

重複測量統計分析
Statistical Analysis for Repeated Measurements

授課教師：林苑俞

課程助教：翁偉倫 (r08849007@ntu.edu.tw)

上課時間：星期四 67 (下午 1:20-3:10)

上課地點：公衛 214 (電腦教室)

Ceiba 網址：https://ceiba.ntu.edu.tw/1092EPM5003_Repeated

課程概述：本門課將介紹常用的重複測量分析方法，包括：隨機效應模式 (random effects models)、廣義估計方程式 (generalized estimating equations, GEE) 與轉銜模式 (transition models)。授課時將涵蓋應用與理論兩個部分，以實際案例分析來介紹統計理論模式，且搭配使用統計軟體 SAS 或 R 來處理重複測量的資料。

可參考書目：

1. Analysis of Correlated Data with SAS and R, 3rd edition. Shoukri, M. M. and Chaudhary M. A. (2007). Chapman & Hall/CRC.
2. Analysis of Longitudinal Data, 2nd edition. Diggle, P.J., Heagerty, P.J., Liang, K.Y., & Zeger, S.L. (2002). Oxford University Press.

評量方式：

1. 作業 (50%)

作業繳交規定：

- 於指派作業的下一次上課**中午十二點前上傳至 ceiba 作業區**。為維持公平性，**遲交作業以零分計**。
- 請假同學需透過 myntu 申請，但該週應交作業仍須於規定時間內上傳，故請儘早完成當週的指派作業。
- **作業抄襲者與被抄者皆以零分計。同學間可討論答題概念，但作業需由自己獨力完成，請勿將自己的作業予他人參考。**

2. 期末報告 (50%)

6/10 (單數號) 與 6/17 (雙數號) 報告請於當天中午 12:00 前，以 pdf 形式，上傳至 ceiba 作業區 (逾時以零分計)，內容為：25 分鐘報告的投影片。本報告佔學期成績的 50%

(抽中報告的同學 25% 投影片內容 + 25% 課堂口頭報告。 未中籤者 50% 皆取決於投影片內容)。

請單數號 (6/10) 雙數號 (6/17) 同學 1:22 前上線 (Microsoft Teams) 或抵達公衛大樓教室，將當場隨機抽出報告人選，中籤同學若不在場則 25% 學期成績以零分計。單/雙數號需等加退選後名單確定方能公布。

重複測量與相依性資料導論

描述性統計

重複測量資料：在生醫領域，研究者常需分析重複測量的資料，例如：分析病人每次回診的血壓值(連續性資料)、分析病人每次回診的病況(0 vs. 1: 沒病 vs. 有病)、或分析病人每兩週的癲癇發作次數(次數資料)等，這類相依性的資料需要進階的統計方法才能分析。

相依性資料：重複測量資料彼此有相依性，有些資料即使非出自於個體的重複測量，但仍有相依性的問題待解決，例如：分析家族成員罹病與否和基因的相關性。

Cluster 是什麼？人或家庭

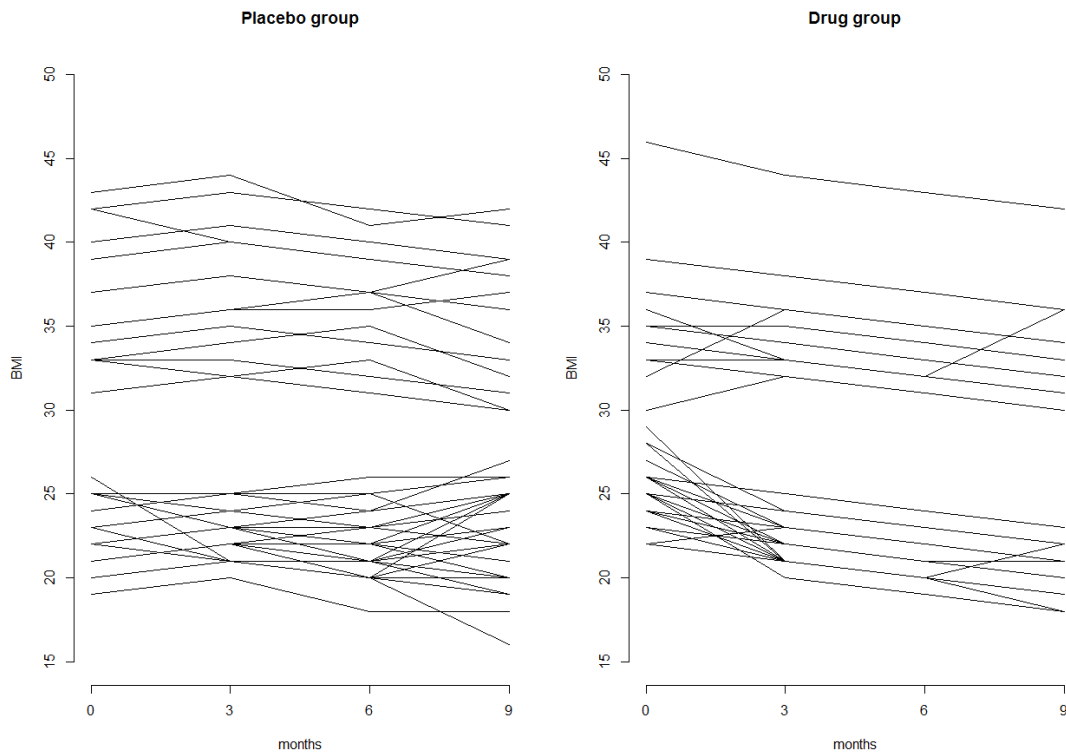
例 1：在課程網頁上有一個 data set，名為 BMIrepeated.csv，內有 100 人參加減肥藥臨床試驗的資料，在試驗開始前，我們先量測受試者的 BMI，變項名稱為 BMI0。然後隨機分派 50 位服用減肥藥(Treatment=1)，另 50 位服用安慰劑(Treatment=0)。之後每三個月再量測一次 BMI，變項名稱為 BMI1, BMI2, BMI3。

例 2：在課程網頁上有一個 data set，名為 Seizure.txt，內有 58 位癲癇患者在五個時段內(8 weeks, 2 weeks, 2 weeks, 2 weeks, 2 weeks)癲癇次數的紀錄，患者分為服用抗癲癇藥物 Progabide(普羅加比)與安慰劑(placebo)兩組。

Step 1 : Graphical presentation of longitudinal data

先以圖形來描述兩組的 BMI / 癲癇次數，可獲得初步的瞭解，但藥效是否顯著，則需要執行統計檢定(本門課後續內容)，而非單由描述性統計可得知。

課堂練習：請以時間為橫軸，BMI / 癲癇次數為縱軸，每一個人以一條線來表示，以任一個統計軟體來繪圖，給讀者一個概括的資料描述。



```
bmi <- read.csv('D:/Repeated/1Introduction/BMIrepeated.csv')
```

```
x <- seq(0,9,3)
```

```
y <- cbind(bmi$BMI0, bmi$BMI1, bmi$BMI2, bmi$BMI3)
```

```
par(mfrow = c(1,2)) # parameter # multiple figures by rows
```

```
plot(x,y[1,],type="l",lwd=1,col=1,lty=1,ylim=c(15,50),axes =  
FALSE,xlab="months",ylab="BMI",main="Placebo group")
```

```
axis(1, at = x, labels = seq(0,9,3))
```

```
axis(2)
```

```
for(subj in 2:50){
```

```
  lines(x,y[subj,],lty=1,lwd=1,col=1,type="l") # line type # line width # color # line type
```

```
}
```

```
plot(x,y[51,],type="l",lwd=1,col=1,lty=1,ylim=c(15,50),axes =  
FALSE,xlab="months",ylab="BMI",main="Drug group")
```

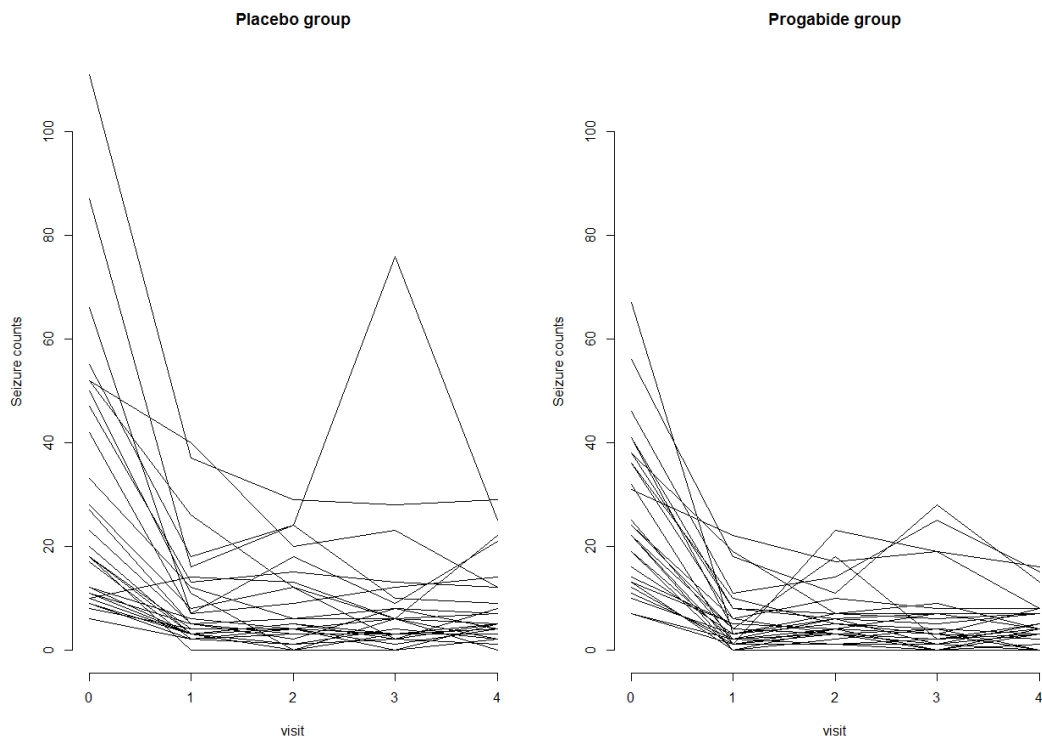
```
axis(1, at = x, labels = seq(0,9,3))
```

```
axis(2)
```

```
for(subj in 52:100){
```

```
  lines(x,y[subj,],lty=1,lwd=1,col=1,type="l")
```

```
}
```



```
seizure <- read.table('D:/Repeated/1Introduction/Seizure.txt')
```

```
x <- seq(0,4,1)
```

```
placebo <- seizure[seizure[,1]==0,2:6]
```

```
progabide <- seizure[seizure[,1]==1,2:6]
```

```
par(mfrow = c(1,2))
```

```
plot(x,placebo[1,],type="l",lwd=1,col=1,lty=1,ylim=c(0,max(placebo,progabide)),axes =  
FALSE,xlab="visit",ylab="Seizure counts",main="Placebo group")
```

```
axis(1, at = x, labels = seq(0,4,1))
```

```
axis(2)
```

```
for(subj in 2:nrow(placebo)){
```

```
  lines(x,placebo[subj,],lty=1,lwd=1,col=1,type="l")
```

```
}
```

```
plot(x,progabide[1,],type="l",lwd=1,col=1,lty=1,ylim=c(0,max(placebo,progabide)),axes =  
FALSE,xlab="visit",ylab="Seizure counts",main="Progabide group")
```

```
axis(1, at = x, labels = seq(0,4,1))
```

```
axis(2)
```

```
for(subj in 2:nrow(progabide)){
```

```
  lines(x,progabide[subj,],lty=1,lwd=1,col=1,type="l")
```

```
}
```

A motivating example for **random effect model**:

Suppose statistical grades can be modelled as $E(Y_{ij}) = 100 + 1 \cdot \text{gender}_i - 1 \cdot \text{age}_{ij}$, related to a student's gender and age. For two male students, if their grades and age are

	statistical grades	gender	age
Student A, 1 st exam	90	1	30
Student A, 2 nd exam	92	1	31
Student A, 3 rd exam	88	1	32
Student A, 4 th exam	90	1	33
Student B, 1 st exam	10	1	30
Student B, 2 nd exam	8	1	31
Student B, 3 rd exam	11	1	32
Student B, 4 th exam	9	1	33

Please calculate the residuals of these eight observations.

Subject-specific effect

Example 1.2 (Shoukri and Chaudhary) (milk.sas)

There are 10 farms. Each farm provided 12 observations representing the average milk yield per cow per day for each month.

Model: $Y_{ij} = \mu + b_i + e_{ij}$, study自己的抽樣分布的誤差

不同study的effect呈現常態分布(random effect) 非唯一真值，每個study有自己的效果值

where Y_{ij} is the milk of the j th observation from the i th farm, μ is the overall mean, b_i is the

effect of the i th farm, and e_{ij} is the random error term. Assume that $b_i \sim Normal(0, \sigma_b^2)$ and

$e_{ij} \sim Normal(0, \sigma_e^2)$, and b_i and e_{ij} are independent of each other.

Under this setup,

$$V(Y_{ij}) = V(\mu + b_i + e_{ij}) = V(b_i + e_{ij}) = V(b_i) + V(e_{ij}) = \sigma_b^2 + \sigma_e^2$$

$$Cov(Y_{ij}, Y_{il}) = Cov(\mu + b_i + e_{ij}, \mu + b_i + e_{il})$$

$$= Cov(b_i + e_{ij}, b_i + e_{il})$$

$$= Cov(b_i, b_i) + Cov(b_i, e_{il}) + Cov(e_{ij}, b_i) + Cov(e_{ij}, e_{il})$$

$$= \sigma_b^2 + 0 + 0 + 0$$

$$= \sigma_b^2, \text{ when } j \neq l$$

Covariance: 共變異數

Therefore, the correlation between any pair of observations within a cluster (here, farm) is

$$Corr(Y_{ij}, Y_{il}) = \frac{Cov(Y_{ij}, Y_{il})}{\sqrt{Var(Y_{ij})Var(Y_{il})}} = \frac{\sigma_b^2}{\sqrt{(\sigma_b^2 + \sigma_e^2)(\sigma_b^2 + \sigma_e^2)}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}, \text{ when } j \neq l.$$

This correlation is known as the intraclass correlation (intraclass correlation) (ICC, 組內相關係數). (intra- 表示內部之意) 衡量信度的方法

Homework (5 分) : (1) ICC 範圍為何？

(2) ICC=0 及 1 時各是什麼意思？

(3) 請計算本例中的 ICC？