Statistics by Hand
An Introductory Course for Psychologists

Relationships

Version 3.0

Lovely, lovely maths

- BEDMAS
  - Order of operations
    - **B**rackets first
    - **E**xponents next (e.g. 2^6)
    - Then **D**ivision and **M**ultiplication
    - Then **A**ddition and **S**ubtraction
  - 3 x 2 + 6 = 12, but 3 x (2 + 6) = 24
- Sigma (Σ)
  - X = [1 2 3]
  - ΣX = 6
- Multiplication is default operation
  - (5-2)(4-1) = 9

- Finger tapping!

Types of inference

- Do two groups of people differ **significantly** on a particular measure? Lecture 2
- Do a group of people differ **significantly** on a particular measure collected under two different conditions? Lecture 3
- Is there a **significant** relationship between two different measures of the same group of people? This lecture

New test: Contingency chi-square

For when you want to know whether two different variables are significantly related

and

- each participant contributes a two pieces of information (one for each variable) *or*
- a <u>single</u> participant contributes, on multiple occasions, two pieces of information (one for each variable)

Contingency chi-square: Prerequisites

- Each variable must be treated as **categorical** rather than **continuous**.

Continuous variable: A variable that can take any value, within certain bounds.

e.g. Height … 178.872cm, but not -26cm.

Categorical variable: A variable that takes only certain defined values – generally speaking, a smallish (2-6) set of values.

e.g. Gender: Male / Female (mainly).

- Note: Continuous variables can be categorized to produce categorical variables
  - e.g. A particular pre-school class could be classified as either 3-years old or 4-years old, on the basis of a continuous variable (e.g. date and time of birth)

Contingency chi-square: Pre-requisites

- The data must be expressed in the form of **counts**, not in the form of, for example, percentages, probabilities etc.
- For example:
  - 10 three-year olds, 10 four-year olds
  - 5 males; 15 females:
- **NOT**
  - 50% three-year olds; 50% four-year olds
  - P(Male) = 0.25; P(female) = 0.75
- If you have data in these latter forms, you must convert it to **counts** (using the sample size) before you can use it in a chi-square.

An example data set

	3 years	4 years
Co-operative	12	3
Competitve	6	27

- These are described as the **observed** values.
  - i.e. what you actually measured/observed in your study.
- It looks like age and play style are related. How can we investigate this?
- As in previous lectures, we set a **null hypothesis** and attempt to reject it.
- The null hypothesis is the age and play style are **independent**.

Null hypothesis: Independence

	3 years	4 years
Co-operative	12	3
Competitve	6	27

- If age and play style were independent, what would be expect this table to look like?
- Probability theory (see lecture 1), specifically the multiplicative law, gives us the answer.

Multiplicative Law

The probability of a set of independent events all happening is the product of the individual probabilities.

P(H)=0.5 P(H)=0.5

0.5 x 0.5 = 0.25

Expected values under independence

	3 years	4 years	**TOTAL**
Co-operative	12	3	**15**
Competitve	6	27	**33**
TOTAL	**18**	**30**	***48***

- P(co-operative) = 15/48 = 0.31
- P(3 years) = 18/48 = 0.38
- P(3 years + co-operative) = 0.31 x 0.38 = 0.117
- No. of co-operative 3 year olds expected = 0.1172 x 48 = 5.625

Chi-square contingency

	3 years	4 years	TOTAL			3 years	4 years
Co-operative	12	3	**15**		Co-operative	5.625	9.375
Competitve	6	27	**33**		Competitve	12.375	20.625
TOTAL	**18**	**30**	***48***				

Short-cut:

E = (Row total x Column total) / Grand total

E = (15 x 18) / 48 = 5.625

This is what the table would look like if age and play-style were independent.

Chi-square calculation

- Basically, this expresses the extent to which the **observed** values differ from the values **expected** under the null hypothesis.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

O = Observed

E = Expected

Chi-square calculation

	3 years	4 years			
Co-operative	12	3	15	5.625	9.375
Competitve	6	27	33	12.375	20.625
	18	30	48		

$$\chi^2 = (12 - 5.625)^2 / 5.625 + (3 - 9.375)^2 / 9.375 + (6-12.375)^2 / 12.375 + (27-20.625)^2 / 20.625$$

$$\chi^2 = 7.225 + 4.335 + 3.284 + 1.970 = 16.8$$

Significance of χ^2

- Use a chi-square table.
- Unlike Wilcoxon, result is significant if your chi-square value **exceeds** the value in the table.
- d.f. = (rows-1)(columns-1) = (2-1)(2-1) = 1
- Chi-square is *multi-tailed*
  - There are many directions in which data could differ from the expecteds.
  - Chi-square tables are appropriate for all common uses.
- χ^2 = 16.8, d.f. = 1. Result is significant.

Caveat!

- The calculation of the chi-square table assumes that **all expected values** are at least 5.
- If this is not true, the test for significance may be invalid.
- However, generally speaking $E < 5$ leads to a loss of power, rather than increases the chances of incorrectly concluding the relationship exists.

Process

	3 years	4 years	
Co-operative	12	3	15
Competitve	6	27	33
	18	30	48

Process

	3 years	4 years	
Co-operative	12	3	15
Competitve	6	27	33
	18	30	48

5.625	9.375
12.375	20.625

E = (15 x 18) / 48 = 5.625

E = (15 x 30) / 48 = 9.375

E = (33 x 18) / 48 = 12.375

E = (33 x 30) / 48 = 20.625

Chi-square calculation

	3 years	4 years			
Co-operative	12	3	15	5.625	9.375
Competitve	6	27	33	12.375	20.625
	18	30	48		

$$\chi^2 = (12 - 5.625)^2 / 5.625 + (3 - 9.375)^2 / 9.375 + (6-12.375)^2 / 12.375 + (27-20.625)^2 / 20.625$$

$$\chi^2 = 7.225 + 4.335 + 3.284 + 1.970 = 16.8$$