MONASH University

MONASH BUSINESS SCHOOL

Statistical Thinking (ETC2420/ETC5242)

Associate Professor Catherine Forbes

Week 3: Randomisation and simulation for testing proportions

Learning Goals for Week 3

- Explain terms relevant to statistical hypothesis testing and inference problems
- Demonstrate the sampling distribution of a statistic
- Construct a randomisation test for independence of two binary variables
- Build parametric tests for one and two proportions using the Central Limit Theorem
- Reiterate the framework for frequentist inference

Assigned reading for Week 3:

- Chapter 2 and Sections 3.1 and 3.2 in ISRS

Randomization case study: gender discrimination

Do males and females have the same chance of being promoted?

- Population proportion of **males** promoted: p_M
- Population proportion of **females** promoted: p_F

- Test **Null hypothesis** H_0: $p_M = p_F$
  - (Gender has no effect on promotion decision)
- Against **Alternative hypothesis** H_1: $p_M > p_F$
  - Gender has an effect on promotion decision, with a man more likely to be promoted than a woman

- A controlled experiment: all other attributes of job applicants identical
- Outcomes:

		decision		
		promoted	not promoted	Total
gender	male	21	3	24
	female	14	10	24
	Total	35	13	48

Table 2.1: Summary results for the gender discrimination study.

Statistical test: Are two population proportions equal?

- Use the sample proportions as **point estimates** of the "true" p_M and p_F
  - **observed** $\hat{p}_M = \dfrac{\text{\# males promoted}}{\text{\# males considered for promotion}}$
  - **observed** $\hat{p}_F = \dfrac{\text{\# females promoted}}{\text{\# females considered for promotion}}$
- Do **observed** values satisfy $\hat{p}_M > \hat{p}_F$?
  - Equivalently, is $x_{obs} =$ **observed** $\hat{p}_M - \hat{p}_F > 0$?
- We take x_{obs} is our **point estimate** for $p_M - p_F$
- **Could $x_{obs} > 0$ be due to "chance"?**
- YES. Even when $p_M = p_F$ we can get $x_{obs} > 0$

Hypothesis test

- Restating the hypotheses of interest:
  - H_0: $p_M - p_F = 0$ (Null hypothesis)
  - H_1: $p_M - p_F > 0$ (Alternative hypothesis)

- Need the **decision rule** to decide whether to reject H_0
- We want to reject H_0 when x_{obs} is far from zero
  - zero is the value of the parameter (here $p_M - p_F$) under H_0

- $\Rightarrow$ Choose the **decision rule**:
  - Reject H_0 when $x_{obs} \geq x^{crit}$
  - Otherwise: Do not Reject H_0

- Here x^{crit} is the **critical value**
  - how is it set?

The Significance level

- Critical value determined by desired to control **Type I error**

		Decision	
		Do not reject H_0	Reject H_0
Truth	H_0 true	no error	**Type I Error**
	H_1 true	**Type II Error**	no error

Table 1: Decision errors from an hypothesis test

- Fix $\Pr(\text{Type I error}) = \alpha$, the **significance level**
  - choose α to be 'small' (e.g. $\alpha = 0.05$)
- Given *alpha*, find x^{crit} to solve:
  - $\Pr\left(\hat{p}_M - \hat{p}_F \geq \ x^{crit} \ \mid H_0 \text{ is true}\right) = 0.05$
  - 'Probability' for point estimate ($\hat{p}_M - \hat{p}_F$) not yet observed (under H_0)

P-values

- *"If H_0 is true and we repeated the experiment, what's the chance we would observed a value of $\hat{p}_M - \hat{p}_F$ that is '**as or more extreme**' than we already have observed with our data?"*

- The "chance" is a probability known as a **p-value**

$$p\text{-value} = \Pr\left(\hat{p}_M - \hat{p}_F \geq \ x_{obs} \ \middle|\ H_0 \text{ is true}\right)$$

  - A one-sided test: 'as or more extreme' implies $\geq x_{obs}$ values
  - 'Probability' for a (hypothetical) repeated experiment (under H_0)

- $\Rightarrow$ p-value approach yields same conclusion if use the same α
- $\Rightarrow$ decision rule for p-value approach
  - If p-value $< \alpha$: Reject H_0
  - Otherwise: Do not reject H_0

Two approximate tests

Probability evaluations typically not feasible $\Rightarrow$ Use an approximate test

1. A **randomisation test**: Use variability in observed data
   - **NEW**: A **modern computational** approach
2. A test based on the **Central Limit Theorem (CLT)**
   - **OLD**: This should be review

Our focus:

- Explain rationale for each test (i.e. in relatively non-technical terms)
- Describe the relative strengths and weaknesses of each test
- Execute the tests using **R**
- Interpret the results and draw a conclusion
- Report so that all steps are reproducible

CLT test for equality of two population proportions

Let $X = \hat{p}_M - \hat{p}_F$ denote the unobserved (random variable)

- Either before the data is collected
- Or from a hypothetical repeated experiment

Under the CLT:

$$X \overset{approx}{\sim} N\left(\mu_X,\ \sigma_X^2\right)$$

- $\mu_X = p_M - p_F$ is the **mean** of X
- σ_X^2 is (an appropriate) **variance** of X
- $\Rightarrow \sigma_X = \sqrt{\sigma_X^2}$ is the **standard error (SE)**
  - but must be estimated since it will depend upon p_M and p_F

- **See Section 3.2.1 in ISRS**
  - for required assumptions and additional formulae
  - test for a given (single) proportion
  - confidence intervals

Normal test for proportions in R

Use **prop.test()** in **R** for **test of equal proportions**

- One-sided test $H_0 : p_1 - p_2 = 0$
  - vs. $H_1 : p_1 - p_2 > 0$ (upper-tailed)
  - vs. $H_1 : p_1 - p_2 < 0$ (lower-tailed)
- Two-sided test $H_0 : p_1 - p_2 = 0$ vs. $H_1 : p_1 - p_2 \neq 0$

Use **prop.test()** in **R** for **test of given proportions**

- One-sided test $H_0 : p = p_0$ (single proportion)
  - vs. $H_1 : p > p_0$ (upper-tailed)
  - vs. $H_1 : p < p_0$ (lower-tailed)
- Two-sided test $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$

Use **prop.test()** in **R** to produce **confidence intervals**

- Difference: $p_1 - p_2$
- Single proportion: p

Randomisation test of equal proportions

Idea: Use the variability in the data to approximate the p-value

- **Shuffle = Randomly permute** resumes (gender) assigned to supervisor (promotion outcome)
    - to simulate X under H_0
    - by breaking association (if present) between gender and promotion outcome

$\longrightarrow$ SHUFFLE RESUMES $\longrightarrow$

Approximate p-value

- Randomly permute for $r = 1, 2, \ldots R$, each time calculating an $x_{obs}^{[r]}$
  - $\Rightarrow$ approximate the **sampling distribution of** X
  - under hypothetical repeated experiments

- Use simulated hypothetical sampling distribution: $\{x_{obs}^{[r]},\ \text{for } r = 1, 2, \ldots, R\}$
- To **estimate the p-value**:

$$\tilde{\text{p}}\text{-value} = \frac{\left(\text{number of } x_{obs}^{[r]} \geq x_{obs}\right)}{R}$$

  - the proportion of the R permuted samples where $x_{obs}^{[r]} \geq x_{obs}$
- Use $\tilde{\text{p}}$-value
  - to find the conclusion of the test
  - to interpret the outcome
  - then report so that all steps are reproducible

Randomisation test for Gender discrimination study

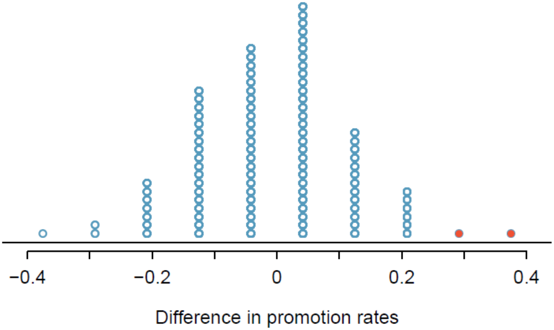

Figure 2.3: A stacked dot plot of differences from 100 simulations produced under the null hypothesis, H_0, where `gender_simulated` and `decision` are independent. Two of the 100 simulations had a difference of at least 29.2%, the difference observed in the study, and are shown as solid dots.

Frequentist testing

- Under a frequentist approach
  - inferential procedures are developed and evaluated
  - in terms of their performance under hypothetical repeated sampling
  - with 'true' parameters fix and unknown
- $\Rightarrow$ *Frequentists treat data as random and parameters as fixed*

- We have considered a Randomisation test and the CLT-based test
  - for equality of two proportions
  - both developed from the same **frequentist** testing framework

- Each test approximate probabilities
  - about a test statistic ($X = \hat{p}_M - \hat{p}_F$)
  - under its **sampling distribution**
  - when $H_0 : p_M = p_F$ is true

Review materials and the Week 3 Lab

- Textbook and videos are for review and practice

Week 3 Lab:

- Build the randomisation test for the Gender discrimination case study data
- Conduct CLT-based tests for
  - single proportion
  - difference between two proportions
- in **R**
  - more plots
  - do some (light) wrangling
  - write functions
  - sample
  - iterate using a for-loop
- . . . more