Project on ricci data

Dean Nasa 209381367

28 9 2022

In this project I analyzed the "RICCI" DATA - Data on firefighter promotion exams as part of the Ricci vs DeStafano court case to find if there is any racial discrimination between 3 types of firefighters (whites, Hispanics, blacks).

We would like to know if the expected weighted score (Combined) is different between the position groups (Position) and the race (Race). I Checked this using two-way analysis of variance, with integration (only if there is evidence in the data that it is needed).

*all tests were conducted with 0.05 confidence

```
library(lawstat)
```

```
## Warning: package 'lawstat' was built under R version 4.0.5
```

```
library(Stat2Data)
library(multcomp)
```

```
## Warning: package 'multcomp' was built under R version 4.0.5
## Loading required package: mvtnorm
## Warning: package 'mvtnorm' was built under R version 4.0.5
## Loading required package: survival
## Loading required package: TH.data
## Warning: package 'TH.data' was built under R version 4.0.5
## Loading required package: MASS
##
## Attaching package: 'TH.data'
## The following object is masked from 'package:MASS':
##
##     geyser
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.5
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 4.0.5
## Warning: The package `vctrs` (>= 0.3.8) is required as of rlang 1.0.0.
## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'tibble'
```

```
## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'pillar'
```

```
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.0.5
```

```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## v purrr   0.3.4
```

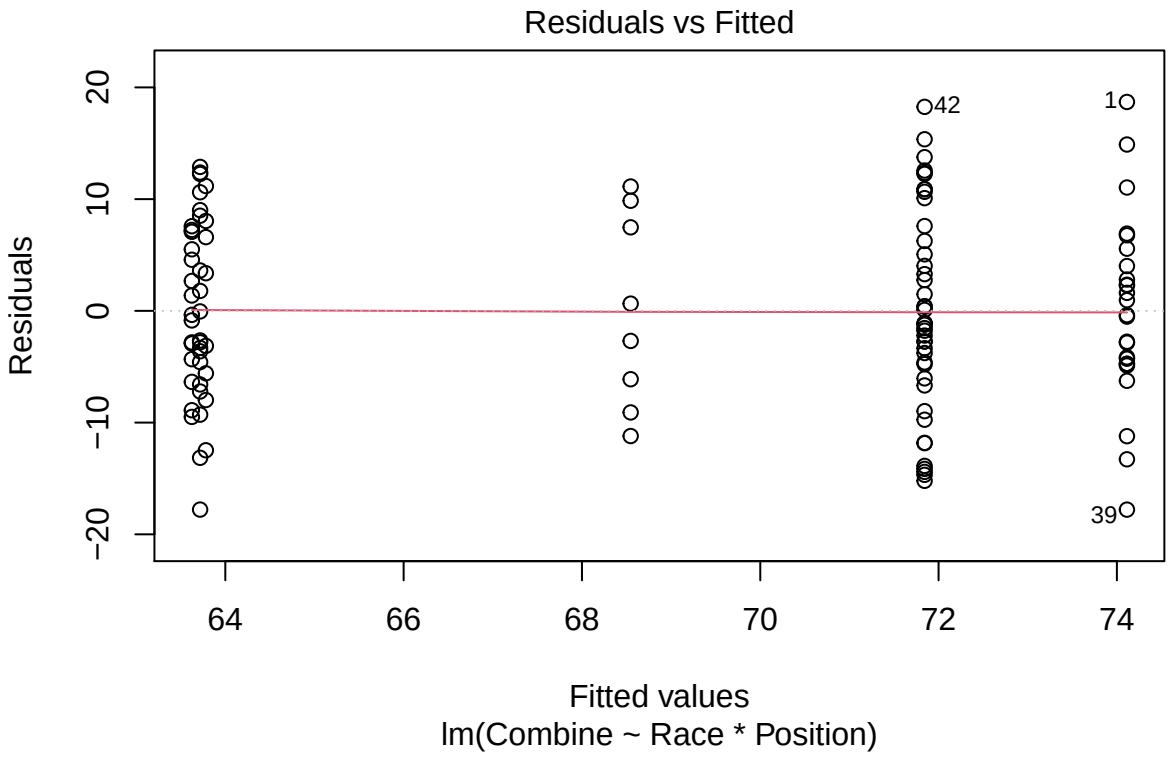
```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks rstatix::filter(), stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks rstatix::select(), MASS::select()
```

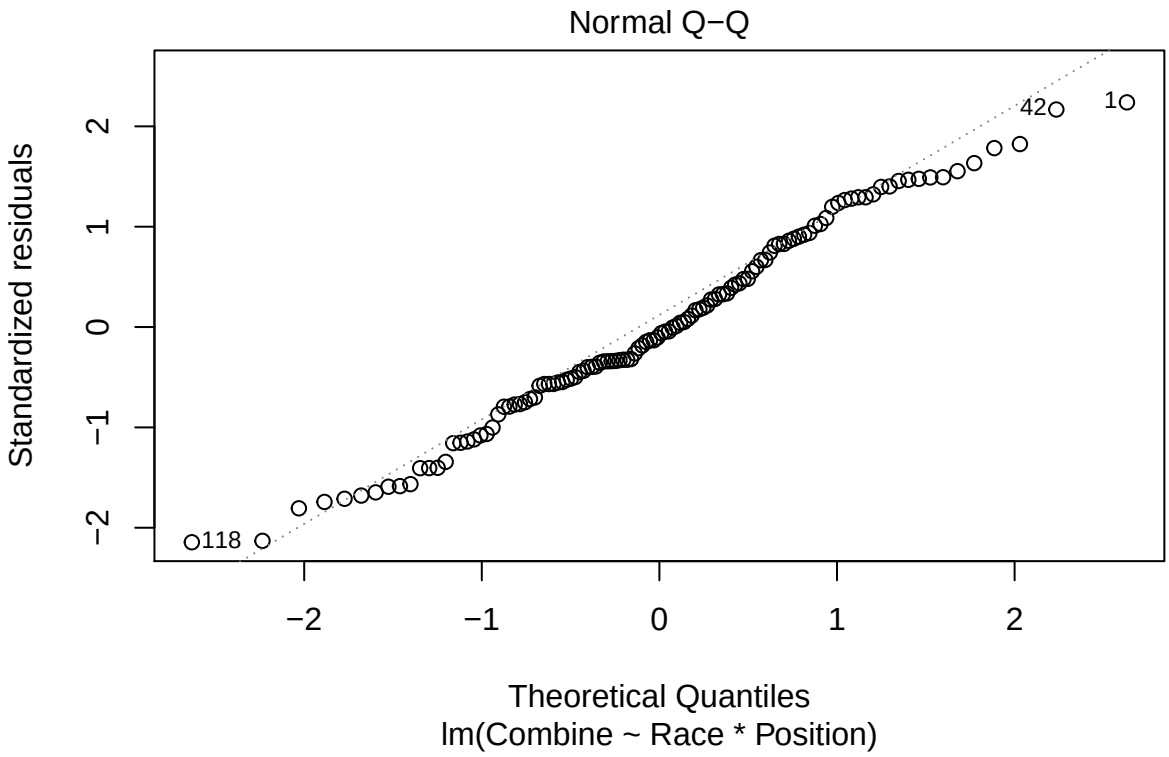
```
data(Ricci)
str(Ricci)
```

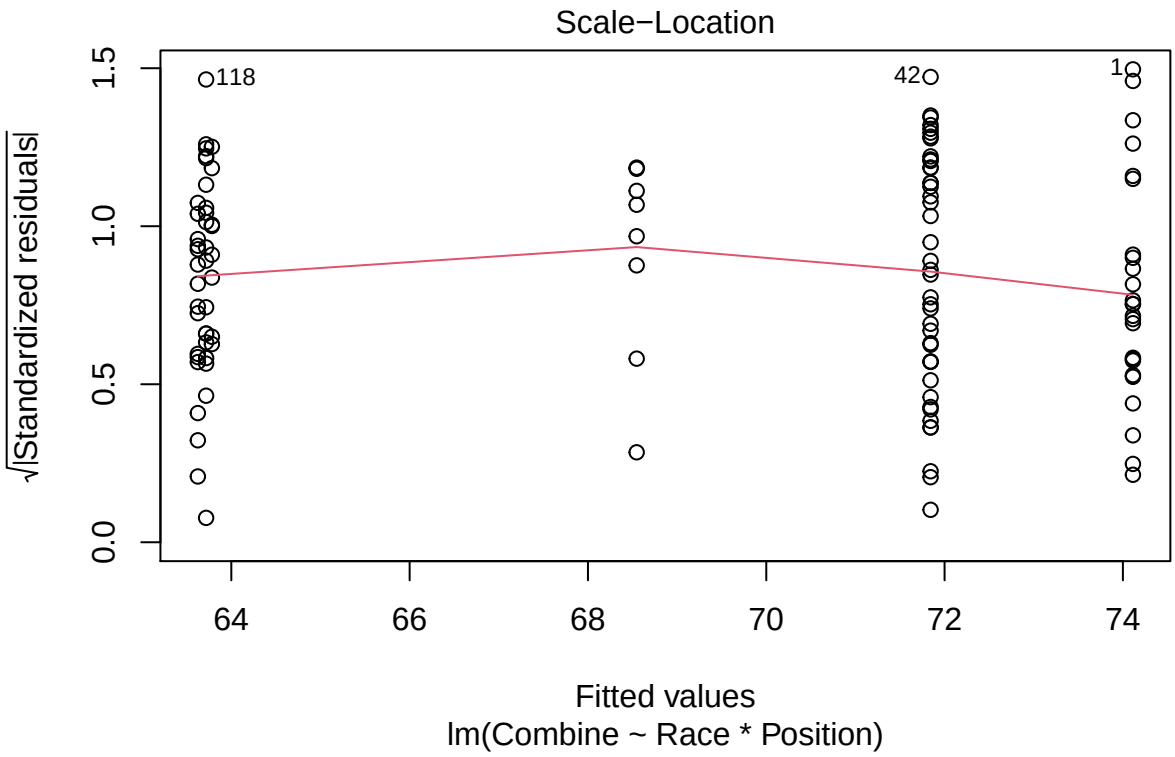
```
## 'data.frame':    118 obs. of  5 variables:
##  $ Race    : Factor w/ 3 levels "B","H","W": 3 3 3 3 3 2 3 2 3 3 ...
##  $ Position: Factor w/ 2 levels "Captain","Lieutenant": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Oral    : num  89.5 80 82.4 88.6 76.2 ...
##  $ Written : int  95 95 87 76 84 82 82 84 81 72 ...
##  $ Combine : num  92.8 89 85.2 81 80.9 ...
```

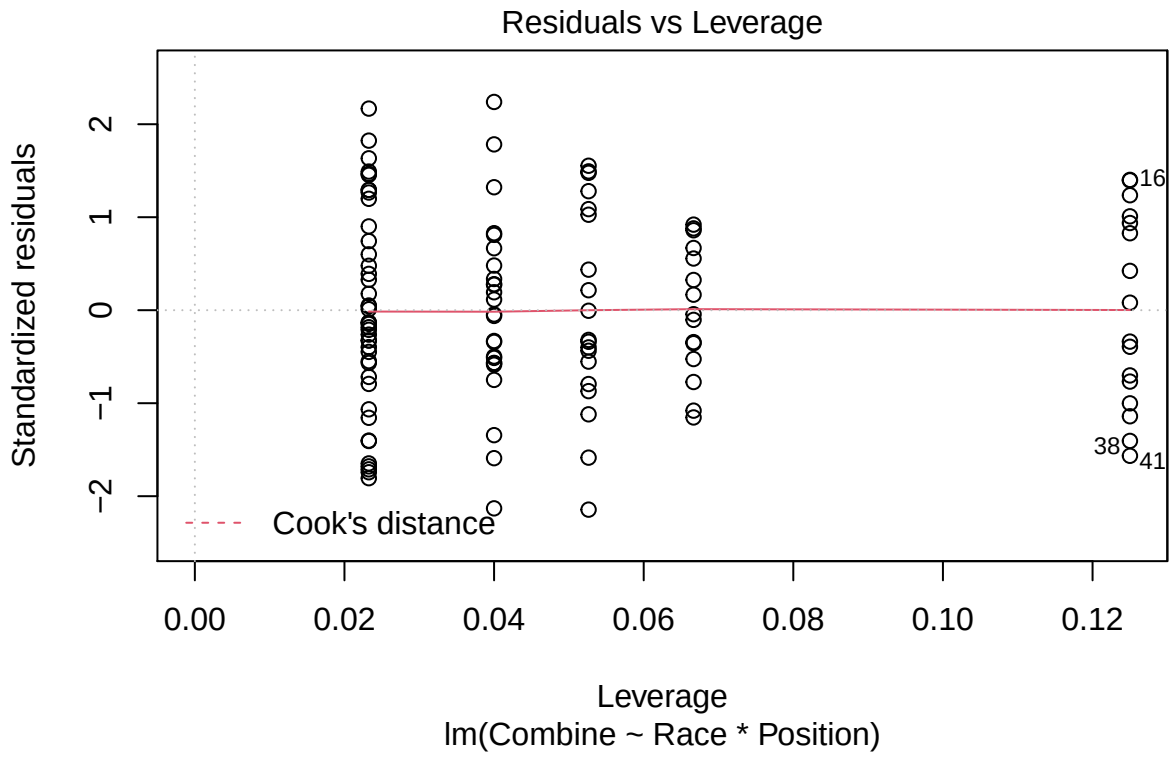
```
lm1 <- lm(Combine ~ Race*Position, Ricci)
# my Assumption

# Normality of errors
plot(lm1)
```


Residuals vs Fitted
Residuals
20
10
0
−10
−20
42
1
39
64
66
68
70
72
74
Fitted values
lm(Combine ~ Race * Position)

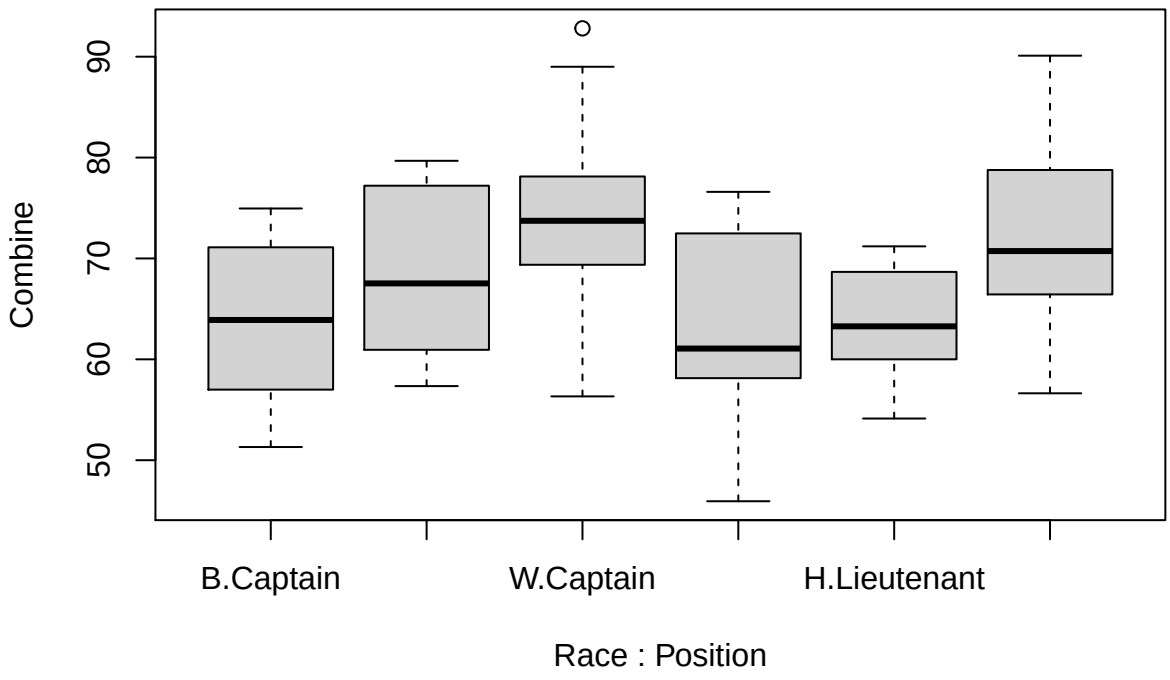

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Combine ~ Race * Position)
118
42
1
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
118
42
1
64
66
68
70
72
74
Fitted values
lm(Combine ~ Race * Position)




```
# Heteroscedasticity

boxplot(Combine ~ Race*Position, Ricci)
```




```
# check for us - variance equality

levene.test(Ricci$Combine, interaction(Ricci$Position ,Ricci$Race), location = "mean")
```

```
##
##  Classical Levene's test based on the absolute deviations from the mean
##  ( none not applied because the location is not set to median )
##
## data:  Ricci$Combine
## Test Statistic = 0.82228, p-value = 0.5363
```

```
# check if the groups are balanced

table(Ricci$Position, Ricci$Race)
```

```
##
##               B  H  W
##   Captain     8  8 25
##   Lieutenant 19 15 43
```

```
#we can see that the table is not balanced


#means of combination of groups

sum_of_groub_ab <- function(a,b){
  c <-0
  for(i in 1:length(Ricci$Race)){
```

```
    if((Ricci$Race[i] == a)&(Ricci$Position[i]== b)) {
      c<- c+1
    }
  }
  return(c)
}


mean_of_groub_ab <- function(a,b){
  s <-0
  for(i in 1:length(Ricci$Race)){
    if((Ricci$Race[i] == a)&(Ricci$Position[i]== b)) {
      s<- s +Ricci$Combine[i]
    }
  }
  return(s/sum_of_groub_ab(a,b))
}


mean(Ricci$Combine)
```

```
## [1] 69.20088
```

```
mean_of_groub_ab('W',"Captain")
```

```
## [1] 74.1128
```

```
mean_of_groub_ab('W',"Lieutenant")
```

```
## [1] 71.84326
```

```
mean_of_groub_ab('B',"Captain")
```

```
## [1] 63.782
```

```
mean_of_groub_ab('B',"Lieutenant")
```

```
## [1] 63.71726
```

```
mean_of_groub_ab('H',"Captain")
```

```
## [1] 68.5465
```

```
mean_of_groub_ab('H',"Lieutenant")
```

```
## [1] 63.62453
```

```
aov_ric <-aov(Combine ~ Position * Race, data = Ricci)
summary(aov_ric)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## Position        1    206   205.9   2.835    0.095 .
## Race            2   1910   954.8  13.152 7.41e-06 ***
## Position:Race   2     64    32.0   0.441    0.645
## Residuals     112   8131    72.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Only the race factor is significance.

# post hoc analysis

#bonforroni correction
pairwise <- glht(aov_ric, linfct = mcp(Race = 'Tukey'))
```

```
## Warning in mcp2matrix(model, linfct = linfct): covariate interactions found --
## default contrast might be inappropriate
```

```r
summary(pairwise, test = adjusted(type = "bonf"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = Combine ~ Position * Race, data = Ricci)
##
## Linear Hypotheses:
##            Estimate Std. Error t value Pr(>|t|)
## H - B == 0    4.765      4.260   1.118   0.7974
## W - B == 0   10.331      3.461   2.985   0.0105 *
## W - H == 0    5.566      3.461   1.608   0.3318
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- bonferroni method)
```

```r
#tokey correction
tukey_hsd(aov_ric)
```

```
## # A tibble: 19 x 9
##    term  group1 group2 null.value estimate conf.low conf.high   p.adj
##  * <chr> <chr>  <chr>       <dbl>    <dbl>    <dbl>     <dbl>   <dbl>
##  1 Posi~ Capta~ Lieut~          0  -2.77     -6.04      0.490 9.50e-2
##  2 Race  B      H               0   1.46     -4.29      7.20  8.19e-1
##  3 Race  B      W               0   8.74      4.14     13.3   4.73e-5
##  4 Race  H      W               0   7.29      2.40     12.2   1.65e-3
##  5 Posi~ Capta~ Lieut~          0  -0.0647  -10.5       10.3   1.00e+0
##  6 Posi~ Capta~ Capta~          0   4.76     -7.59      17.1   8.73e-1
##  7 Posi~ Capta~ Lieut~          0  -0.157   -11.0       10.7   1.00e+0
##  8 Posi~ Capta~ Capta~          0  10.3       0.295    20.4   3.97e-2
##  9 Posi~ Capta~ Lieut~          0   8.06     -1.45      17.6   1.46e-1
## 10 Posi~ Lieut~ Capta~          0   4.83     -5.58      15.2   7.59e-1
## 11 Posi~ Lieut~ Lieut~          0  -0.0927   -8.63       8.44  1.00e+0
## 12 Posi~ Lieut~ Capta~          0  10.4       2.88      17.9   1.51e-3
## 13 Posi~ Lieut~ Lieut~          0   8.13      1.32      14.9   9.66e-3
## 14 Posi~ Capta~ Lieut~          0  -4.92    -15.7        5.89  7.74e-1
## 15 Posi~ Capta~ Capta~          0   5.57     -4.47      15.6   5.95e-1
## 16 Posi~ Capta~ Lieut~          0   3.30     -6.22      12.8   9.15e-1
## 17 Posi~ Lieut~ Capta~          0  10.5       2.42      18.6   3.51e-3
## 18 Posi~ Lieut~ Lieut~          0   8.22      0.810     15.6   2.05e-2
## 19 Posi~ Capta~ Lieut~          0  -2.27     -8.48       3.94  8.96e-1
## # ... with 1 more variable: p.adj.signif <chr>
```

```r
aov_ric2 <- aov(Combine ~ Race, data = Ricci)
summary(aov_ric2)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Race          2   1972   985.8    13.6 5.01e-06 ***
## Residuals   115   8339    72.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# We can see that the difference between Black and Hispanic is not significance.
```

In the above we focused on the sub-sample of the observations for which Lieutenant = Position holds and checked whether there is evidence that the assumption - Equality of differences does not hold for the weighted score (Combined) between the race groups. now I will perform a one-way analysis of variance and then also post hoc analyses.

```r
library(knitr)
library(lawstat)
library(Stat2Data)
library(dplyr)
data(Ricci)
str(Ricci)
```

```
## 'data.frame':    118 obs. of  5 variables:
##  $ Race    : Factor w/ 3 levels "B","H","W": 3 3 3 3 3 2 3 2 3 3 ...
##  $ Position: Factor w/ 2 levels "Captain","Lieutenant": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Oral    : num  89.5 80 82.4 88.6 76.2 ...
##  $ Written : int  95 95 87 76 84 82 82 84 81 72 ...
##  $ Combine : num  92.8 89 85.2 81 80.9 ...
```

```r
Ricci_l <- Ricci[Ricci$Position == "Lieutenant",]
aov_l<-aov(Ricci_l$Combine~Ricci_l$Race)

# one-way analysis of variance examining whether in this sub-sample, Combined differs between the race

# pairwise comparison with bonferroni correction

pairwise.t.test(Ricci_l$Combine,Ricci_l$Race, p.adjust.method="bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  Ricci_l$Combine and Ricci_l$Race
##
##   B      H
## H 1.0000 -
## W 0.0029 0.0063
##
## P value adjustment method: bonferroni
```

```r
# we will reject h0 for white and black

# anova with tokey correction

TukeyHSD(aov(lm(Ricci_l$Combine ~ Ricci_l$Race)))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = lm(Ricci_l$Combine ~ Ricci_l$Race))
##
## $`Ricci_l$Race`
##            diff       lwr       upr     p adj
## H-B -0.09272982 -7.190168  7.004708 0.9994618
## W-B  8.12599266  2.465302 13.786683 0.0027861
## W-H  8.21872248  2.056765 14.380680 0.0058494
```

```
# we will reject h0 for white and black


# From the results in this part we concluded that there is discrimination against blacks in comparison
```

I can see that there is a difference between blacks, Hispanics and whites. and there is a difference between whites and blacks and there is a difference between whites and Hispanics. However there can be other explanations (for example, they come from different neighborhoods with different schools).

To conclude there is no doubt that there is a big problem with firefighter promotion exams. we cannot say precisely what is the source of the problem but the test discriminates hispanics and blacks in favor of the whites.