normsANOVA

Aaron Dantzler

2023-07-18

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.1

## Warning: package 'ggplot2' was built under R version 4.3.1

## Warning: package 'tidyr' was built under R version 4.3.1

## Warning: package 'stringr' was built under R version 4.3.1

## Warning: package 'forcats' was built under R version 4.3.1

## Warning: package 'lubridate' was built under R version 4.3.1

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.3.1
```

```
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.3.1

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

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
data <- read.csv(file = "D:\\Princeton\\BSPL\\norms.csv")
head(data)
```

```
##                   prolific  control    treatment frq_topic_t1
## 1 546e3778fdf99b2bc7ebcff6  climate     evidence            6
## 2 55519750fdf99b7f2114cc3e   health normevidence            4
## 3 55a29659fdf99b5ff49937d3  climate         norm            4
## 4 55b9a9b0fdf99b6906d2aba4 politics         norm            1
## 5 55c43918fdf99b080551e044 politics normevidence            4
## 6 55d0945934e9060005e57258 politics     evidence            4
##
## 1
## 2
## 3
## 4
## 5 I believe access to mental health matters because a lot of people have mental issues, but they don
## 6
##   frq_t1_character_count frq_t1_word_count importance_t1_1 importance_t1_2
## 1                    329                69               3               3
## 2                    396                70               3               3
## 3                    397                65               3               2
## 4                    328                57               4               4
## 5                    694               117               3               4
## 6                    290                54               1               1
##   importance_t1_3 importance_t1_4 importance_t1_5 importance_t1_6 frq_topic_t2
## 1               5               5               4               5            6
## 2               5               5               2               5            6
## 3               3               4               2               3            3
## 4               4               3               3               3            3
## 5               3               4               3               4            6
## 6               1               4               1               1            3
##
## 1
## 2
## 3
## 4
## 5 It is important that everyone be able to get paid enough to live. It's ridiculous that people can
```

```
## 6
##   frq_t2_character_count frq_t2_word_count importance_t2_1 importance_t2_2
## 1                    395               79               3               4
## 2                    488               88               2               5
## 3                    371               65               3               3
## 4                    342               51               4               4
## 5                    761              134               4               4
## 6                    287               49               4               5
##   importance_t2_3 importance_t2_4 importance_t2_5 importance_t2_6 frq_topic_t3
## 1               5               5               4               5            3
## 2               5               5               5               5            4
## 3               4               3               3               2            4
## 4               5               4               4               3            1
## 5               3               5               3               5            6
## 6               5               5               1               1            3
##
## 1
## 2
## 3
## 4
## 5 Increasing the federal minimum wage is extremely important since it's ridiculous that people can't
## 6
##   frq_t3_character_count frq_t3_word_count importance_t3_1 importance_t3_2
## 1                    289               59               3               4
## 2                    385               71               2               3
## 3                    442               79               3               3
## 4                    330               55               4               4
## 5                    799              134               4               3
## 6                    301               51               3               3
##   importance_t3_3 importance_t3_4 importance_t3_5 importance_t3_6
## 1               5               3               3               4
## 2               5               5               3               5
## 3               4               5               3               3
## 4               4               3               3               3
## 5               3               4               3               5
## 6               1               3               1               3
```

```
data <- data[-nrow(data), ]
```

```
new_df <- data.frame(
  prolific = data$prolific,
  treatment = data$treatment
)
```

```
new_df$t1_control <- ifelse(data$control == "climate", (data$importance_t1_1 +
                            data$importance_t1_2) / 2,
                     ifelse(data$control == "health", (data$importance_t1_3 +
                            data$importance_t1_4) / 2,
                      ifelse(data$control == "politics", (data$importance_t1_5 +
                            data$importance_t1_6) / 2, NA)))
```

```
new_df$t1_treat <- ifelse(data$control == "climate", (data$importance_t1_3 +
                            data$importance_t1_4 + data$importance_t1_5 +
                              data$importance_t1_6) / 4,
                    ifelse(data$control == "health", (data$importance_t1_1 +
                            data$importance_t1_2 + data$importance_t1_5 +
                              data$importance_t1_6) / 4,
                     ifelse(data$control == "politics", (data$importance_t1_1 +
                            data$importance_t1_2 + data$importance_t1_3 +
                              data$importance_t1_4) / 4, NA)))
```

```
new_df$t2_control <- ifelse(data$control == "climate", (data$importance_t2_1 +
                            data$importance_t2_2) / 2,
                    ifelse(data$control == "health", (data$importance_t2_3 +
                            data$importance_t2_4) / 2,
                     ifelse(data$control == "politics", (data$importance_t2_5 +
                            data$importance_t2_6) / 2, NA)))
```

```
new_df$t2_treat <- ifelse(data$control == "climate", (data$importance_t2_3 +
                            data$importance_t2_4 + data$importance_t2_5 +
                              data$importance_t2_6) / 4,
                    ifelse(data$control == "health", (data$importance_t2_1 +
                            data$importance_t2_2 + data$importance_t2_5 +
                              data$importance_t2_6) / 4,
                     ifelse(data$control == "politics", (data$importance_t2_1 +
                            data$importance_t2_2 + data$importance_t2_3 +
                              data$importance_t2_4) / 4, NA)))
```

```
new_df$t3_control <- ifelse(data$control == "climate", (data$importance_t3_1 +
                            data$importance_t3_2) / 2,
                    ifelse(data$control == "health", (data$importance_t3_3 +
                            data$importance_t3_4) / 2,
                     ifelse(data$control == "politics", (data$importance_t3_5 +
                            data$importance_t3_6) / 2, NA)))
```

```
new_df$t3_treat <- ifelse(data$control == "climate", (data$importance_t3_3 +
                            data$importance_t3_4 + data$importance_t3_5 +
                              data$importance_t3_6) / 4,
                    ifelse(data$control == "health", (data$importance_t3_1 +
                            data$importance_t3_2 + data$importance_t3_5 +
                              data$importance_t3_6) / 4,
                     ifelse(data$control == "politics", (data$importance_t3_1 +
                            data$importance_t3_2 + data$importance_t3_3 +
                              data$importance_t3_4) / 4, NA)))
```

```
data <- new_df
```

```
# Convert to long-form (prolific, treatment, time {t1, t2, t3}, condition {control, treat})

df <- data %>%
  pivot_longer(c('t1_control', 't2_control', 't3_control',
                 't1_treat', 't2_treat', 't3_treat'), names_to = 'tmp') %>%
```

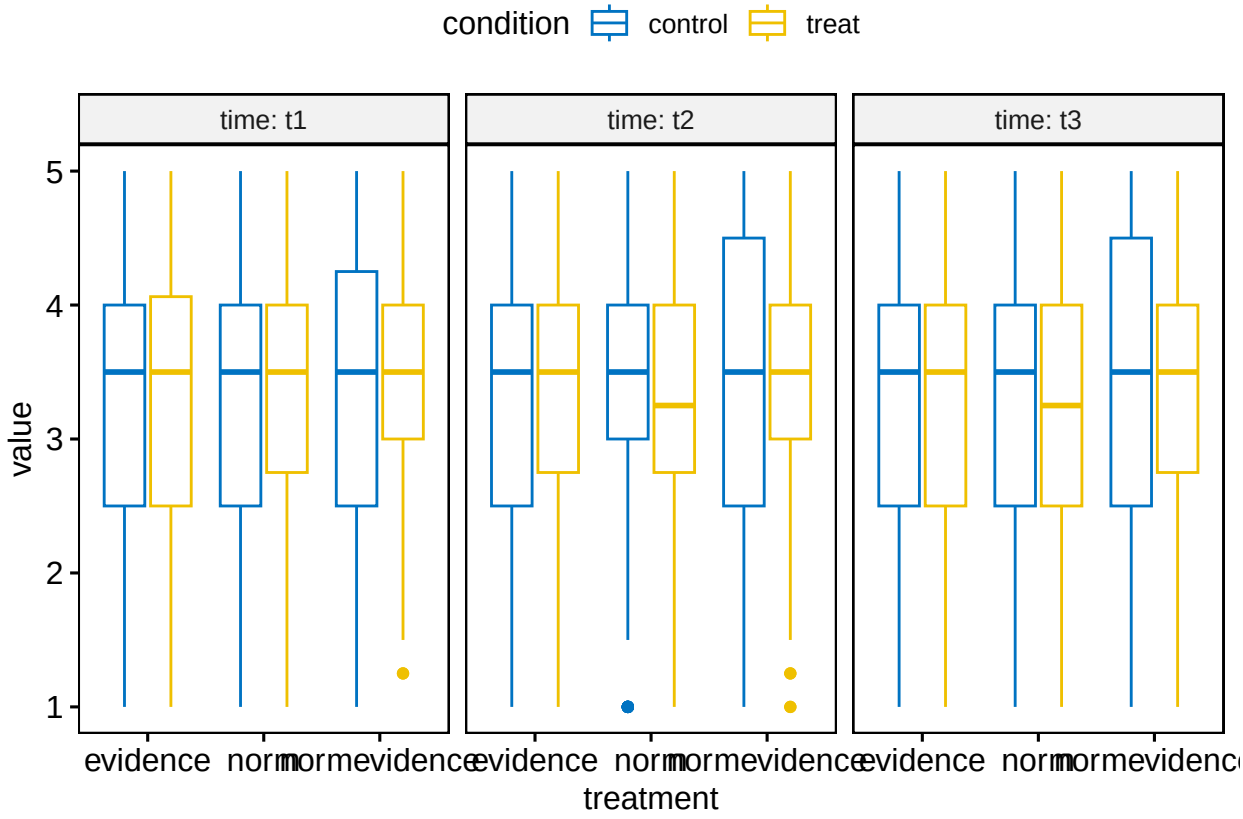
```
  group_by(prolific) %>%
  mutate('condition' = ifelse(grepl(pattern = 'control', x = tmp), 'control', 'treat')) %>%
  mutate('time' = str_extract(pattern = 't\\d', string = tmp)) %>%
  dplyr::select(-tmp)
```

```
# Convert to factor variables

df$prolific <- as.factor(df$prolific)
df$treatment <- as.factor(df$treatment)
df$condition <- as.factor(df$condition)
df$time <- as.factor(df$time)

data <- data.frame(df)
```

```
# Summary statistics
stats <- df %>%
  group_by(condition, time) %>%
  get_summary_stats(value, type = "mean_sd")
stats
```

```
## # A tibble: 6 x 6
##   condition time  variable     n  mean    sd
##   <fct>     <fct> <fct>    <dbl> <dbl> <dbl>
## 1 control   t1    value      616  3.36 1.12
## 2 control   t2    value      616  3.34 1.13
## 3 control   t3    value      616  3.34 1.12
## 4 treat     t1    value      616  3.37 0.931
## 5 treat     t2    value      616  3.34 0.923
## 6 treat     t3    value      616  3.33 0.953
```

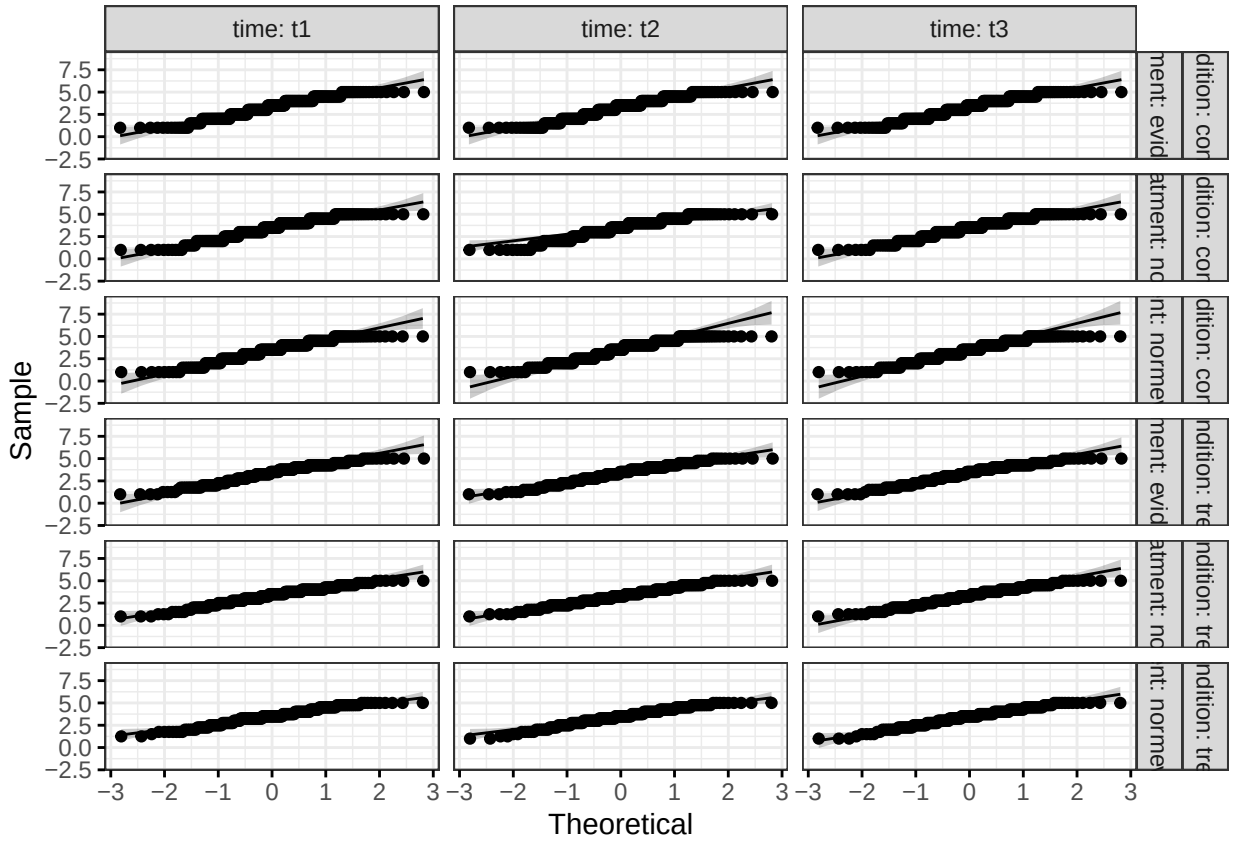
```
# Box plot
bxp <- ggboxplot(
  df, x = "treatment", y = "value",
  color = "condition", palette = "jco",
  facet.by = "time", short.panel.labs = FALSE
)
bxp
```




```
# Outliers
outliers <- df %>%
  group_by(treatment, condition, time) %>%
  identify_outliers(value)
outliers
```

```
## # A tibble: 16 x 7
##    treatment    condition time  prolific             value is.outlier is.extreme
##    <fct>        <fct>     <fct> <fct>                <dbl> <lgl>      <lgl>
##  1 norm         control   t2    5be1ac50ca095c0001b~  1    TRUE       FALSE
##  2 norm         control   t2    5ca3d57fa2941200168~  1    TRUE       FALSE
##  3 norm         control   t2    5d615f7632cfbd001b1~  1    TRUE       FALSE
##  4 norm         control   t2    5df1822669a11709f14~  1    TRUE       FALSE
##  5 norm         control   t2    5ff4997ab11116b6e41~  1    TRUE       FALSE
##  6 norm         control   t2    60712d937752fb8780e~  1    TRUE       FALSE
##  7 norm         control   t2    60df4e94499bfbcaa37~  1    TRUE       FALSE
##  8 norm         control   t2    615327ab3d238b03236~  1    TRUE       FALSE
##  9 norm         control   t2    6298df7da7b45df1b73~  1    TRUE       FALSE
## 10 norm         control   t2    6334b0fc574064de4e6~  1    TRUE       FALSE
## 11 normevidence treat     t1    55e5d5fdc70c7a000b2~  1.25 TRUE       FALSE
## 12 normevidence treat     t1    5d1511f1e814b800016~  1.25 TRUE       FALSE
## 13 normevidence treat     t2    55e5d5fdc70c7a000b2~  1.25 TRUE       FALSE
## 14 normevidence treat     t2    56e6a66af6ed900006a~  1.25 TRUE       FALSE
## 15 normevidence treat     t2    5d1511f1e814b800016~  1    TRUE       FALSE
## 16 normevidence treat     t2    62dc2f45630616c5920~  1    TRUE       FALSE
```

```
# Normality
ggqqplot(df, "value", ggtheme = theme_bw()) +
  facet_grid(condition + treatment ~ time, labeller = "label_both")
```



Within Condition and Time

```
# ANOVA
res.aov <- anova_test(
  data = data, dv = value, wid = prolific,
  between = treatment, within = c(condition, time)
)
get_anova_table(res.aov)
```

```
## ANOVA Table (type III tests)
##
##                      Effect DFn  DFd     F     p p<.05      ges
## 1                 treatment   2  613 1.513 0.221       3.00e-03
## 2                 condition   1  613 0.004 0.952       1.72e-06
## 3                      time   2 1226 1.089 0.337       1.25e-04
## 4       treatment:condition   2  613 1.528 0.218       1.00e-03
## 5            treatment:time   4 1226 0.381 0.823       8.77e-05
## 6            condition:time   2 1226 0.471 0.625       2.96e-05
## 7 treatment:condition:time   4 1226 1.137 0.338       1.43e-04
```

```
# head(df)
```

The results indicate there there is not a statistically significant difference between the means of all of these groups.