

實驗設計與分析 HW3 solution

Problem 1.

Do Problem 1 in Chapter 2 of textbook.

For the pulp experiment obtain the 95% simultaneous confidence intervals for the six pairs of treatment differences using the Bonferroni method and the Tukey method. Which gives shorter intervals?

```
kable(cbind(dat1[1:5,], dat1[6:10,], dat1[11:15,], dat1[16:20,]),
      format = "markdown", caption = "Pulp Experiment Data", align = 'c') %>%
  kable_styling(position = "center")
```

Table 1: Pulp Experiment Data

bright	operator	bright	operator	bright	operator	bright	operator
59.8	a	59.8	b	60.7	c	61.0	d
60.0	a	60.2	b	60.7	c	60.8	d
60.8	a	60.4	b	60.5	c	60.6	d
60.8	a	59.9	b	60.9	c	60.5	d
59.8	a	60.0	b	60.3	c	60.5	d

首先配適以下模型

$$y_{ij} = \eta + \tau_i + \epsilon_{ij}, \quad \text{for } i = 1, \dots, 4, \quad j = 1, \dots, 5.$$

其中 η 代表 overall mean, τ_i 代表 operator effect, $\epsilon_{ij} \sim N(0, \sigma^2)$ 代表誤差項, i 代表 a、b、c、d 四個 operators, k 則代表每個 operators 的第幾次實驗。

建立以下假設檢定

$$H_0^{ij} : \tau_i = \tau_j \quad v.s. \quad H_1^{ij} : \tau_i \neq \tau_j \quad \text{for } i, j = 1, \dots, 4, i \neq j$$

其中 t-statistic 為

$$t_{ij} = \frac{\bar{y}_i - \bar{y}_j}{\hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

根據題意使用下列兩個方法建立信賴區間:

1. Bonferroni method:

根據 Bonferroni method 及 t-statistic 我們能有以下信賴區間

$$\left[t_{ij} - t_{N-k, \frac{\alpha}{2k'}}, t_{ij} + t_{N-k, \frac{\alpha}{2k'}} \right]$$

其中 $k' = \binom{k}{2}$, k 為 treatment 個數, k' 則為 test 的次數, 根據計算我們能有以下結果

```
attach(dat1)
bonf1 = bonfCI(bright, operator)
kable(bonf1$summary, format='markdown', caption = "95% Bonferroni confidence intervals")
```

Table 2: 95% Bonferroni confidence intervals

	Diff	Lower	Upper	Decision	Adj. p-value
mua-mub	0.18	-0.44018	0.80018	FTR H0	1
mua-muc	-0.38	-1.00018	0.24018	FTR H0	0.503359
mub-muc	-0.56	-1.18018	0.06018	FTR H0	0.091504
mua-mud	-0.44	-1.06018	0.18018	FTR H0	0.291823
mub-mud	-0.62	-1.24018	0.00018	FTR H0	0.050093
muc-mud	-0.06	-0.68018	0.56018	FTR H0	1

2. Tukey method:

Tukey method 及 t-statistic 我們能有以下信賴區間

$$\left[t_{ij} - \frac{1}{\sqrt{2}} q_{k, N-k, \alpha}, t_{ij} + \frac{1}{\sqrt{2}} q_{k, N-k, \alpha} \right]$$

其中 k 為 treatment 個數， $q_{k, N-k, \alpha}$ 則為 student range 的 upper α -quantile，根據計算我們能有以下結果

```
attach(dat1)
tukey1 = tukeyCI(bright, operator)
kable(tukey1$summary, format='markdown', caption = "95% Tukey-Kramer confidence intervals")
```

Table 3: 95% Tukey-Kramer confidence intervals

	Diff	Lower	Upper	Decision	Adj. p-value
mua-mub	0.18	-0.40981	0.76981	FTR H0	0.818543
mua-muc	-0.38	-0.96981	0.20981	FTR H0	0.290304
mub-muc	-0.56	-1.14981	0.02981	FTR H0	0.065794
mua-mud	-0.44	-1.02981	0.14981	FTR H0	0.184479
mub-mud	-0.62	-1.20981	-0.03019	Reject H0	0.037669
muc-mud	-0.06	-0.64981	0.52981	FTR H0	0.991078

比較兩者的信賴區間半徑，因為 $t_{16, \frac{0.05}{12}} = 3.008334 > 2.86102 = \frac{1}{\sqrt{2}} q_{4, 16, 0.05}$ ，因此 Tukey method 的信賴區間會比 Bonferroni method 的信賴區間短。

Problem 2.

Do Problem 2 in Chapter 2 of textbook.

(a) For the pulp experiment show that neither the Bonferroni nor the Tukey method declares any pair of treatments as different at the 0.01 level.

根據 Problem 1. 中的結果可以發現，若設定 $\alpha = 0.01$ ，則所有 p-value 皆大於 $\alpha = 0.01$ 不拒絕 H_0 ，故沒有足夠證據說明 operators 間的實驗結果存在差異。

(b) How do you reconcile the finding in (a) with the result in Section 2.1 that the F test rejects the null hypothesis H_0 at the 0.05 level? After rejecting the null hypothesis, do you expect the multiple comparison method to identify at least one pair of treatments as different? (Hint: One is at the 0.01 level while the other is at the 0.05 level.)

Section 2.1 中選用的顯著水準 α 與 Problem 1. 中使用的顯著水準不同，導致結果不同，加上檢定的方法不同，所以 Section 2.1 與 Problem 1. 結論不同是合理的。

(c) Recall that the p value for the observed F statistic value 4.20 is 0.02. How can you use this fact to reach the same conclusion in (a) without actually performing the multiple comparisons? (Hint: Use the relationship between the p value and the significance level of the F test.)

根據題目所述，Problem 1. 中使用 F-test 的 p-value=0.02，所以若設定 $\alpha = 0.01$ ，則最後結論不拒絕 H_0 ，故沒有足夠證據說明 operators 間的實驗結果存在差異，得到與 (a) 相同結論。

Problem 3.

Do Problem 14 in Chapter 2 of textbook.

Data from a one-way lay out are given in Table 2.8. The response is the muzzle velocity (in feet per second) of mortar-like antipersonnel weapon. The quantitative factor is the discharge hole area (in square inches), which has four levels in the experiment. An inverse relationship between muzzle velocity and discharge hole area was expected because a smaller hole would increase the pressure pulse of the propellant gases. Analyze the data in two ways:

```
kable(dat3,format = "markdown",caption = "Adapted Muzzle Velocity Data")
```

Table 4: Adapted Muzzle Velocity Data

A.0.016.	B.0.030.	C.0.044.	D.0.058.
294.9	295.0	270.5	258.6
294.1	301.1	263.2	255.9
301.7	293.1	278.6	257.1
307.9	300.6	267.9	263.6
285.5	285.0	269.6	262.6
298.6	289.1	269.1	260.3
303.1	277.8	262.2	305.3
305.3	266.4	263.2	304.9
264.9	248.1	224.2	216.0
262.9	255.7	227.9	216.0
256.0	245.7	217.7	210.6
255.3	251.0	219.6	207.4
256.3	254.9	228.5	214.6
258.2	254.5	230.9	214.3
243.6	246.3	227.6	222.1
250.1	246.9	228.6	222.2

(a) by treating it as a one-way layout and using an F test and multiple comparisons;

首先配適以下模型

$$y_{ij} = \eta + \tau_i + \epsilon_{ij}, \text{ for } i = A, B, C, D, \quad j = 1, \dots, 16$$

其中 η 代表 overall mean, τ_i 代表 treatment effect(the discharge hole area), $\epsilon_{ir} \sim N(0, \sigma^2)$ 代表誤差項
建立以下假設檢定

$$H_0 : \tau_A = \tau_B = \tau_C = \tau_D \quad v.s. \quad H_1 : \text{not } H_0$$

接著在下方做 F-test

```
dat3_block=stack(dat3)
treatment=rep(c(0.016,0.030,0.044,0.058),each=16)
dat3_block=cbind(dat3_block,treatment)
aovtab3 = tidy(aov(values~ind,data = dat3_block))
options(knitr.kable.NA = '')
kable(aovtab3,format='markdown',caption = "ANOVA table")
```

Table 5: ANOVA table

term	df	sumsq	meansq	statistic	p.value
ind	3	13514.98	4504.9937	7.066609	0.0003813
Residuals	60	38250.26	637.5043		

經過計算 $p\text{-value} = 0.0003813$ ，因此推論口徑大小對射速是有影響。
接著建立以下假設檢定，做 multiple comparisons

$$H_0^{ij} : \tau_i = \tau_j \quad v.s. \quad H_1^{ij} : \tau_i \neq \tau_j \quad \text{for } i, j = 1, \dots, 4, i \neq j$$

1. Bonferroni method

根據 Bonferroni method 及 t-statistic 我們能有以下 rejection region

$$|t_{ij}| > t_{60, 0.05/12} = 2.728552$$

```
bonf3 = bonfCI(dat3_block$values, dat3_block$ind)
kable(bonf3$summary, format='markdown', caption = "95% Bonferroni confidence intervals")
```

Table 6: 95% Bonferroni confidence intervals

	Diff	Lower	Upper	Decision	Adj. p-value
muA.0.016.-muB.0.030.	7.95	-16.40728	32.30728	FTR H0	1
muA.0.016.-muC.0.044.	30.56875	6.21147	54.92603	Reject H0	0.0067
muB.0.030.-muC.0.044.	22.61875	-1.73853	46.97603	FTR H0	0.083478
muA.0.016.-muD.0.058.	34.18125	9.82397	58.53853	Reject H0	0.001859
muB.0.030.-muD.0.058.	26.23125	1.87397	50.58853	Reject H0	0.028051
muC.0.044.-muD.0.058.	3.6125	-20.74478	27.96978	FTR H0	1

從上表可以發現，A-C、A-D、B-D 的 p-value 皆小於 0.05 拒絕 H_0 ，此三組之間效應具有顯著差異。

2. Tukey method

根據 Bonferroni method 及 t-statistic 我們能有以下 rejection region

$$|t_{ij}| > \frac{1}{\sqrt{2}} q_{4, 60, 0.05} = 2.642521$$

```
tukey3 = tukeyCI(dat3_block$values, dat3_block$ind)
kable(tukey3$summary, format='markdown', caption = "95% Tukey-Kramer confidence intervals")
```

Table 7: 95% Tukey-Kramer confidence intervals

	Diff	Lower	Upper	Decision	Adj. p-value
muA.0.016.-muB.0.030.	7.95	-15.6393	31.5393	FTR H0	0.809747
muA.0.016.-muC.0.044.	30.56875	6.97945	54.15805	Reject H0	0.005991
muB.0.030.-muC.0.044.	22.61875	-0.97055	46.20805	FTR H0	0.064808
muA.0.016.-muD.0.058.	34.18125	10.59195	57.77055	Reject H0	0.00172
muB.0.030.-muD.0.058.	26.23125	2.64195	49.82055	Reject H0	0.023551
muC.0.044.-muD.0.058.	3.6125	-19.9768	27.2018	FTR H0	0.977394

從上表可以發現，A-C、A-D、B-D 的 p-value 皆小於 0.05 拒絕 H_0 ，此三組之間效應具有顯著差異。

(b) by using orthogonal polynomials to model the linear and quadratic effects. (Note: These data are obtained by collapsing and adapting three-way layout data given in an exercise in Chapter 3.)

根據題意運用 orthogonal polynomials 配適以下模型

$$y = \beta_0 + \beta_1 P_1(x) + \beta_2 P_2(x) + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

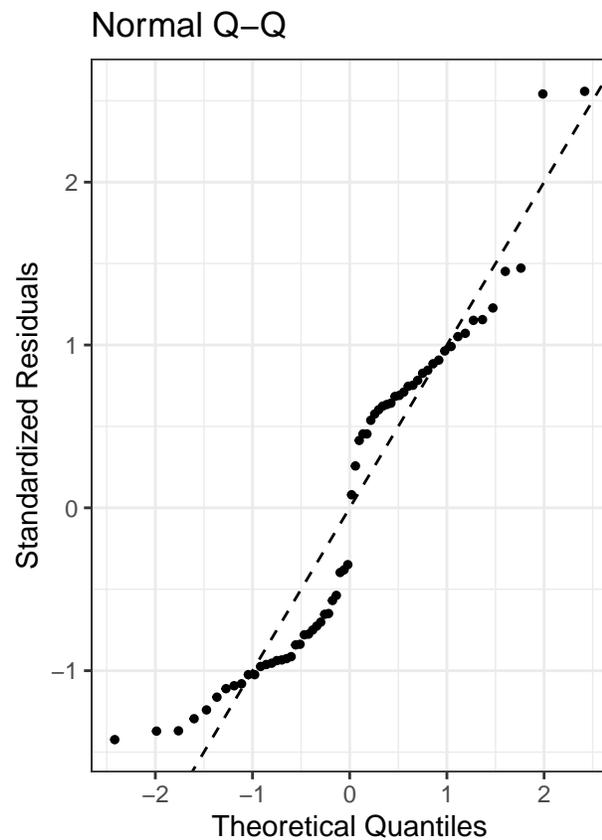
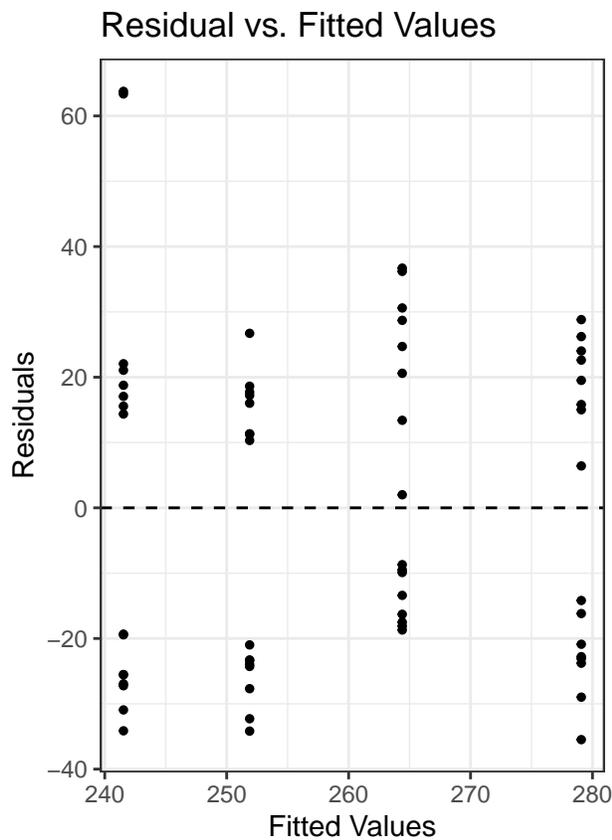
```
lm.fit3 <- lm(values~poly(treatment,2),data = dat3_block)
sumtab3 = summary.lm(lm.fit3)
kable(sumtab3$coefficients,format='markdown',caption = "model fitting table")
```

Table 8: model fitting table

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	259.2250	3.167031	81.8510987	0.0000000
poly(treatment, 2)1	-111.9487	25.336251	-4.4185204	0.0000415
poly(treatment, 2)2	8.6750	25.336251	0.3423948	0.7332308

從上表可以發現 linear effect 顯著拒絕於 0，且 $\hat{\beta}_1 = -111.949$ ，因此當口徑越大，速度越慢。

```
resplot = ggplot(lm.fit3, aes(x = .fitted, y = .resid)) +
  geom_point(color='black',size=1) +
  geom_hline(yintercept = 0,linetype = 2) +
  # geom_smooth(se=FALSE,color='darkred',linewidth=.5) +
  labs(title='Residual vs. Fitted Values', x='Fitted Values', y='Residuals') +
  theme_bw()
residuals = residuals(lm.fit3) %>% scale()
qqplot = ggplot(data.frame(residuals = residuals), aes(sample = residuals)) +
  geom_qq(color='black',size=1) +
  geom_abline(intercept = mean(residuals),slope = sd(residuals),linetype = 2) +
  labs(title='Normal Q-Q', x='Theoretical Quantiles', y='Standardized Residuals') +
  theme_bw()
grid.arrange(resplot, qqplot, ncol=2)
```



然而觀察上方的 residual plot，可以發現結果明顯分上與下兩群，觀察 normal Q-Q 也能發現 residual 並不完全服從常態分配，透過題目給的 note 得知，因為資料為 three-way layout data，本題只放入一個 factor，因此模型是配適效果不佳。

Problem 4.

Do Problem 17 in Chapter 2 of textbook.

A new electronic device for measuring blood pressure is introduced in the market. An experiment has been conducted to compare the precision (measurement variation) of measurements taken by the new device with those taken by doctors with existing devices. Three devices are randomly selected from store and three doctors (with their own existing devices) are randomly chosen from all available doctors. One patient's blood pressure has been monitored for 15 days. On each day, the blood pressure was read by all three devices and three doctors. The readings are listed in Table 2.11.

```
kable(dat4,format = "markdown",caption = "Blood Pressure Data")
```

Table 9: Blood Pressure Data

Day	Dev1	Dev2	Dev3	Doc1	Doc2	Doc3
1	133.34	133.36	133.45	126.54	127.36	131.88
2	110.94	110.85	110.92	124.69	128.86	132.39
3	118.54	118.56	118.67	125.46	129.43	134.43
4	137.94	137.80	137.77	125.95	130.72	134.28
5	139.52	139.62	139.59	125.90	130.13	134.44
6	139.23	139.11	139.36	127.85	132.03	137.37
7	117.96	117.81	117.85	125.55	132.05	132.17
8	119.59	119.42	119.48	125.80	129.87	134.97
9	116.12	116.00	115.93	125.11	128.09	133.97
10	128.38	128.48	128.41	125.75	131.94	132.68
11	125.17	125.25	125.34	128.77	130.05	134.75
12	134.62	134.41	134.55	125.26	131.13	134.29
13	136.14	136.07	136.22	126.26	130.91	133.38
14	131.21	131.03	130.96	125.68	128.83	135.67
15	132.51	132.86	132.65	124.47	129.46	134.39

(a) Analyze the blood pressures measured by devices and by doctors separately using a one-way random effects model. Your work should include two ANOVA tables, one for devices and one for doctors. You should also include F tests and estimates of variance components.

```
dat4_dev = dat4[,2:4] %>% gather(dev,y)
dat4_doc = dat4[,5:7] %>% gather(doc,y)
```

首先我們想探討三台不同的新測量儀器之間是否存在差異，故我們需要建立一個 one-way random effects model 來做分析，模型假設如下：

$$y_{ij} = \mu + \tau_i^{\text{dev}} + \epsilon_{ij}, \quad \text{for } i = 1, 2, 3, \quad j = 1, \dots, 15.$$

其中 $\epsilon_{ij} \sim N(0, \sigma^2)$ ， $\tau_i^{\text{dev}} \sim N(0, \sigma_{\tau^{\text{dev}}}^2)$ ，接著觀察其 ANOVA table

```
aovtab4_dev = tidy(aov(y ~ dev, data = dat4_dev))
options(knitr.kable.NA = '')
kable(aovtab4_dev,format='markdown',caption = "ANOVA table")
```

Table 10: ANOVA table

term	df	sumsq	meansq	statistic	p.value
dev	2	0.0135644	0.0067822	7.7e-05	0.999923

term	df	sumsq	meansq	statistic	p.value
Residuals	42	3699.4370667	88.0818349		

分別估計未知參數 σ^2 與 $\sigma_{\tau^{\text{dev}}}^2$ ，

- $\hat{\sigma}^2 = MSE = 88.08$
- $\hat{\sigma}_{\tau^{\text{dev}}}^2 = \frac{MSTr - MSE}{n} = \frac{0.01 - 88.08}{15} = -5.87 \Rightarrow 0$

由此結果我們可以發現不同台測量儀器之間的效應其解釋的 sum of square 幾乎等於零，且 ANOVA table 中 F test 的 p value 也很大，故其效應不顯著，因此沒有足夠證據證明不同台新儀器之間存在顯著差異。接著想探討三個不同的醫師之間的測量值是否存在差異，我們建立另一個 one-way random effects model 來做分析，模型假設如下：

$$y_{ij} = \mu + \tau_i^{\text{doc}} + \epsilon_{ij}, \quad \text{for } i = 1, 2, 3, \quad j = 1, \dots, 15.$$

其中 $\epsilon_{ij} \sim N(0, \sigma^2)$ ， $\tau_i^{\text{doc}} \sim N(0, \sigma_{\tau^{\text{doc}}}^2)$ ，接著觀察其 ANOVA table

```
aovtab4_doc = tidy(aov(y ~ doc, data = dat4_doc))
options(knitr.kable.NA = '')
kable(aovtab4_doc, format='markdown', caption = "ANOVA table")
```

Table 11: ANOVA table

term	df	sumsq	meansq	statistic	p.value
doc	2	496.32517	248.162587	139.0973	0
Residuals	42	74.93195	1.784094		

分別估計未知參數 σ^2 與 $\sigma_{\tau^{\text{doc}}}^2$ 得到

$$\hat{\sigma}^2 = MSE = 1.78 \quad \hat{\sigma}_{\tau^{\text{doc}}}^2 = \frac{MSTr - MSE}{n} = \frac{248.16 - 1.78}{15} = 16.425$$

由於 ANOVA table 中 F test 的 p value 極小，說明不同醫師之間測量值的效應顯著，因此是我們有足夠證據證明不同醫師之間測量值存在顯著差異。

(b) What can you conclude from (a)?

以 (a) 小題得到的結果而言，若使用新儀器測量，新儀器之間的測量差異不大，而不同醫生之間使用舊儀器測量的差異較大。

比較同一台新儀器所測量出的血壓值以及同一位醫生所測出的血壓值，我們可以發現同一台新儀器所測量出的血壓值變異較大，但對於這筆資料而言，我們沒有真實血壓來判斷 $\sigma_{\tau^{\text{dev}}}^2$ 以及 $\sigma_{\tau^{\text{doc}}}^2$ 何者估計較合理，故我們無法藉由此 ANOVA 下新舊儀器哪個具有較好測量效果的結論。

(c) Find 95% confidence intervals for the mean blood pressure measured by devices and the mean blood pressure measured by doctors.

μ 之 95% 信賴區間如下：

$$\left[\hat{\mu} - t_{2,0.975} \sqrt{\frac{MSTr}{15 \times 3}}, \hat{\mu} + t_{2,0.975} \sqrt{\frac{MSTr}{15 \times 3}} \right]$$

因此新儀器的 95% 信賴區間為

$$\left[128.0664 - 4.3027 \sqrt{\frac{0.01}{45}}, 128.0664 + 4.3027 \sqrt{\frac{0.01}{45}} \right] = [128.0023, 128.1305]$$

醫生的 95% 信賴區間為

$$\left[130.0213 - 4.3027\sqrt{\frac{248.16}{45}}, 130.0213 + 4.3027\sqrt{\frac{248.16}{45}} \right] = [119.9172, 140.1254]$$

Problem 5.

Do Problem 18 in Chapter 2 of textbook.

(a) Explain why the parameter ϕ in the one-way random-effects model in (2.39) is called the population mean.

- one-way random effects model:

$$y_{ij} = \eta + \tau_i + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i. \quad (2.39)$$

treatment effect τ_i 是從 treatment 的母體隨機抽取，隨後取 y_{ij} 期望值

$$E(y_{ij}) = E(\eta + \tau_i + \epsilon_{ij}) = \eta$$

由上式我們可知 η 為 response 的 population mean， η 考慮了 treatment population 中所不同 treatment 及所有的 experimental units 得到的平均數。

(b) Argue, by using a concrete example, that the ϕ in (a) is generally of interest (and that is why we gave a formula for its confidence interval) while the grand mean parameter ϕ in the one-way fixed-effects model in (2.1) is generally not of interest.

- fixed effect model: treatment 的水準是先由實驗者主觀挑選，為比較選定的水準對 response 是否產生顯著差異，由於 treatment 的水準並不是隨機選取，所以其無法代表整個 treatment 的母體，在 fixed effect model 中得到的結論都只能應用在實驗者選定的水準上，沒辦法應用在整個 treatment 的母體。
- random effect model: treatment 的水準都是隨機選取的，因此可以代表整個 treatment 的母體，因為隨機選取，所以選出水準之間的差異涵蓋了其母體中 σ_τ^2 的資訊。

以課本 Pulp experiment 為例，treatment 為操作員，在有眾多操作員的情況下，若我們想知道操作員間是否存在差異，就須使用 random effect model，從所有操作員中隨機抽取四個操作員來做實驗；若只想知道某四個操作員之間是否有差異，則可以選用 fixed effect model，並讓四個操作員進行實驗，假如剛好選取的四名操作員都較為熟練，則 fixed effect model 估計出的 $\hat{\eta}$ 會高於真實的 η ；反之選取出來的都是資淺的操作員時，得到的 $\hat{\eta}$ 會低於真實的 η ，由此可知，在 fixed effect model 中， $\hat{\eta}$ 不太能代表 response 的母體平均數，但是在 random effect model 中，每個操作員都有相同的機會被抽到，在這個情況下的 $\hat{\eta}$ 就平均而言較能反映出母體平均數 η 的數值。