Special Case: 2×2 Tables → 2 binary variables

2/1/22

		Y		
		1	2	
X	1	n_{11}		n_1
	2	n_{21}		n_2
				n

"Success" (response)

2 groups want to compare

★ Some books etc. put expl. variable as columns.

Interested in inference about:

$$\pi_1 = P(Y=1 \mid X=1) \qquad \hat{\pi}_1 = \frac{n_{11}}{n_{1+}}$$

$$\pi_2 = P(Y=1 \mid X=2) \qquad \hat{\pi}_2 = \frac{n_{21}}{n_{2+}}$$

Goal: Comparing π_1 & π_2 — Single parameter

① Difference in proportions: $\pi_1 - \pi_2 = \theta$

② Relative risk (risk ratio): $\dfrac{\pi_1}{\pi_2}$

("risk" ≈ "probability")

★ ③ Odds ratio: $\dfrac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$ } "odds for group 1" } "odds for group 2"

Example: Advertising MSU

X = Type of Brochure: 1 = Skier, 2 = Snowboarder

Y = Enrolled?

X	1 = Yes	2 = No	
1 = Skier	17	58	75
2 = Snowboarder	14	61	75
	31	119	150

"binomial" Sampling → row totals fixed

Interested in comparing:

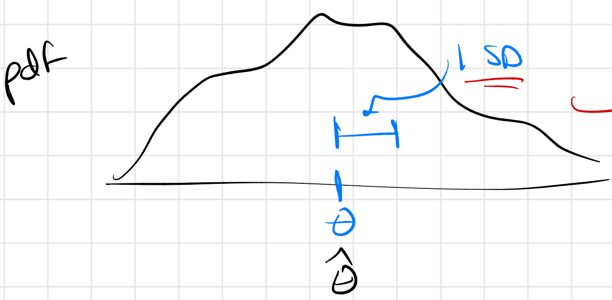
$\pi_1 = P(\text{enroll} \mid \text{skier})$

$\pi_2 = P(\text{enroll} \mid \text{snowboarder})$

Statistical Inference on a single parameter $\theta = f(\pi_1, \pi_2)$:

1. Estimate: $\hat{\theta}$ (Point estimate)
2. Standard error of estimate: $SE(\hat{\theta})$ →
   - how close your would expect your observed estimate $\hat{\theta}$ is to the true θ, on average, over many samples

Sampling distribution of $\hat{\theta}$ → distribution of possible values of $\hat{\theta}$ over all possible samples



estimate of std. dev of $\hat{\theta}$ = Standard error

1. $\theta = \pi_1 - \pi_2$ — Difference in probability of success

Estimate: $\hat{\theta} = \hat{\pi}_1 - \hat{\pi}_2$

$$SE(\hat{\theta}) = SE(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

__If__ assuming $H_0: \pi_1 - \pi_2 = 0 \Rightarrow \pi_1 = \pi_2$

$$\text{Pooled } SE(\hat{\theta}) = \sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{\pi} = \frac{n_{11} + n_{21}}{n}$$ (pooled prop. of successes)

Central Limit Theorem $\Rightarrow$

$$\underbrace{\hat{\pi}_1 - \hat{\pi}_2}_{\hat{\theta}} \;\dot{\sim}\; N\left(\pi_1 - \pi_2, \overbrace{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}^{SD(\hat{\theta})}\right)$$

for large n_1 & n_2

Rule of Thumb $\rightarrow$ ≥ 5 or 10 successes & failures in each group.

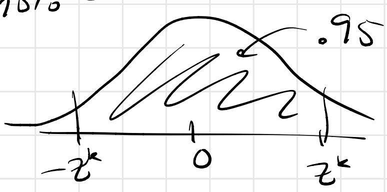
$(1-\alpha)100\%$

$\Rightarrow$ Conf. Interval for θ:

$$\hat{\theta} \pm z^* SE(\hat{\theta})$$

z^* ↳ $(1-\frac{\alpha}{2})^{th} \times 100\%$ percentile of std. normal

95% CI



$H_0: \theta = 0$ $\rightarrow$ Test stat: $$Z = \frac{\hat{\theta} - 0}{SE(\hat{\theta})}$$

Under H_0 $\dot{\sim}$ $N(0,1)$

Ex: $\hat{\pi}_1 = \frac{17}{75} \approx 0.227$ $\quad \hat{\pi}_2 = \frac{14}{75} \approx 0.187$ $\rightarrow$ R

$\hat{\theta} = 0.04$ $\longrightarrow$ Interpret: $\hat{\theta}$ $\longrightarrow$ See R file.

$SE(\hat{\theta}) \simeq 0.066$

② $\theta = \pi_1 / \pi_2$ — risk ratio

Estimate: $\hat{\theta} = \hat{\pi}_1 / \hat{\pi}_2$ $\longrightarrow$ Sampling distn. of this is highly skewed.

$\Rightarrow$ Inference on log-scale.

Large Samples $\Rightarrow$

$$\log\hat{\theta} \overset{\sim}{\sim} N\left(\log\theta, \sqrt{\frac{1-\pi_1}{n_1\pi_1} + \frac{1-\pi_2}{n_2\pi_2}}\right)$$

Estimate $\log\theta$: $\log(\hat{\theta})$

$$SE(\log\hat{\theta}) = \sqrt{\frac{1-\hat{\pi}_1}{n_1\hat{\pi}_1} + \frac{1-\hat{\pi}_2}{n_2\hat{\pi}_2}}$$

Approximate (large sample) $100(1-\alpha)\%$ CI for $\log\theta$:

$$\log\hat{\theta} \pm z^* \cdot SE(\log\hat{\theta}) = (L, U)$$

CI for θ : (e^L, e^U)

Ex: $\frac{\hat{\pi}_1}{\hat{\pi}_2} = 1.214$ $\longrightarrow$ Interpret: (R file)

$\left(\hat{\pi}_1/\hat{\pi}_2 - 1\right) \times 100\%$ — increase (positive)
decrease (negative)

1.214 $\longrightarrow$ 21.4% increase

0.8 $\longrightarrow$ 20% decrease

(3) Odds ratio: $\theta = \dfrac{\pi_1 / (1-\pi_1)}{\pi_2 / (1-\pi_2)}$

Estimate: $\hat{\theta} = \dfrac{\hat{\pi}_1 / 1-\hat{\pi}_1}{\hat{\pi}_2 / 1-\hat{\pi}_2}$ $\longrightarrow$ Inference on log-scale

$$SE(\log \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

17	58
14	61

$\longrightarrow$ $\hat{\theta} = \dfrac{17/58}{14/61} = \dfrac{17(61)}{14(58)}$

Sample odds ratio