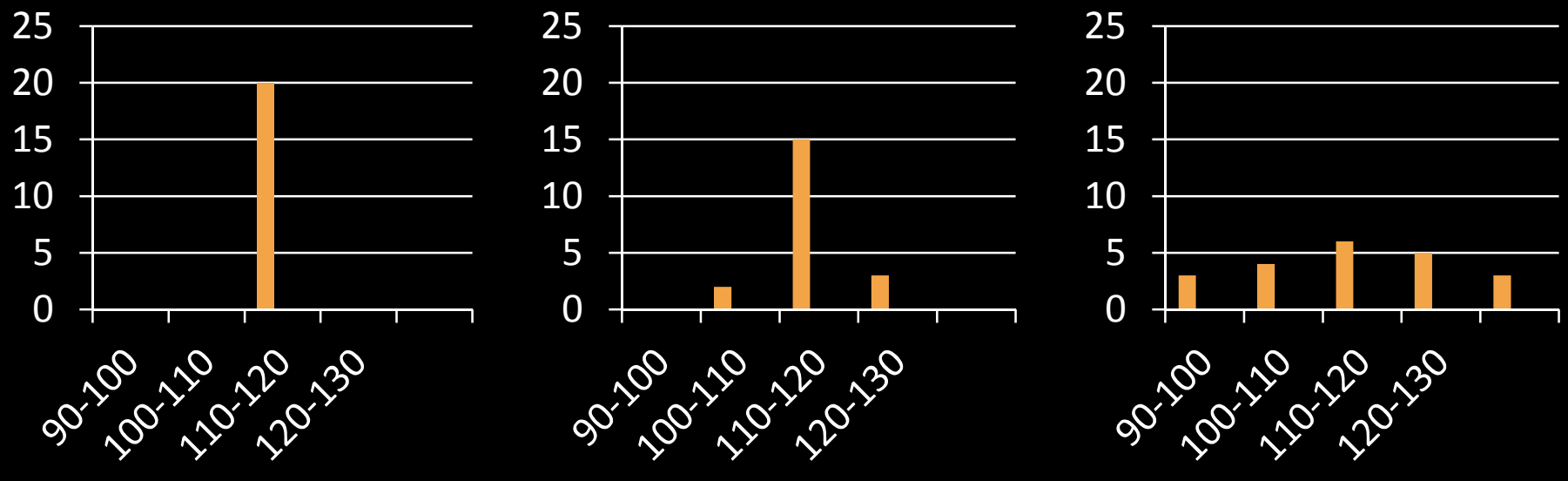
- The mean
  - A measure of “central tendency”
  - Also referred to as the “average”

  - Equation: Mean = sum of all entries / number of entries
  - Excel function: =AVERAGE(cell coordinates)

- 1, 2, 3, 4, 3, 2, 5, 4
  - Sum = 24
  - Number = 8
  - Mean = 3

Variability

- How much variation there is between scores (the spread of scores) - IQ in a classroom
  - No variability = all have the exact same score
  - Small amount of variability = all have roughly the same score
  - Large amount of variability= pupils' scores vary a lot



Measures of variability

- Range
  - Highest score – lowest score
    - 1 3 6 7 9, range = ?
    - 8
  - Problem: doesn't tell us much about what happens between the two extremes
    - 1 1 2 3 5 7 8 9 9, 1 1 1 1 1 9 9 9 9, 1 9 9 9 9 9 9 9 9: same range

- Variance
  - Sigma means Sum
  - s^2 is the variance (s is the Standard Deviation)
  - X is each entry
  - X with bar on top is the mean
  - N is the number of entries

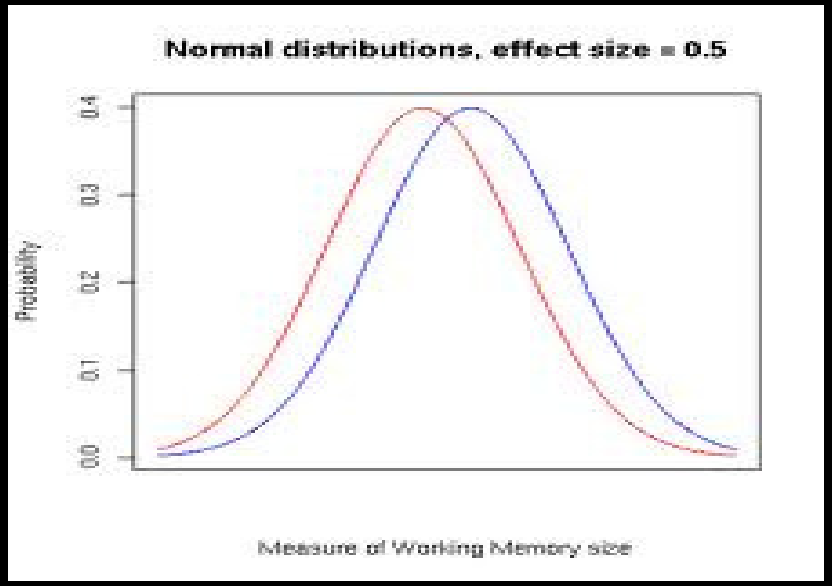
$$s^2 = \frac{\sum \left(X - \overline{X}\right)^2}{N}$$

Variance and standard deviation

- All you need to remember: Excel function STDEV
- Mathematical relationship to the normal distribution:
  - 95% of data points are within Mean +/- (1.96 x SD)
- Example: IQ scores
  - Should always be standardised
  - Mean = 100, SD = 16
  - What range of IQ scores comprises 95% of the population?
    - 68.64 to 131.36

Variance and standard deviation

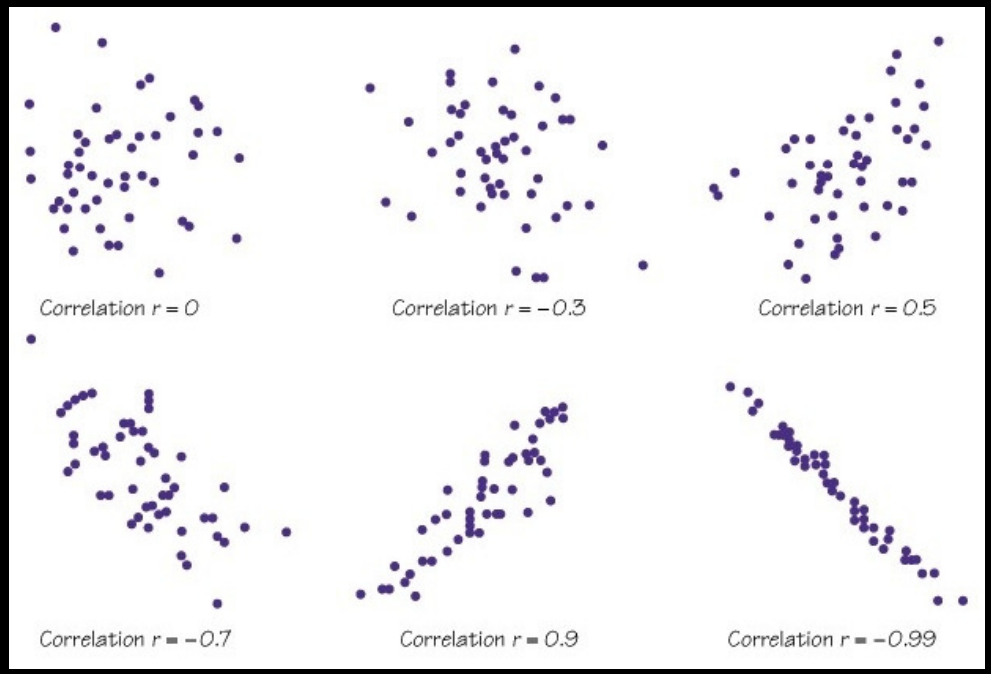
- Why is this of any interest?
  - “Musicians have better memory than non-musicians”*
  - Difference in means but real difference between groups?



- *Warning: Completely fictitious data used here!

Variance and standard deviation

- Why is this of any interest?
  - “Mathematical abilities correlate with musical abilities”*



- *Warning: Completely fictitious data used here!

Inferential statistics

- Help us determine to what extent the results we get could be obtained by chance

- Erroneous statistical reporting is unethical (dangerous?)

http://www.ted.com/talks/ben_goldacre_battling_bad_science.html

Inferential statistics

- The probability of the data patterns occurring by chance
- Groups have “significantly different” means if the probability of this difference happening by chance is less than $p = 0.05$.
- A correlation is “statistically significant” if the probability of the trend happening by chance is less than $p = 0.05$.

Where is this all going?

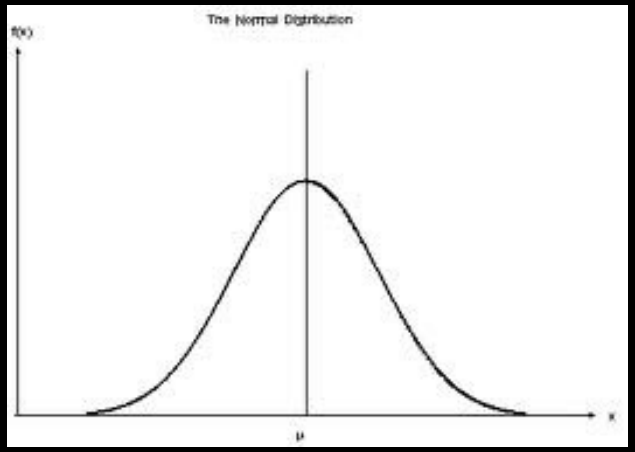
- Background – done!
- Next
  - Comparing means – t-tests
  - Comparing trends – correlations
- Before we do that: Break time!

Comparing two means: T-test

- Question 1:
  - Between groups or within group?
  - Musicians compared to non-musicians
    - Between-groups = independent groups
  - Classroom pupils before and after musical intervention
    - Within group = repeated measures (= paired groups)

Comparing two means: T-test

- Question 2: One-tailed or two-tailed?
  - "Musicians are more extroverted than non-musicians"
    - One-tailed
    - Based on expectations derived from the literature and/or prior studies
  - "There is a difference between musicians and non-musicians in extroversion"
    - Two-tailed
    - "Exploratory", when don't know what to expect



T-tests

- Predict that children's ability to tap along to a rhythm will improve after 6 months of musical training
  - Within (Repeated measures)
  - One-tailed

$$t = \frac{\overline{D} - 0}{\tilde{\sigma}_{\overline{D}}} = \frac{\overline{D}}{\frac{\sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}}}{\sqrt{N}}}$$

  - Hurrah for Excel

T-test, repeated measures

- Put scores from pre-intervention in one column
- Put scores from post-intervention in another column

- Insert Formula, and follow instructions
- =TTEST(
  - Array1 = first column of data (pre-intervention)
  - Array2 = second column of data (post-intervention)
  - Tails = 1 because we have a one-tailed hypothesis
  - Type = 1 (paired = repeated measures)

T-test, repeated measures

- See Excel file
- Look at descriptives, report significance of the difference

T-tests

- Difference between mathematical skills of musicians and non-musicians
  - Two-tailed
  - Between groups

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left[\frac{\left(\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N_1}\right) + \left(\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N_2}\right)}{N_1 + N_2 - 2}\right]\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

  - Once again, Hurrah for Excel!

T-test, independent groups

- Put scores from musicians in one column
- Put scores from non-musicians in another column

- Insert Formula, and follow instructions
- =TTEST(
  - Array1 = first column of data (musicians)
  - Array2 = second column of data (non-musicians)
  - Tails = 2 because we have a two-tailed hypothesis
  - Type = 2 (independent (assume equal variance))

T-test, independent groups

- See Excel file
- Look at descriptives, report significance of the difference

Correlation

- Measures how much two variables “covary”
  - Is change in one variable met with similar change in the other variable?
  - Musical skill and mathematical skill
  - Year of publication and frequency of word “neoclassical”
- Positive (when one goes up, the other goes up too), or Negative (when one goes up the other goes down)
- Always between -1 and 1

Correlation

- Warning: Correlation NOT EQUAL TO Causation
  - Bananas and male life expectancy

- Correlation coefficient = r

$$r = \frac{\mathrm{cov}_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y}$$

Correlation

- Hurrah for Excel, but not as much as with T-tests...
  - Gives us r, we then need to figure out if that means anything
  - R^2 = amount of variability in variable B that is explained by variability in variable A
    - Transform into a percentage to make it clearer
  - How much variability is accounted for if...
    - r = 0.4?
    - r = 0.1?
    - r = -0.2?

Correlation

- Hurrah for Excel, but not as much as with T-tests...
  - Gives us r, we then need to figure out if that means anything
  - To find out if the correlation is significant, we need to look up the value in significance tables
    - Handout
    - Find the right *df* (degrees of freedom = number of pairs – 2)
    - Find the probability you want (0.05)
    - Is your r bigger than the one at that point?

Correlation

- Is there a significant correlation between years of musical training and mathematical abilities?
  - Scores in maths test
  - Years of musical training

- Enter data into Excel, one column for each variable
  - =CORREL(
  - Array 1 = the first column
  - Array 2 = the second column
  - r is outputted → is it significant?

Which test?

- Length of name and time spent in the charts
- Effect of practicing piano scales on ability to play piano
- Men and women's propensity to be swayed by reviews
- Influence of age on quality of composition
- Ability to discriminate pitches in musicians and non-musicians
- Positive reviews and number of downloads...

Reporting statistics

- Make a figure of the data or a table (means and SDs)
- Comment on the figure or table
- Report the statistical analysis
  - Does it confirm your description?

- Look at published papers and copy their reporting style
  - The results showed that musicians had a higher mean score than non-musicians in a test of mathematical abilities (8.12 compared to 7.32), but the variability, as shown by the standard deviations was higher in musicians than in non-musicians (2.34 compared to 1.21). The results of the two-tailed independent groups t-test showed that this difference was not significant ($p > 0.05$). We therefore cannot conclude that musicians are significantly better at maths than non-musicians.

In this lecture

- Principles of quantitative analysis
  - Variance
  - T-tests
  - Correlations
  - When to use which test
- Practice at each analysis
- Matching analysis to research question

Between now and next time

- Think of how these statistical analyses could be used in relation to your research interests

- Play around with the Excel file (if you break it, just download a new copy of it)

Next time

- Qualitative approaches
- c.r.featherstone@leeds.ac.uk