Assignment 4

Jiarong Ye

October 1, 2018

Q1

Consider a completely randomized design with observations on three treatments coded 1,2,3. For the one-way ANOVA model, determine which of the following are estimable. For those that are estimable, write out the estimable function as $\sum_i^3 = b_i(\mu + \tau_i)$ and clearly state b_1, b_2, b_3. Finally, for those that are estimable, state the least squares estimator.

- a) $\tau_1 + \tau_2 - 2\tau_3$

Suppose it is estimable, then it should be represented as

$$\tau_1 + \tau_2 - 2\tau_3 = b_1(\mu + \tau_1) + b_2(\mu + \tau_2) + b_3(\mu + \tau_3)$$

$$= \mu(b_1 + b_2 + b_3) + b_1\tau_1 + b_2\tau_2 + b_3\tau_3$$

$$\begin{cases} (b_1 + b_2 + b_3) = 0 \\ b_1 = 1 \\ b_2 = 1 \\ b_3 = -2 \end{cases} \tag{1}$$

Therefore, $\tau_1 + \tau_2 - 2\tau_3 = 1 \cdot (\mu + \tau_1) + 1 \cdot (\mu + \tau_2) + (-2) \cdot (\mu + \tau_3)$, it is estimable.

- b) $\mu + \tau_3$

Suppose it is estimable, then it should be represented as

$$\mu + \tau_3 = b_1(\mu + \tau_1) + b_2(\mu + \tau_2) + b_3(\mu + \tau_3)$$

$$= \mu(b_1 + b_2 + b_3) + b_1\tau_1 + b_2\tau_2 + b_3\tau_3$$

$$\begin{cases} (b_1 + b_2 + b_3) = 1 \\ b_1 = 0 \\ b_2 = 0 \\ b_3 = 1 \end{cases} \tag{2}$$

Therefore, $\mu + \tau_3 = 0 \cdot (\mu + \tau_1) + 0 \cdot (\mu + \tau_2) + 1 \cdot (\mu + \tau_3)$, it is estimable.

- c) $\tau_1 - \tau_2 - \tau_3$

Suppose it is estimable, then it should be represented as

$$\tau_1 - \tau_2 - \tau_3 = b_1(\mu + \tau_1) + b_2(\mu + \tau_2) + b_3(\mu + \tau_3)$$

$$= \mu(b_1 + b_2 + b_3) + b_1\tau_1 + b_2\tau_2 + b_3\tau_3$$

$$\begin{cases} (b_1 + b_2 + b_3) = 0 \\ b_1 = 1 \\ b_2 = -1 \\ b_3 = -1 \end{cases} \tag{3}$$

And here, if $b_1 = 1, b_2 = -1, b_3 = -1$, then $b_1 + b_2 + b_3 = -1 \neq 0$, it contradicts with $(b_1 + b_2 + b_3) = 0$

Therefore, $\tau_1 - \tau_2 - \tau_3$ is not estimable.

- d) $\mu + (\tau_1 + \tau_2 + \tau_3)/3$

Suppose it is estimable, then it should be represented as

$$\mu + (\tau_1 + \tau_2 + \tau_3)/3 = b_1(\mu + \tau_1) + b_2(\mu + \tau_2) + b_3(\mu + \tau_3)$$

$$= \mu(b_1 + b_2 + b_3) + b_1\tau_1 + b_2\tau_2 + b_3\tau_3$$

$$\begin{cases} (b_1 + b_2 + b_3) = 1 \\ b_1 = \frac{1}{3} \\ b_2 = \frac{1}{3} \\ b_3 = \frac{1}{3} \end{cases} \tag{4}$$

Therefore, $\mu + (\tau_1 + \tau_2 + \tau_3)/3 = \frac{1}{3} \cdot (\mu + \tau_1) + \frac{1}{3} \cdot (\mu + \tau_2) + \frac{1}{3} \cdot (\mu + \tau_3)$, it is estimable.

Q2

Recall the soap experiment from Homework 1. Look back at Homework 1 for an explanation of the experiment. The data are the weight lost over 24 hours by different types of soap.

Cube	Regular	Deodorant	Moisturizing
1	-0.30	2.63	1.86
2	-0.10	2.61	2.03
3	-0.14	2.41	2.26
4	0.40	3.15	1.82

- (a) Write out the one-way ANOVA model for this experiment.

```
In [11]: library(lsmeans)
         types = c('Regular', 'Deodorant', 'Moisturizing')
         soap_type = c(rep(types[1], 4),
                       rep(types[2], 4),
                       rep(types[3], 4))
         weight_loss = c(-0.30, -0.10, -0.14, 0.40, 2.63, 2.61, 2.41, 3.15, 1.86, 2.03, 2.26, 1.
         soap_data = data.frame(soap_type, weight_loss)
         soap_data
```

soap_type	weight_loss
Regular	-0.30
Regular	-0.10
Regular	-0.14
Regular	0.40
Deodorant	2.63
Deodorant	2.61
Deodorant	2.41
Deodorant	3.15
Moisturizing	1.86
Moisturizing	2.03
Moisturizing	2.26
Moisturizing	1.82

```
In [12]: aov.soap=aov(weight_loss~soap_type)
         aov.soap
```

```
Call:
   aov(formula = weight_loss ~ soap_type)

Terms:
                soap_type Residuals
Sum of Squares  16.122050  0.694575
Deg. of Freedom         2         9

Residual standard error: 0.2778039
Estimated effects may be unbalanced
```

- (b) By hand or calculator (without using R), obtain the LS estimate for the mean weight lost by a cube of deodorant soap. Show all calculations.

Since the LS estimate of an estimable function is obtained by replacing each model treatment mean $(\mu + \tau_i)$ with it's corresponding treatment sample mean $Y_i\cdot$

Thus the LS estimate for each type of soap:

- Regular : $\frac{-0.30-0.10-0.14+0.40}{4} = -0.035$
- Deodorant: $\frac{2.63+2.61+2.41+3.15}{4} = 2.700$

- Moisturizing: $\frac{1.86+2.03+2.26+1.82}{4} = 1.9925$

- (c) Consider estimating the difference in weight loss between regular soap and any other type of soap. That is, consider estimating $\tau_{regular} - (\tau_{deodorant} + \tau_{moisturizing})/2$. Show that this is estimable, and find the LS estimate by hand or calculator. Show all calculations.

Suppose it is estimable, then it should be represented as

$$\tau_{regular} - (\tau_{deodorant} + \tau_{moisturizing})/2 = b_{regular}(\mu + \tau_{regular}) + b_{deodorant}(\mu + \tau_{deodorant}) + b_{moisturizing}(\mu + \tau_{moisturizing})$$

$$= \mu(b_{regular} + b_{deodorant} + b_{moisturizing}) + b_{regular}\tau_{regular} + b_{deodorant}\tau_{deodorant} + b_{moisturizing}\tau_{moisturizing}$$

$$\begin{cases} (b_{regular} + b_{deodorant} + b_{moisturizing}) = 0 \\ b_{regular} = 1 \\ b_{deodorant} = -\frac{1}{2} \\ b_{moisturizing} = -\frac{1}{2} \end{cases} \tag{5}$$

Therefore, $\tau_{regular} - (\tau_{deodorant} + \tau_{moisturizing})/2 = 1 \cdot (\mu + \tau_{regular}) + (-\frac{1}{2}) \cdot (\mu + \tau_{deodorant}) + (-\frac{1}{2}) \cdot (\mu + \tau_{moisturizing})$, it is estimable.

The LS estimate of $\tau_{regular} - (\tau_{deodorant} + \tau_{moisturizing})/2$:

$$1 \cdot \hat{\tau}_{regular} - \frac{1}{2} \cdot \hat{\tau}_{deodorant} - \frac{1}{2}\hat{\tau}_{moisturizing} = 1 \cdot \bar{Y}_{regular\cdot} - \frac{1}{2} \cdot \bar{Y}_{deodorant\cdot} - \frac{1}{2} \cdot \bar{Y}_{moisturizing\cdot}$$

where as calculated in part (b):

$$\bar{Y}_{regular\cdot} = -0.035$$

$$\bar{Y}_{deodorant\cdot} = 2.700$$

$$\bar{Y}_{moisturizing\cdot} = 1.9925$$

So

$$1 \cdot \bar{Y}_{regular\cdot} - \frac{1}{2}\bar{Y}_{deodorant\cdot} - \frac{1}{2} \cdot \bar{Y}_{moisturizing\cdot} = 1 \cdot (-0.035) - \frac{1}{2}(2.700) - \frac{1}{2} \cdot (1.9925) = -2.38125$$

- (d) Now use R to obtain the LS estimates in parts (b) and (c). Include your R code and the relevant output in your homework.

```
In [17]: lsm.soap = lsmeans(aov.soap,"soap_type")
         lsm.soap
```

```
 soap_type     lsmean        SE df  lower.CL upper.CL
 Deodorant     2.7000 0.1389019  9  2.385782 3.014218
 Moisturizing  1.9925 0.1389019  9  1.678282 2.306718
 Regular      -0.0350 0.1389019  9 -0.349218 0.279218

Confidence level used: 0.95
```

```
In [22]: contrast(lsm.soap,list("regular.minus.(deodorant + moisturizing)/2"=c(-0.5,-0.5, 1)))
```

```
 contrast                                   estimate        SE df t.ratio
 regular.minus.(deodorant + moisturizing)/2 -2.38125 0.1701194  9 -13.998
 p.value
  <.0001
```

Q3

Pedestrian light experiment (Larry Lesher, 1985) This experiment questions whether pushing a certain pedestrian light button had an effect on the wait time before the pedestrian light showed "walk." The treatment factor of interest was the number of pushes of the button, and 32 observations were taken with a mix of 0, 1, 2, and 3 pushes of the button. The waiting times for the "walk" sign are shown in the following table, with $r0 = 7, r1 = r2 = 10, r3 = 5$ (where the levels of the treatment factor are coded as $0, 1, 2, 3$ for simplicity).

0	**1**	**2**	**3**
38.14	38.28	38.17	38.14
38.20	38.17	38.13	38.30
38.31	38.08	38.16	38.21
38.14	38.25	38.30	38.04
38.29	38.18	38.34	38.37
38.17	38.03	38.34	
38.20	37.95	38.17	
	38.26	38.18	
	38.30	38.09	
	38.21	38.06	

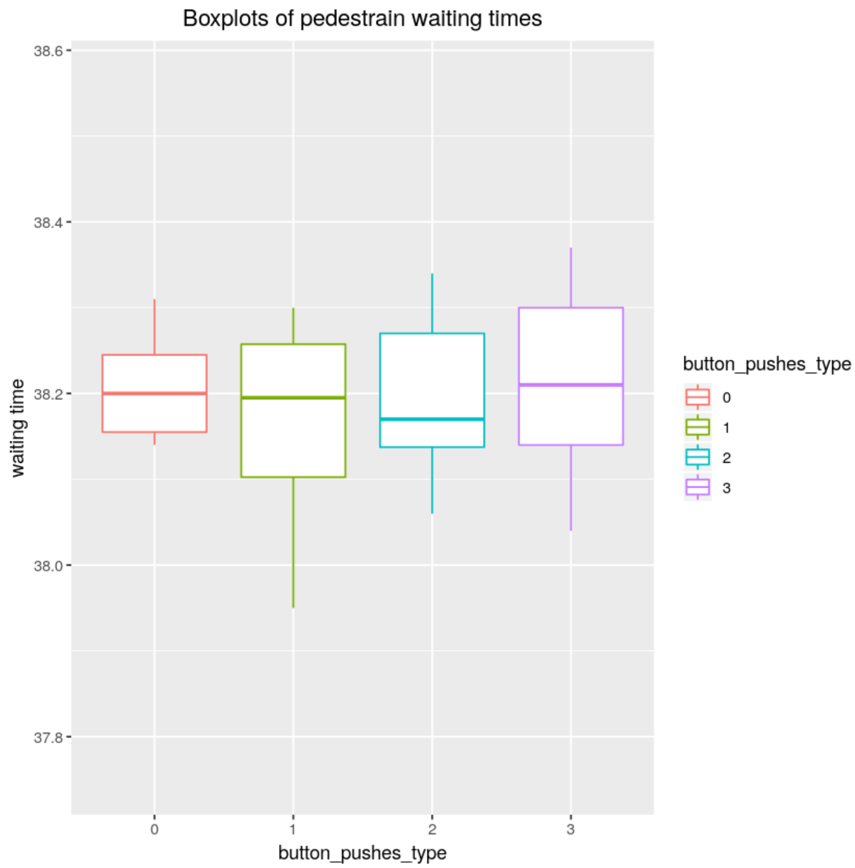
```
In [54]: r = c(7, 10, 10, 5)
         types = c(0, 1, 2, 3)
         button_pushes_type = as.factor(c(rep(types[1], r[1]),
                                          rep(types[2], r[2]),
                                          rep(types[3], r[3]),
                                          rep(types[4], r[4])))
         waiting_time = c(38.14, 38.20, 38.31, 38.14, 38.29, 38.17, 38.20,
                          38.28, 38.17, 38.08, 38.25, 38.18, 38.03, 37.95, 38.26, 38.30, 38.21,
                          38.17, 38.13, 38.16, 38.30, 38.34, 38.34, 38.17, 38.18, 38.09, 38.06,
```

```
                  38.14, 38.30, 38.21, 38.04, 38.37)
pedestrian = data.frame(button_pushes_type, waiting_time)
pedestrian[sample(nrow(pedestrian), 10), ]
```

	button_pushes_type	waiting_time
14	1	37.95
11	1	38.25
25	2	38.18
27	2	38.06
23	2	38.34
2	0	38.20
10	1	38.08
17	1	38.21
20	2	38.16
13	1	38.03

- (a) Plot the waiting times against the number of pushes of the button. What does the plot show?

In [55]:
```
library(ggplot2)
ggplot(pedestrian, aes(x=button_pushes_type, y=waiting_time, color=button_pushes_type))
            geom_boxplot() +
            ylab('waiting time') +
            ggtitle('Boxplots of pedestrain waiting times') +
            theme(plot.title = element_text(hjust = 0.5)) +
            ylim(min(pedestrian$waiting_time)-0.2, max(pedestrian$waiting_time)+0.2)
```



- the mean of $0, 1, 2, 3$ button pushes are quite similar to each other, stable at around 38.2, except for the mean of 2 pushes being slightly smaller;
- the shortest waiting time falls in the group where $pushes = 1$, the longest waiting time occurs when $pushes = 3$;
- the boxplots of these 3 button pushes share significant resemblance;

- (b) Write out the one-way ANOVA model for this experiment.

```
In [56]: aov.pedestrian=aov(waiting_time~button_pushes_type)
         aov.pedestrian
```

```
Call:
   aov(formula = waiting_time ~ button_pushes_type)

Terms:
                button_pushes_type  Residuals
Sum of Squares          0.00804714 0.30595286
Deg. of Freedom                  3         28

Residual standard error: 0.1045318
Estimated effects may be unbalanced
```

- (c) Use R to estimate the mean waiting time for each number of pushes.

```
In [61]: lsm.pedestrian = lsmeans(aov.pedestrian, 'button_pushes_type')
         lsm.pedestrian
```

```
 button_pushes_type   lsmean         SE df lower.CL upper.CL
 0                  38.20714 0.03950929 28 38.12621 38.28807
 1                  38.17100 0.03305584 28 38.10329 38.23871
 2                  38.19400 0.03305584 28 38.12629 38.26171
 3                  38.21200 0.04674802 28 38.11624 38.30776

Confidence level used: 0.95
```

- (d) Show that the contrast $\tau_1 - \tau_0$ is estimable, and use R to find it's LS estimate. This contrast compares the effect of no pushes of the button with the effect of pushing the button once.

Suppose it is estimable, then it should be represented as

$$\tau_1 - \tau_0 = b_0(\mu + \tau_0) + b_1(\mu + \tau_1) + b_2(\mu + \tau_2) + b_3(\mu + \tau_3)$$

$$= \mu(b_0 + b_1 + b_2 + b_3) + b_0\tau_0 + b_1\tau_1 + b_2\tau_2 + b_3\tau_3$$

$$\begin{cases} (b_0 + b_1 + b_2 + b_3) = 0 \\ b_0 = -1 \\ b_1 = 1 \\ b_2 = 0 \\ b_3 = 0 \end{cases} \tag{6}$$

Therefore, $\tau_1 - \tau_2 = (-1) \cdot (\mu + \tau_0) + 1 \cdot (\mu + \tau_1) + 0 \cdot (\mu + \tau_2) + 0 \cdot (\mu + \tau_3)$, it is estimable.

```
In [63]: contrast(lsm.pedestrian,list("1.minus.0"=c(-1, 1, 0, 0)))
```

```
 contrast      estimate         SE df t.ratio p.value
 1.minus.0 -0.03614286 0.05151381 28  -0.702  0.4887
```

- (e) Show that the contrast $(\tau_1 + \tau_2 + \tau_3)/3 - \tau_0$ is estimable, and use R to find it's LS estimate. This contrast compares the effect of no pushes of the button with the effect of pushing the button at least once.

Suppose it is estimable, then it should be represented as

$$(\tau_1 + \tau_2 + \tau_3)/3 - \tau_0 = b_0(\mu + \tau_0) + b_1(\mu + \tau_1) + b_2(\mu + \tau_2) + b_3(\mu + \tau_3)$$

$$= \mu(b_0 + b_1 + b_2 + b_3) + b_0\tau_0 + b_1\tau_1 + b_2\tau_2 + b_3\tau_3$$

$$
\begin{cases}
(b_0 + b_1 + b_2 + b_3) = 0 \\
b_0 = -1 \\
b_1 = \frac{1}{3} \\
b_2 = \frac{1}{3} \\
b_3 = \frac{1}{3}
\end{cases}
\tag{7}
$$

Therefore, $(\tau_1 + \tau_2 + \tau_3)/3 - \tau_0 = (-1) \cdot (\mu + \tau_0) + \frac{1}{3} \cdot (\mu + \tau_1) + \frac{1}{3} \cdot (\mu + \tau_2) + \frac{1}{3} \cdot (\mu + \tau_3)$, it is estimable.

```
In [66]: contrast(lsm.pedestrian,list("(1+2+3)/3.minus.0"=c(-1, 1/3, 1/3, 1/3)))
```

```
 contrast             estimate        SE df t.ratio p.value
 (1+2+3)/3.minus.0 -0.01480952 0.04523962 28  -0.327  0.7458
```