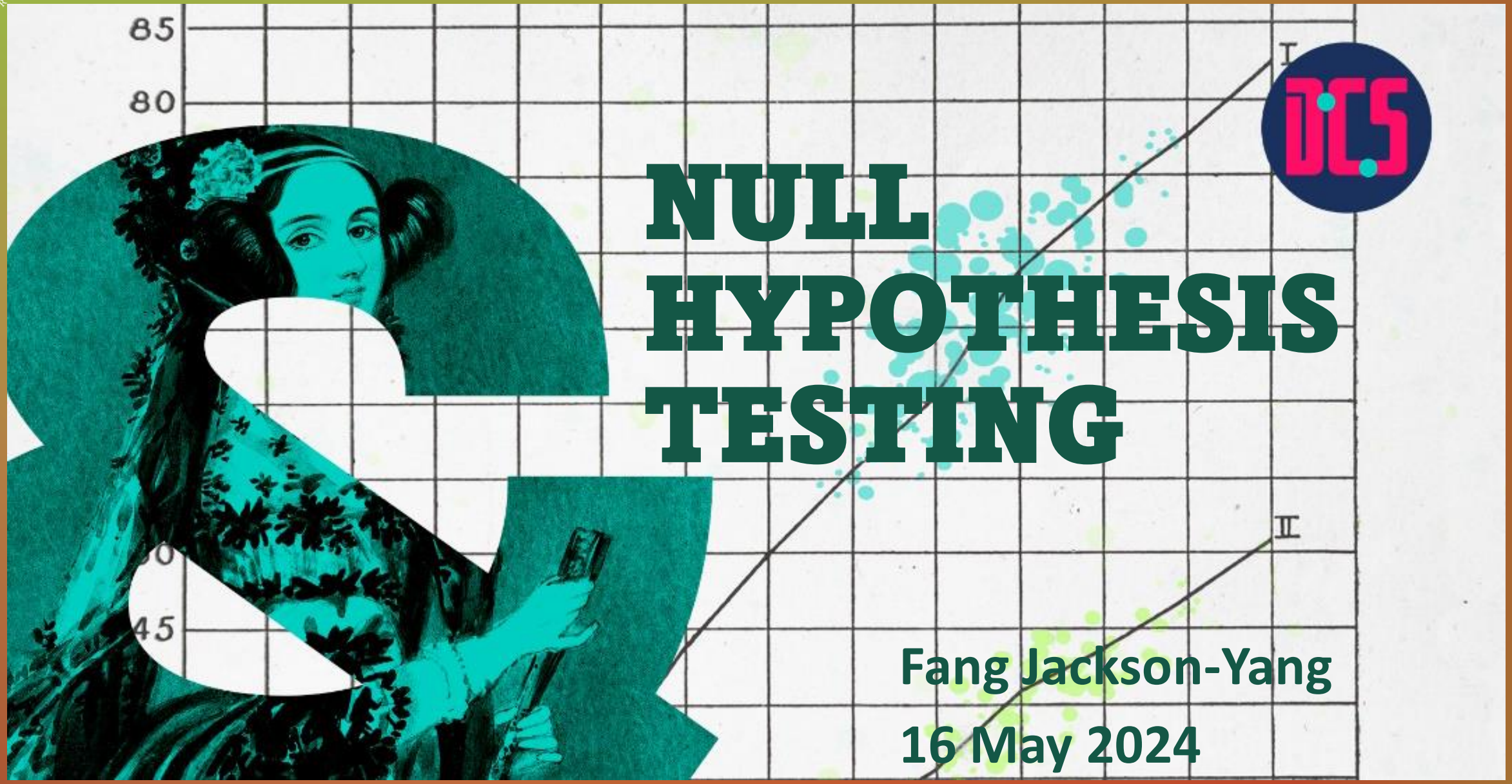
THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society



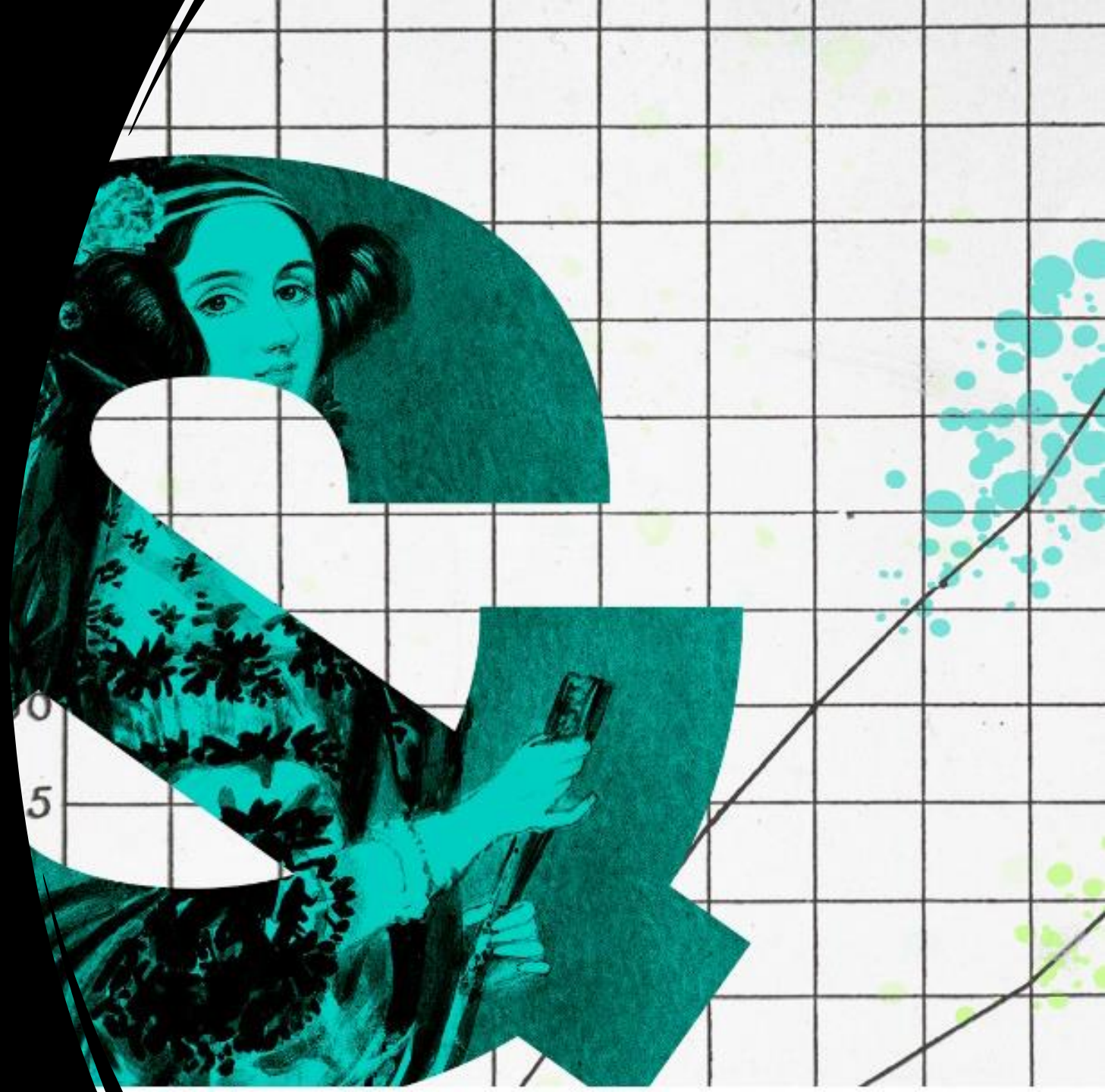
NULL HYPOTHESIS TESTING

Fang Jackson-Yang

16 May 2024

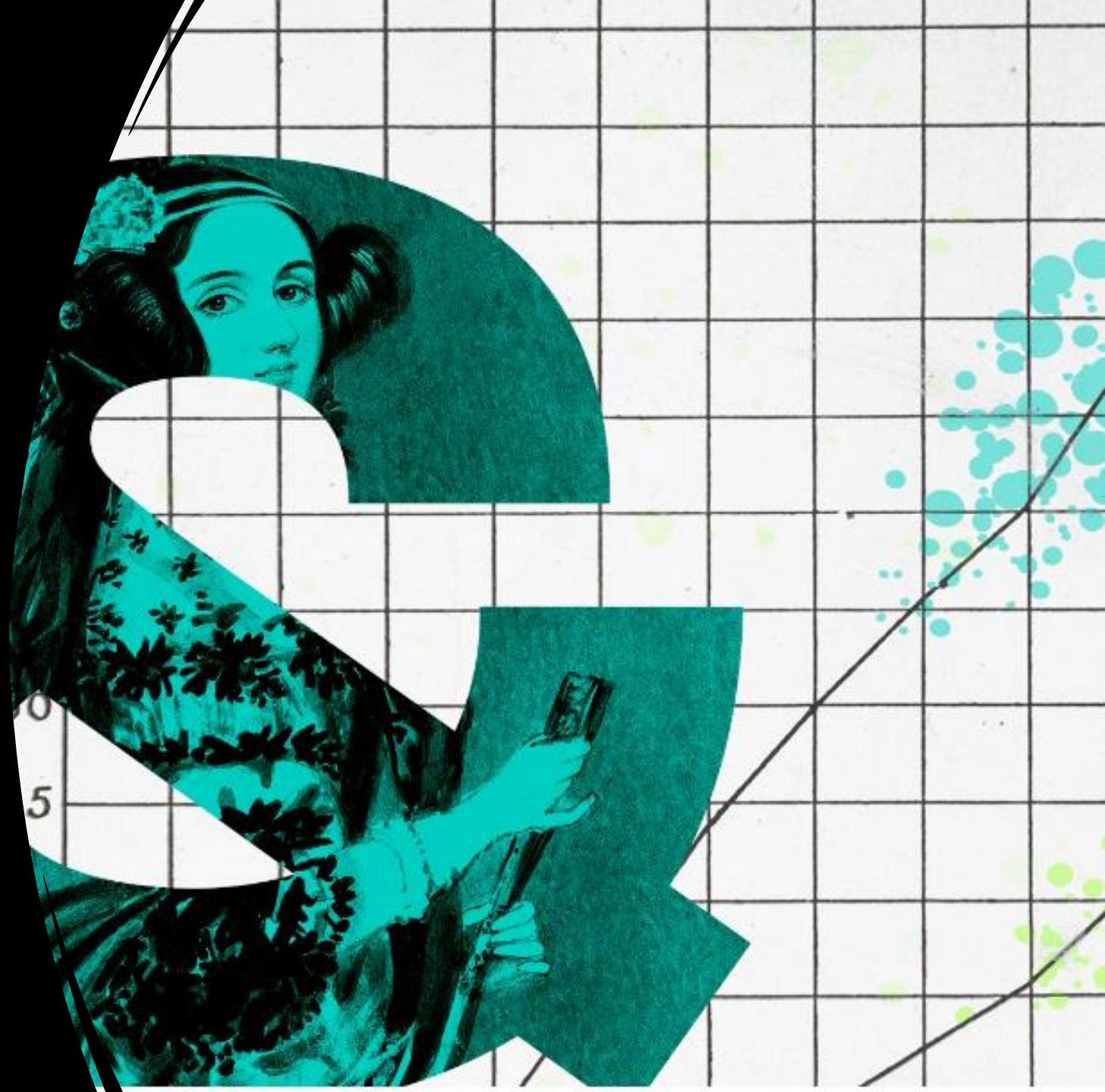
Course outline

- Session 1. Concepts & two-sample t-test
- Session 2. Power, effect size & other example tests for NHT



Session 2 Roadmap (today)

- Statistical Significance & Practical Importance
- Assumptions Check
- Paired t-test



Statistical Significance & Practical Importance

Recall: Errors in Hypothesis Testing

- Type 1 Error : Reject a true null hypothesis
- Type 2 Error : Fail to reject a false null hypothesis

		Your Decision	
		Retain H_0	**Reject H_0**
$H0$ **Truth**	True	1 - α = probability of correct retention	**α (Type I error rate)**
	False	**β (Type II error rate)**	1 - β = **power** of the test

Significance vs Importance

- Errors in decision making
- Power (Probability of rejecting $H0$ given that $H0$ is false)
- Effect size ($Cohen's\ D$) : the difference between the sample and the hypothesised mean, measured in units of the standard deviation

Effect Size ($Cohen's\ D$)

- the magnitude of the difference between the **true** population mean μ and the hypothesised mean μ
- We do not know the TRUE population mean, thus we report its estimate
- The difference between them can be statistically significant but not of practical importance

Statistical Significance ≠ Practical Importance

Effect Size ($Cohen's\ D$)

Magnitude of $Cohen's\ D$ in absolute value:

- ≤ 0.20 Small (or weak)
- ≈ 0.50 Medium (or moderate)
- ≥ 0.80 Large (or strong)

What about in between the categories?

- Use "small to medium" or "medium to large" effect size.

Assumptions Check

Assumption Check

For t-test:

- **Independence**: Randomly obtained sample
- **Normality**: The sampling distribution of the sample mean is normally distributed; or sample size sufficiently large ($n > 30$)
  - qqplot
  - Shapiro-Wilk test

Two-Sample T-test

- **Equality of Variance between groups**

$$H0: \mu_{\text{uni\_student}} = \mu_{FT\_employee}$$
$$H1: \mu_{\text{uni\_student}} > \mu_{FT\_employee}$$

var.test() in *R*

Paired t-test

Paired T-test

- Simulated data based on a real case
- A secondary school in Glasgow were considering purchasing a new online course for teaching vocabulary in language classes.
- 120 randomly selected pupils had exams before and after taking the online course.
- How effective was the course? Worthy investing or not?

Paired T-test

- $H0$: No improvement between pupils scores in vocabular exams before and after taking the online course
- $H1$: Pupils performed better in vocabulary exams after taking the online course

- $H0$: $\mu_{\text{after}} = \mu_{before}$
- $H1$: $\mu_{\text{after}} > \mu_{before}$

Paired T-test

Questions for you:

- Why is it called “paired” t-test?
- How does it differ from two-sample t-test?
- How do we do perform such tests using *r* ?

Specify ‘paired = TRUE’ in t.test()

Exercise

- Try to run a paired t-test to analyse the vocab-score data

THANK YOU

THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society