

Chapter 2: Homework

711378912 蔡宜誼

3/19/2025

作業題目：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。以下簡稱：超過安全標準的酒精為“超標”，未超過安全標準的酒精為“未超標”。

	超標 Yes	未超標 No	總計 Total
男性 Male	90	10	100
女性 Female	20	60	80
總計 Total	110	70	180

```
A <- matrix(c(90,20,10,60), 2, 2); B <- A
rownames(B) <- c("Male", "Female")
colnames(B) <- c("Yes", "No")
B
```

```
##      Yes No
## Male   90 10
## Female 20 60
```

✓ 觀察：

- 超標的機率為何？
 $\Pr(\text{超標}) = 110 / 180 = 0.611$
- 男性中，超標的機率為何：
 - $\pi_1 = \Pr(\text{男性且超標} \mid \text{男性})$
 - $\widehat{\pi}_1 = 90/100 \approx 0.9$
- 女性中，超標的機率為何：
 - $\pi_2 = \Pr(\text{女性且超標} \mid \text{女性})$
 - $\widehat{\pi}_2 = 20/80 \approx 0.25$

👍 問題：「性別」是否影響「超過安全標準的酒精攝取」？？？

📊 回答：

以下將以四個角度：差異檢定 **Difference of Proportions** 來比較不同性別的超標機率、相對風險 **RR** 來衡量風險差距、勝算比 **OR** 來評估相對機率、獨立性檢定 **Independence Test** 來檢測是否有統計上的顯著關聯，來回答這個問題：

一、差異檢定 **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] 0.9
```

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

```
## [1] 0.25
```

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

```
## [1] 0.65
```

```
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 0.65
```

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

```
## The corresponding 95 CI for diff is [ 0.5383724 , 0.7616276 ]
```

- 結論：
 - 男性的超標酒精攝取機率為 0.9，而女性的超標機率為 0.25。
 - 由於上述 95% 信賴區間 (**0.5383724, 0.7616276**) 不包含 0，因此拒絕虛無假設。
 - 由於 $\hat{\pi}_1 - \hat{\pi}_2 > 0$ ，我們的結論為，男性的超標機率顯著高於女性，這顯示性別對超標酒精攝取行為有統計上的顯著影響。

二、相對風險 Relative Risk $H_0 : \pi_1 / \pi_2 = 1$

- 公式
 - $RR = \pi_1 / \pi_2$ and $\hat{RR} = \hat{\pi}_1 / \hat{\pi}_2$
 - $\log(\hat{RR}) \sim N\left(\log(RR), \hat{Var}(\log(\hat{RR}))\right)$ where $\hat{Var}(\log(\hat{RR})) = \frac{1 - \hat{\pi}_1}{n_{1+} \hat{\pi}_1} + \frac{1 - \hat{\pi}_2}{n_{2+} \hat{\pi}_2}$
 - $(1 - \alpha) \times 100\%$ 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] 3.6
```

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

```
## [1] 1.280934
```

```
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 3.6
```

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

```
## The corresponding 95 CI for RR is [ 2.449311 , 5.291285 ]
```

- 結論：
 - 男性的超標酒精攝取機率是女性的 3.6 (\hat{RR}) 倍。
 - 由於上述 95% 信賴區間不包含 1，因此拒絕虛無假設。
 - 由於 $\hat{\pi}_1 / \hat{\pi}_2 > 1$ ，我們的結論為，男性的超標機率顯著高於女性，這表示性別對超標酒精攝取行為有統計上的顯著影響。

三、勝算比 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

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

- 計算過程

```
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] 27
```

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

```
## [1] 3.295837
```

```
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 27
```

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

```
## The corresponding 95 CI for the odds ratio is [ 11.81587 , 61.69669 ]
```

- 結論：
 - 男性的超標與未超標的比例是女性 (超標與未超標比例) 的 $27 (\hat{\theta})$ 倍。
 - 由於上述 95% 信賴區間不包含 1，因此拒絕虛無假設。
 - 由於 $\hat{\theta} > 1$ ，我們的結論為，男性的超標與未超標比例顯著高於女性，這進一步顯示性別對超標酒精攝取行為有顯著影響。

四、獨立性檢定 Independence Test: H_0 : no association vs H_a : some association

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

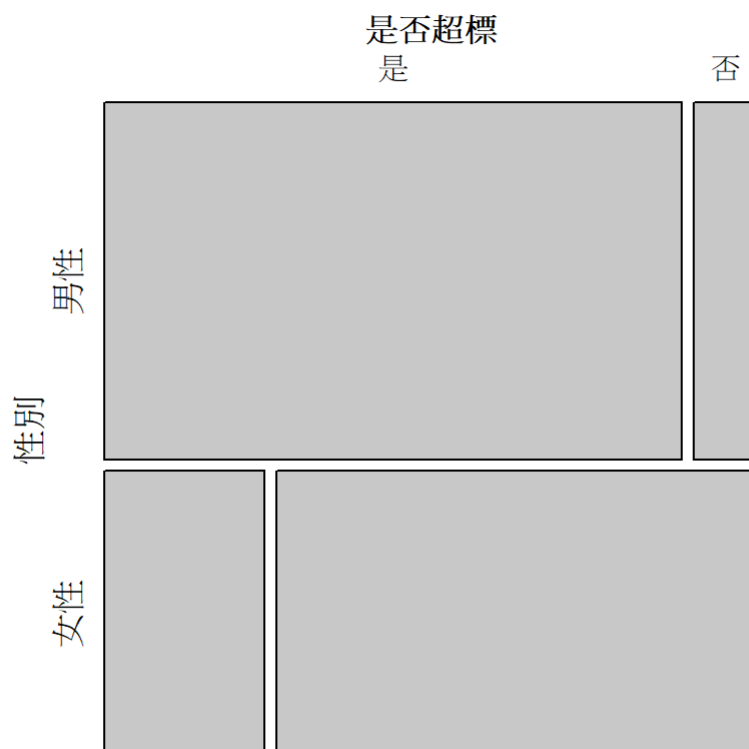
$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \stackrel{H_0}{\sim} \chi^2_{(I-1)(J-1)}$$

- 計算過程

```
library(vcd)
```

```
## Warning: 套件 'vcd' 是用 R 版本 4.3.3 來建造的
```

```
dimnames(B) <- list("性別" = c("男性", "女性"), "是否超標" = c("是", "否"))  
mosaic(B, main = "")
```



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

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: B
## X-squared = 76.302, df = 1, p-value < 2.2e-16
```

```
ct$residuals
```

```
##      是否超標
## 性別      是      否
##  男性  3.695479 -4.632528
##  女性 -4.131671  5.179324
```

- 結論：
 - 根據卡方檢定，性別與超過安全標準的酒精攝取之間存在顯著關聯。
 - 根據標準化殘差分析，男性且超標的組別比期望次數多，而女性且超標的組別比期望次數少，因此與獨立（無關聯）的假設相比，本資料顯示男性更可能超標飲酒，進一步證明性別對酒精攝取行為有統計上的影響。



Summary

- 本研究顯示，男性的超標酒精攝取機率為 **0.9**，而女性的超標機率為 **0.25**，兩者之間存在顯著差異。各項統計檢定結果如下：

檢定方法	結果
差異檢定	95% 信賴區間不包含 0，拒絕虛無假設，表明男性的超標機率顯著高於女性。
相對風險	男性的超標風險是女性的 3.6 倍 ，說明性別對酒精攝取行為有重要影響。
勝算比	男性的超標與未超標比例是女性的 27 倍 ，進一步證實性別對超標酒精攝取的影響顯著。
獨立性檢定	p 值極小，拒絕性別與超標酒精攝取無關的假設，證明兩者之間確實存在顯著關聯。

- 綜合以上分析，**性別確實對超過安全標準的酒精攝取行為產生顯著影響，男性比女性更可能超標**。這些發現對於公共衛生政策具有重要意義，未來可針對不同性別群體制定更有效的飲酒教育與預防措施，以降低酒精相關的健康風險。