

# Additional data analysis

Andreas Gammelgaard Damsbo

Knitted: 06 January, 2022

## Import

```
dta_all<-read.csv("/Volumes/Data/depression/dep_dataset.csv")
```

## Defining patients to include for analysis

Only including cases with complete pase\_0 and MDI at 1 & 6 months

```
dta<-dta_all[!is.na(dta_all$pase_0),]  
# &!is.na(dta$mdi_1)&!is.na(dta$mdi_6)
```

## Formatting

## Summaries

Fraction with inc\_time of at least 166 days

```
dt<-as.numeric(dta[dta$excluded_6%in%c("dt_6","en_6"),c("inc_time")])>=166  
summary(dt)
```

```
##      Mode  FALSE   TRUE  
## logical      74   434
```

```
length(dt[dt==TRUE])/length(dt)*100 # Percent after 166 days
```

```
## [1] 85.43307
```

5% percentiler

```
quantile(dt, probs = seq(0, 1, 0.05), names = TRUE)
```

```
##   0%   5%  10%  15%  20%  25%  30%  35%  40%  45%  50%  55%  60%  65%  70%  75%  
##   0    0    0    1    1    1    1    1    1    1    1    1    1    1    1    1  
##  80%  85%  90%  95% 100%  
##   1    1    1    1    1
```

Base version

```
aggregate(pase_6 ~ rtreat, data = dta, summary)
```

```
##   rtreat pase_6.Min. pase_6.1st Qu. pase_6.Median pase_6.Mean pase_6.3rd Qu.
## 1 Active      0.0000      71.4800      139.5100      146.1080      206.0500
## 2 Placebo     0.0000      79.5700      136.0000      151.6013      211.0000
##   pase_6.Max.
## 1      492.3800
## 2      506.3500
```

Fancy version

```
psych::describeBy(dta$pase_6, dta$rtreat,mat=T)
```

```
##   item group1 vars  n   mean      sd median trimmed   mad min  max
## X11   1 Active   1 253 146.1080 90.80133 139.51 139.7366 98.66703  0 492.38
## X12   2 Placebo  1 269 151.6013 94.85593 136.00 144.5774 93.53723  0 506.35
##   range      skew kurtosis      se
## X11 492.38 0.7409903 0.6931755 5.708631
## X12 506.35 0.8200586 0.7402124 5.783468
```

## Mann-Whitney U test

See: <https://stat-methods.com/home/mann-whitney-u-r/>

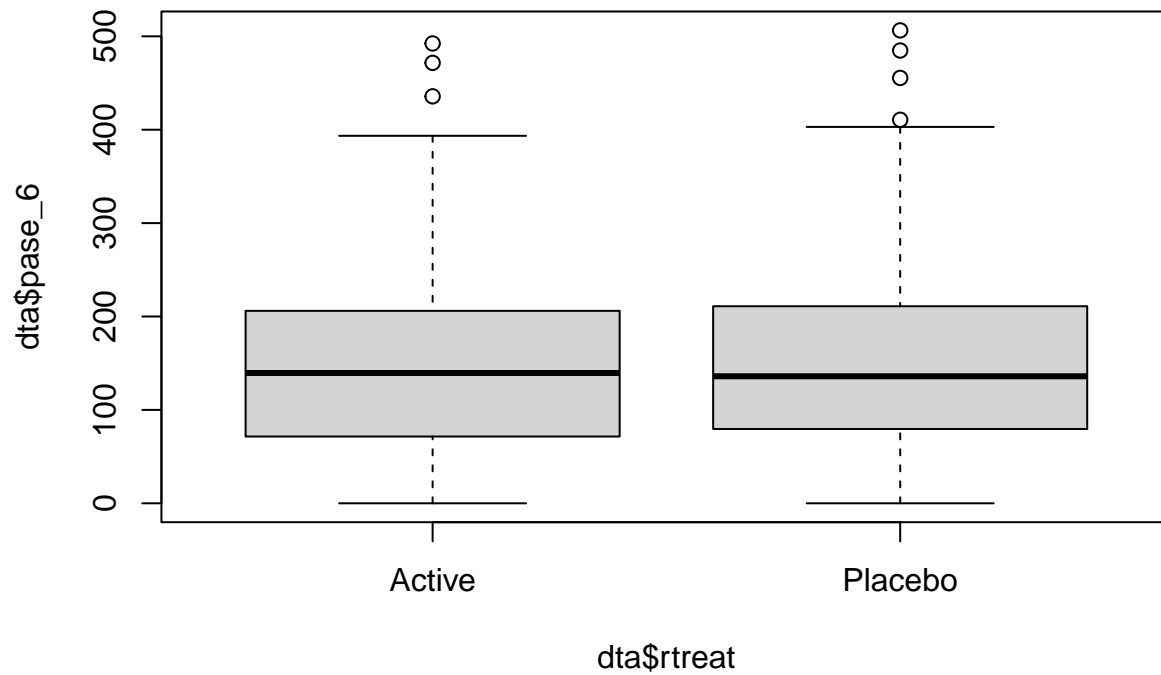
```
#Perform the Mann-Whitney U test
m1<-wilcox.test(pase_6 ~ rtreat, data=dta, na.rm=TRUE,
                paired=FALSE, exact=FALSE, conf.int=TRUE)
print(m1)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  pase_6 by rtreat
## W = 33188, p-value = 0.6259
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -19.67994 11.20995
## sample estimates:
## difference in location
## -3.790064
```

## Boxplot

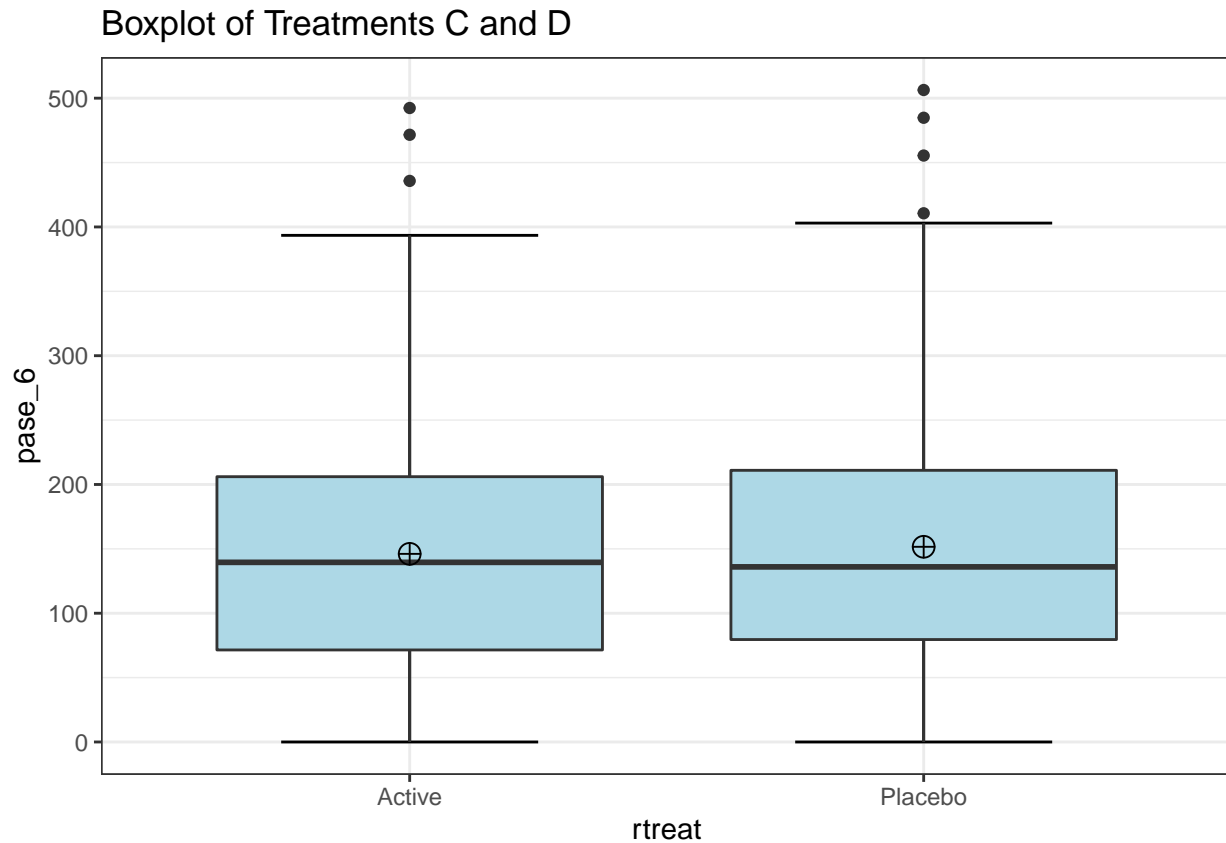
Base function - simple

```
boxplot(dta$pase_6 ~ dta$rtreat)
```



### ggplot2 - fancy version

```
library(ggplot2)
ggplot(dta, aes(x = rtreat, y = pase_6, fill = rtreat)) +
  stat_boxplot(geom = "errorbar", width = 0.5) +
  geom_boxplot(fill = "light blue") +
  stat_summary(fun.y=mean, geom="point", shape=10, size=3.5, color="black") +
  # Point symbol is mean value
  ggtitle("Boxplot of Treatments C and D") +
  theme_bw() + theme(legend.position="none")
```



### Bonus: QQ plots

```
library(qqplotr)
ggplot(data = dta, mapping = aes(sample = pase_6, color = rtreat, fill = rtreat)) +
  stat_qq_band(alpha=0.5, conf=0.95, qtype=1, bandType = "boot") +
  stat_qq_line(identity=TRUE) +
  stat_qq_point(col="black") +
  facet_wrap(~ rtreat, scales = "free") +
  labs(x = "Theoretical Quantiles", y = "Sample Quantiles") + theme_bw()
```

