R Notes:

o Repeated Measures ANOVA

Repeated measures ANOVA is the equivalent of the one-way ANOVA, but for related, not independent groups, and is the extension of the dependent t-test. We can analyze data using a repeated measures ANOVA for two types of study design. Studies that investigate either (1) changes in mean scores over three or more time points, or (2) differences in mean scores under three or more different conditions. For example, for (1), you might be investigating the effect of a 6-month exercise training program on blood pressure and want to measure blood pressure at 3 separate time points (pre-, midway and post-exercise intervention), which would allow you to develop a time-course for any exercise effect. For (2), you might get the same subjects to eat different types of cake (chocolate, caramel and lemon) and rate each one for taste, rather than having different people taste each different cake. The important point with these two study designs is that the same people are being measured more than once on the same dependent variable (i.e., why it is called repeated measures).

The null hypothesis (H$_0$) states that the means are equal: $H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ where μ = population mean and k = number of related groups. The alternative hypothesis (H$_A$) states that the related population means are not equal (at least one mean is different to another mean): H_A: *at least two means are significantly different.*

In your base (Repeated Measures ANOVA) heart rate was measured before and shortly after administration of enalapril, an angiotensin converting enzyme inhibitor in nine patients with congestive heart failure. You want to investigate if the average heart rate changes over time.

Load and have a look at the data

Import the file "Repeated Measures ANOVA.xlsx" you will work with and inspect the data frame:

The following output is obtained:

```
Classes 'tbl_df', 'tbl' and 'data.frame':      9 obs. of   5 variables:
 $ Subjects: num   1 2 3 4 5 6 7 8 9
 $ t0      : num   96 110 89 95 128 100 72 79 100
 $ t30     : num   92 106 86 78 124 98 68 75 106
 $ t60     : num   86 108 85 78 118 100 67 74 104
 $ t120    : num   92 114 83 83 118 94 71 74 102
```

Data preparation

```
# Data preparation
# Wide format
# Gather columns t0, t30, t60 and t120 into long format
# Convert Subjects and time into factor variables

data <- data %>%
    gather(key = "time", value = "heart_rate", t0, t30, t60, t120) %>%
    convert_as_factor(Subjects, time)
 head(data)
```

The following output is obtained:

```
# A tibble: 6 x 3
  Subjects time  heart_rate
  <fct>    <fct>      <dbl>
1 1        t0            96
2 2        t0           110
3 3        t0            89
4 4        t0            95
5 5        t0           128
6 6        t0           100
```

Summary statistics

Compute some summary statistics of the heart rate by groups (time): mean and sd (standard deviation)

```
data %>%
  group_by(time) %>%
  get_summary_stats(heart_rate, type = "mean_sd")
```
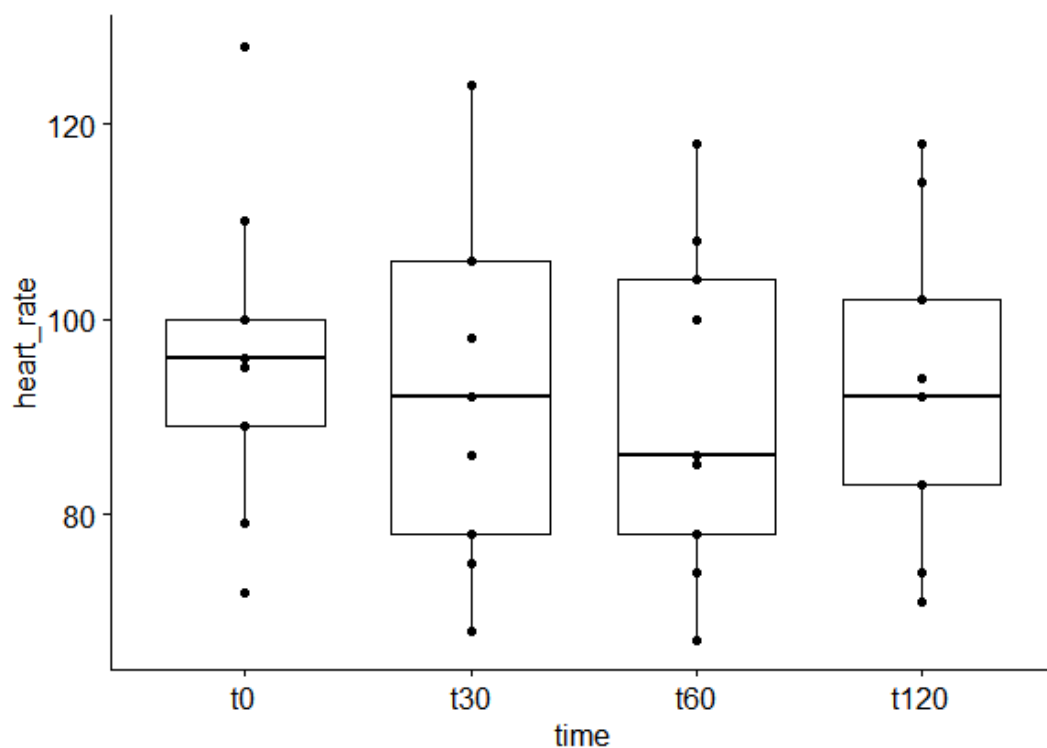
The following output is obtained:

```
# A tibble: 4 x 5
  time  variable      n  mean    sd
  <fct> <chr>      <dbl> <dbl> <dbl>
1 t0    heart_rate     9  96.6  16.4
2 t120  heart_rate     9  92.3  16.5
3 t30   heart_rate     9  92.6  17.8
4 t60   heart_rate     9  91.1  17.2
```

<u>**Visualization**</u>

Create a box plot and add points corresponding to individual values:

```
data$time <- factor(data$time , levels=c("t0", "t30", "t60", "t120"))
bxp <- ggboxplot(data, x = "time", y = "heart_rate", add = "point")
bxp
```


<u>**Check assumptions**</u>

1) **Outliers**

Outliers can be easily identified using box plot methods, implemented in the R function identify_outliers() [rstatix package].

```
data %>%
   group_by(time) %>%
   identify_outliers(heart_rate)
```

The following output is obtained:

```
# A tibble: 2 x 5
  time  Subjects heart_rate is.outlier is.extreme
  <fct> <fct>         <dbl> <lgl>      <lgl>
1 t0    5               128 TRUE       FALSE
2 t0    7                72 TRUE       FALSE
```

There were no extreme outliers.

Note that, in the situation where you have extreme outliers, this can be due to: 1) data entry errors, measurement errors or unusual values.

You can include the outlier in the analysis anyway if you do not believe the result will be substantially affected. This can be evaluated by comparing the result of the ANOVA with and without the outlier.

2) **Normality assumption**

The normality assumption can be checked by computing Shapiro-Wilk test for each time point. If the data is normally distributed, the p-value should be greater than 0.05.

```
data %>%
   group_by(time) %>%
   shapiro_test(heart_rate)
```

The following output is obtained:

```
# A tibble: 4 x 4
  time  variable   statistic     p
  <fct> <chr>          <dbl> <dbl>
1 t0    heart_rate     0.964 0.838
2 t120  heart_rate     0.943 0.616
3 t30   heart_rate     0.970 0.898
4 t60   heart_rate     0.960 0.798
```

The heart rates were normally distributed at each time point, as assessed by Shapiro-Wilk's test (p > 0.05).

3) Assumption of sphericity

The assumption of sphericity will be automatically checked during the computation of the ANOVA test using the R function anova_test() [rstatix package]. The Mauchly's test is internally used to assess the sphericity assumption.

By using the function get_anova_table() [rstatix] to extract the ANOVA table, the Greenhouse-Geisser sphericity correction is automatically applied to factors violating the sphericity assumption.

Computation

```
res.aov <- anova_test(data = data, dv = heart_rate, wid = Subjects, within = time)
res.aov
get_anova_table(res.aov)
```

The following output is obtained:

```
$ANOVA
  Effect DFn DFd   F     p p<.05   ges
1   time   3  24 4.07 0.018     * 0.016

$`Mauchly's Test for Sphericity`
  Effect     W     p p<.05
1   time 0.471 0.412

$`Sphericity Corrections`
  Effect   GGe       DF[GG] p[GG] p[GG]<.05    HFe      DF[HF] p[HF] p[HF]<.05
1   time 0.707 2.12, 16.96 0.034           * 0.968 2.9, 23.23 0.019         *

> get_anova_table(res.aov)
ANOVA Table (type III tests)

  Effect DFn DFd   F     p p<.05   ges
1   time   3  24 4.07 0.018     * 0.016
```

The assumption of sphericity is not violated (0.412>0.05)

The heart rates were statistically significantly different at the different time points, $F(3,24) = 4.07$, $p=0.018$, eta2[g] = 0.02.

where,

- F Indicates that we are comparing to an F-distribution (F-test); (3,24) indicates the degrees of freedom in the numerator (DFn) and the denominator (DFd), respectively; 4.07 indicates the obtained F-statistic value
- p specifies the p-value
- ges is the generalized effect size (amount of variability due to the within-subjects factor)

Post-hoc tests

You can perform multiple **pairwise paired t-tests** between the levels of the within-subjects factor (here time). P-values are adjusted using the Bonferroni multiple testing correction method.

```
# pairwise comparisons
pwc <- data %>%
  pairwise_t_test(
    heart_rate ~ time, paired = TRUE,
    p.adjust.method = "bonferroni"
  )
pwc
```

The following output is obtained:

```
# A tibble: 6 x 10
  .y.            group1 group2    n1    n2 statistic    df     p p.adj p.adj.signif
* <chr>          <chr>  <chr>  <int> <int>     <dbl> <dbl> <dbl> <dbl> <chr>
1 heart_rate t0  t120          9     9      2.43      8 0.041 0.248 ns
2 heart_rate t0  t30           9     9      2.05      8 0.074 0.447 ns
3 heart_rate t0  t60           9     9      2.63      8 0.03  0.181 ns
4 heart_rate t120 t30          9     9     -0.142     8 0.89  1     ns
5 heart_rate t120 t60          9     9      0.866     8 0.412 1     ns
6 heart_rate t30  t60          9     9      1.48      8 0.176 1     ns
```

All the pairwise differences are not statistically significant using the Bonferroni multiple testing correction method.

o Friedman's test

The **Friedman test** is a non-parametric alternative to the one-way repeated measures ANOVA test. It extends the *Wilcoxon Signed-Rank test* in the situation where there are more than two groups to compare.

Friedman test is used to assess whether there are any statistically significant differences between the distributions of three or more paired groups. It's recommended when the normality assumptions of the one-way repeated measures ANOVA test is not met or when the dependent variable is measured on an ordinal scale.

We'll use the self esteem score dataset (Friedman test.xlsx) measured over three time points.

	Subject	t1	t2	t3
1	1	4.01	5.18	7.11
2	2	2.56	6.91	6.31
3	3	3.24	4.44	9.78
4	4	3.42	4.71	8.35
5	5	2.87	3.91	6.46
6	6	2.05	5.34	6.65
7	7	3.53	5.58	6.84
8	8	3.18	4.37	7.82
9	9	3.51	4.40	8.47
10	10	3.04	4.49	8.58

Gather columns t1, t2 and t3 into long format. Convert Subject and time variables into factor (or grouping) variables:

```
1   library(tidyverse)
2   library(ggpubr)
3   library(rstatix)
4
5   data <- data%>%
6     gather(key = "time", value = "score", t1, t2, t3) %>%
7     convert_as_factor(Subject, time)
8   head(data, 3)
```

The following output is obtained:

```
# A tibble: 3 x 3
  Subject time  score
  <fct>   <fct> <dbl>
1 1       t1     4.01
2 2       t1     2.56
3 3       t1     3.24
```

Summary statistics

Compute some summary statistics of the self-esteem score by groups (time):

```
data %>%
  group_by(time) %>%
  get_summary_stats(score, type = "common")
```

The following output is obtained:
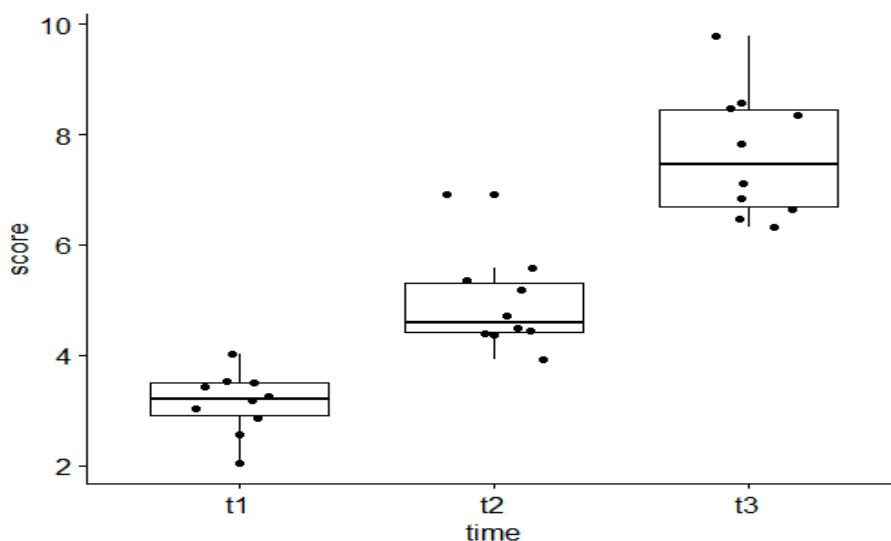
```
# A tibble: 3 x 11
  time  variable     n   min   max median   iqr  mean    sd    se    ci
  <fct> <chr>    <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 t1    score       10  2.05  4.01   3.21 0.575  3.14 0.552 0.175 0.395
2 t2    score       10  3.91  6.91   4.6  0.89   4.93 0.862 0.273 0.617
3 t3    score       10  6.31  9.78   7.46 1.74   7.64 1.14  0.361 0.817
```

Visualization

Create a box plot and add points corresponding to individual values

```
ggboxplot(data, x = "time", y = "score", add = "jitter")
```

Normality assumption

From the plot above, as the median at time 2 is not exactly in the middle the data may not be normally distributed.

Computation

We'll use the pipe-friendly friedman_test() function [rstatix package], a wrapper around the R base function friedman.test().

```
res.fried <- data %>% friedman_test(score ~ time |Subject)
res.fried
```

The following output is obtained:

```
# A tibble: 1 x 6
  .y.          n statistic    df          p method
* <chr> <int>      <dbl> <dbl>      <dbl> <chr>
1 score     10       18.2     2 0.000112 Friedman test
```

The self esteem score was statistically significantly different at the different time points during the diet, p <0.0001.

Multiple pairwise-comparisons

From the output of the Friedman test, we know that there is a significant difference between groups, but we don't know which pairs of groups are different.

A significant Friedman test can be followed up by pairwise **Wilcoxon signed-rank tests** for identifying which groups are different.

Note that, the data must be correctly ordered by the blocking variable (id) so that the first observation for time t1 will be paired with the first observation for time t2, and so on.

Pairwise comparisons using paired Wilcoxon signed-rank test. P-values are adjusted using the Bonferroni multiple testing correction method.

```
# pairwise comparisons
pwc <- data %>%
  wilcox_test(score ~ time, paired = TRUE, p.adjust.method = "bonferroni")
pwc
```

The following output is obtained:

```
# A tibble: 3 x 9
  .y.   group1 group2    n1     n2 statistic       p p.adj p.adj.signif
* <chr> <chr>  <chr>  <int> <int>      <dbl>   <dbl> <dbl> <chr>
1 score t1     t2        10    10          0   0.002 0.006 **
2 score t1     t3        10    10          0   0.002 0.006 **
3 score t2     t3        10    10          1   0.004 0.012 *
```

All the pairwise differences are statistically significant.