

POST VR5-KET Image Data ANOVA Mean Cell ns

Jonathan Ramos

2024-03-12

```
library(ggplot2)
library(ggpubr)
library(car) # For levene.test() function
```

```
## Loading required package: carData
```

```
library(emmeans)
```

In this rmd file I will simply repeat the statistical procedures that were performed on the PRE-VR5 ketamine image data on this set of POST-VR5 ketamine image data.

EDA and ANOVA function

This function performs the same type of ANOVA as performed in graphpad prism. In addition, performs some exploratory data analysis to assess normality and homogeneity of variances (both quantitatively and qualitatively)

```
Sidak <- function(pvals)
  # takes a vector of p-values and corrects p-values according to
  # Sidaks method for multiple comparisons (1967)
  #
  # Jonathan Ramos 3/12/2024
  {
    adjusted <- c()
    j <- length(pvals)

    for (i in 1:j){
      adj_p <- 1-(1-pvals[i])^j
      adjusted <- c(adjusted, adj_p)
    }
    return(adjusted)
  }

eda_anova <- function(fname)
  # takes a filename, loads data from csv; data 4 columns:
  # react_treat, react, treat, and norm_int (response var)
  # react_treat is just react and treat in one string separated by "_"
  # builds factor cols for categorical cols (norm_int is numeric, all others are categorical)
  # then performs the following tasks:
  # checks assumptions of normality with qqplot and shapiro wilk tests
  # checks assumptions of equal variances with box plot and levene test
  # performs 2way ANOVA (2 by 2, react by treat)
```

```

# performs post hoc pairwise comparisons (emmeans of levels of react by treat
# and emmeans of levels of treat by react)
# prints out all statistical test results and returns plot objects
# for the two plots: the qqplots and the box plots
#
# Jonathan Ramos 3/12/2024

{
df <- read.csv(fname, header=TRUE, sep=",")
df$react_treat_factor <- as.factor(df$react_treat)
df$react_factor <- as.factor(df$react)
df$treat_factor <- as.factor(df$treat)

### check assumption of normality
# quantitative assessment
print(tapply(df$norm_int, df$react_treat_factor, shapiro.test))

# qualitative assessment
g <- ggqqplot(df, x="norm_int", facet.by=c("treat_factor", "react_factor"))

### check assumption of equal variances
# quantitative assessment
print(leveneTest(y = df$norm_int, group=df$react_treat_factor, center='mean'))

# qualitative assessment
f <- ggplot(df, aes(x=treat_factor, y=norm_int)) + geom_boxplot(aes(fill=treat_factor), alpha=0.5) +
  #geom_dotplot(binaxis = "y", stackdir = "center", dotsize=0.5) +
  facet_wrap(~react_factor) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))

# run the ANOVA, display summary
df.lm <- lm(norm_int ~ treat_factor + react_factor + treat_factor*react_factor, contrasts=list(treat_
df.III.aov <- car::Anova(df.lm, type = 3)
print(df.III.aov)

# post hoc pairwise comparisons
emm <- emmeans(df.lm, ~ treat_factor * react_factor)
p1 <- pairs(emm, simple="treat_factor", adjust="tukey")
p2 <- pairs(emm, simple="react_factor", adjust="tukey")

# add col to summary dataframe containing sidak adjusted p-values
adjusted_p.value1 <- Sidak(summary(p1, adjust="tukey")$p.value)
s1 <- summary(p1)
s1['adjusted_p.value'] <- adjusted_p.value1

adjusted_p.value2 <- Sidak(summary(p2, adjust="tukey")$p.value)
s2 <- summary(p2)
s2['adjusted_p.value'] <- adjusted_p.value2

# display results
print(s1)
print(s2)

```

```

    return(list(g, f))
}

eda_anova_1way <- function(fname)
# doc
{
  df <- read.csv(fname, header=TRUE, sep=",")
  df$treat_factor <- as.factor(df$treat)

  ### check assumption of normality
  # quantitative assessment
  print(tapply(df$norm_int, df$treat_factor, shapiro.test))

  # qualitative assessment
  g <- ggqqplot(df, x="norm_int", facet.by=c("treat_factor"))
  g

  ### check assumption of equal variances
  # quantitative assessment
  print(leveneTest(y=df$norm_int, group=df$treat_factor, center='mean'))

  # qualitative assessment
  f <- ggplot(df, aes(x=treat_factor, y=norm_int)) + geom_boxplot(aes(fill=treat_factor), alpha=0.5) +
  geom_dotplot(binaxis = "y", stackdir = "center", dotsize=0.5)
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
  f

  # run the ANOVA, display summary
  # since this is just a one way ANOVA with only 2 levels, this is essentially just a t-test
  # ANOVA is performed here for consistency
  df.lm <- lm(norm_int ~ treat_factor, data=df)
  df.III.aov <- car::Anova(df.lm, type = 3)
  print(df.III.aov)

  # no post hoc is required for 1way ANOVA with 2 levels
  # return figure objects only
  return(list(g, f))
}

```

pulling out filenames

```

fnames <- list.files(pattern='NORM')
fnames

## [1] "KET-POSTVR5_PV_coloc_w_WFA_NORM_Rsubset.csv"
## [2] "KET-POSTVR5_single_PV_NORM_Rsubset.csv"
## [3] "KET-POSTVR5_single_WFA_NORM_Rsubset.csv"
## [4] "KET-POSTVR5_WFA_coloc_w_PV_NORM_Rsubset.csv"

```

POST-VR5 single PV

```
f <- fnames[2]
print(f)

## [1] "KET-POSTVR5_single_PV_NORM_Rsubset.csv"
figs <- eda_anova_1way(f)

## $`6mg/kg`
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
## W = 0.83731, p-value < 2.2e-16
##
##
## $Saline
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
## W = 0.74231, p-value < 2.2e-16
##
##
## Levene's Test for Homogeneity of Variance (center = "mean")
##      Df F value Pr(>F)
## group  1  1e-04  0.993
##      541
## Anova Table (Type III tests)
##
## Response: norm_int
##              Sum Sq Df F value Pr(>F)
## (Intercept)  326.21  1 418.8176 <2e-16 ***
## treat_factor    0.00  1   0.0002  0.9875
## Residuals    421.38 541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

POST-VR5 single WFA

```
f <- fnames[3]
print(f)

## [1] "KET-POSTVR5_single_WFA_NORM_Rsubset.csv"
figs <- eda_anova_1way(f)

## $`6mg/kg`
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
## W = 0.84039, p-value = 6.36e-13
##
## $Saline
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
## W = 0.72447, p-value = 3.581e-15
##
##
## Levene's Test for Homogeneity of Variance (center = "mean")
##      Df F value    Pr(>F)
## group  1 37.249 2.941e-09 ***
##      327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Anova Table (Type III tests)
##
## Response: norm_int
##              Sum Sq Df F value    Pr(>F)
## (Intercept)  75.041   1 183.88 < 2.2e-16 ***
## treat_factor  10.591   1  25.95 5.932e-07 ***
## Residuals    133.452 327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

POST-VR5 PV coloc w/WFA

```
f <- fnames[1]
print(f)

## [1] "KET-POSTVR5_PV_coloc_w_WFA_NORM_Rsubset.csv"
figs <- eda_anova_1way(f)

## $`6mg/kg`
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
## W = 0.90678, p-value = 5.327e-07
##
##
## $Saline
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
## W = 0.74476, p-value = 3.304e-11
##
##
## Levene's Test for Homogeneity of Variance (center = "mean")
##      Df F value Pr(>F)
## group  1  0.2284 0.6332
##      206
## Anova Table (Type III tests)
##
## Response: norm_int
##              Sum Sq Df F value    Pr(>F)
## (Intercept) 224.211  1 284.0096 < 2.2e-16 ***
## treat_factor  7.312  1   9.2625  0.002644 **
## Residuals   162.626 206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

POST-VR5 WFA coloc w/ PV

```
f <- fnames[4]
print(f)

## [1] "KET-POSTVR5_WFA_coloc_w_PV_NORM_Rsubset.csv"
figs <- eda_anova_1way(f)

## $`6mg/kg`
##
## Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.85703, p-value = 2.667e-09
##
##
## $Saline
##
## Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.85476, p-value = 6.257e-08
##
##
## Levene's Test for Homogeneity of Variance (center = "mean")
##      Df F value    Pr(>F)
## group  1  30.27 1.108e-07 ***
##      206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Anova Table (Type III tests)
##
## Response: norm_int
##              Sum Sq Df F value    Pr(>F)
## (Intercept)  60.649   1 188.195 < 2.2e-16 ***
## treat_factor   4.092   1  12.696 0.0004553 ***
## Residuals    66.387 206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```