Anova1.R

acarraro

2023-01-18

```r
# Example 1, One Way ANOVA
# Create Dataframe
foods<-data.frame(milk=c(25.40,26.31,24.10,23.74,25.10,
                         23.40,21.80,23.50,22.75,21.60,
                         20.00,22.20,19.75,20.60,20.40),
                  food=c("Food1","Food1","Food1","Food1","Food1",
                         "Food2","Food2","Food2","Food2","Food2",
                         "Food3","Food3","Food3","Food3","Food3"),
                  stringsAsFactors = TRUE)

# ANOVA test
tm<-lm(milk ~ food, data = foods)  # fit a linear model to the independent variable food
summary(tm)                         # ANOVA table
```

```
##
## Call:
## lm(formula = milk ~ food, data = foods)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -1.19  -0.82   0.01   0.63   1.61
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.9300     0.4292  58.084 4.48e-16 ***
## foodFood2    -2.3200     0.6070  -3.822  0.00243 **
## foodFood3    -4.3400     0.6070  -7.150 1.16e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9597 on 12 degrees of freedom
## Multiple R-squared:  0.8101, Adjusted R-squared:  0.7785
## F-statistic:  25.6 on 2 and 12 DF,  p-value: 4.684e-05
```

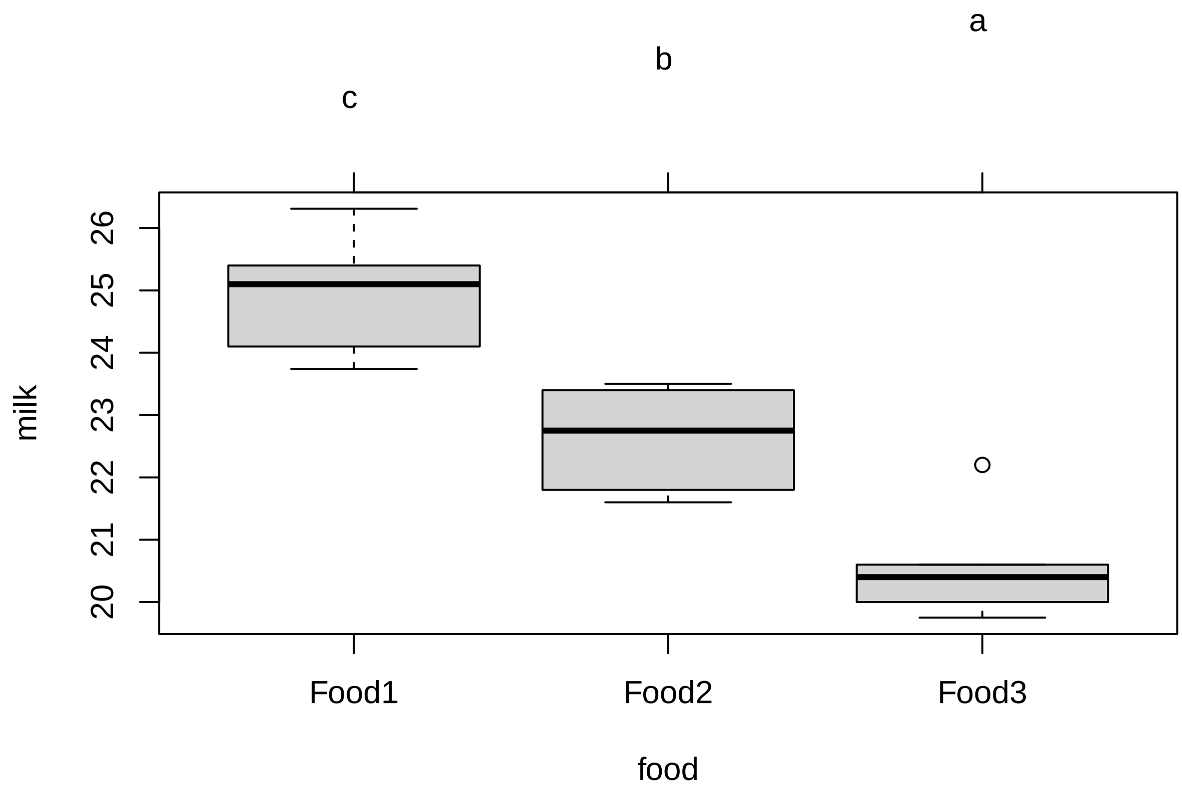
```r
fm<-aov(milk ~ food, data = foods)  # fit the ANOVA (there are different types of ANOVA)
summary(fm)                          #
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## food         2  47.16  23.582    25.6 4.68e-05 ***
## Residuals   12  11.05   0.921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Mean Sq is equal to Sum Sq divided by Df  (degrees of freedom)
# The number Multiple R-squared:  0.8101
# is obtained as 0.8101 = (47.16)/(47.16+11.05)
#F-statistic:  25.6 on 2 and 12 DF,  p-value: 4.684e-05
# the observations are 14 and the groups are 2
# The test reveals that the diet is significant (but we still don't know which diet gives highest
# milk production difference)

# we see that the median is 0.01 (almost 0) and the 1st and 3rd quartiles are almost centered
# around the median


library(multcomp)
```

```
## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'

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

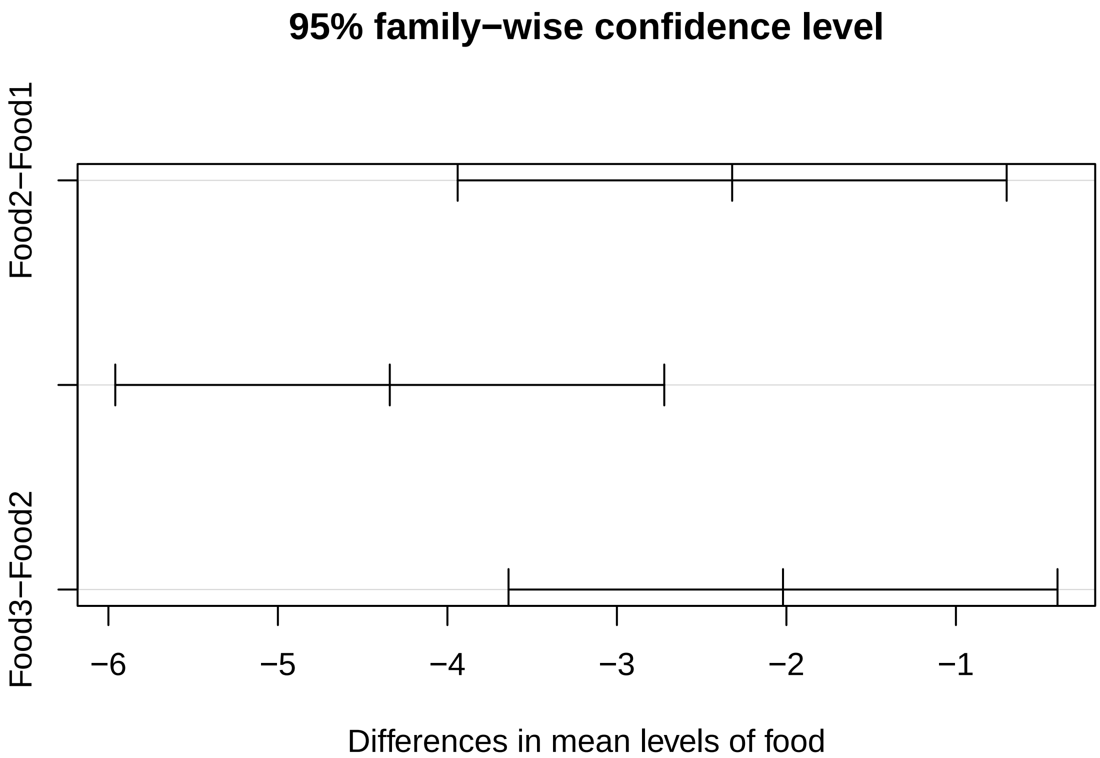
```r
par(mar=c(5,4,6,2))           # Change parameters for the plot margins
tuk <- glht(fm, linfct=mcp(food="Tukey")) # Fit the general Linear Hypotheses
plot(cld(tuk, level=0.05),col="lightgrey")   # Plot the mean differences
```




```
# Multiple comparisons

TukeyHSD(fm)                    # Tukey test for multiple comparisons on the results of the ANOVA
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = milk ~ food, data = foods)
##
## $food
##              diff       lwr        upr     p adj
## Food2-Food1 -2.32 -3.939373 -0.7006265 0.0063517
## Food3-Food1 -4.34 -5.959373 -2.7206265 0.0000322
## Food3-Food2 -2.02 -3.639373 -0.4006265 0.0153900
```

```
plot(TukeyHSD(fm))              # Plot for tukey test
```




```
aggregate(foods$milk, by=list(foods$food), FUN = mean) # Means by group
```

```
##   Group.1     x
## 1   Food1 24.93
## 2   Food2 22.61
## 3   Food3 20.59
```