Chapter 2: Study Sheet

Sheng-Mao Chang

3/13/2025

範例：心肌梗塞與每日使用阿斯匹靈之關聯

實驗：

找一群人分成二組，一組每日給予一顆阿斯匹靈，另一組每日給予一顆安慰劑（但需告訴使用者該藥丸有療效），經過一段時間後觀察使用者是否有得心肌梗塞。結果可整理如下表

	Attack	No Attack	Total
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037
Total	293	21,778	22,071

```
A <- matrix(c(189,104,10845,10933), 2, 2); B <- A[c(2,1),]
rownames(B) <- c("Aspirin", "Placebo")
colnames(B) <- c("Attack", "non-Attack")
B
```

```
##         Attack non-Attack
## Aspirin    104      10933
## Placebo    189      10845
```

觀察：

- 得心肌梗塞的機率為何？

Pr（得心肌梗塞）= 293 / 11034 = 0.027

- 每日吃一顆阿斯匹靈的人中，得心肌梗塞的機率為何：
    - π_1 = Pr（得心肌梗塞｜每日吃一顆阿斯匹靈）
    - $\widehat{\pi_1} = 104/11037 \approx 0.009$
- 每日吃一顆安慰劑的人中，得心肌梗塞的機率為何：
    - π_2 = Pr（得心肌梗塞｜每日吃一顆安慰劑）
    - $\hat{\pi}_2 = 189/11034 \approx 0.017$

問題：吃阿斯匹靈是否「有效」降低得心肌梗塞之「風險」？

回答：

以下將以四個角度來回答這個問題：

Difference of Proportions $H_0 : \pi_1 - \pi_2 = 0$ vs $H_a : \pi_1 - \pi_2 \neq 0$

- 公式
    - $d = \hat{\pi_1} - \hat{\pi_2}$
    - $$Var(d) = \frac{\hat{\pi_1}(1-\hat{\pi_1})}{n_1} + \frac{\hat{\pi_2}(1-\hat{\pi_2})}{n_2}$$
    - $(1-\alpha) \times 100\%$ CI of $\pi_1 - \pi_2$ is

$$\left(d - z_{\alpha/2} \times \sqrt{Var(d)}, \quad d + z_{\alpha/2} \times \sqrt{Var(d)}\right)$$

- 計算過程

```
pi_1_hat <- B[1,1]/sum(B[1,]); pi_1_hat
```

```
## [1] 0.00942285
```

```
pi_2_hat <- B[2,1]/sum(B[2,]); pi_2_hat
```

```
## [1] 0.01712887
```

```
diff <- pi_1_hat - pi_2_hat; diff
```

```
## [1] -0.007706024
```

```
var_diff <- pi_1_hat*(1-pi_1_hat)/sum(B[1,]) +
  pi_2_hat*(1-pi_2_hat)/sum(B[2,])
ub <- diff + qnorm(.975)*sqrt(var_diff)
lb <- diff + qnorm(.025)*sqrt(var_diff)
cat("The estimate of pi_1 - pi_2 is", diff, "\n")
```

```
## The estimate of pi_1 - pi_2 is -0.007706024
```

```
cat("The corresponding 95 CI for diff is [", lb, " , ", ub, "]\n")
```

```
## The corresponding 95 CI for diff is [ -0.0107243  ,  -0.004687751 ]
```

- 結論：
    - 吃阿斯匹靈組得心肌梗塞的機率為0.009，而吃安慰劑組得心肌梗塞的機率為0.017。
    - 因為上述95%信賴區間不包含0，所以拒絕虛無假設，又因為 $\hat{p_1} - \hat{p_2} < 0$，所以我們的結論為，吃阿斯匹靈組得心肌梗塞之機率顯著低於吃安慰劑組，故吃阿斯匹靈有效降低得心肌梗塞之風險。

Relative Risk $H_0 : \pi_1 / \pi_2 = 1$

- 公式
    - $RR = \pi_1 / \pi_2$ and $\hat{RR} = \hat{\pi_1} / \hat{\pi_2}$
    - $$\log(\hat{RR}) \sim N\left(\log(RR), \hat{Var}(\log(\hat{RR}))\right) \quad \text{where} \quad \hat{Var}(\log(\hat{RR})) = \frac{1-\hat{\pi_1}}{n_{1+}\hat{\pi_1}} + \frac{1-\hat{\pi_2}}{n_{2+}\hat{\pi_2}}$$

- $(1-\alpha) \times 100\%$ CI of $\log(RR)$ is

$$log(\hat{RR}) \pm z_{\alpha/2} \times \sqrt{\hat{Var}(\log(\hat{RR}))}$$

- 計算過程

```
RR_hat <- pi_1_hat/pi_2_hat; RR_hat
```

```
## [1] 0.550115
```

```
log_RR_hat <- log(RR_hat); log_RR_hat
```

```
## [1] -0.597628
```

```
var_log_RR_hat <- (1-pi_1_hat)/(sum(B[1,])*pi_1_hat) + (1-pi_2_hat)/(sum(B[2,])*pi_2_hat)
ub <- log_RR_hat + qnorm(.975)*sqrt(var_log_RR_hat)
lb <- log_RR_hat + qnorm(.025)*sqrt(var_log_RR_hat)
elb <- exp(lb); eub <- exp(ub)
cat("The estimate of the RR is", RR_hat, "\n")
```

```
## The estimate of the RR is 0.550115
```

```
cat("The corresponding 95 CI for RR is [", elb, " , ", eub, "]\n")
```

```
## The corresponding 95 CI for RR is [ 0.4336731  ,  0.6978217 ]
```

- 結論
  - 吃阿斯匹靈組得心肌梗塞的機率是吃安慰劑組得心肌梗塞的機率的0.55（$\hat{RR}$）倍。
  - 因為上述95%信賴區間不包含1，所以拒絕虛無假設，又因為 $\hat{p_1}/\hat{p_2} < 1$，所以我們的結論為，吃阿斯匹靈組得心肌梗塞之機率顯著低於吃安慰劑組，故吃阿斯匹靈有效降低得心肌梗塞之風險。

Odds Ratio $H_0 : \theta = 1$

- 公式
    - Odds ratio

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

    - Odds ratio estimator
        - If all counts are greater than zero,

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

with

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

        - If any count is equal to zero,

$$\hat{\theta} = \frac{(n_{11} + 0.5)(n_{22} + 0.5)}{(n_{12} + 0.5)(n_{21} + 0.5)}$$

with

$$Var(\log(\hat{\theta})) = \frac{1}{n_{11} + 0.5} + \frac{1}{n_{12} + 0.5} + \frac{1}{n_{21} + 0.5} + \frac{1}{n_{22} + 0.5}$$

    - $(1 - \alpha) \times 100\%$ CI of θ is

$$log(\hat{\theta}) \pm z_{\alpha/2} \times \sqrt{\hat{Var}(\log(\hat{\theta}))}$$

- 計算過程

```r
if(min(B) == 0){
  B2 <- B + 0.5
}else{
  B2 <- B
}
theta_hat <- B2[1,1]*B2[2,2]/B2[1,2]/B2[2,1]; theta_hat
```

```
## [1] 0.5458355
```

```r
log_theta_hat <- log(theta_hat); log_theta_hat
```

```
## [1] -0.6054377
```

```r
var_log_theta_hat <- sum(1/B2)
ub <- log_theta_hat + qnorm(.975)*sqrt(var_log_theta_hat)
lb <- log_theta_hat + qnorm(.025)*sqrt(var_log_theta_hat)
elb <- exp(lb); eub <- exp(ub)
cat("The estimate of the odds ratio is", theta_hat, "\n")
```

```
## The estimate of the odds ratio is 0.5458355
```

```r
cat("The corresponding 95 CI for the odds ratio is [", elb, " , ", eub, "]\n")
```

```
## The corresponding 95 CI for the odds ratio is [ 0.429041  ,  0.694424 ]
```

- 結論
  - 吃阿斯匹靈組得病與不得病的比例是吃安慰劑組(得病與不得病比例)的0.55（$\hat{\theta}$）倍。
  - 因為上述95%信賴區間不包含1，所以拒絕虛無假設，又因為 $\hat{\theta} < 1$，所以我們的結論為，吃阿斯匹靈組得病與不得病之比例顯著低於吃安慰劑組，故吃阿斯匹靈有效降低得心肌梗塞之風險。

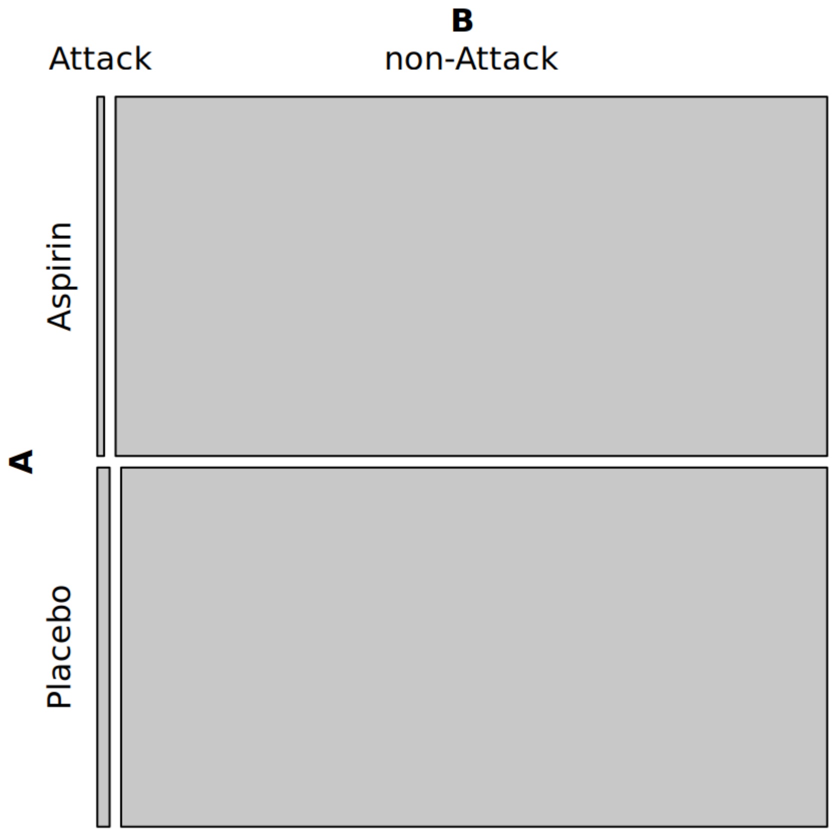
Independence Test: H_0 : no association vs H_a : some association

- 公式
  - Expected count: $E_{ij} = n_{i+} n_{+j}/n$
  - Test statistic

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \overset{H_0}{\sim} \chi^2_{(I-1)(J-1)}$$

- 計算過程

```
library(vcd)
mosaic(B)
```




```
ct <- chisq.test(B); ct
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  B
## X-squared = 24.429, df = 1, p-value = 7.71e-07
```

```
ct$residuals
```

```
##            Attack non-Attack
## Aspirin -3.512724  0.4074449
## Placebo  3.513202 -0.4075003
```

- 結論
    - 根據卡方檢定，是否吃阿斯匹靈與是否得心肌梗塞之間有相關。
    - 根據標準化殘差可知，吃斯匹靈且得病組比期望次數少，吃安慰劑且得病組比期望次數多，所以比起獨立（不相關）的情況，本資料顯示有較少的吃斯匹靈且得病之病人，故吃阿斯匹靈有效降低得心肌梗塞之風險。

作業：性別與酒精攝取之關聯

實驗：

有一個實驗，在某機構的員工中隨機訪問了100名男性與80名女性，其中有90名男性與20名女性在上個週末有超過安全標準的酒精攝取。請問在該機構中，性別是否影響超過安全標準的酒精攝取。