統計諮詢 - 作業 5

國立成功大學統計學系暨數據科學研究所

廖傑恩（RE6094028）

2021-05-21

1 Exercise 12.4

1.1 問題敘述

一位農藝學家想比較 5 種氮肥肥料與沒有肥料對於大麥產量的效果。他採用隨機集區設計（randomized block design）進行實驗，將田地分成 4 個集區（block），每個集區使用不同的土壤型態。每一種土壤型態被隨機分成 6 個 plot，每個 plot 會被隨機分配到一種處置，包含 (NH4)SO4、NH4NO3、CO(NH2)2、Ca(NO3)2、NaNO3 與控制組（沒有使用肥料）。

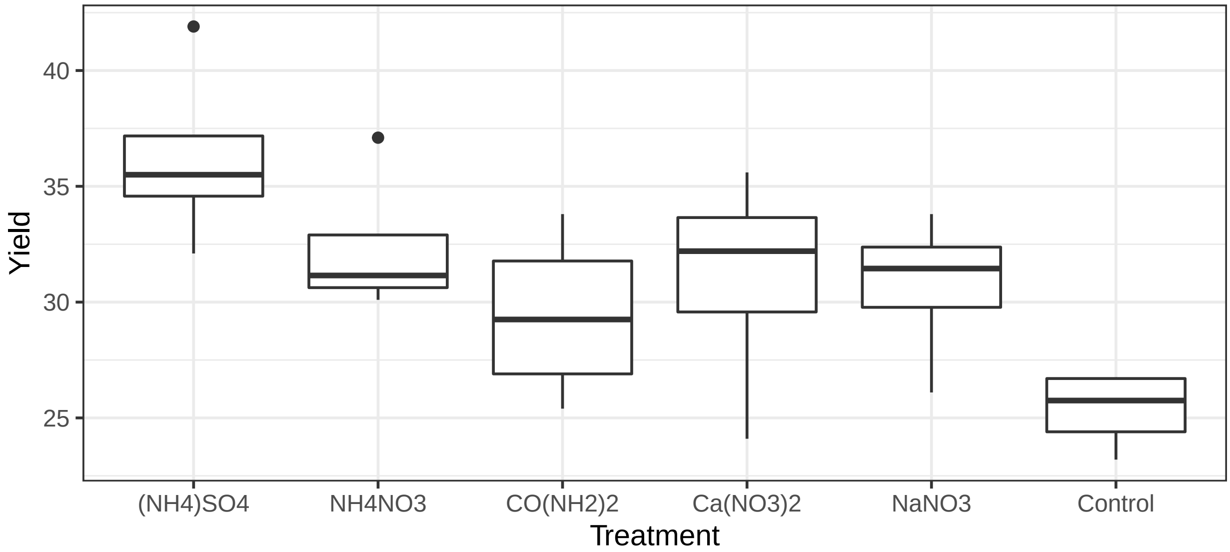
1.2 資料介紹

資料來自 Peterson (1985)。資料共有 24 列、3 個變項，每一列為一個 plot 的大麥產量、處置（肥料種類或沒有肥料）與土壤型態。變項說明如下表：

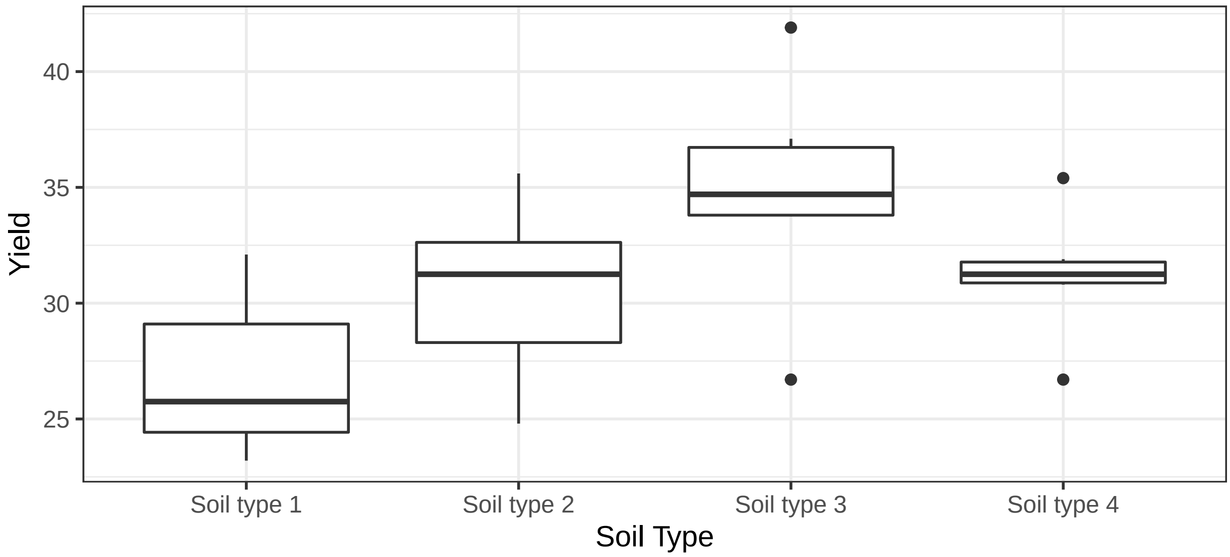
- `Yield`: 大麥產量。
- `Trt`: 處置，包含 (NH4)SO4、NH4NO3、CO(NH2)2、Ca(NO3)2、NaNO3 與控制組（沒有使用肥料）6 種。
- `Soid`: 土壤型態。

1.3 資料探索

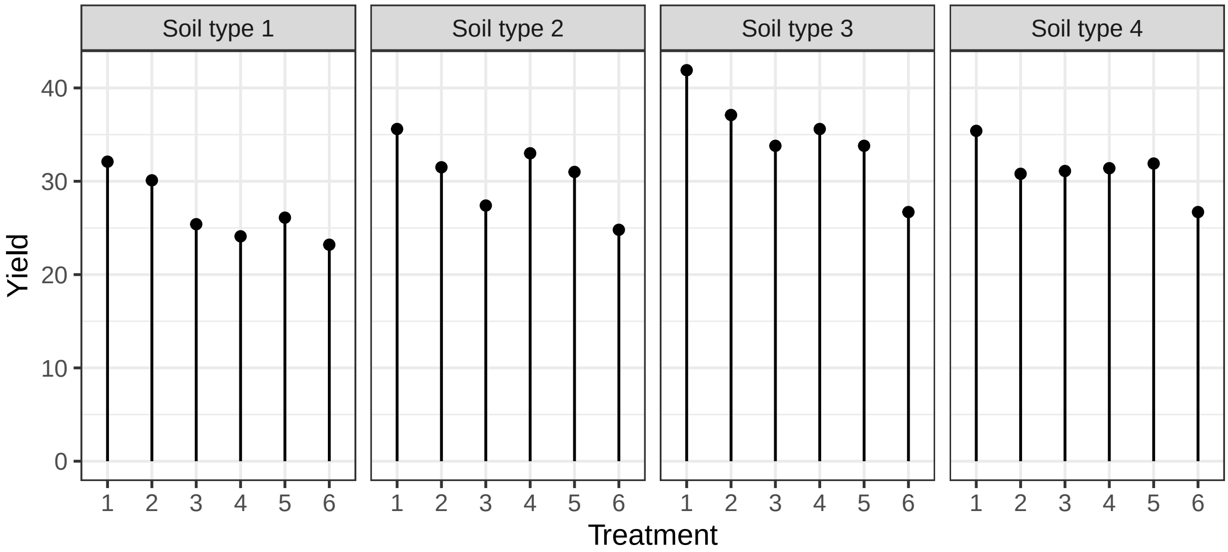
下圖是 6 個處置下大麥產量的盒鬚圖，可以看出不同處置下大麥產量似乎有所差異：處置 1（使用 (NH4)SO4 肥料）下大麥產量明顯較多，處置 6（不使用肥料）下大麥產量明顯較低。



下圖是 4 個土壤型態下大麥產量的盒鬚圖，可以看出不同土壤型態下大麥產量似乎有所差異，因此在之後的分析中，我們應控制住土壤型態對於大麥產量的效果。



下圖是 6 個處置下大麥產量的棒棒糖圖（Lollipop plot），並以土壤型態分層來看。可以看出不論土壤型態為何，不同處置下大麥產量都似乎有所差異：處置 1 下大麥產量明顯較多，處置 6 下大麥產量明顯較低。



1.4 資料分析

我們以二因子變異數分析（analysis of variance, ANOVA）進行分析。

1.4.1 變數與模型定義

定義模型：

$$y_{ij} = \mu + \tau_i + \rho_j + \epsilon_{ij}$$

其中 y_{ij} 是在土壤型態 j 下接受處置 i 的大麥產量；μ 是產量總平均；τ_i 是處置 i 對於產量的效果；ρ_j 是集區 j（土壤型態 j）對於產量的效果；而 ϵ_i 是隨機殘差且 $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$。

1.4.2 研究假設

- 虛無假設（H_0）：$\tau_i = 0,\ \forall i = 1, 2, ..., 6$
- 對立假設（H_1）：$Not\ H_0$

1.4.3 前提檢驗

ANOVA 有若干個前提假設（assumption），在進行正式的分析之前，我們先檢查資料是否符合這些假設。

1. 獨變項須為類別變數（categorical variable），依變項必須是連續變數（continuous variable）

此分析獨變項為處置類型（肥料種類或不用肥料）以及土壤型態，前者為含有 6 個水準（level）的類別變數，後者則為含有 4 個水準的類別變數；依變項為大麥產量，為連續變項。符合。

2. 各組樣本依變項獨立

此分析中，不同處置下的大麥產量互不影響彼此，不同土壤型態（隨機分配到不同塊的集區）下的大麥產量也互不影響彼此，符合前提假設。

3. 變異數同質（homogeneity of variance）：各組依變項的變異數必須相等。

- 針對依變項進行常態檢定

在常態分布下，以 Bartlett 檢定變異數同質性會有較高的統計檢定力（power）（Lim & Loh，1996），因此在進行變異數同質性檢定前，我們先以 Shapiro-Wilk 常態檢定法對依變項（Y）進行常態檢定，檢定的假說如下，顯著水準設定為 0.05。

$$H_0 : Y \sim ND \;\; v.s. \;\; H_1 : Y \; does \; not \sim ND$$

檢定結果：檢定統計量為 0.9656，其 p 值為 0.5606，不小於顯著水準，因此我們不拒絕 H_0，也就是說我們沒有足夠的證據證明母體分配不服從常態分佈。

- 變異數同質檢定

研究假說如下：

$$H_0 : \sigma^2_{i=1} = ... = \sigma^2_{i=6} \; vs \; H_1 : Not \; H_0$$

$$H_0 : \sigma^2_{j=4} = ... = \sigma^2_{j=4} \; vs \; H_1 : Not \; H_0$$

我們同樣令顯著水準為 0.05。檢定結果：針對「處置」與「土壤型態」兩獨變項之檢定統計量 $Barlett's \; k^2$ 分別為 2.8149 與 1.6029，其 p 值分別為 0.7285 與 0.6587，都不小於顯著水準，因此在兩組假設檢定中，我們都不拒絕 H_0，意味著我們無法證明至少有一組母體變異數與其他組不同，也就是說我們沒有足夠的證據證明變異數同質性不存在。

我們再以 Levene 檢定做核對：

同樣令顯著水準為 0.05。針對「處置」與「土壤型態」兩獨變項之檢定 0 的 p 值分別為 0.8833 與 0.7263，都不小於顯著水準，因此在兩組假設檢定中，我們都不拒絕 H_0，意味著我們無法證明至少有一組母體變異數與其他組不同，也就是說我們沒有足夠的證據證明變異數同質性不存在。通過此前提假設。

4. 殘差（residuals）服從常態分配。

待配適完模型後診斷。

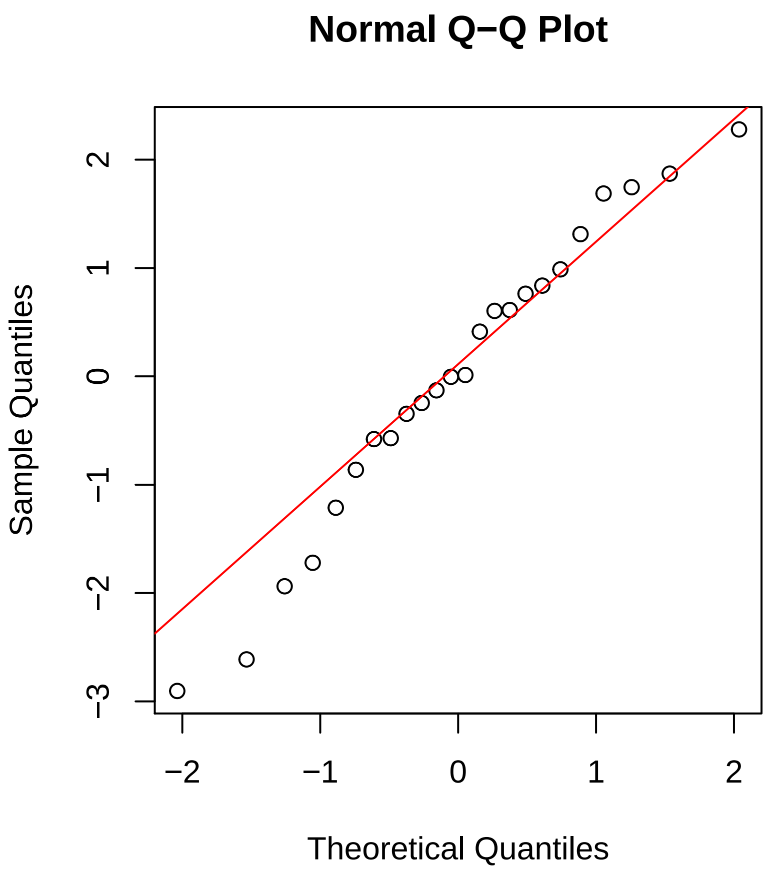
以上步驟顯示，在我們的資料中，ANOVA 的前提假設均滿足（殘差常態假設待檢驗），因此我們可以進行 ANOVA。

1.4.4 分析結果

變異來源	自由度	殘差平方和	均方誤差	檢定統計量	p 值
處置（Trt）	5	255.28	51.06	17.20	9.7e-6
土壤（Soil）	3	192.55	64.18	21.62	1.1e-5
殘差	15	44.53	2.97		

令顯著水準為 0.05，意味著我們在假設檢定中犯型一錯誤（type I error, 當 H_0 為真卻拒絕 H_0）的機率為 5%。二因子 ANOVA 分析結果如上表，兩檢定統計量之 p 值遠小於顯著水準，因此在兩個效果的檢定中，我們拒絕虛無假設，表示我們有充分證據支持並非所有 τ_i 皆為 0，也就是說，在固定集區（土壤型態）的效果下，不同處置（肥料種類或不用肥料）下的大麥產量有顯著差異；我們也有充分證據支持並非所有 ρ_j 皆為 0，也就是說，不同土壤型態下的大麥產量有所差異。

1.4.4.1 模型診斷



配適模型之後，我們應對殘差的常態假設進行診斷。我們先繪製常態 Q-Q 圖（如上圖），可發現資料點沒有明顯偏離 45 度線，顯示殘差可能服從常態分配。我們以 Shapiro-Wilk 檢定檢驗模型殘差 ϵ_i 是否為常態分配，令顯著水準為 0.05，其假設如下:

$$H_0 : \epsilon_{ij} \sim ND \ v.s. \ H_1 : \epsilon_{ij} \ does \ not \sim ND$$

檢定結果：檢定統計量 W 為 0.9707，p 值為 0.6833，大於顯著水準，因此我們不拒絕 H_0，表示我們沒有足夠的證據顯示殘差不服從常態分配。

我們接著對殘差進行顯著水準為 0.05 的 NCV 變異數同質檢定（NCV test），其檢定假設為：H_0: 殘差變異數具有同質性 v.s. H_1: 殘差變異數不具有同質性。檢定結果：檢定統計量為 0.7221，其 p 值為 0.3954，不小於顯著水準，因此我們不拒絕 H_0，表示殘差變異數具有同質性。

以上檢定顯示模型診斷通過。

1.4.4.2 事後比較

由於模型診斷通過，我們可以相信二因子 ANOVA 的分析結果：在固定集區（土壤型態）的效果下，不同處置（肥料種類或不用肥料）下的大麥產量有顯著差異。我們接著進行事後比較，對各種處置下的大麥產量進行與控制組（不施以肥料）比較。因為想要知道肥料是否會影響產量，因此必須要重新調整變數，去除土壤這個區集效應。重新定義的依變數如下：

$$Y'_{ij} = Y_{ij} - \rho_j = \mu + \tau_i + \epsilon_{ij}$$

假設檢定如下：

$$\begin{cases} H_0 : \mu_6 \geq \mu_i \\ H_1 : \mu_6 < \mu_i \end{cases}, \ \forall i = 1, ..., 5$$

Dunnett 檢定的結果如下表，其顯示施以任何種肥料都會得到比控制組顯著多的產量（顯著水準為 0.05）。

比較	差異估計值	95%CI 下界	95%CI 上界	p 值
1-6	10.90	8.2213	Inf	6.6e-9
2-6	7.02	4.3463	Inf	2.6e-5
3-6	4.08	1.4963	Inf	0.0076
4-6	5.68	2.9963	Inf	6.6e-4
5-6	5.35	2.6713	Inf	6.2e-4

1.5 結論

肥料與土壤的型態皆與產量有關，所以對土壤進行區集化是必要的。在固定集區（土壤型態）的效果下，不同處置（肥料種類或不用肥料）下的大麥產量有顯著差異，其中處置 6（不使用肥料，控制組）的大麥產量顯著小於其他處置，也就是說，使用任何肥料後的大麥產量都會顯著高於控制組，顯示各種肥料都是有效的。

2 Exercise 13.1

2.1 問題敘述

研究者操作了一項實驗，想得知在 4 種用於人工心臟的瓣膜中，何種有最大流量梯度（單位：毫米汞柱），也就是探究何種可以最大化血壓控制。每種瓣膜都在相同的 6 個脈衝速率下保持流動。每種瓣膜的實驗都隨機進行 2 回合。回合次別（run）是一個鑲嵌於瓣膜（valve）這個因子內的隨機因子。

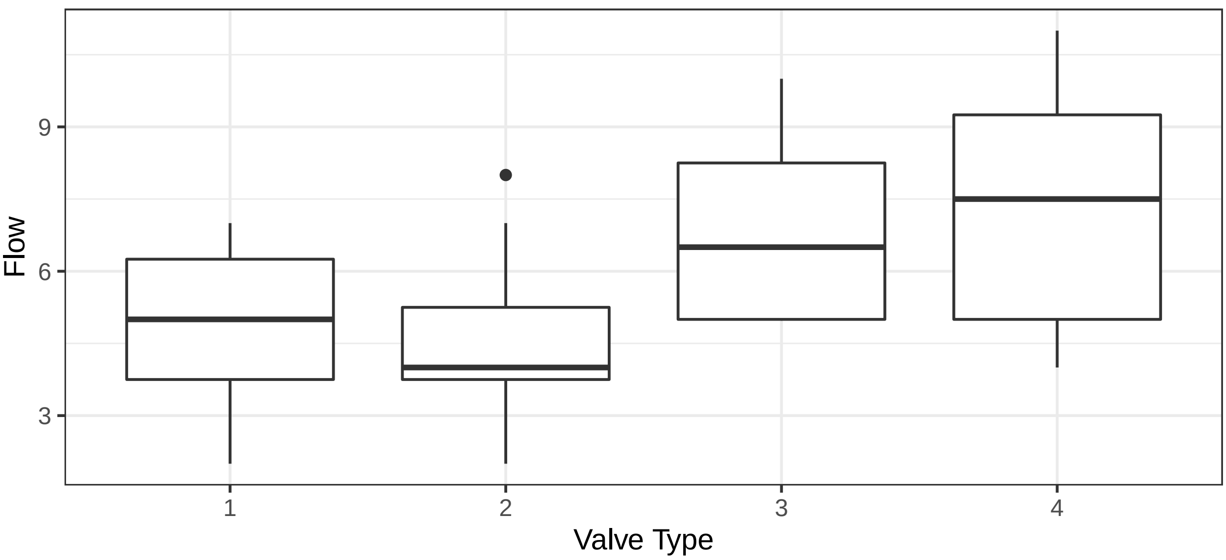
2.2 資料介紹

資料來自於 Anderson 與 McLean（1974），包含 48 列、4 個變項，每一列為每一次試驗（trial）的資訊，變項說明如下：

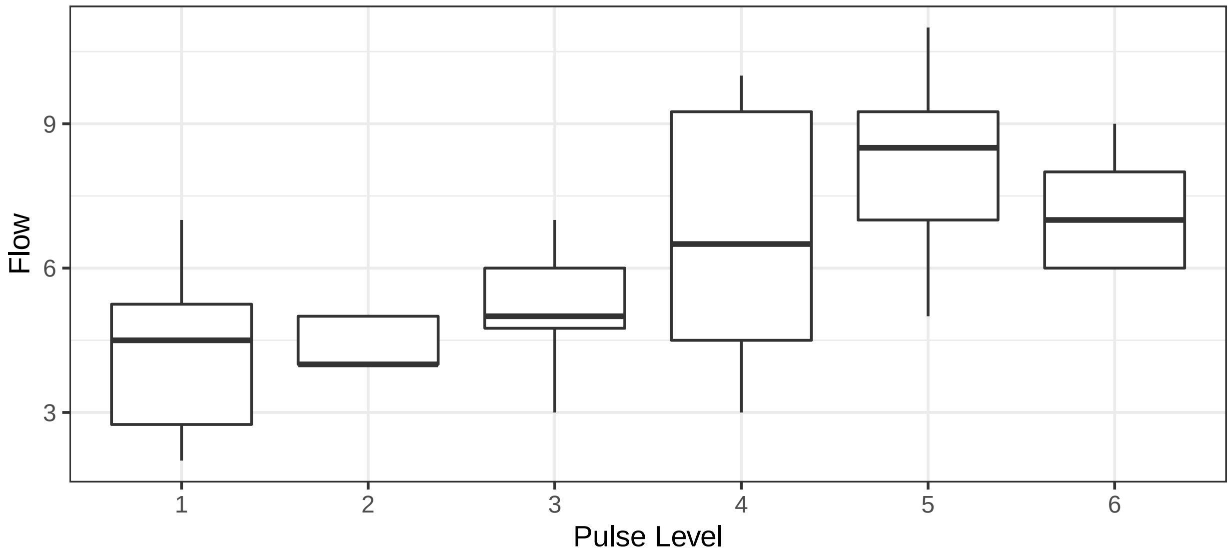
- `Flow`: 最大血流梯度（單位：毫米汞柱）
- `Valve`: 用於人工心臟的瓣膜類型，為含有 4 個水準（level）的類別變項
- `Run`: 回合次別，為含有 8 個水準的類別變項，鑲嵌於 `Valve` 的 4 個小準之中
- `Pulse`: 脈衝速率，為含有 6 個水準、具有順序尺度的類別變項

2.3 資料探索

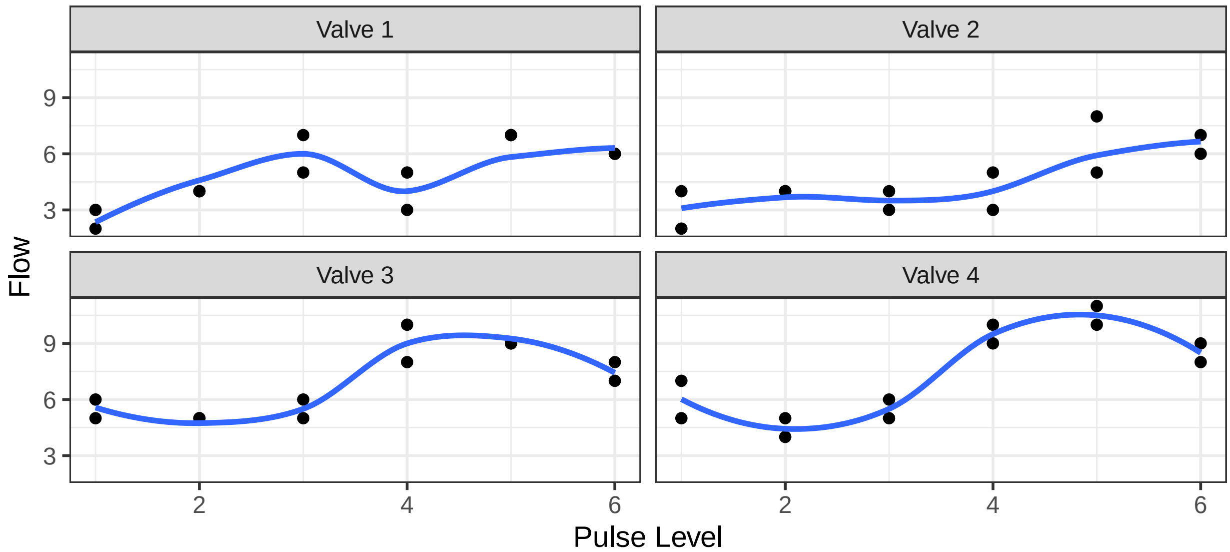
下圖是 4 種瓣膜的最大血流梯度盒鬚圖，可以發現，瓣膜 3 與瓣膜 4 的最大血流梯度中位數較瓣膜 1 與瓣膜 2 的高一點，顯示不同瓣膜的最大血液梯度可能有所差異。



下圖是 6 種脈衝速率水準的最大血流梯度盒鬚圖，可以發現，基本上脈衝速率水準較高者，最大血液梯度的中位數也較高，顯示脈衝速率水準可能與最大血液梯度有關，可能需要將其效果固定住。



下圖是 6 種脈衝速率水準的最大血流梯度點圖，並將 4 種瓣膜的資料分層來看，可以發現，在 4 種瓣膜中，在不同的脈衝速率水準下，最大血液梯度的中位數有所差異。



2.4 資料分析

我們以二因子巢狀變異數分析（Nested analysis of variance, ANOVA）進行分析。

2.4.1 變數與模型定義

我們的模型考慮兩獨變項交互作用：

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(i)} + \epsilon_{ijk}$$

$$i = 1, ..., 4;\ j = 1, ..., 6;\ k = 2(i-1) + l;\ l = 1, 2$$

$$\sum_{1=1}^{4}\alpha_i = \sum_{j=1}^{6}\beta_j = 0$$

其中 y_{ij} 是在脈衝速率為水準 j 下，編號 i 的瓣膜型態的最大血液梯度；μ 是最大血液梯度總平均；α_i 是瓣膜型態 i 對於最大血液梯度的主效果；β_j 是脈衝速率水準 j 對於最大血液梯度的主效果；$(\alpha\beta)_{ij}$ 是瓣膜型態 i 與脈衝速率水準 j 對於最大血液梯度的交互作用效果；γ_k 是回合次數對於最大血液梯度的效果且 $\gamma_{k(i)} \overset{iid}{\sim} N(0,\sigma_\gamma^2)$；而 ϵ_{ijk} 是隨機殘差且 $\epsilon_{ijk} \overset{iid}{\sim} N(0,\sigma_\epsilon^2)$。

2.4.2 研究假設

- 檢定瓣膜型態的效果：$\begin{cases} H_0 : \alpha_1 = ... = \alpha_4 = 0 \\ H_1 : Not\ H_0 \end{cases}$
- 檢定脈衝速率的效果：$\begin{cases} H_0 : \beta_1 = ... = \beta_6 = 0 \\ H_1 : Not\ H_0 \end{cases}$
- 檢定交互作用效果：$\begin{cases} H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{21} = ... = (\alpha\beta)_{46} = 0 \\ H_1 : Not\ H_0 \end{cases}$

2.4.3 前提檢驗

ANOVA 有若干個前提假設，在進行正式的分析之前，我們先檢查資料是否符合這些假設。

1. 獨變項須為類別變數，依變項必須是連續變數

此分析獨變項為瓣膜類型以及脈衝速率水準，前者為含有 4 個水準的類別變數，後者則為含有 6 個水準的類別變數；依變項為最大血液梯度，為連續變項。符合。

2. 各組樣本依變項獨立

此分析中，不同型態的瓣膜量測到的最大血液梯度互不影響彼此，不同脈衝速率水準下的最大血液梯度也互不影響彼此，符合前提假設。

3. 變異數同質（homogeneity of variance）：各組依變項的變異數必須相等。

- 針對依變項進行常態檢定

在常態分布下，以 Bartlett 檢定變異數同質性會有較高的統計檢定力（power）（Lim & Loh，1996），因此在進行變異數同質性檢定前，我們先以 Shapiro-Wilk 常態檢定法對依變項（Y）進行常態檢定，檢定的假說如下，顯著水準設定為 0.05。

$$H_0 : Y \sim ND\ \ v.s.\ \ H_1 : Y\ does\ not \sim ND$$

檢定結果：檢定統計量為 0.958，其 p 值為 0.084，不小於顯著水準，因此我們不拒絕 H_0，也就是說我們沒有足夠的證據證明母體分配不服從常態分佈。

- 變異數同質檢定

研究假說如下：

$$H_0 : \sigma^2_{i=1} = ... = \sigma^2_{i=6} \ vs \ H_1 : Not \ H_0$$

$$H_0 : \sigma^2_{j=4} = ... = \sigma^2_{j=4} \ vs \ H_1 : Not \ H_0$$

我們同樣令顯著水準為 0.05。檢定結果：檢定統計量 $Barlett's\ k^2$ 分別為 1.6566，其 p 值為 0.6466，不小於顯著水準，因此我們不拒絕 H_0，意味著我們無法證明至少有一組母體變異數與其他組不同，也就是說我們沒有足夠的證據證明變異數同質性不存在。

我們再以 Levene 檢定做核對：

同樣令顯著水準為 0.05。檢定的 p 值為 0.2395，同樣不小於顯著水準，因此我們不拒絕 H_0，意味著我們無法證明至少有一組母體變異數與其他組不同，也就是說我們沒有足夠的證據證明變異數同質性不存在。通過此前提假設。

4. 殘差服從常態分配。

待配適完模型後診斷。

以上步驟顯示，在我們的資料中，ANOVA 的前提假設均滿足（殘差常態假設待檢驗），因此我們可以進行 ANOVA。

2.4.4 分析結果

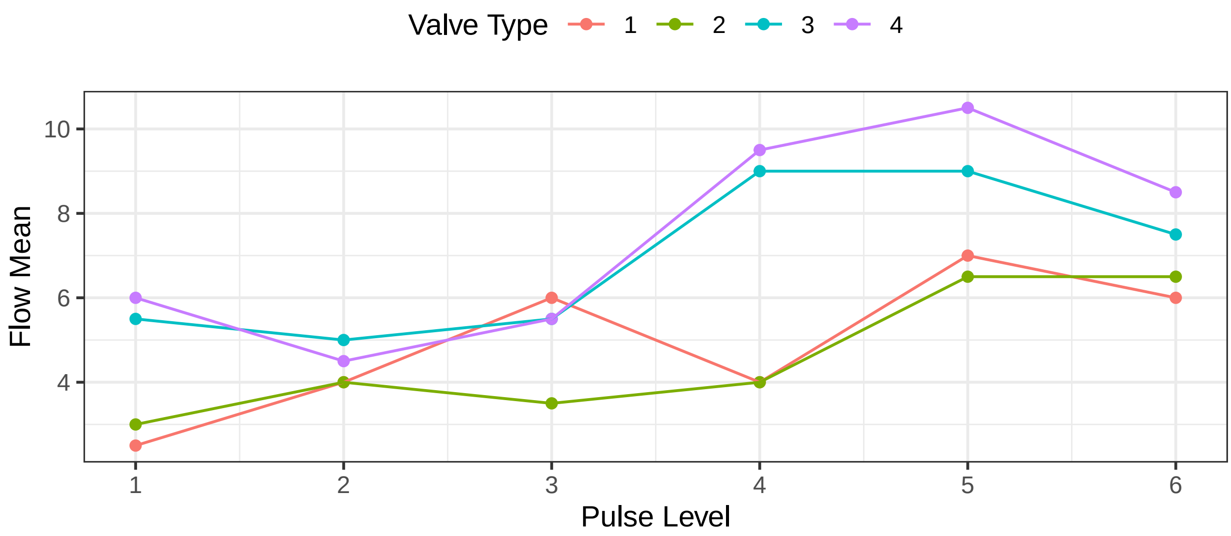
變異來源	自由度 1	自由度 2	殘差平方和	均方誤差	檢定統計量	p 值
瓣膜型態（V）	5	20	105.42	21.08	28.11	2.1e-8
脈衝速率（P）	3	4	30.96	10.32	13.76	0.0142
V 與 P 交互作用	15	20	38.25	2.55	3.40	0.0059
殘差	20	-	15.00	0.75	-	-

令顯著水準為 0.05，意味著我們在假設檢定中犯型一錯誤（type I error, 當 H_0 為真卻拒絕 H_0）的機率為 5%。ANOVA 結果如上表。針對瓣膜型態主效果、脈衝速率主效果與瓣膜型態與脈衝速率交互作用這三個檢定的檢定統計量的 p 值為都小於顯著水準。因此在這三組的檢定中，我們都拒絕虛無假設，表示我們有充分證據支持：

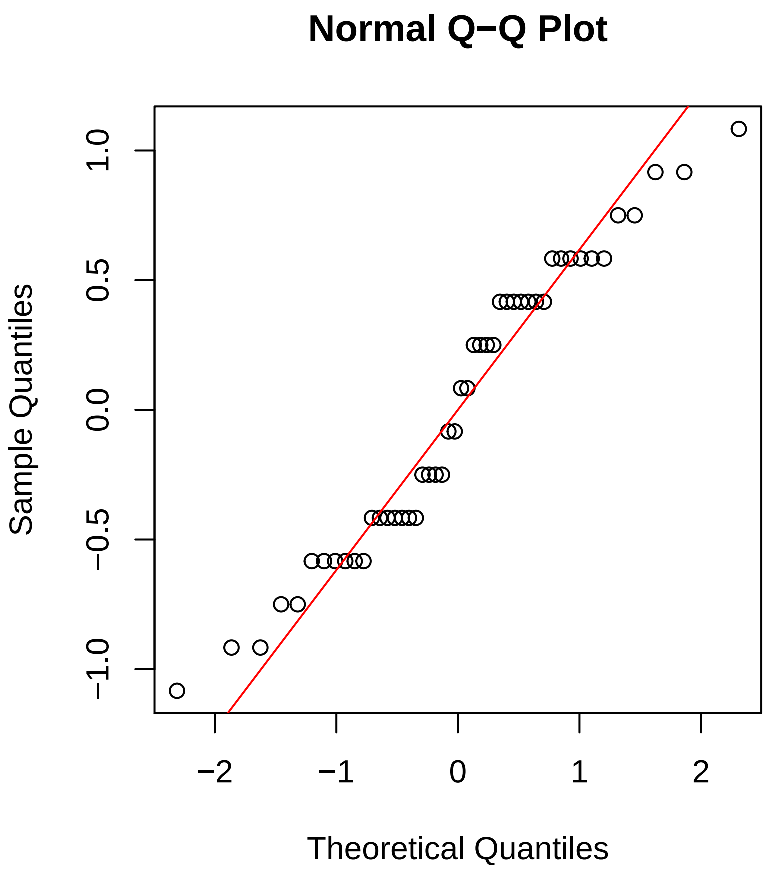
1. 並非所有 α_i 皆為 0，也就是說，瓣膜型態的主效果顯著，使用包含不同瓣膜的人工心臟，最大血液梯度有所差異。

2. 並非所有 β_j 皆為 0，也就是說，脈衝速率的主效果顯著，在不同脈衝速率水準下，最大血液梯度有所差異。

3. 並非所有 $(\alpha\beta)_{ij}$ 皆為 0，也就是說，瓣膜型態與脈衝速率的交互作用顯著，使用包含不同瓣膜的人工心臟，脈衝速率與最大血液梯度的關聯有所差異，如下圖所示。



2.4.4.1 模型診斷



配適模型之後，我們應對殘差的常態假設進行診斷。我們先繪製常態 Q-Q 圖（如上圖），可發現資料點沒有很明顯偏離 45 度線，顯示殘差可能服從常態分配。我們以 Shapiro-Wilk 檢定檢驗模型殘差 ϵ_i 是否為常態分配，令顯著水準為 0.01，其假設如下:

$$H_0 : \epsilon_{ijk} \sim ND \ v.s. \ H_1 : \epsilon_{ijk} \ does \ not \sim ND$$

檢定結果：檢定統計量 W 為 0.9506，p 值為 0.0423，大於顯著水準，因此我們不拒絕 H_0，表示我們沒有足夠的證據顯示殘差不服從常態分配。

對殘差進行顯著水準為 0.05 的 NCV 變異數同質檢定 (NCV test)，其檢定假設為: H_0: 殘差變異數具有同質性 v.s. H_1: 殘差變異數不具有同質性。檢定結果：檢定統計量為 0.1819，其 p 值為 0.6697，不小於顯著水準，因此我們不拒絕 H_0，表示殘差變異數具有同質性。

以上檢定顯示模型診斷通過。

2.4.4.2 事後比較

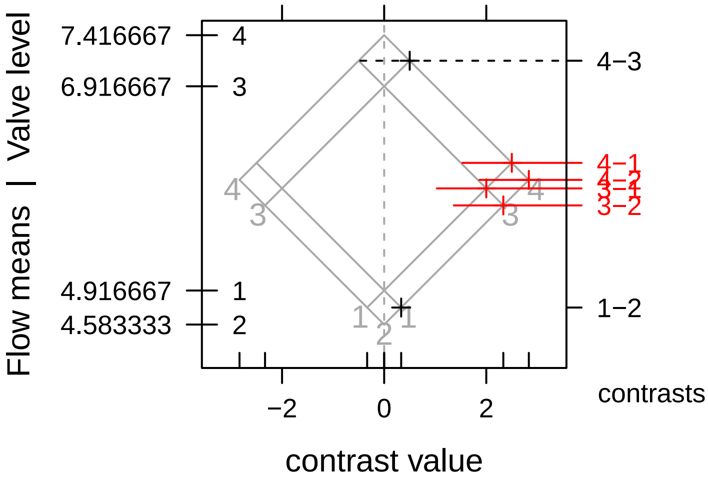
由於模型診斷通過，我們可以相信多因子 ANOVA 的分析結果：使用包含不同瓣膜的人工心臟，最大血液梯度有所差異。我們想得知道使用哪種瓣膜的人工心臟，其最大血液梯度最高，所以接著進行事後比較，對各種處置下的大麥產量進行兩兩比較，假設檢定如下:

$$\begin{cases} H_0 : \mu_i \geq \mu_{i'} \\ H_1 : \mu_i < \mu_{i'} \end{cases} , \ \forall i = 1, ..., 4, \ i' = 1, ..., 4, \ i > i'$$

因為各處置樣本大小相同，因此我們選用 Tukey method 進行事後比較，並將整體的信心水準控制在 95%，以避免型一誤差發生機率大幅膨脹。下表為各型態的瓣膜兩兩配對比較的差異估計值，其中「組間差異平均估計值」為-前面組別之平均值減去-後面組別之平均值，例如「4-1」的「組間差異平均估計值」為瓣膜 4 組別之平均值減去瓣膜 1 組別之平均值。

	組間差異平均估計值	CI 下界	CI 上界
4-3	0.5000	-0.4685	Inf
4-1	2.5000	1.5315	Inf
4-2	2.8333	1.8648	Inf
3-1	2.0000	1.0315	Inf
3-2	2.3333	1.3648	Inf
1-2	0.3333	-Inf	1.3019

下圖則為將上表視覺化的結果。每一個 bar 為一兩兩組別配對之信賴區間，標註紅色者表示其信賴區間沒包含 0 值，也就是說在整體信賴區間為 95% 下，這些配對的組別間的大麥產量有顯著差異。



由圖表可知:

- 使用瓣膜 4 的人工心臟的最大血液梯度顯著大於使用瓣膜 1 的人工心臟
- 使用瓣膜 4 的人工心臟的最大血液梯度顯著大於使用瓣膜 2 的人工心臟
- 使用瓣膜 3 的人工心臟的最大血液梯度顯著大於使用瓣膜 1 的人工心臟
- 使用瓣膜 3 的人工心臟的最大血液梯度顯著大於使用瓣膜 2 的人工心臟
- 使用瓣膜 3 的人工心臟的最大血液梯度與使用瓣膜 4 的人工心臟並無顯著差異
- 使用瓣膜 1 的人工心臟的最大血液梯度與使用瓣膜 2 的人工心臟並無顯著差異

2.5 結論

在不同的脈衝速率下，最大血液梯度有所差異。且使用不同瓣膜的人工心臟之最大血液梯度有所差異，其比較如下:

$$\mu_3 \approx \mu_4 > \mu_1 \approx \mu_2$$

相較於使用瓣膜 1 或瓣膜 2 的人工心臟，使用瓣膜 3 或瓣膜 4 的人工心臟可以得到較佳的血壓控制。