260762008 Tianze Lin **MATH 204 HW2**

10.78

a.

H_0: $\tau_1=\tau_2=\tau_3=\tau_4=\tau_5$

H_a: At least one mean is different.

b. There are 5 reviews, and each review is a treatment. There are 11 blocks, and each item is a block. Each block receives all the 5 treatments, so it is a randomized block design. Therefore, a randomized block ANOVA is appropriate to apply.

c. Assume $\alpha=0.05$. The p-value for REVIEW is 0.3186, and 0.3186 is greater than 0.05. This means the means of each treatment are not significantly different when $\alpha=0.05$, so our null hypothesis in part a is not rejected. The p-value for ITEM is smaller than 0.0001, so it is smaller than 0.05. This means at least one pair of means of each block are significantly different when $\alpha=0.05$.

d. All the five review means have same Tukey Grouping "A", which means all pairs of means are not significantly different. These results agree with my conclusion in c.

e. This value means the probability that making at least one Type 1 Error is 0.05. In the words, the probability that we reject the null hypothesis when in fact it is true is 0.05.

10.80

a. This experiment could be analysed using an ANOVA for a randomized design since all the data points can be grouped into blocks, and each block receives all the treatments. In this experiment, there are 5 treatments, and they are "BC", "CC", "GF", "OJ", "PP". There are 200 blocks, and each taster is a block. The dependent variable is the average 9-point taste rating.

b. Assume $\alpha=0.05$. In the table above, the p-value of PRODUCT is smaller than 0.0001, so it is also smaller than α. This implies that there are at least one pair of block means are significantly different. The table below shows means of each block, and some of them are significantly different from others. Specifically, except BB-GF and PP-OG, all other pairs of means are different from each other. This agrees with our previous conclusion.

c.

```
> taste = read_csv("D:/学习/McGill/U0/Winter/Math204/Assignment/A2/TASTE3.csv")
Parsed with column specification:
cols(
  Taster = col_integer(),
  Group = col_character(),
  OJ = col_integer(),
  CC = col_integer(),
  PP = col_integer(),
  GF = col_integer(),
  BC = col_integer()
)
```

```
> head(taste)
# A tibble: 6 x 7
  Taster Group    OJ    CC    PP    GF    BC
   <int> <chr> <int> <int> <int> <int> <int>
1      1 gLMS     40    53    57   -32   -18
2      2 gLMS    - 7   - 2   -22   -27   -49
3      3 gLMS     20    16   -18   -46    22
4      4 gLMS     46    19    55   -41   -92
5      5 gLMS    - 2    36    69   -51     9
6      6 gLMS    -13   - 3   -34   -32   - 3
```

```
> taste <- taste %>% gather (key=Treat, value=Point, -Taster, -Group)
```

```
> taste$Taster <- as.factor(taste$Taster)
> head(taste)
# A tibble: 6 x 4
  Taster Group Product Point
  <fctr> <chr> <chr>   <int>
1 1      gLMS  OJ         40
2 2      gLMS  OJ        - 7
3 3      gLMS  OJ         20
4 4      gLMS  OJ         46
5 5      gLMS  OJ        - 2
6 6      gLMS  OJ        -13
> dim(taste)
[1] 1000    4
> taste.aov <- aov(Point~Product+Taster, data=taste)
> summary(taste.aov)
             Df Sum Sq Mean Sq F value Pr(>F)
Product       4 579034  144759 165.508 <2e-16 ***
Taster      199 165962     834   0.954  0.655
Residuals   796 696205     875
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assume α=0.05. In the table above, the p-value of PRODUCT is smaller than 2^{-16}, so it is also smaller than α. This implies that there are at least one pair of means of gLMS rating are significantly different. Next, we can conduct a Tukey test to determine which pairs of means are significantly different.

```
> tukey.test <- TukeyHSD(taste.aov)
> tukey.test$Product
         diff        lwr        upr        p adj
CC-BC  46.655  38.569313  54.740687 0.000000e+00
GF-BC   1.075  -7.010687   9.160687 9.962720e-01
OJ-BC  56.735  48.649313  64.820687 0.000000e+00
PP-BC  42.950  34.864313  51.035687 0.000000e+00
GF-CC -45.580 -53.665687 -37.494313 0.000000e+00
OJ-CC  10.080   1.994313  18.165687 6.161548e-03
PP-CC  -3.705 -11.790687   4.380687 7.202068e-01
OJ-GF  55.660  47.574313  63.745687 0.000000e+00
PP-GF  41.875  33.789313  49.960687 0.000000e+00
PP-OJ -13.785 -21.870687  -5.699313 3.620792e-05
```

From the table above, all pairs of means are different (since their confidence intervals do not contain zero) except GF-BC and PP-CC. This result agrees with our previous conclusion.

10.82

a.

H_0: $\tau_1=\tau_2=\tau_3=\tau_4$

H_a: At least one mean is different.

b.

There are four treatments:

(M1) in round 1, following a pre-bout sports massage; (R1) in round 1, following a period of rest; (M5) in round 5, following a sports massage between rounds; (R5) in round 5, following a period of rest between rounds.

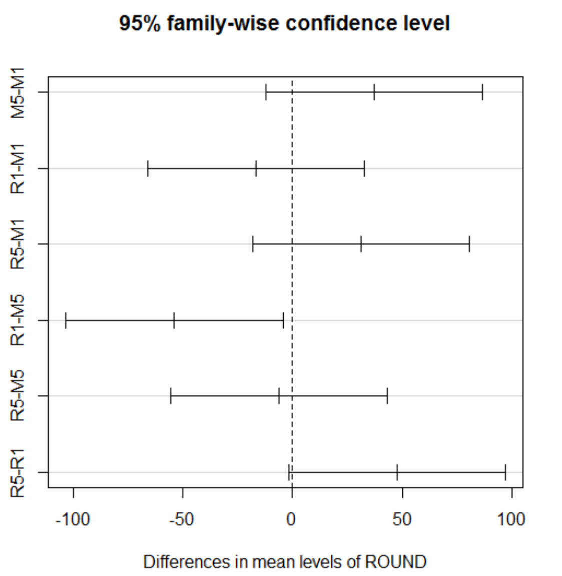
There are 8 blocks, and the blocks are 8 boxers.

c.

```
> boxing <- read_csv("D:/学习/McGill/U0/Winter/Math204/Assignment/A2/BOXING.csv")
Parsed with column specification:
cols(
  BOXER = col_integer(),
  ROUND = col_character(),
  POWER = col_integer()
)
> boxing$BOXER <- as.factor(boxing$BOXER)
> boxing_two <- aov(POWER~BOXER+ROUND,data=boxing)
> summary(boxing_two)
            Df Sum Sq Mean Sq F value   Pr(>F)    
BOXER        7 117044   16721  13.238 1.98e-06 ***
ROUND        3  15754    5251   4.158   0.0185 *  
Residuals   21  26525    1263                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Choose α=0.05. The p-value is 0.0185, which is smaller than our α. Therefore, we reject the null hypothesis when α=0.05. Next, we can conduct a Tukey test to determine which pairs of means are significantly different.

```
> tukey.test <- TukeyHSD(boxing_two)
> plot(tukey.test)
```



The confidence interval of R1-M5 does not contain zero, so the means of these two groups are significantly different. Finally, we could conclude that at this significance level, the punching powder means of the four interventions of massage are different. Specifically, the means of R1 and M5 are significantly different.

10.84

a.

Yes, I agree with this data analysis method. There are two treatments, and they are consistent sentences and inconsistent sentences. There are 16 blocks, and the blocks are 16 infants. Each block receives all the treatments, so the data could be analysed as a randomized block design.

b.

Under null hypothesis, F~F(k-1, n-k-B+1). Assume α=0.05. Under H_0, F (0.05) =4.54.

```
> qf(0.95,1,32-2-16+1)
[1] 4.543077
```

Since our test statistic (F=25.7) is greater than F(0.05), we reject the null hypothesis when α=0.05. α=0.05 is greater than p-value ($p<0.001$), which agrees with our conclusion. Therefore, at this significance level, the mean listening times of consistent sentences and inconsistent sentences are different.

c.

We can also use a t-test since t-test and F-test require same assumptions. If we want to use either one of them, we must assume: 1. The probability distribution of each treatment must be normal with equal variance. 2. Each experimental unit is independent. Therefore, when we compare two independent samples, these two tests are equivalent. The only difference is when the number of samples is greater than 2, the F-test can be used, but t-test cannot.

d.

combined mean: x=(6.3*16+9*16)/32=244.8/32=7.65

SST: $16*(6.3-7.65)^2+16*(9-7.65)^2=58.32$

MST: 58.32/(2-1)=58.32

SSE: $15*2.6^2+15*2.16^2=171.384$

MSE: 171.384/(32-2-16+1)=11.4256

F: MST/MSE=58.32/11.4256=5.10

This test statistic provides a weaker evidence of difference between means since comparing to F in b, this test statistic is much smaller and closer to F(0.05). This may be caused by the difference between completely randomized design and randomized block design. In completely randomized design, we only consider the treatment effects. In randomized block design, we consider treatment effects AND the effects of the blocks.

e.

We don’t need to control the experiment-wise error rate since there is only one experiment-wise comparison in this experiment.

10.100

a. Since the p-value (0.247) is greater than the significance level α, there is no evidence to prove the interaction between two factors. Therefore, the effect of Group on test results doesn't depend on Load.

b. Since the p-value (0.449) is greater than the significance level α, there is no evidence of the effects of playing or not playing video games on test results.

c. Since the p-value (<0.0005) is smaller than the significance level, there is evidence of the impacts of Load on test results.

d. No, I don't. From a, b and c, there is no evidence to prove video game players have superior visual attention skills.

10.102

Assume α=0.05.

H_{01}: There is no effect of bait type on the mean number of beetles captured.

H_{a1}: There is effect of bait type on the mean number of beetles captured.

H_{02}: There is no effect of colour on the mean number of beetles captured.

H_{a2}: There is effect of colour on the mean number of beetles captured.

H_{03}: The effect of bait type on the mean number of beetles captured does not depend on colour.

H_{a3}: The effect of bait type on the mean number of beetles captured depends on colour.

We need to test H_{03} first. If we fail to reject H_{03}, then we need to do hypothesis testing on H_{01} and H_{02}.

```
> bee <- read_csv("D:/学习/McGill/U0/Winter/Math204/Assignment/A2/BEETLES.csv")
Parsed with column specification:
cols(
  NUMBER = col_integer(),
  TRAP = col_character(),
  COLOR = col_character()
)
> bee_two_way <- aov(NUMBER~TRAP*COLOR,data=bee)

> anova(bee_two_way) %>% kable(.)


|           | Df|      Sum Sq|     Mean Sq|    F value|    Pr(>F)|
|:----------|--:|-----------:|-----------:|----------:|---------:|
|TRAP       |  1|   46.285714|   46.285714|  2.2951594| 0.1428384|
|COLOR      |  1| 1106.285714| 1106.285714| 54.8571429| 0.0000001|
|TRAP:COLOR |  1|    5.142857|    5.142857|  0.2550177| 0.6181698|
|Residuals  | 24|  484.000000|   20.166667|         NA|        NA|
```

Since the p-value (0.618) is greater than the significance level α, there is no evidence to prove the interaction between two factors. Therefore, we fail to reject H_{03}.

Since the p-value (0.143) is greater than the significance level α, there is no evidence of the effects of bait type on the mean number of beetles captured. Therefore, we fail to reject the null hypothesis H_{01}.

Since the p-value (0.0000001) is smallerer than the significance level, there is evidence of the impacts of colour on the mean number of beetles captured. Therefore, we reject the null hypothesis H_{02}.

10.106

Assume α=0.05.

H_{01}: There is no effect of SONG type on the aggressive recognition score.
H_{a1}: There is effect of SONG type on the aggressive recognition score.
H_{02}: There is no effect of POOL type on the aggressive recognition score.
H_{a2}: There is effect of POOL type on the aggressive recognition score.
H_{03}: The effect of SONG type on the aggressive recognition score does not depend on POOL.
H_{a3}: The effect of bait type on the aggressive recognition score depends on POOL.
We need to test H_{03} first. If we fail to reject H_{03}, then we need to do hypothesis testing on H_{01} and H_{02}.

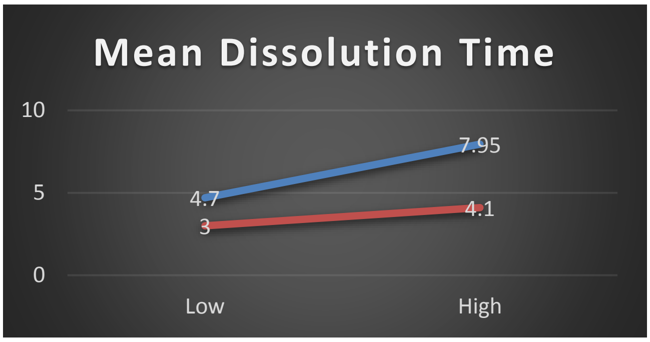
```
> lyric <- read_csv(("D:/学习/McGill/U0/Winter/Math204/Assignment/A2/LYRICS.csv")
+ )
Parsed with column specification:
cols(
  SUBJECT = col_integer(),
  SONG = col_character(),
  POOL = col_character(),
  SCORE = col_double()
)

> two_way_lyric <- aov(lyric$SCORE~lyric$SONG*lyric$POOL, data=lyric)
> symmary(two_way_lyric)
Error in symmary(two_way_lyric) : 没有"symmary"这个函数
> summary(two_way_lyric)
                        Df Sum Sq Mean Sq F value   Pr(>F)    
lyric$SONG               1  5.891   5.891  26.114 4.03e-06 ***
lyric$POOL               1  0.131   0.131   0.579    0.450    
lyric$SONG:lyric$POOL    1  0.353   0.353   1.563    0.216    
Residuals               56 12.632   0.226                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value (0.216) is greater than the significance level α, there is no evidence to prove the interactions between two factors. Therefore, we fail to reject the null hypothesis H_{03}.
Since the p-value (4.03E-6) is smaller than the significance level α, there is evidence to prove the impacts of SONG on the score. Therefore, we reject the null hypothesis H_{01}.
Since the p-value (0.45) is greater than the significance level α, there is no evidence to prove the impacts of POOL on the score. Therefore, we fail to reject the null hypothesis H_{02}.

10.108

From the given data, we can construct the following table and graph.

	Low	High
Gum	4.7	7.95
PVP	3	4.1



The blue line is Gum, and the red line is PVP.

There is no intersection between two lines, and both lines have positive slopes and are approximately parallel. Therefore, we could conclude that there is no interaction between binding agent and relative density.