

Lecture 27 Compare Two Populations - Proportion

BIO210 Biostatistics

Xi Chen

Spring, 2025

School of Life Sciences

Southern University of Science and Technology



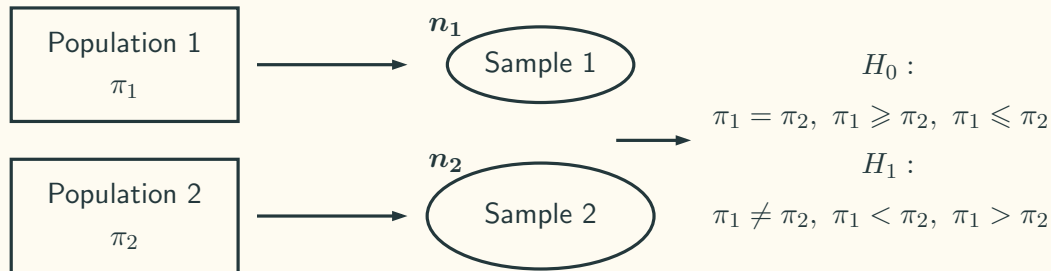
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Compare two proportions

Whether the proportions of colour blindness are the same in two different populations (e.g. male vs female, Asian vs European) ?

Whether chemical A is better than chemical B for culturing cells in petri dishes (can be measured by percentage of cells that express *Pou5f1*) ?

Whether drug A is more efficient than drug B in terms of curing a certain disease (can be measured by percentage of cured patients) ?



Clinical Infectious Diseases

BRIEF REPORT

Relationship Between the ABO Blood Group and the Coronavirus Disease 2019 (COVID-19) Susceptibility

Jiao Zhao,^{1,a} Yan Yang,^{2,a} Hanping Huang,^{3,a} Dong Li,^{4,a} Dongfeng Gu,¹ Xiangfeng Lu,⁵ Zheng Zhang,² Lei Liu,² Ting Liu,³ Yukun Liu,⁶ Yunjiao He,¹ Bin Sun,¹ Meilan Wei,¹ Guangyu Yang,^{7,b} Xinghuan Wang,^{8,b} Li Zhang,^{3,b} Xiaoyang Zhou,^{4,b} Mingzhao Xing,^{1,b} and Peng George Wang^{1,b}

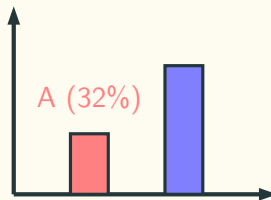
¹School of Medicine, The Southern University of Science and Technology, Shenzhen,

Type A blood in normal people and COVID-19 patients

Normal Population
 π_1



$n_1 = 3694$

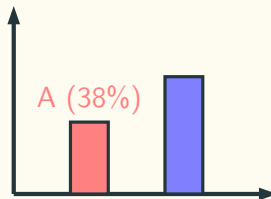


Is π_1 different from π_2 ?

COVID-19 Population
 π_2



$n_2 = 1775$



$$H_0 : \pi_1 = \pi_2$$

$$H_1 : \pi_1 \neq \pi_2$$

Strategy 1: Use One-sample Hypothesis Testing ??

Two choices:

- $H_0 : \pi_1 = 0.38$

$$H_1 : \pi_1 \neq 0.38$$

- $H_0 : \pi_2 = 0.32$

$$H_1 : \pi_2 \neq 0.32$$

Two answers:

- $z = -7.5$

$$p = 6.4 \times 10^{-14}$$

- $z = 4.4$

$$p = 1.1 \times 10^{-5}$$

Strategy 2: Figure Out The Sampling Distribution of The Difference

- Let the random variable P_1 represent the proportion of blood type A in a sample ($n_1 = 3694$) drawn from normal people.
- Let the random variable P_2 represent the proportion of blood type A in a sample ($n_2 = 1775$) drawn from COVID-19 patients.

Normal

π_1

$$P_1 \sim \mathcal{N}\left(\mu_P = \pi_1, \sigma_P^2 = \frac{\pi_1(1 - \pi_1)}{n_1}\right)$$

$$\delta = \pi_1 - \pi_2$$

COVID-19

π_2

$$P_2 \sim \mathcal{N}\left(\mu_P = \pi_2, \sigma_P^2 = \frac{\pi_2(1 - \pi_2)}{n_2}\right)$$

$$D = P_1 - P_2$$

$$D \sim ?$$

Sampling Distribution of The Difference of The Sample Proportion

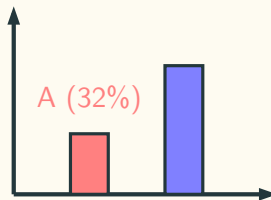
- $D \sim \mathcal{N}\left(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}\right)$
- $D = P_1 - P_2$ and $d = p_1 - p_2$ are the point estimator/estimate of δ
- 95% CI: $(p_1 - p_2) \pm 1.96\sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$

Type A blood in normal people and COVID-19 patients

Normal Population
 π_1



$n_1 = 3694$



Is π_1 different from π_2 ?

$$H_0 : \pi_1 = \pi_2$$

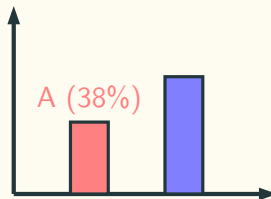
$$H_1 : \pi_1 \neq \pi_2$$



COVID-19 Population
 π_2



$n_2 = 1775$



$$H_0 : \delta = \pi_1 - \pi_2 = 0$$

$$H_1 : \delta = \pi_1 - \pi_2 \neq 0$$

Two-sample Hypothesis Testing For Proportion

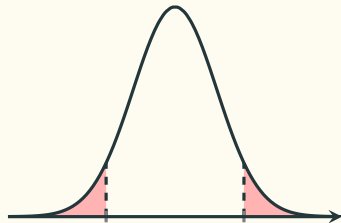
$$H_0 : \delta = \pi_1 - \pi_2 = 0$$

$$H_1 : \delta = \pi_1 - \pi_2 \neq 0$$

$$D \sim \mathcal{N} \left(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2} \right) \xrightarrow[\text{were true}]{\text{if } H_0} D \sim \mathcal{N} \left(0, \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \pi(1 - \pi) \right)$$

1. What we observe is: $d = p_1 - p_2$
2. What is the probability of observing d or more extreme?

$$z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \pi(1 - \pi)}} = \frac{p_1 - p_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \pi(1 - \pi)}}$$



What is the best estimate for π ?

Two-sample Hypothesis Testing For Proportion

Sample size: bigger is always better:

	Normal	COVID-19
A	a	b
Non-A	c	d
Total	n_1	n_2

$$\pi : \frac{a + b}{n_1 + n_2} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = p$$

The test statistic:

$$z = \frac{p_1 - p_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) p(1 - p)}}$$

The test statistic:

$$p = \frac{1188 + 670}{3694 + 1775} = 0.34, z = \frac{0.32 - 0.38}{\sqrt{\left(\frac{1}{3694} + \frac{1}{1775}\right) \times 0.34 \times 0.66}} = -4.4$$

Example: Two-sample Hypothesis Testing For Proportion

Myopia: Researchers suspect that **myopia**, or nearsightedness, is becoming more common over time. A study from the year 2000 showed 139 cases of myopia in 420 randomly selected people. A separate study from 2015 showed 228 cases in 600 randomly selected people. Perform a hypothesis testing to see if the researchers' suspicion is true or not.

Sample statistics: $n_1 = 420$, $p_1 = \frac{139}{420} = 0.33$, $n_2 = 600$, $p_2 = \frac{228}{600} = 0.38$

Pooled estimate for π : $p = \frac{139 + 228}{420 + 600} = 0.36$

The test statistics: $z = \frac{p_1 - p_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) p(1 - p)}} = \frac{0.33 - 0.38}{\sqrt{\left(\frac{1}{420} + \frac{1}{600}\right) \times 0.36 \times 0.64}}$