Statistics by Hand

An Introductory Course for Psychologists

Statistical Inference

Version 3.0

Examples…

- Attempt problems PCB-1 and PCB-2 in your booklet (p.16-17)
- Remember:
  - The probability of a set of independent events all happening is the product of the individual probabilities.

$$C_r^N = \frac{N!}{r!(N-r)!}$$

Statistical inference

- Last week, we discussed *mean*, *median,* and *distributions.*
- These are **descriptive** statistics.
- The rest of the course is largely concerned with **inferential** statistics.
- Inferential statistics help us decide whether the patterns we see in descriptive statistics are true of the **population**, or whether they are an illusion created by the particular **sample** we have collected.

The problem of inference

"Women have a larger vocabulary than men",
says leading scientist.

- Clearly, this scientist has not tested the whole **population.**
- She or he has tested a **sample** of the population.

Example experiment

- Ask for definitions of a list of words

  Bed, Ship, Penny, Winter, Repair, Breakfast, Fabric, Slice, Assemble, Conceal, Enormous, Hasten, Sentence, Regulate, Commence, Ponder, Cavern, Designate, Domestic, Consume, Terminate, Obstruct, Remorse, Sanctuary, Matchless, Reluctant, Calamity, Fortitude, Tranquil, Edifice, Compassion, Tangible, Perimeter, Audacious, Ominous, Tirade, Encumber, Plagiarise, Impale, Travesty

- Collect scores from a *random sample* of men and women.

Data

	1	2	3	4	5	6
Women	34	29	40	16	38	31
Men	32	35	24	18	12	22

Averages (Means)

	1	2	3	4	5	6
Women	34	29	40	16	38	31
Men	32	35	24	18	12	22

Women $(34 + 29 + 40 + 16 + 38 + 31)/6 = 31.3$

Men $(32 + 35 + 24 + 18 + 12 + 22) / 6 = 21.0$

- Is this difference true for the **population**?
- Or a fluke of the particular **sample** collected?

Basis of statistical inference

- Inferential statistics, generally speaking:
  - Assume there is **no difference** in the population
    - e.g. no difference between men and women on this test.
    - This is called the **null hypothesis**.
  - Work out the **probability** that the **null hypothesis** is true, given the observed difference in the descriptive statistics for the **sample.**
    - e.g. In the sample, women get 10.3 more words correct, on average.
    - What's the probability this would happen in a **sample** of 12 people if, in the population they're drawn from, men and women are equal on this measure?
  - Compare that **probability** to some "criterion of truth"
    - This criterion is called **statistical significance.**
    - In psychology, things for which the probability of the null hypothesis being true are less than 1 in 20 ($P < 0.05$) are **statistically significant**.

Statistical tests, statistical tables

- "Work out the **probability** that the **null hypothesis** is true"
  - This is done by calculating a **statistical test**
  - A **statistical test** usually results in a single number.
  - That number has a number of names, depending on the test
  - This number is converted into a probability by a **statistical table.**
  - Very often, this table just tells you whether a test is significant or not, rather than giving a precise probability.

Statistical tests, statistical tables

- "Work out the **probability** that the **null hypothesis** is true"
  - This is done by calculating a **statistical test**
  - A **statistical test** usually results in a single number.
    - Central to this course is to help you understand how these tests work, when to apply them, and how to calculate them.
  - That number has a number of names, depending on the test
    - e.g. a **t-value,** an **F ratio**, and so on…
  - This number is converted into a probability by a **statistical table.**
    - There are many different statistical tables. The choice of table is determined by the choice of test.
    - Statistical tables are calculated by mathematicians. It is not the primary purpose of this course to explain how they are calculated.

How many tails?

- For most **statistical tests**, you need to know whether the question you are asking is **one-tailed** or **two-tailed**.
- A **two-tailed** question is non-directional
  - e.g. Do men and women differ in vocabulary score?
- A **one-tailed** question is directional
  - e.g. Do women have a higher vocabulary score than men?

How many tails?

- For any given sample, a **one-tailed** question is more likely to result in **statistical significance** than the corresponding **two-tailed** question.
- Statisticians are generally conservative. They will therefore generally only pose a **one-tailed** question is there is previous, scientific, evidence that the difference is likely to be in the proposed direction.
- **You are expected to follow this convention in this course** (in fact, throughout psychology).

Wilcoxon Rank-Sum test

- Wilcoxon rank-sum is a test for a significant difference between two different groups (e.g. of people).
- It is basically an assessment of how much the samples for the two groups overlap:

Men 12 14 16 18 20
Women 22 24 26 28 30 **Low overlap**

Men 12 16 20 24 28
Women 14 18 22 26 30 **High overlap**

- **High overlap** – difference between groups more likely to be due to chance.
- **Low overlap** - difference between groups less likely to be due to chance.

Recipes:
1. Wilcoxon Rank-Sum test

- 1. Rank all the data

	1	2	3	4	5	6
Women	38 (11)	31 (7)	32 (8)	40 (12)	29 (6)	35 (10)
Men	16 (2)	24 (5)	34 (9)	18 (3)	12 (1)	22 (4)

Recipes:
1. Wilcoxon Rank-Sum test

- 2. Calculate the **sum of the ranks** of the group with lower *n.*
  - If groups are of equal *n,* calculate the sum of the ranks for each group, and take the smaller.

- **High overlap** – this number will be larger.
- **Low overlap –** this number will be smaller.

Recipes:
1. Wilcoxon Rank-Sum test

	1	2	3	4	5	6
Women	38 (11)	31 (7)	32 (8)	40 (12)	29 (6)	35 (10)
Men	16 (2)	24 (5)	34 (9)	18 (3)	12 (1)	22 (4)

Women, $W = 11 + 7 + 8 + 12 + 6 + 10 = 54$

Men, $W = 2 + 5 + 9 + 3 + 1 + 4 = 24$

$W = 24$

Recipes:
1. Wilcoxon Rank-Sum test

- The result is *significant* if W is smaller or equal to the appropriate value in a Wilcoxon Rank-sum *table.*
- $N_1 = n$ for the smaller group
- One or two tailed?

Examples

- Attempt problems W-1 and W-2.

Tied ranks:

Numbers:	6	8	8	9
Ranks:	1	2.5	2.5	4