

# ANOVA

Jincheng Yu

2024-04-23

```
Mdata <- read.csv('Mario Kart Data - Merged.csv')
colnames(Mdata) <- c('Round', 'Race', 'Race Name', 'Vehicle', 'Type(C/B)', 'Place',
                    'Completion Time', 'Player', 'Completion Time(s)')
Mdata$Round <- as.factor(Mdata$Round)
Mdata$Race <- ordered(as.factor(Mdata$Race))
```

Here, I tried to make the Race as an ordered categorical data.

Check if there's round effect in the game

```
round_aov <- aov(`Completion Time(s)` ~ Round, data=Mdata)
summary(round_aov)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Round      5      95    18.91    0.118  0.988
## Residuals  66   10575    160.23
```

The result shows that there's no significant difference only considering the round and the player effect

Check if there's race # effect in the game

```
race_no_aov <- aov(`Completion Time(s)` ~ Race, data=Mdata)
summary(race_no_aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Race      17    6925    407.4    5.874 2.73e-07 ***
## Residuals  54    3745     69.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(race_no_aov))
```

```
##
## Call:
## lm(formula = race_no_aov)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.732  -4.976  -1.044   4.427  21.905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 131.6125     0.9814 134.107 < 2e-16 ***
## Race.L       2.5026     4.1637   0.601 0.55033
## Race.Q       3.1241     4.1637   0.750 0.45631
## Race.C      -2.3216     4.1637  -0.558 0.57943
## Race^4       6.4103     4.1637   1.540 0.12951
## Race^5       3.7108     4.1637   0.891 0.37677
## Race^6       4.0287     4.1637   0.968 0.33757
## Race^7      10.6546     4.1637   2.559 0.01334 *
## Race^8       0.4296     4.1637   0.103 0.91820
## Race^9      11.8432     4.1637   2.844 0.00627 **
## Race^10      2.6623     4.1637   0.639 0.52527
## Race^11     20.4962     4.1637   4.923 8.43e-06 ***
## Race^12     -4.1603     4.1637  -0.999 0.32217
## Race^13     -6.9000     4.1637  -1.657 0.10328
## Race^14    -22.6336     4.1637  -5.436 1.34e-06 ***
## Race^15     10.2721     4.1637   2.467 0.01683 *
## Race^16     -9.0701     4.1637  -2.178 0.03376 *
## Race^17     13.8811     4.1637   3.334 0.00155 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.327 on 54 degrees of freedom
## Multiple R-squared:  0.649, Adjusted R-squared:  0.5386
## F-statistic: 5.874 on 17 and 54 DF, p-value: 2.732e-07
```

The result shows that race number has significant effect, this is intuitive.

*Comment* \* Here the baseline is race 18. So positive means the time significant higher than race 18. race 18 is MC \* When we treat time series as round, the difference is between round is not significant but if we treat single race as a ordered categorical data, then the series effect exist. \* When we look into the ordered categorical result, it shows that race 1-6 have no significant improvement but in 7-9 positive effect appears and so did in race 11. With time going by we can see concentration goes down and negative effect shows up. But meanwhile there's improvement in the tail. But then we can look into each race significant effect. \* 7, 9, 11, 15, 17 ~ RR, MMM, RR, MMM, RR. These race makes the time goes down. \* 14, 16 ~ RR, MMM. These race makes the time goes up. \* Maybe the race RR and MMM is more difficult or easier?

```
race_no_aov <- aov(`Completion Time(s)` ~ `Race Name`, data=Mdata)
summary(race_no_aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## 'Race Name'   2    6722     3361   58.74 1.27e-15 ***
## Residuals    69     3948        57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(race_no_aov))
```

```
##
## Call:
## lm(formula = race_no_aov)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.730  -4.750  -1.537   4.884  24.770
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    141.164     1.544   91.427 < 2e-16 ***
## 'Race Name'MMM     -5.864     2.184   -2.685  0.00907 **
## 'Race Name'RR    -22.790     2.184  -10.437 7.77e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.564 on 69 degrees of freedom
## Multiple R-squared:  0.63, Adjusted R-squared:  0.6193
## F-statistic: 58.74 on 2 and 69 DF, p-value: 1.267e-15
```

MMM and RR seems easier.

Consider the player effect

```
race_no_aov <- aov(`Completion Time(s)` ~ Round + Player, data=Mdata)
summary(race_no_aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Round          5      95    18.9    0.147 0.980322
## Player          3    2455    818.4    6.350 0.000787 ***
## Residuals      63    8120    128.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(race_no_aov))
```

```
##
## Call:
## lm(formula = race_no_aov)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.633  -9.448   2.222   9.111  17.473
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    126.9283     4.0139  31.622 < 2e-16 ***
## Round2          2.9150     4.6349   0.629 0.531670
## Round3          0.1392     4.6349   0.030 0.976141
## Round4          0.5575     4.6349   0.120 0.904641
## Round5          1.7850     4.6349   0.385 0.701443
```

```
## Round6          2.5033      4.6349    0.540 0.591026
## PlayerJincheng   2.2017      3.7844    0.582 0.562790
## PlayerJosephine -1.8967      3.7844   -0.501 0.617987
## PlayerSungmin    13.1650      3.7844    3.479 0.000919 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.35 on 63 degrees of freedom
## Multiple R-squared:  0.239, Adjusted R-squared:  0.1423
## F-statistic: 2.473 on 8 and 63 DF,  p-value: 0.02127
```

When consider the player effect, the completion time still not relevant to the round. But we can see due to the experience, Sungmin played not well in the test.

### Check race# and player effect

```
race_no_aov <- aov(`Completion Time(s)` ~ Race + Player, data=Mdata)
summary(race_no_aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Race          17   6925   407.4    16.11 7.18e-15 ***
## Player         3   2455   818.4    32.37 7.38e-12 ***
## Residuals     51   1289    25.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(race_no_aov))
```

```
##
## Call:
## lm(formula = race_no_aov)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.4683 -3.6312  0.1467  2.6319 12.1075
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   128.2450     1.1852  108.208 < 2e-16 ***
## Race.L         2.5026     2.5141    0.995 0.324242
## Race.Q         3.1241     2.5141    1.243 0.219687
## Race.C        -2.3216     2.5141   -0.923 0.360134
## Race^4         6.4103     2.5141    2.550 0.013830 *
## Race^5         3.7108     2.5141    1.476 0.146101
## Race^6         4.0287     2.5141    1.602 0.115235
## Race^7        10.6546     2.5141    4.238 9.47e-05 ***
## Race^8         0.4296     2.5141    0.171 0.864991
## Race^9        11.8432     2.5141    4.711 1.94e-05 ***
## Race^10        2.6623     2.5141    1.059 0.294623
## Race^11       20.4962     2.5141    8.152 8.40e-11 ***
## Race^12       -4.1603     2.5141   -1.655 0.104118
```

```
## Race^13          -6.9000      2.5141  -2.744  0.008347 **
## Race^14         -22.6336      2.5141  -9.003  4.09e-12 ***
## Race^15          10.2721      2.5141   4.086  0.000156 ***
## Race^16          -9.0701      2.5141  -3.608  0.000703 ***
## Race^17          13.8811      2.5141   5.521  1.14e-06 ***
## PlayerJincheng    2.2017      1.6761   1.314  0.194869
## PlayerJosephine  -1.8967      1.6761  -1.132  0.263095
## PlayerSungmin     13.1650      1.6761   7.855  2.45e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.028 on 51 degrees of freedom
## Multiple R-squared:  0.8792, Adjusted R-squared:  0.8318
## F-statistic: 18.55 on 20 and 51 DF,  p-value: < 2.2e-16
```

- The conclusion is similar to before, but more race shows significant effect.
  - 4, 7, 9, 11, 15, 17 ~ MMM, RR, MMM, RR, MMM, RR. These race makes the time goes down.
  - 13, 14, 16 ~ MC, RR, MMM. These race makes the time goes up.

## Check the whole model

```
whole_aov <- aov(`Completion Time(s)` ~ Round + Vehicle + `Race Name` + Player, data=Mdata)
summary(whole_aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Round          5      95      19   0.783  0.56635
## Vehicle         5     429      86   3.552  0.00736 **
## 'Race Name'     2    6339    3169 131.213 < 2e-16 ***
## Player          3    2455      818  33.883 1.27e-12 ***
## Residuals      56    1353       24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Interaction between cars and blocks

```
whole_inter_aov <- aov(`Completion Time(s)` ~ Vehicle * (as.factor(Round) + Player + `Race Name`), data=Mdata)
summary(whole_inter_aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Vehicle          5      430      86   3.193  0.070440 .
## as.factor(Round)  5       93      19   0.694  0.642961
## Player            3    2455      818  30.384  0.000101 ***
## 'Race Name'       2    6339    3169 117.661 1.17e-06 ***
## Vehicle:as.factor(Round) 24    552      23   0.854  0.644295
## Vehicle:Player     14     393      28   1.043  0.496960
## Vehicle:'Race Name' 10     192      19   0.713  0.697654
## Residuals          8      215      27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction is not significant but the Player and Race Name play a role in the response variable.