Course: Multivariate statistics (AUT23)

Chapter 5: Repeated measurements

5.13 Time to practice on your own

5.13.1 Exercise 1: strengthening environmental protection over time

Use the data from the Selects 2019 Panel Survey and assess whether respondents' stance towards strengthening environmental protection has increased over the first three waves (before, during and after the campaign).

Start by downloading the data and by selecting the variables.

- Show the code

Next, reshape the data so that there are in a long format.

- Show the code

Then, check the normality of the dependent variable and the sphericity (variances of the differences must be homogeneous):

- Show the code

The assumption of sphericity will be automatically checked during the computation of the ANOVA test using the R function anova_test(). By using the function get_anova_table() to extract the ANOVA table, the Greenhouse-Geisser sphericity correction is automatically applied to factors violating the sphericity assumption. Now, we can check whether there are group differences:

- Show the code

Finally, we can assess which group (or time) differences are statistically significant:

- Show the code

5.13.2 Exercise 2: fictive score over time

Let's create a dataset containing a score measured at three points in time. In a second step, we will investigate if (frequently) working in group can induce a significant increase of the score over time.

- Show the code

Now, we will test whether there is significant interaction between working in group and time on the score. We can use boxplots of the score colored by working in group:

- Show the code

We can check whether there are outliers:

- Show the code

We next compute Shapiro-Wilk test to test for the normality assumption for each combinations of factor levels:

- Show the code

We can assess whether there is a statistically significant two-way interactions between group work and time:

- Show the code

Chapter 5: Repeated measurements (*answers)*

5.13 Time to practice on your own

5.13.1 Exercise 1: strengthening environmental protection over time

Use the data from the Selects 2019 Panel Survey and assess whether respondents' stance towards strengthening environmental protection has increased over the first three waves (before, during and after the campaign).

Start by downloading the data and by selecting the variables.

➢ Show the code

```
library(foreign)
db <- read.spss(file=paste0(getwd(),
                "/data/1184_Selects2019_Panel_Data_v4.0.sav"),
                use.value.labels = F,
                to.data.frame = T)
sel <- db |>
  dplyr::select(id,
    # wave 1
    W1_f15340d,
    # wave 2
    W2_f15340d,
    # wave 3
    W3_f15340d) |>
  stats::na.omit()
# inverse the scale
sel$W1_f15340d=(sel$W1_f15340d-6)*(-1)
sel$W2_f15340d=(sel$W2_f15340d-6)*(-1)
sel$W3_f15340d=(sel$W3_f15340d-6)*(-1)
```

Next, reshape the data so that there are in a long format.

➢ Show the code

```
long <- reshape(as.data.frame(sel),
```

```
                direction="long",
                varying = c("W1_f15340d","W2_f15340d","W3_f15340d"),
                v.names = "pro_env",
                times =c("wave1","wave2","wave3"))
```

Then, check the normality of the dependent variable and the sphericity (variances of the differences must be homogeneous):

➢ Show the code

```
# Shapiro-Wilk test
long |>
  dplyr::group_by(time) |>
  rstatix::shapiro_test(pro_env)
## # A tibble: 3 × 4
##   time  variable statistic       p
##   <chr> <chr>       <dbl>   <dbl>
## 1 wave1 pro_env      0.773 8.31e-45
## 2 wave2 pro_env      0.743 8.74e-47
## 3 wave3 pro_env      0.786 6.03e-44
```

The assumption of sphericity will be automatically checked during the computation of the ANOVA test using the R function anova_test(). By using the function get_anova_table() to extract the ANOVA table, the Greenhouse-Geisser sphericity correction is automatically applied to factors violating the sphericity assumption. Now, we can check whether there are group differences:

➢ Show the code

```
# group differences
res.aov <- rstatix::anova_test(data = long,
                dv = pro_env,
                wid = id,
                within = time)
rstatix::get_anova_table(res.aov)
## ANOVA Table (type III tests)
##
```

```
##   Effect  DFn     DFd      F       p p<.05   ges
## 1   time 1.99 3655.01 37.718 7.76e-17     * 0.004
# non-parametric test
# kruskal.test(long$pro_env, long$time)
```

Finally, we can assess which group (or time) differences are statistically significant:

➢ Show the code

```
# Post-hoc test to assess differences
pwc <- long |>
  rstatix::pairwise_t_test(
    pro_env ~ time,
    paired = TRUE,
    p.adjust.method = "bonferroni"
    )
pwc[,c(2,3,6,8,10)]
## # A tibble: 3 × 5
##   group1 group2 statistic       p p.adj.signif
##   <chr>  <chr>      <dbl>    <dbl> <chr>
## 1 wave1  wave2     -3.15 2   e- 3 **
## 2 wave1  wave3      5.26 1.58e- 7 ****
## 3 wave2  wave3      8.80 3.19e-18 ****
```

5.13.2 Exercise 2: fictive score over time

Let's create a dataset containing a score measured at three points in time. In a second step, we will investigate if (frequently) working in group can induce a significant increase of the score over time.

➢ Show the code

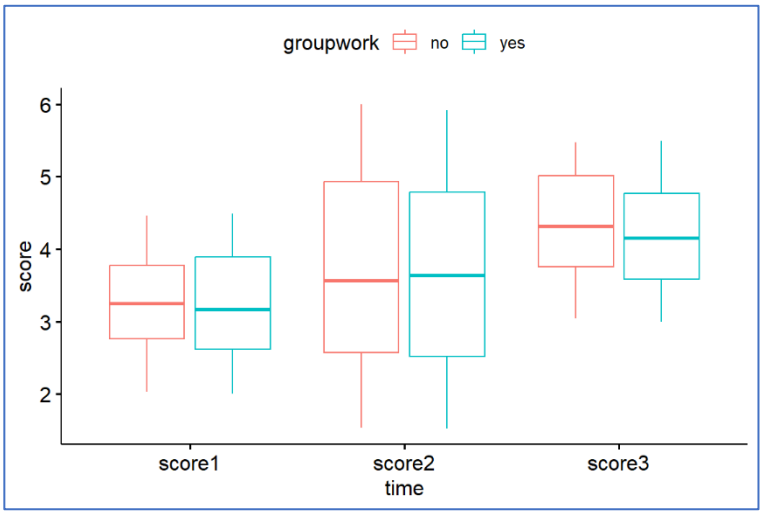
```
data <- data.frame(matrix(nrow = 200, ncol = 0))
set.seed(123)
data$score1 <- runif(nrow(data), min=2, max=4.5)
data$score2 <- runif(nrow(data), min=1.5, max=6)
```

```
data$score3 <- runif(nrow(data), min=3, max=5.5)
# assign id
data$id = rep(seq(1:100),2)
# assign group work variable
data$groupwork = c(rep(c("yes"),100), rep(c("no"),100))
# copy of the data
copy = data
# re-arrange the data
data <- data |>
 tidyr::gather(key = "time", value = "score", score1, score2, score3) |>
 rstatix::convert_as_factor(id, time)
```

Now, we will test whether there is significant interaction between working in group and time on the score. We can use boxplots of the score colored by working in group:

➢ Show the code

```
ggpubr::ggboxplot(
 data, x = "time",
 y = "score",
 color = "groupwork"
 )
```



We can check whether there are outliers:

- Show the code

```
data |>
  dplyr::group_by(groupwork, time) |>
  rstatix::identify_outliers(score)
## [1] groupwork  time      id        score     is.outlier is.extreme
## <0 lignes> (ou 'row.names' de longueur nulle)
```

We next compute Shapiro-Wilk test to test for the normality assumption for each combinations of factor levels:

- Show the code

```
# Shapiro
data |>
  dplyr::group_by(groupwork, time) |>
  rstatix::shapiro_test(score)
## # A tibble: 6 × 5
##   groupwork time   variable statistic       p
##   <chr>     <fct>  <chr>       <dbl>   <dbl>
## 1 no        score1 score       0.964 0.00736
## 2 no        score2 score       0.950 0.000789
## 3 no        score3 score       0.945 0.000379
## 4 yes       score1 score       0.952 0.00119
## 5 yes       score2 score       0.945 0.000418
## 6 yes       score3 score       0.948 0.000592
```

We can assess whether there is a statistically significant two-way interactions between group work and time:

- Show the code

```
# We also need to convert id and time into factor variables
# data$groupwork <- as.factor(data$groupwork)
data$time <- as.factor(data$time)
```

```
data$id <- as.factor(data$id)
res.aov <- rstatix::anova_test(
  data = data,
  dv = score,
  wid = id,
  within = c(groupwork, time)
  )
rstatix::get_anova_table(res.aov)
## ANOVA Table (type III tests)
##
##          Effect  DFn    DFd      F       p p<.05      ges
## 1     groupwork 1.00  99.00  0.454 5.02e-01       0.000696
## 2          time 1.65 163.74 46.979 7.74e-15     * 0.153000
## 3 groupwork:time 1.68 166.31  0.050 9.28e-01       0.000163
```