



112-2 生物統計學一

變異數分析

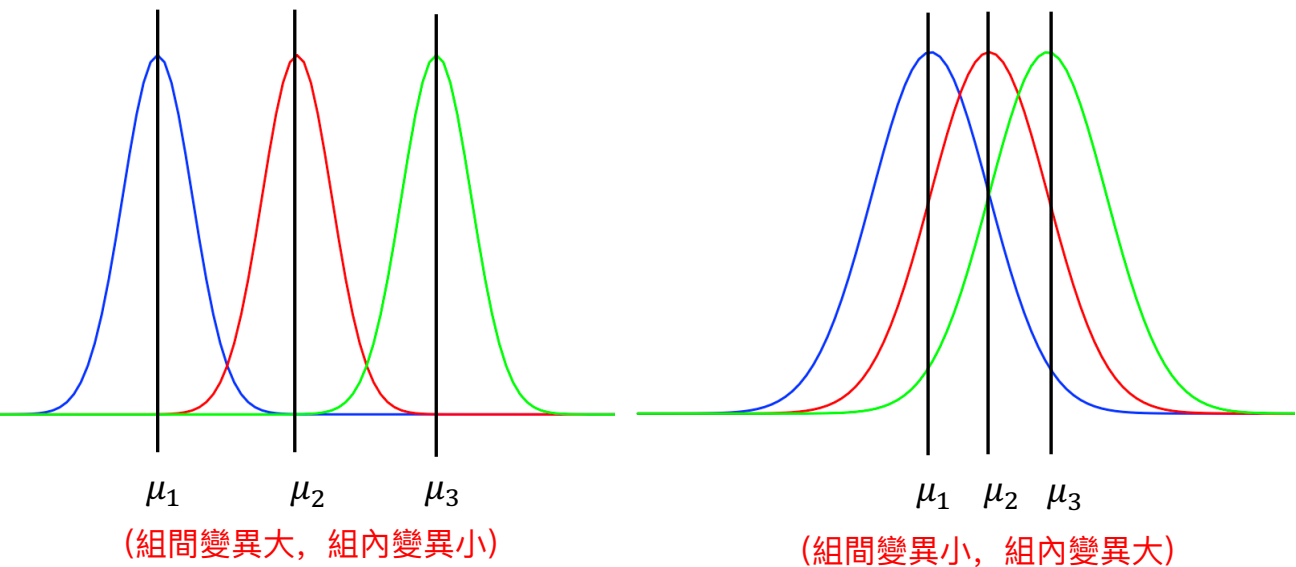
2024/05/07

助教: 廖振博

變異數分析(Analysis of variance, ANOVA)

- 適用情境：檢定三組或以上母體平均值是否相同
- 當組間變異越大，組內變異越小，則代表各組間平均值差異越明顯 → 都是右尾檢定
- 總變異(SST) = 組間變異(SSB) + 組內變異(SSW)

$$F = \frac{\frac{SSB}{k - 1}}{\frac{SSW}{n - k}} = \frac{MS_B}{MS_W}$$



Source of Variation	Sum of Squares	DF	Mean Squares	F ratio
Between groups	SSB	$k - 1$	$MSB = \frac{SSB}{k - 1}$	$F = \frac{MS_B}{MS_W}$
Within groups	SSW	$n - k$	$MSW = \frac{SSW}{n - k}$	
Total	SST	$n - 1$		

變異數分析(Analysis of variance, ANOVA)

- 適用情境：檢定三組或以上母體平均值是否相同

1. 連續型的隨機樣本
2. 母體為常態分佈
3. 各組樣本互相獨立
4. 各組母體變異數同質
(利用巴特力檢定Bartlett test)

```
> bartlett.test(y ~ x)
```

- 假說：

H_0 : 各組母體平均值均相等 ($\mu_1 = \mu_2 = \mu_3 = \dots$)

H_1 : 各組母體平均值**不完全相等** (At least two μ_i 's are not equal)

$$F = \frac{\frac{SSB}{k-1}}{\frac{SSB}{n-k}} = \frac{MS_B}{MS_W}$$

```
> anova_test <- aov(y ~ as.factor(x), data)
```

```
> summary(anova_test)
```

若x為數字，需要將它轉換成因子(factor)，否則會被R當成數值，並執行迴歸分析

多重檢定(Multiple testing)

- 又稱事後比較(post-hoc comparison)
- 當 ANOVA 結果顯示各組平均不完全相同時，則可進一步以事後比較方式，找出差異來自於哪幾組 以兩兩互相比較方式進行 T 檢定
- 但重複進行多個檢定時，整體犯行一錯誤的機率也會提高
- 例如做了十次檢定，顯著水準 α 皆設為0.05，則整體犯型一錯誤的機率為

$$P(\text{type I error}) = 1 - (1 - 0.05)^{10}$$

10次檢定皆正確的機率

- 故每次檢定的顯著水準要調整控制整體犯錯機率(常用的方法為Bonferroni correction):
 - 將每次個別檢定的顯著水準 α 調整為較保守的 α/n
 - R的指令是將算出來的p-value乘上 n 後與 α 比較，兩者等價

```
> pairwise.t.test(y, x, p.adjust.method = "bonf")
```

練習

請先匯入Studata.csv，並命名為studata，並進行以下檢定：

(假設本學期生統一修課同學是一組由台大學生抽出的隨機樣本)

- 計算每個人的BMI，並命名為bmi新增到studata1中
- 若以不同體態滿意度 (satisfy_body) 分組，請問各組同學的平均 bmi 是否相同
- 承上，如果有不同，請問是哪幾組不同

本日作業

- 生統實習課__作業4 (作業區)
- 所需資料檔： hwddata.csv
- 繳交期限： 2024.05.14 (二) 23:59 前上傳至COOL作業區
- 遲交一天扣10分，請務必確認是否上傳成功
- 繳交格式： pdf 檔