Assignment 6

Jiarong Ye

October 11, 2018

Q1

Pedestrian light experiment (Larry Lesher, 1985) Recall the pedestrian light experiment from Homework 4. 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 $r_0 = 7, r_1 = r_2 = 10, r_3 = 5$ (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	-

Answer the question: "Does pushing the button make the light change sooner?". Clearly state the null and alternative hypotheses, the model used, and all assumptions in the model. Obtain a test statistic (show all code and R output), and interpret the results of the test.

```
In [1]: 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
23	2	38.34
6	0	38.17
16	1	38.30
19	2	38.13
10	1	38.08
27	2	38.06
24	2	38.17
17	1	38.21
11	1	38.25
1	0	38.14

The first hypothesis tested in the experiment would be that if there is any difference in the waiting time when pushing different buttons. So if the waiting time of n times of button pushing turns out to be the same, that it would indicate that fact that pushing buttons does not actually shorten the waiting time, thus not making the light changing faster.

So denote the waiting time as τ, this is expressed as

the null hypothesis:

$$H_0 : \tau_0 = \tau_1 = \tau_2 = \tau_3$$

the alternative hypothesis:

$$H_1 : \text{at least there's one among the 4 } \tau_i \text{ that's different from the others}$$

Then we apply the AVOVA model and do pairwise contrasts among each two treatments, calculate the p values adjusted by the tukey method from each contrast.

```
In [5]: library(lsmeans)
        aov.pedestrian = aov(waiting_time~button_pushes_type, data = pedestrian)
        lsm.pedestrian=lsmeans(aov.pedestrian, ~ button_pushes_type)
        summary(contrast(lsm.pedestrian, method="pairwise", adjust="tukey"),
                        infer=c(T,T), level=0.95, side="two-sided")
```

contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
0 - 1	0.036142857	0.05151381	28	-0.1045059	0.1767916	0.70161489	0.8955744
0 - 2	0.013142857	0.05151381	28	-0.1275059	0.1537916	0.25513269	0.9940396
0 - 3	-0.004857143	0.06120753	28	-0.1719728	0.1622585	-0.07935532	0.9998162
1 - 2	-0.023000000	0.04674802	28	-0.1506367	0.1046367	-0.49199943	0.9602303
1 - 3	-0.041000000	0.05725440	28	-0.1973224	0.1153224	-0.71610217	0.8898585
2 - 3	-0.018000000	0.05725440	28	-0.1743224	0.1383224	-0.31438632	0.9890012

Alternatively

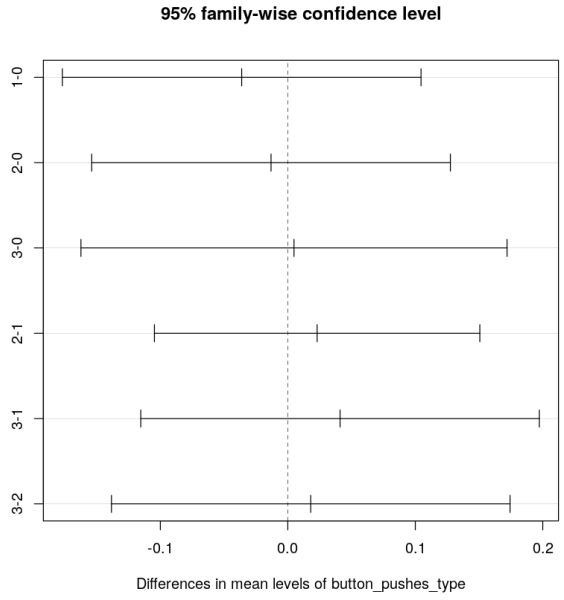
```
In [6]: TukeyHSD(aov.pedestrian, "button_pushes_type")
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = waiting_time ~ button_pushes_type, data = pedestrian)

$button_pushes_type
            diff        lwr       upr     p adj
1-0 -0.036142857 -0.1767916 0.1045059 0.8955744
2-0 -0.013142857 -0.1537916 0.1275059 0.9940396
3-0  0.004857143 -0.1622585 0.1719728 0.9998162
2-1  0.023000000 -0.1046367 0.1506367 0.9602303
3-1  0.041000000 -0.1153224 0.1973224 0.8898585
3-2  0.018000000 -0.1383224 0.1743224 0.9890012
```

```
In [7]: plot(TukeyHSD(aov.pedestrian))
```



Here we can see that the p values of all the pairwise difference are all $\gg 0.05$, thus it does not fall into the rejection region. So with the significance level of 95%, we could reach the conclusion that the waiting time of pushing any buttons, no matter how many pushes of the button, is not significantly different from each other, indicating we have no confidence to state that pushing the button will make the light change sooner.

Q2

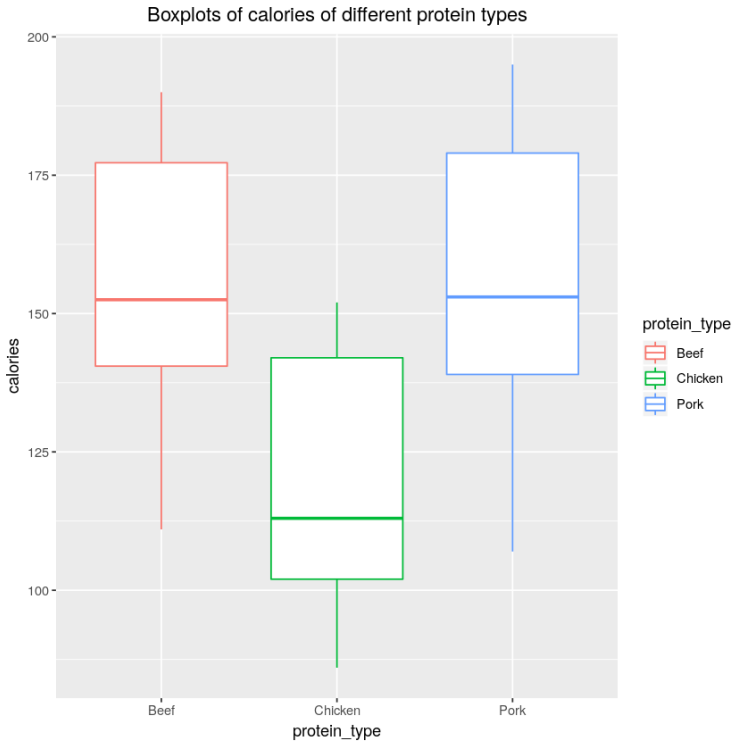
Hot Dogs A study was conducted to compare the calories and sodium in hot dogs made with different types of meat.

(a) Read the data into R and plot calories as a response variable with the type of meat on the x-axis. Your plot could be either a boxplot or a plot with one dot for each hot dog.

```
In [2]: r = c(20, 17, 17)
        types = c('Beef', 'Pork', 'Chicken')
        protein_type = as.factor(c(rep(types[1], r[1]),
                                   rep(types[2], r[2]),
                                   rep(types[3], r[3])))
        calories = c(186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190,
                     173, 191, 182, 190, 172, 147, 146, 139, 175, 136, 179, 153, 107, 195, 135, 1
                     129, 132, 102, 106, 94, 102, 87, 99, 107, 113, 135, 142, 86, 143, 152, 146,
        sodium = c(495, 477, 425, 322, 482, 587, 370, 322, 479, 375, 330, 300, 386, 401, 645, 44
                   458, 506, 473, 545, 496, 360, 387, 386, 507, 393, 405, 372, 144, 511, 405, 42
                   430, 375, 396, 383, 387, 542, 359, 357, 528, 513, 426, 513, 358, 581, 588, 52
        hot_dog = data.frame(protein_type, calories, sodium)
        hot_dog[sample(nrow(hot_dog), 10), ]
```

	protein_type	calories	sodium
52	Chicken	152	588
43	Chicken	102	542
21	Pork	173	458
13	Beef	141	386
50	Chicken	86	358
39	Chicken	132	375
6	Beef	190	587
5	Beef	184	482
35	Pork	135	405
22	Pork	191	506

```
In [15]: library(ggplot2)
         ggplot(hot_dog, aes(x=protein_type, y=calories, color=protein_type)) +
                    geom_boxplot() +
                    ylab('calories') +
                    ggtitle('Boxplots of calories of different protein types') +
                    theme(plot.title = element_text(hjust = 0.5)) +
                    ylim(min(hot_dog$calories)-0.05, max(hot_dog$calories)+0.05)
```



(b) Answer the following question: "Are there differences in the average calories of hot dogs made with different kinds of meat?". To answer this question, write down a statistical model (clearly state the response variable, treatment levels, number of replicates, . . .), express the above question as a testable null hypothesis, and report the p-value of the test statistic under the null hypothesis. Your answer should include all R code used, and the important R output.

Calories			
	Beef	Pork	Chicken
0	186	173	129
1	181	191	132
2	176	182	102
3	149	190	106
4	184	172	94
5	190	147	102
6	158	146	87
7	139	139	99
8	175	175	107
9	148	136	113
10	152	179	135
11	111	153	142
12	141	107	86
13	153	195	143
14	190	135	152
15	157	140	146
16	131	138	144
17	149		
18	135		
19	132		

Model:

- Treatment levels:
  - Beef
  - Pork
  - Chicken
- Response variable:
  - Calories
- number of replicates
  - Beef: 20
  - Pork: 17
  - Chicken: 17

In [14]:
```
library(knitr)
library(lsmeans)
aov.hotdog = aov(calories~protein_type, data = hot_dog)
kable(anova(aov.hotdog), format='markdown')
```

```
|             | Df|   Sum Sq|  Mean Sq|  F value|  Pr(>F)|
|:------------|--:|--------:|--------:|--------:|-------:|
|protein_type |  2| 17692.20| 8846.098| 16.07399| 3.9e-06|
|Residuals    | 51| 28067.14|  550.336|       NA|      NA|
```

Denote calories as c,

- Null hypothesis:

$$H_0 : c_{beef} = c_{pork} = c_{chicken}$$

- Alternative hypothesis:

H_1 : among the three protein types, at least the calories of one is different from the others

```
In [15]: lsm.hotdog=lsmeans(aov.hotdog, ~ protein_type)
         summary(contrast(lsm.hotdog, method="pairwise", adjust="tukey"),
                         infer=c(T,T), level=0.95, side="two-sided")
```

contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Beef - Chicken	38.085294	7.738831	51	19.40391	56.76667	4.9213237	2.767694e-05
Beef - Pork	-1.855882	7.738831	51	-20.53726	16.82550	-0.2398143	9.688129e-01
Chicken - Pork	-39.941176	8.046454	51	-59.36515	-20.51720	-4.9638236	2.390087e-05

```
In [18]: beef_chicken_pvalue = 2.767694e-05
         beef_pork_pvalue = 9.688129e-01
         chicken_pork_pvalue = 2.390087e-05
```

(c) Are there significant differences in mean calories between Beef and Pork hot dogs? What about between Beef and Chicken hot dogs? What about between Pork and Chicken hot dogs?

```
In [19]: beef_pork_pvalue<0.05
         beef_chicken_pvalue<0.05
         chicken_pork_pvalue<0.05
```

FALSE
TRUE
TRUE
We can interpret the results of these tests with the following statements:

- 1. The calories of beef is not significantly different from the calories of pork;
- 2. The calories of beef is significantly more than the calories of chicken;
- 3. The calories of chicken is significantly less than the calories of pork;