Mixed ANOVA in R

Mixed ANOVA is used to compare the means of groups cross-classified by two different types of factor variables, including:

- **between-subjects factors**, which have independent categories (e.g., gender: male/female)
- **within-subjects factors**, which have related categories also known as repeated measures (e.g., time: before/after treatment).

The mixed ANOVA test is also referred as *mixed design ANOVA* and *mixed measures ANOVA*.

This chapter describes different types of mixed ANOVA, including:

- **two-way mixed ANOVA**, used to compare the means of groups cross-classified by two independent categorical variables, including one between-subjects and one within-subjects factors.

- three-way mixed ANOVA

  , used to evaluate if there is a three-way interaction between three independent variables, including between-subjects and within-subjects factors. You can have two different designs for three-way mixed ANOVA:

  1. one between-subjects factor and two within-subjects factors
  2. two between-subjects factor and one within-subjects factor

You will learn how to:

- **Compute and interpret the different mixed ANOVA tests in R**.
- **Check mixed ANOVA test assumptions**
- **Perform post-hoc tests**, multiple pairwise comparisons between groups to identify which groups are different
- **Visualize the data** using box plots, add ANOVA and pairwise comparisons p-values to the plot

Contents:

Related Book

Practical Statistics in R II - Comparing Groups: Numerical Variables

Assumptions

The mixed ANOVA makes the following assumptions about the data:

- **No significant outliers** in any cell of the design. This can be checked by visualizing the data using box plot methods and by using the function `identify_outliers()` [rstatix package].
- **Normality**: the outcome (or dependent) variable should be approximately normally distributed in each cell of the design. This can be checked using the **Shapiro-Wilk normality test** (`shapiro_test()` [rstatix]) or by visual inspection using **QQ plot** (`ggqqplot()` [ggpubr package]).
- **Homogeneity of variances**: the variance of the outcome variable should be equal between the groups of the between-subjects factors. This can be assessed using the **Levene's test for equality of variances** (`levene_test()` [rstatix]).
- **Assumption of sphericity**: the variance of the differences between within-subjects groups should be equal. This can be checked using the **Mauchly's test of sphericity**, which is automatically reported when using the `anova_test()` R function.
- **Homogeneity of covariances** tested by Box's M. The covariance matrices should be equal across the cells formed by the between-subjects factors.

Before computing mixed ANOVA test, you need to perform some preliminary tests to check if the assumptions are met.

Prerequisites

Make sure that you have installed the following R packages:

- `tidyverse` for data manipulation and visualization
- `ggpubr` for creating easily publication ready plots
- `rstatix` provides pipe-friendly R functions for easy statistical analyses
- `datarium`: contains required data sets for this chapter

Start by loading the following R packages:

```
1  library(tidyverse)
2  library(ggpubr)
3  library(rstatix)
```

Key R functions:

-
  ```
  1  anova_test()
  ```

  [rstatix package], a wrapper around

  ```
  1  car::Anova()
  ```

  for making easy the computation of repeated measures ANOVA. Key arguments for performing repeated measures ANOVA:
    - `data` : data frame
    - `dv` : (numeric) the dependent (or outcome) variable name.
    - `wid` : variable name specifying the case/sample identifier.
    - `between` : between-subjects factor or grouping variable
    - `within` : within-subjects factor or grouping variable
- `get_anova_table()` [rstatix package]. Extracts the ANOVA table from the output of `anova_test()` . It returns ANOVA table that is automatically corrected for eventual deviation from the sphericity assumption. The default is to apply automatically the Greenhouse-Geisser sphericity correction to only within-subject factors violating the sphericity assumption (i.e., Mauchly's test p-value is significant, p <= 0.05). Read more in Chapter @ref(mauchly-s-test-of-sphericity-in-r).

Two-way mixed ANOVA

Data preparation

We'll use the `anxiety` dataset [in the datarium package], which contains the anxiety score, measured at three time points (t1, t2 and t3), of three groups of individuals practicing physical exercises at different levels (grp1: basal, grp2: moderate and grp3: high)

Two-way mixed ANOVA can be used to evaluate if there is interaction between group and time in explaining the anxiety score.

Load and show one random row by group:

```
1  # Wide format
2  set.seed(123)
3  data("anxiety", package = "datarium")
4  anxiety %>% sample_n_by(group, size = 1)
5  ## # A tibble: 3 x 5
6  ##   id    group    t1     t2     t3
7  ##   <fct> <fct> <dbl>  <dbl>  <dbl>
8  ## 1 5     grp1   16.5   15.8   15.7
```

```
 9  ## 2 27    grp2   17.8  17.7  16.9
10  ## 3 37    grp3   17.1  15.6  14.3
11  # Gather the columns t1, t2 and t3 into long format.
12  # Convert id and time into factor variables
13  anxiety <- anxiety %>%
14    gather(key = "time", value = "score", t1, t2, t3) %>%
15    convert_as_factor(id, time)
16  # Inspect some random rows of the data by groups
17  set.seed(123)
18  anxiety %>% sample_n_by(group, time, size = 1)
19  ## # A tibble: 9 x 4
20  ##   id     group time   score
21  ##   <fct> <fct> <fct> <dbl>
22  ## 1 5      grp1  t1     16.5
23  ## 2 12     grp1  t2     17.7
24  ## 3 7      grp1  t3     16.5
25  ## 4 29     grp2  t1     18.4
26  ## 5 30     grp2  t2     18.9
27  ## 6 16     grp2  t3     12.7
28  ## # … with 3 more rows
```

Summary statistics

Group the data by `time` and `group`, and then compute some summary statistics of the `score` variable: mean and sd (standard deviation)

```
 1  anxiety %>%
 2    group_by(time, group) %>%
 3    get_summary_stats(score, type = "mean_sd")
 4  ## # A tibble: 9 x 6
 5  ##   group time   variable     n  mean    sd
 6  ##   <fct> <fct> <chr>     <dbl> <dbl> <dbl>
 7  ## 1 grp1  t1    score        15  17.1  1.63
 8  ## 2 grp2  t1    score        15  16.6  1.57
 9  ## 3 grp3  t1    score        15  17.0  1.32
10  ## 4 grp1  t2    score        15  16.9  1.70
11  ## 5 grp2  t2    score        15  16.5  1.70
12  ## 6 grp3  t2    score        15  15.0  1.39
13  ## # … with 3 more rows
```
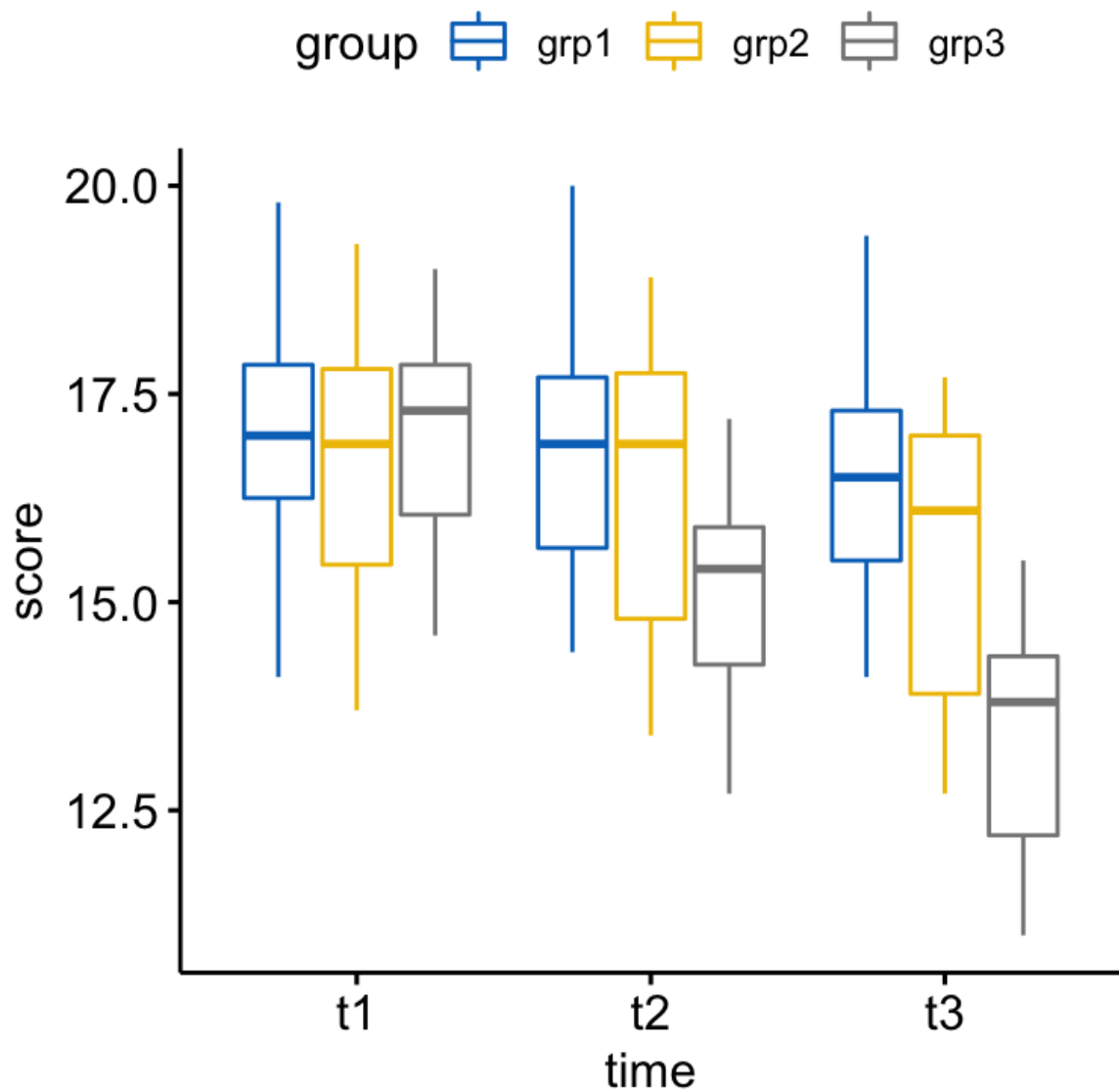
Visualization

Create a box plots:

```
 1  bxp <- ggboxplot(
 2    anxiety, x = "time", y = "score",
 3    color = "group", palette = "jco"
 4    )
 5  bxp
```

Check assumptions

Outliers

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

```
1  anxiety %>%
2    group_by(time, group) %>%
3    identify_outliers(score)
4  ## [1] group     time      id        score     is.outlier is.extreme
5  ## <0 rows> (or 0-length row.names)
```

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.

Yo 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.

It's also possible to keep the outliers in the data and perform robust ANOVA test using the WRS2 package.

Normality assumption

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

```
 1   anxiety %>%
 2     group_by(time, group) %>%
 3     shapiro_test(score)
 4   ## # A tibble: 9 x 5
 5   ##   group time  variable statistic     p
 6   ##   <fct> <fct> <chr>        <dbl> <dbl>
 7   ## 1 grp1  t1    score        0.964 0.769
 8   ## 2 grp2  t1    score        0.977 0.949
 9   ## 3 grp3  t1    score        0.954 0.588
10   ## 4 grp1  t2    score        0.956 0.624
11   ## 5 grp2  t2    score        0.935 0.328
12   ## 6 grp3  t2    score        0.952 0.558
13   ## # … with 3 more rows
```

The score were normally distributed (p > 0.05) for each cell, as assessed by Shapiro-Wilk's test of normality.

Note that, if your sample size is greater than 50, the normal QQ plot is preferred because at larger sample sizes the Shapiro-Wilk test becomes very sensitive even to a minor deviation from normality.

QQ plot draws the correlation between a given data and the normal distribution.

```
 1   ggqqplot(anxiety, "score", ggtheme = theme_bw()) +
 2     facet_grid(time ~ group)
```
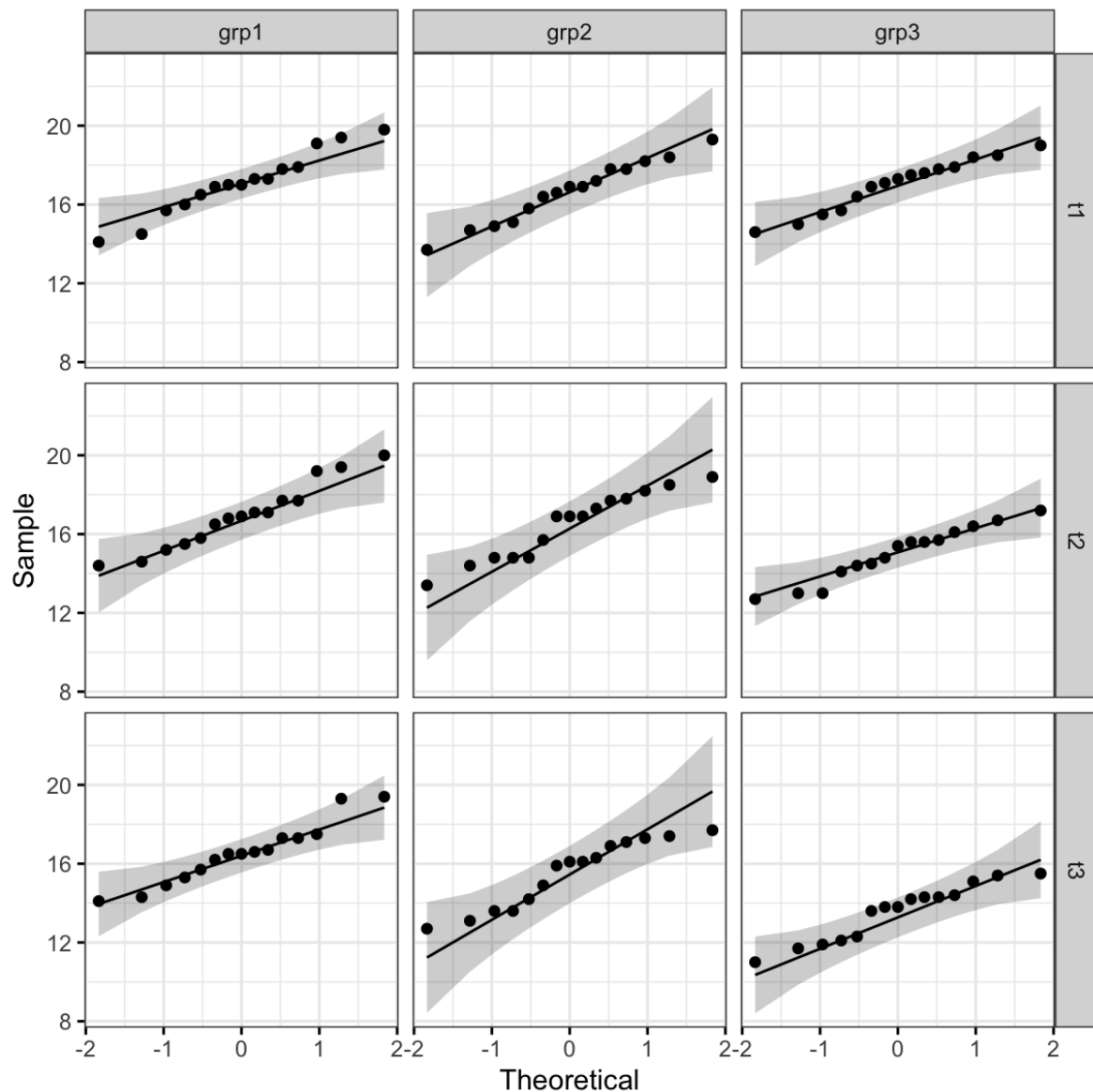
All the points fall approximately along the reference line, for each cell. So we can assume normality of the data.

In the situation where the assumptions are not met, you could consider running the two-way repeated measures ANOVA on the transformed or performing a robust ANOVA test using the WRS2 R package.

Homogneity of variance assumption

The homogeneity of variance assumption of the between-subject factor (group) can be checked using the Levene's test. The test is performed at each level of time variable:

```
anxiety %>%
  group_by(time) %>%
  levene_test(score ~ group)
## # A tibble: 3 x 5
##   time    df1   df2 statistic     p
##   <fct> <int> <int>     <dbl> <dbl>
## 1 t1        2    42     0.176 0.839
## 2 t2        2    42     0.249 0.781
## 3 t3        2    42     0.335 0.717
```

There was homogeneity of variances, as assessed by Levene's test (p > 0.05).

Note that, if you do not have homogeneity of variances, you can try to transform the outcome (dependent) variable to correct for the unequal variances.

It's also possible to perform robust ANOVA test using the WRS2 R package.

Homogeneity of covariances assumption

The homogeneity of covariances of the between-subject factor (`group`) can be evaluated using the **Box's M-test** implemented in the `rstatix` package. If this test is statistically significant (i.e., p < 0.001), you do not have equal covariances, but if the test is not statistically significant (i.e., p > 0.001), you have equal covariances and you have not violated the assumption of homogeneity of covariances.

Note that, the Box's M is highly sensitive, so unless p < 0.001 and your sample sizes are unequal, ignore it. However, if significant and you have unequal sample sizes, the test is not robust (https://en.wikiversity.org/wiki/Box%27s_M, Tabachnick & Fidell, 2001).

Compute Box's M-test:

```
1  box_m(anxiety[, "score", drop = FALSE], anxiety$group)
2  ## # A tibble: 1 x 4
3  ##    statistic p.value parameter method
4  ##        <dbl>   <dbl>     <dbl> <chr>
5  ## 1      1.93   0.381         2 Box's M-test for Homogeneity of Covariance
   Matrices
```

There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices (p > 0.001).

Note that, if you do not have homogeneity of covariances, you could consider separating your analyses into distinct repeated measures ANOVAs for each group. Alternatively, you could omit the interpretation of the interaction term.

Unfortunately, it is difficult to remedy a failure of this assumption. Often, a mixed ANOVA is run anyway and the violation noted.

Assumption of sphericity

As mentioned in previous sections, 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

```
1   # Two-way mixed ANOVA test
2   res.aov <- anova_test(
3     data = anxiety, dv = score, wid = id,
4     between = group, within = time
5     )
6   get_anova_table(res.aov)
7   ## ANOVA Table (type II tests)
8   ##
9   ##        Effect DFn DFd      F        p p<.05   ges
10  ## 1      group   2  42   4.35 1.90e-02     * 0.168
11  ## 2       time   2  84 394.91 1.91e-43     * 0.179
12  ## 3 group:time   4  84 110.19 1.38e-32     * 0.108
```

From the output above, it can be seen that, there is a statistically significant two-way interactions between group and time on anxiety score, F(4, 84) = 110.18, p < 0.0001.

Post-hoc tests

A **significant two-way interaction** indicates that the impact that one factor has on the outcome variable depends on the level of the other factor (and vice versa). So, you can decompose a significant two-way interaction into:

- **Simple main effect**: run one-way model of the first variable (factor A) at each level of the second variable (factor B),
- **Simple pairwise comparisons**: if the simple main effect is significant, run multiple pairwise comparisons to determine which groups are different.

For a **non-significant two-way interaction**, you need to determine whether you have any statistically significant **main effects** from the ANOVA output.

Procedure for a significant two-way interaction

Simple main effect of group variable. In our example, we'll investigate the effect of the between-subject factor `group` on anxiety score at every `time` point.

```
1   # Effect of group at each time point
2   one.way <- anxiety %>%
3     group_by(time) %>%
4     anova_test(dv = score, wid = id, between = group) %>%
5     get_anova_table() %>%
6     adjust_pvalue(method = "bonferroni")
7   one.way
8   ## # A tibble: 3 x 9
9   ##   time  Effect   DFn   DFd     F         p `p<.05`   ges     p.adj
10  ##   <fct> <chr>  <dbl> <dbl> <dbl>     <dbl> <chr>   <dbl>     <dbl>
11  ## 1 t1    group      2    42 0.365 0.696       ""    0.017 1
12  ## 2 t2    group      2    42  5.84 0.006        *    0.218 0.018
13  ## 3 t3    group      2    42  13.8 0.0000248 *       0.396 0.0000744
14  # Pairwise comparisons between group levels
15  pwc <- anxiety %>%
16    group_by(time) %>%
17    pairwise_t_test(score ~ group, p.adjust.method = "bonferroni")
18  pwc
19  ## # A tibble: 9 x 10
```

```
20  ##   time  .y.    group1 group2   n1    n2       p p.signif    p.adj
    p.adj.signif
21  ## * <fct> <chr> <chr>  <chr>  <int> <int>   <dbl> <chr>       <dbl> <chr>

22  ## 1 t1    score grp1   grp2     15    15 0.43     ns         1       ns

23  ## 2 t1    score grp1   grp3     15    15 0.895    ns         1       ns

24  ## 3 t1    score grp2   grp3     15    15 0.51     ns         1       ns

25  ## 4 t2    score grp1   grp2     15    15 0.435    ns         1       ns

26  ## 5 t2    score grp1   grp3     15    15 0.00212  **         0.00636 **

27  ## 6 t2    score grp2   grp3     15    15 0.0169   *          0.0507  ns

28  ## # … with 3 more rows
```

Considering the Bonferroni adjusted p-value (p.adj), it can be seen that the simple main effect of group was significant at t2 (p = 0.018) and t3 (p < 0.0001) but not at t1 (p = 1).

Pairwise comparisons show that the mean anxiety score was significantly different in grp1 vs grp3 comparison at t2 (p = 0.0063); in grp1 vs grp3 (p < 0.0001) and in grp2 vs grp3 (p = 0.0013) at t3.

Simple main effects of time variable. It's also possible to perform the same analyze for the within-subject `time` variable at each level of `group` as shown in the following R code. You don't necessarily need to do this analysis.

```
1   # Effect of time at each level of exercises group
2   one.way2 <- anxiety %>%
3     group_by(group) %>%
4     anova_test(dv = score, wid = id, within = time) %>%
5     get_anova_table() %>%
6     adjust_pvalue(method = "bonferroni")
7   one.way2
8   ## # A tibble: 3 x 9
9   ##   group Effect   DFn   DFd    F        p `p<.05`   ges    p.adj
10  ##   <fct> <chr>  <dbl> <dbl> <dbl>    <dbl> <chr>    <dbl>    <dbl>
11  ## 1 grp1  time       2    28  14.8 4.05e- 5 *        0.024 1.21e- 4
12  ## 2 grp2  time       2    28  77.5 3.88e-12 *        0.086 1.16e-11
13  ## 3 grp3  time       2    28 490.  1.64e-22 *        0.531 4.92e-22
14  # Pairwise comparisons between time points at each group levels
15  # Paired t-test is used because we have repeated measures by time
16  pwc2 <- anxiety %>%
17    group_by(group) %>%
18    pairwise_t_test(
19      score ~ time, paired = TRUE,
20      p.adjust.method = "bonferroni"
21      ) %>%
22    select(-df, -statistic, -p) # Remove details
23  pwc2
24  ## # A tibble: 9 x 8
25  ##   group .y.    group1 group2   n1    n2       p.adj p.adj.signif
26  ## * <fct> <chr> <chr>  <chr>  <int> <int>       <dbl> <chr>
27  ## 1 grp1  score t1     t2       15    15 0.194       ns
```

```
28   ## 2 grp1   score t1     t3          15     15 0.002          **
29   ## 3 grp1   score t2     t3          15     15 0.006          **
30   ## 4 grp2   score t1     t2          15     15 0.268          ns
31   ## 5 grp2   score t1     t3          15     15 0.000000151   ****
32   ## 6 grp2   score t2     t3          15     15 0.0000000612  ****
33   ## # … with 3 more rows
```

There was a statistically significant effect of time on anxiety score for each of the three groups. Using pairwise paired t-test comparisons, it can be seen that for grp1 and grp2, the mean anxiety score was not statistically significantly different between t1 and t2 time points.

The pairwise comparisons t1 vs t3 and t2 vs t3 were statistically significantly different for all groups.

Procedure for non-significant two-way interaction

If the interaction is not significant, you need to interpret the main effects for each of the two variables: `group` and `time. A significant main effect can be followed up with pairwise comparisons.

In our example, there was a statistically significant main effects of group (F(2, 42) = 4.35, p = 0.02) and time (F(2, 84) = 394.91, p < 0.0001) on the anxiety score.

Perform multiple pairwise paired t-tests for the `time` variable, ignoring group. P-values are adjusted using the Bonferroni multiple testing correction method.

```
1   anxiety %>%
2     pairwise_t_test(
3       score ~ time, paired = TRUE,
4       p.adjust.method = "bonferroni"
5     )
```

All pairwise comparisons are significant.

You can perform a similar analysis for the `group` variable.

```
1   anxiety %>%
2     pairwise_t_test(
3       score ~ group,
4       p.adjust.method = "bonferroni"
5     )
```

All pairwise comparisons are significant except grp1 vs grp2.

Report

There was a statistically significant interaction between exercises group and time in explaining the anxiety score, F(4, 84) = 110.19, p < 0.0001.

Considering the Bonferroni adjusted p-value, the simple main effect of exercises group was significant at t2 (p = 0.018) and t3 (p < 0.0001) but not at t1 (p = 1).

Pairwise comparisons show that the mean anxiety score was significantly different in grp1 vs grp3 comparison at t2 (p = 0.0063); in grp1 vs grp3 (p < 0.0001) and in grp2 vs grp3 (p = 0.0013) at t3.

Note that, for the plot below, we only need the pairwise comparison results for t2 and t3 but not for t1 (because the simple main effect of exercises group was not significant at this time point). We'll filter the comparison results accordingly.

```
# Visualization: boxplots with p-values
pwc <- pwc %>% add_xy_position(x = "time")
pwc.filtered <- pwc %>% filter(time != "t1")
bxp +
  stat_pvalue_manual(pwc.filtered, tip.length = 0, hide.ns = TRUE) +
  labs(
    subtitle = get_test_label(res.aov, detailed = TRUE),
    caption = get_pwc_label(pwc)
  )
```

Anova, $F(4,84) = 110.19$, p = <0.0001, $\eta^2_g = 0.11$

pwc: **T test**; p.adjust: **Bonferroni**
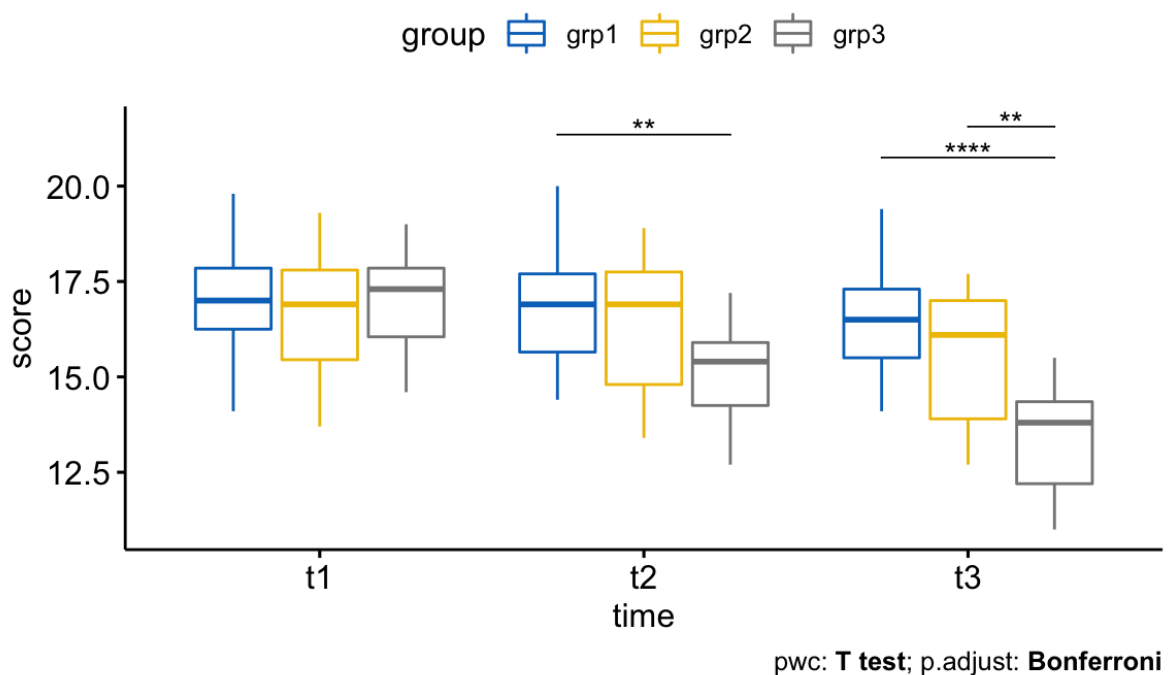
Three-way mixed ANOVA: 2 between- and 1 within-subjects factors

This section describes how to compute the three-way mixed ANOVA, in R, for a situation where you have **two between-subjects factors and one within-subjects factor**.

This setting is for investigating group differences over time (i.e., the within-subjects factor) where groups are formed by the combination of two between-subjects factors. For example, you might want to understand how performance score changes over time (e.g., 0, 4 and 8 months) depending on gender (i.e., male/female) and stress (low, moderate and high stress).

Data preparation

We'll use `performance` dataset [datarium package] containing the performance score measures of participants at two time points. The aim of this study is to evaluate the effect of gender and stress on performance score.

The data contains the following variables:

1. Performance score (outcome or dependent variable) measured at two time points, `t1` and `t2`.
2. Two between-subjects factors: `gender` (levels: male and female) and `stress` (low, moderate, high)

1. One within-subjects factor, `time`, which has two time points: `t1` and `t2`.

Load and inspect the data by showing one random row by group:

```
1   # Load and inspect the data
2   # Wide format
3   set.seed(123)
4   data("performance", package = "datarium")
5   performance %>% sample_n_by(gender, stress, size = 1)
6   ## # A tibble: 6 x 5
7   ##       id gender stress       t1    t2
8   ##    <int> <fct>  <fct>     <dbl> <dbl>
9   ## 1      3 male   low        5.63  5.47
10  ## 2     18 male   moderate   5.57  5.78
11  ## 3     25 male   high       5.48  5.74
12  ## 4     39 female low        5.50  5.66
13  ## 5     50 female moderate   5.96  5.32
14  ## 6     51 female high       5.59  5.06
15  # Gather the columns t1, t2 and t3 into long format.
16  # Convert id and time into factor variables
17  performance <- performance %>%
18    gather(key = "time", value = "score", t1, t2) %>%
19    convert_as_factor(id, time)
20  # Inspect some random rows of the data by groups
21  set.seed(123)
22  performance %>% sample_n_by(gender, stress, time, size = 1)
23  ## # A tibble: 12 x 5
24  ##    id    gender stress   time  score
25  ##    <fct> <fct>  <fct>    <fct> <dbl>
26  ## 1 3      male   low      t1     5.63
27  ## 2 8      male   low      t2     5.92
28  ## 3 15     male   moderate t1     5.96
29  ## 4 19     male   moderate t2     5.76
30  ## 5 30     male   high     t1     5.38
31  ## 6 21     male   high     t2     5.64
32  ## # … with 6 more rows
```

Summary statistics

Group the data by `gender`, `stress` and `time`, and then compute some summary statistics of the `score` variable: mean and sd (standard deviation)

```
1   performance %>%
2     group_by(gender, stress, time ) %>%
3     get_summary_stats(score, type = "mean_sd")
4   ## # A tibble: 12 x 7
5   ##   gender stress   time  variable     n  mean     sd
6   ##   <fct>  <fct>    <fct> <chr>     <dbl> <dbl>  <dbl>
7   ## 1 male   low      t1    score        10  5.72  0.19
8   ## 2 male   low      t2    score        10  5.70  0.143
9   ## 3 male   moderate t1    score        10  5.72  0.193
10  ## 4 male   moderate t2    score        10  5.77  0.155
11  ## 5 male   high     t1    score        10  5.48  0.121
12  ## 6 male   high     t2    score        10  5.64  0.195
13  ## # … with 6 more rows
```
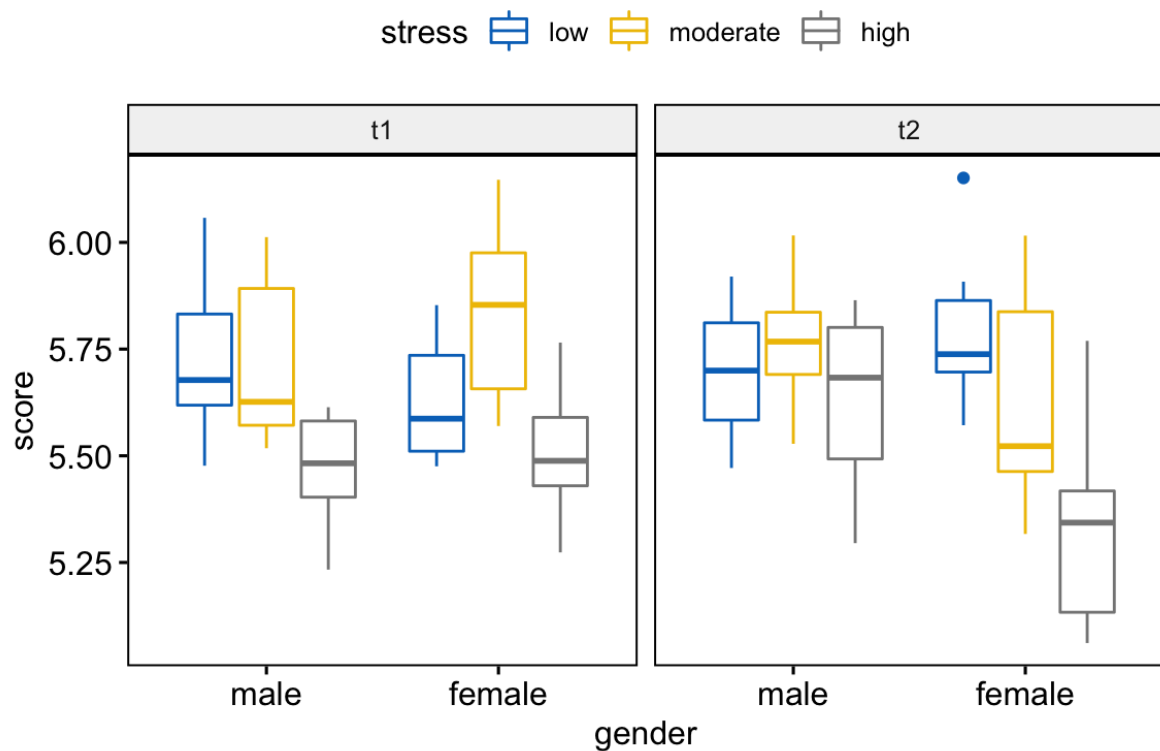
Visualization

Create box plots of performance `score` by `gender` colored by `stress` levels and faceted by `time` :

```
1   bxp <- ggboxplot(
2     performance, x = "gender", y = "score",
3     color = "stress", palette = "jco",
4     facet.by =  "time"
5     )
6   bxp
```

Check assumptions

Outliers

```
1  performance %>%
2    group_by(gender, stress, time) %>%
3    identify_outliers(score)
4  ## # A tibble: 1 x 7
5  ##   gender stress time  id    score is.outlier is.extreme
6  ##   <fct>  <fct>  <fct> <fct> <dbl> <lgl>      <lgl>
7  ## 1 female low    t2    36     6.15 TRUE       FALSE
```

There were no extreme outliers.
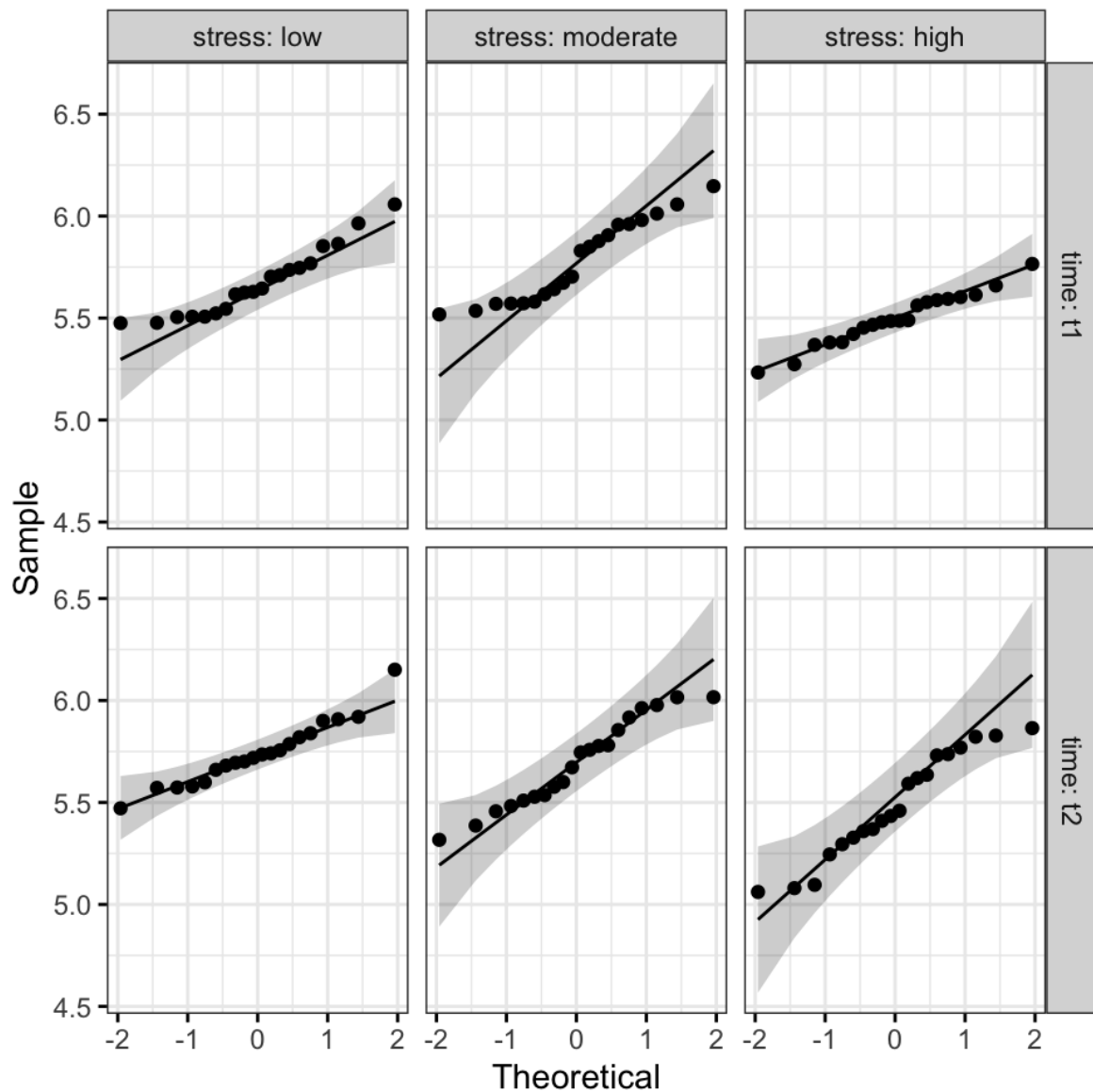
Normality assumption

Compute Shapiro-Wilk test for each combinations of factor levels:

```
1   performance %>%
2     group_by(gender, stress, time ) %>%
3     shapiro_test(score)
4   ## # A tibble: 12 x 6
5   ##   gender stress   time  variable statistic      p
6   ##   <fct>  <fct>    <fct> <chr>        <dbl>  <dbl>
7   ## 1 male   low      t1    score        0.942 0.574
8   ## 2 male   low      t2    score        0.966 0.849
9   ## 3 male   moderate t1    score        0.848 0.0547
10  ## 4 male   moderate t2    score        0.958 0.761
11  ## 5 male   high     t1    score        0.915 0.319
12  ## 6 male   high     t2    score        0.925 0.403
13  ## # … with 6 more rows
```

The score were normally distributed (p > 0.05) for each cell, as assessed by Shapiro-Wilk's test of normality.

Create QQ plot for each cell of design:

```
1   ggqqplot(performance, "score", ggtheme = theme_bw()) +
2     facet_grid(time ~ stress, labeller = "label_both")
```

All the points fall approximately along the reference line, for each cell. So we can assume normality of the data.

Homogneity of variance assumption

Compute the Levene's test at each level of the within-subjects factor, here `time` variable:

```
1  performance %>%
2    group_by(time) %>%
3    levene_test(score ~ gender*stress)
4  ## # A tibble: 2 x 5
5  ##   time    df1   df2 statistic     p
6  ##   <fct> <int> <int>     <dbl> <dbl>
7  ## 1 t1        5    54     0.974 0.442
8  ## 2 t2        5    54     0.722 0.610
```

There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance (p > .05).

Assumption of sphericity

As mentioned in the two-way mixed ANOVA section, the Mauchly's test of sphericity and the sphericity corrections are internally done using the R function `anova_test()` and `get_anova_table()` [rstatix package].

Computation

```
1  res.aov <- anova_test(
2    data = performance, dv = score, wid = id,
3    within = time, between = c(gender, stress)
4    )
5  get_anova_table(res.aov)
6  ## ANOVA Table (type II tests)
7  ##
8  ##                 Effect DFn DFd      F       p p<.05      ges
9  ## 1              gender   1  54  2.406 1.27e-01       0.023000
10 ## 2              stress   2  54 21.166 1.63e-07     * 0.288000
11 ## 3                time   1  54  0.063 8.03e-01       0.000564
12 ## 4       gender:stress   2  54  1.554 2.21e-01       0.029000
13 ## 5         gender:time   1  54  4.730 3.40e-02     * 0.041000
14 ## 6         stress:time   2  54  1.821 1.72e-01       0.032000
15 ## 7 gender:stress:time   2  54  6.101 4.00e-03     * 0.098000
```

There was a statistically significant three-way interaction between time, gender, and stress $F(2, 54)$ = 6.10, $p = 0.004$.

Post-hoc tests

If there is a significant three-way interaction effect, you can decompose it into:

- **Simple two-way interaction**: run two-way interaction at each level of third variable,
- **Simple simple main effect**: run one-way model at each level of second variable, and
- **simple simple pairwise comparisons**: run pairwise or other post-hoc comparisons if necessary.

If you do not have a statistically significant three-way interaction, you need to determine whether you have any statistically significant two-way interaction from the ANOVA output. A significant two-way interaction can be followed up by a simple main effect analysis, which can be followed up by simple pairwise comparisons if significant.

In this section we'll describe the procedure for a significant three-way interaction.

Compute simple two-way interaction

You are free to decide which two variables will form the simple two-way interactions and which variable will act as the third variable.

In the following R code, we have considered the simple two-way interaction of `gender*stress` at each level of `time`.

Group the data by `time` (the within-subject factors) and analyze the simple two-way interaction between `gender` and `stress`, which are the between-subjects factors.

```
1   # two-way interaction at each time levels
2   two.way <- performance %>%
3     group_by(time) %>%
4     anova_test(dv = score, wid = id, between = c(gender, stress))
5   two.way
6   ## # A tibble: 6 x 8
7   ##    time  Effect          DFn   DFd     F          p `p<.05`   ges
8   ##    <fct> <chr>         <dbl> <dbl> <dbl>      <dbl> <chr>    <dbl>
9   ## 1 t1    gender            1    54  0.186 0.668        ""       0.003
10  ## 2 t1    stress            2    54 14.9   0.00000723 *        0.355
11  ## 3 t1    gender:stress     2    54  2.12  0.131        ""       0.073
12  ## 4 t2    gender            1    54  5.97  0.018        *        0.1
13  ## 5 t2    stress            2    54  9.60  0.000271     *        0.262
14  ## 6 t2    gender:stress     2    54  4.95  0.011        *        0.155
```

There was a statistically significant simple two-way interaction of gender and stress at t2, F(2, 54) = 4.95, p = 0.011, but not at t1, F(2, 54) = 2.12, p = 0.13.

Note that, statistical significance of a simple two-way interaction was accepted at a Bonferroni-adjusted alpha level of 0.025. This corresponds to the current level you declare statistical significance at (i.e., p < 0.05) divided by the number of simple two-way interaction you are computing (i.e., 2).

Compute simple simple main effects

A statistically significant simple two-way interaction can be followed up with **simple simple main effects**.

In our example, you could therefore investigate the effect of `stress` on the performance score at every level of `gender` or investigate the effect of `gender` at every level of `stress`.

Note that, you will only need to do this for the simple two-way interaction for "t2" as this was the only simple two-way interaction that was statistically significant.

Group the data by `time` and `gender`, and analyze the simple main effect of `stress` on performance score:

```
1   stress.effect <- performance %>%
2     group_by(time, gender) %>%
3     anova_test(dv = score, wid = id, between = stress)
4   stress.effect %>% filter(time == "t2")
5   ## # A tibble: 2 x 9
6   ##    gender time  Effect   DFn   DFd     F        p `p<.05`   ges
7   ##    <fct>  <fct> <chr>  <dbl> <dbl> <dbl>    <dbl> <chr>    <dbl>
8   ## 1 male   t2    stress     2    27  1.57 0.227      ""       0.104
9   ## 2 female t2    stress     2    27 10.5  0.000416 *        0.438
```

In the table above, we only need the results for `time = t2`. Statistical significance of a simple simple main effect was accepted at a Bonferroni-adjusted alpha level of 0.025, that is 0.05 divided y the number of simple simple main effects you are computing (i.e., 2).

There was a statistically significant simple simple main effect of stress on the performance score for female at t2 time point, F(2, 27) = 10.5, p = 0.0004, but not for males, F(2, 27) = 1.57, p = 0.23.

Compute simple simple comparisons

A statistically significant simple simple main effect can be followed up by **multiple pairwise comparisons** to determine which group means are different.

Note that, you will only need to concentrate on the pairwise comparison results for female, because the effect of stress was significant for female only in the previous section.

Group the data by `time` and `gender`, and perform pairwise comparisons between `stress` levels with Bonferroni adjustment:

```
# Fit pairwise comparisons
pwc <- performance %>%
  group_by(time, gender) %>%
  pairwise_t_test(score ~ stress, p.adjust.method = "bonferroni") %>%
  select(-p, -p.signif) # Remove details
# Focus on the results of "female" at t2
pwc %>% filter(time == "t2", gender == "female")
## # A tibble: 3 x 9
##   gender time  .y.   group1   group2      n1    n2    p.adj p.adj.signif
##   <fct>  <fct> <chr> <chr>    <chr>    <int> <int>    <dbl> <chr>
## 1 female t2    score low      moderate    10    10 0.323    ns
## 2 female t2    score low      high        10    10 0.000318 ***
## 3 female t2    score moderate high        10    10 0.0235   *
```

For female, the mean performance score was statistically significantly different between low and high stress levels (p < 0.001) and between moderate and high stress levels (p = 0.023).

There was no significant difference between low and moderate stress groups (p = 0.32)

Report

A three-way mixed ANOVA was performed to evaluate the effects of gender, stress and time on performance score.

There were no extreme outliers, as assessed by box plot method. The data was normally distributed, as assessed by Shapiro-Wilk's test of normality (p > 0.05). There was homogeneity of variances (p > 0.05) as assessed by Levene's test of homogeneity of variances.

There was a statistically significant three-way interaction between gender, stress and time, F(2, 54) = 6.10, p = 0.004.

For the simple two-way interactions and simple simple main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the p < 0.025 level.

There was a statistically significant simple two-way interaction between gender and stress at time point t2, F(2, 54) = 4.95, p = 0.011, but not at t1, F(2, 54) = 2.12, p = 0.13.

There was a statistically significant simple simple main effect of stress on the performance score for female at t2 time point, F(2, 27) = 10.5, p = 0.0004, but not for males, F(2, 27) = 1.57, p = 0.23.

All simple simple pairwise comparisons were run between the different stress groups for female at time point t2. A Bonferroni adjustment was applied.

The mean performance score was statistically significantly different between low and high stress levels (p < 0.001) and between moderate and high stress levels (p = 0.024). There was no significant difference between low and moderate stress groups (p = 0.32).

```
1  # Visualization: box plots with p-values
2  pwc <- pwc %>% add_xy_position(x = "gender")
3  pwc.filtered <- pwc %>% filter(time == "t2", gender == "female")
4  bxp +
5    stat_pvalue_manual(pwc.filtered, tip.length = 0, hide.ns = TRUE) +
6    labs(
7      subtitle = get_test_label(res.aov, detailed = TRUE),
8      caption = get_pwc_label(pwc)
9    )
```

Anova, $F(2,54) = 6.1$, $p = 0.004$, $\eta_g^2 = 0.1$

pwc: **T test**; p.adjust: **Bonferroni**
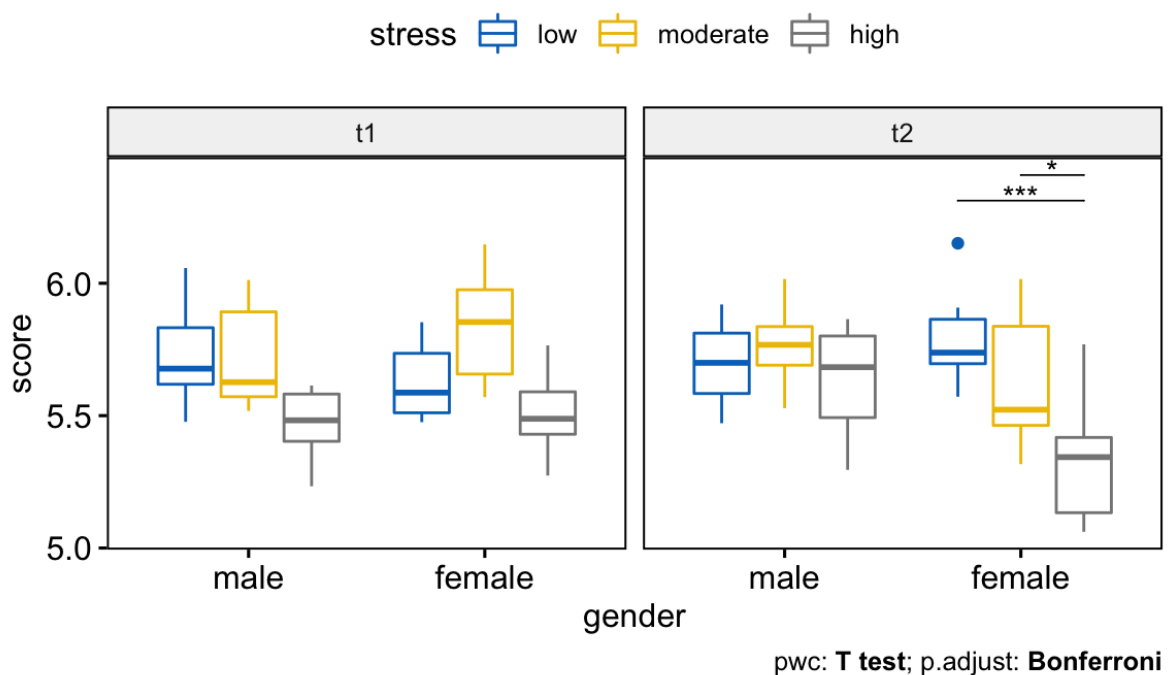
Three-way Mixed ANOVA: 1 between- and 2 within-subjects factors

This section describes how to compute the three-way mixed ANOVA, in R, for a situation where you have **one between-subjects factor and two within-subjects factors**. For example, you might want to understand how weight loss score differs in individuals doing exercises vs not doing exercises over three `time` points (t1, t2, t3) depending on participant diets (`diet:no` and `diet:yes`).

Data preparation

We'll use the `weightloss` dataset available in the datarium package. This dataset was originally created for three-way repeated measures ANOVA. However, for our example in this article, we'll modify slightly the data so that it corresponds to a three-way mixed design.

A researcher wanted to assess the effect of `time` on weight loss score depending on `exercises` programs and `diet`.

The weight loss score was measured in two different groups: a group of participants doing exercises (`exercises:yes`) and in another group not doing exercises (`excises:no`).

Each participant was also enrolled in two trials: (1) no diet and (2) diet. The order of the trials was counterbalanced and sufficient time was allowed between trials to allow any effects of previous trials to have dissipated.

Each trial lasted 9 weeks and the weight loss score was measured at the beginning of each trial (t1), at the midpoint of each trial (t2) and at the end of each trial (t3).

In this study design, 24 individuals were recruited. Of these 24 participants, 12 belongs to the `exercises:no` group and 12 were in `exercises:yes` group. The 24 participants were enrolled in two successive trials (`diet:no` and `diet:yes`) and the weight loss `score` was repeatedly measured at three time points.

In this setting, we have:

- one dependent (or outcome) variable: `score`
- One between-subjects factor: `exercises`
- two within-subjects factors: `diet` end `time`

Three-way mixed ANOVA can be performed in order to determine whether there is a significant interaction between diet, exercises and time on the weight loss score.

Load the data and inspect some random rows by group:

```
1   # Load the original data
2   # Wide format
3   data("weightloss", package = "datarium")
4   # Modify it to have three-way mixed design
5   weightloss <- weightloss %>%
6     mutate(id = rep(1:24, 2)) # two trials
7   # Show one random row by group
8   set.seed(123)
9   weightloss %>% sample_n_by(diet, exercises, size = 1)
10  ## # A tibble: 4 x 6
11  ##      id diet  exercises    t1     t2     t3
12  ##   <int> <fct> <fct>     <dbl>  <dbl>  <dbl>
13  ## 1     4 no    no         11.1    9.5   11.1
14  ## 2    22 no    yes        10.2   11.8   17.4
15  ## 3     5 yes   no         11.6   13.4   13.9
16  ## 4    23 yes   yes        12.7   12.7   15.1
17  # Gather the columns t1, t2 and t3 into long format.
18  # Convert id and time into factor variables
19  weightloss <- weightloss %>%
20    gather(key = "time", value = "score", t1, t2, t3) %>%
21    convert_as_factor(id, time)
22  # Inspect some random rows of the data by groups
23  set.seed(123)
24  weightloss %>% sample_n_by(diet, exercises, time, size = 1)
25  ## # A tibble: 12 x 5
26  ##   id    diet  exercises time  score
27  ##   <fct> <fct> <fct>     <fct> <dbl>
28  ## 1 4     no    no        t1     11.1
29  ## 2 10    no    no        t2     10.7
30  ## 3 5     no    no        t3     12.3
```

```
31  ## 4 23    no    yes       t1    10.2
32  ## 5 24    no    yes       t2    13.2
33  ## 6 13    no    yes       t3    15.8
34  ## # … with 6 more rows
```

Summary statistics

Group the data by `exercises`, `diet` and `time`, and then compute some summary statistics of the `score` variable: mean and sd (standard deviation)

```
1   weightloss %>%
2     group_by(exercises, diet, time) %>%
3     get_summary_stats(score, type = "mean_sd")
4   ## # A tibble: 12 x 7
5   ##   diet  exercises time  variable     n  mean    sd
6   ##   <fct> <fct>     <fct> <chr>    <dbl> <dbl> <dbl>
7   ## 1 no    no        t1    score       12  10.9 0.868
8   ## 2 no    no        t2    score       12  11.6 1.30
9   ## 3 no    no        t3    score       12  11.4 0.935
10  ## 4 yes   no        t1    score       12  11.7 0.938
11  ## 5 yes   no        t2    score       12  12.4 1.42
12  ## 6 yes   no        t3    score       12  13.8 1.43
13  ## # … with 6 more rows
```
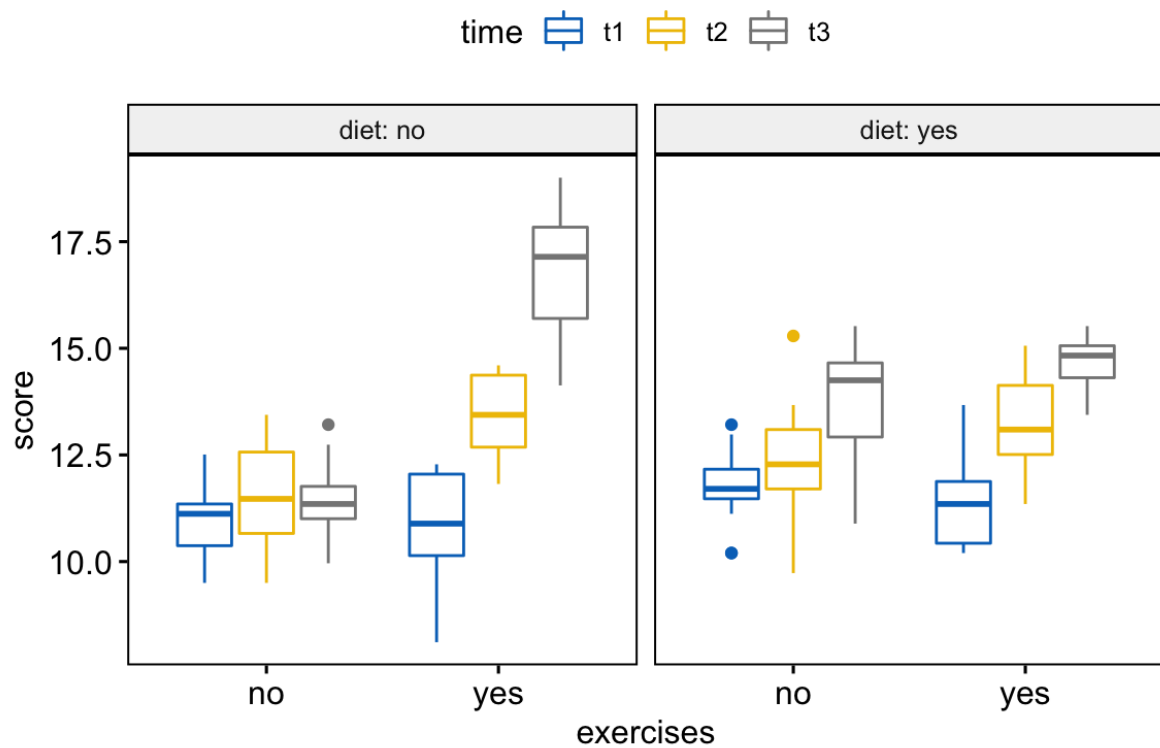
Visualization

Create box plots of weightloss `score` by `exercises` groups, colored by `time` points and faceted by `diet` trials:

```
1   bxp <- ggboxplot(
2     weightloss, x = "exercises", y = "score",
3     color = "time", palette = "jco",
4     facet.by = "diet", short.panel.labs = FALSE
5     )
6   bxp
```

Check assumptions

Outliers

```
 1  weightloss %>%
 2    group_by(diet, exercises, time) %>%
 3    identify_outliers(score)
 4  ## # A tibble: 5 x 7
 5  ##   diet  exercises time  id    score is.outlier is.extreme
 6  ##   <fct> <fct>     <fct> <fct> <dbl> <lgl>      <lgl>
 7  ## 1 no    no        t3    2      13.2 TRUE       FALSE
 8  ## 2 yes   no        t1    1      10.2 TRUE       FALSE
 9  ## 3 yes   no        t1    3      13.2 TRUE       FALSE
10  ## 4 yes   no        t1    4      10.2 TRUE       FALSE
11  ## 5 yes   no        t2    10     15.3 TRUE       FALSE
```

There were no extreme outliers.

Normality assumption

Compute Shapiro-Wilk test for each combinations of factor levels:

```
1   weightloss %>%
2     group_by(diet, exercises, time) %>%
3     shapiro_test(score)
4   ## # A tibble: 12 x 6
5   ##   diet  exercises time  variable statistic     p
6   ##   <fct> <fct>     <fct> <chr>        <dbl> <dbl>
7   ## 1 no    no        t1    score        0.917 0.264
8   ## 2 no    no        t2    score        0.957 0.743
9   ## 3 no    no        t3    score        0.965 0.851
10  ## 4 no    yes       t1    score        0.922 0.306
11  ## 5 no    yes       t2    score        0.912 0.229
12  ## 6 no    yes       t3    score        0.953 0.674
13  ## # … with 6 more rows
```
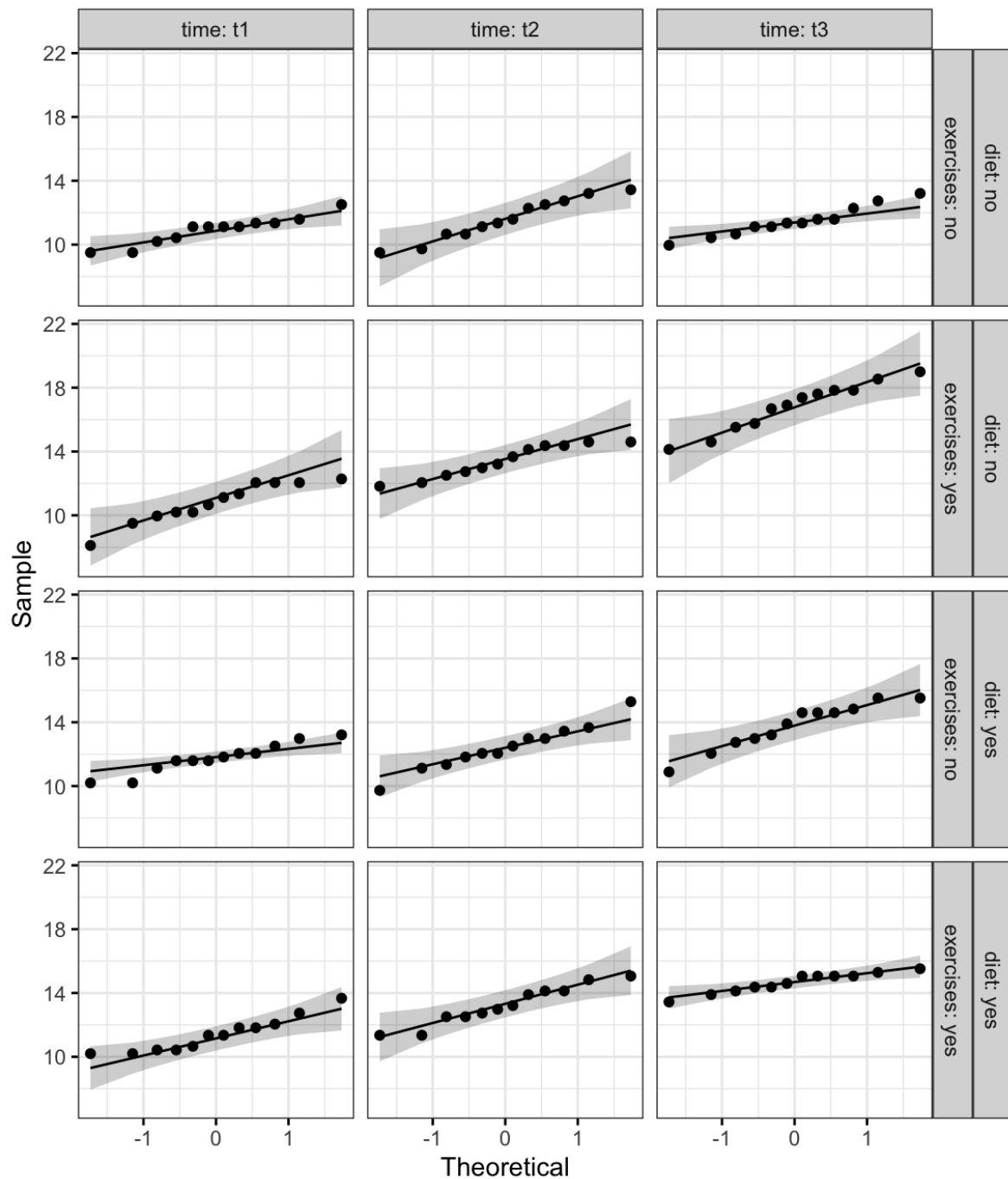
The weight loss score was normally distributed ($p > 0.05$), as assessed by Shapiro-Wilk's test of normality.

Create QQ plot for each cell of design:

```
1   ggqqplot(weightloss, "score", ggtheme = theme_bw()) +
2     facet_grid(diet + exercises ~ time, labeller = "label_both")
```

From the plot above, as all the points fall approximately along this reference line, we can assume normality.

Homogneity of variance assumption

Compute the Levene's test after grouping the data by `diet` and `time` categories:

```
 1   weightloss %>%
 2     group_by(diet, time) %>%
 3     levene_test(score ~ exercises)
 4   ## # A tibble: 6 x 6
 5   ##   diet  time   df1   df2 statistic      p
 6   ##   <fct> <fct> <int> <int>    <dbl>  <dbl>
 7   ## 1 no    t1      1    22     2.44   0.132
 8   ## 2 no    t2      1    22     0.691  0.415
 9   ## 3 no    t3      1    22     2.87   0.105
10   ## 4 yes   t1      1    22     0.376  0.546
11   ## 5 yes   t2      1    22     0.0574 0.813
12   ## 6 yes   t3      1    22     5.14   0.0336
```

There was homogeneity of variances for all cells (p > 0.05), except for the condition diet:yes at time:t3 (p = 0.034), as assessed by Levene's test of homogeneity of variance.

Note that, if you do not have homogeneity of variances, you can try to transform the outcome (dependent) variable to correct for the unequal variances.

If group sample sizes are (approximately) equal, run the three-way mixed ANOVA anyway because it is somewhat robust to heterogeneity of variance in these circumstances.

It's also possible to perform robust ANOVA test using the WRS2 R package.

Assumption of sphericity

As mentioned in the two-way mixed ANOVA section, the Mauchly's test of sphericity and the sphericity corrections are internally done using the R function `anova_test()` and `get_anova_table()` [rstatix package].

Computation

```
 1   res.aov <- anova_test(
 2     data = weightloss, dv = score, wid = id,
 3     between = exercises, within = c(diet, time)
 4     )
 5   get_anova_table(res.aov)
 6   ## ANOVA Table (type II tests)
 7   ##
 8   ##                   Effect DFn DFd     F        p p<.05   ges
 9   ## 1              exercises   1  22 38.771 2.88e-06     * 0.284
10   ## 2                   diet   1  22  7.912 1.00e-02     * 0.028
11   ## 3                   time   2  44 82.199 1.38e-15     * 0.541
12   ## 4          exercises:diet   1  22 51.698 3.31e-07     * 0.157
13   ## 5          exercises:time   2  44 26.222 3.18e-08     * 0.274
14   ## 6               diet:time   2  44  0.784 4.63e-01       0.013
15   ## 7 exercises:diet:time   2  44  9.966 2.69e-04     * 0.147
```

From the output above, it can be seen that there is a statistically significant three-way interactions between exercises, diet and time F(2, 44) = 9.96, p = 0.00027.

Note that, if the three-way interaction is not statistically significant, you need to consult the two-way interactions in the output.

In our example, there was a statistically significant two-way exercises*diet interaction (p < 0.0001), and two-way exercises*time (p < 0.0001). The two-way diet*time interaction was not statistically significant (p = 0.46).

Post-hoc tests

If there is a significant three-way interaction effect, you can decompose it into:

- **Simple two-way interaction**: run two-way interaction at each level of third variable,
- **Simple simple main effect**: run one-way model at each level of second variable, and
- **simple simple pairwise comparisons**: run pairwise or other post-hoc comparisons if necessary.

If you do not have a statistically significant three-way interaction, you need to determine whether you have any statistically significant two-way interaction from the ANOVA output. You can follow up a significant two-way interaction by simple main effects analyses and pairwise comparisons between groups if necessary.

In this section we'll describe the procedure for a significant three-way interaction.

Compute simple two-way interaction

In this example, we'll consider the `diet*time` interaction at each level of `exercises`. Group the data by `exercises` and analyze the simple two-way interaction between `diet` and `time`:

```
1   # Two-way ANOVA at each exercises group level
2   two.way <- weightloss %>%
3     group_by(exercises) %>%
4     anova_test(dv = score, wid = id, within = c(diet, time))
5   two.way
6   ## # A tibble: 2 x 2
7   ##   exercises anova
8   ##   <fct>     <list>
9   ## 1 no        <anov_tst>
10  ## 2 yes       <anov_tst>
11  # Extract anova table
12  get_anova_table(two.way)
13  ## # A tibble: 6 x 8
14  ##   exercises Effect      DFn   DFd      F         p `p<.05`    ges
15  ##   <fct>     <chr>     <dbl> <dbl>  <dbl>     <dbl> <chr>     <dbl>
16  ## 1 no        diet          1    11   56.4  1.18e- 5 *         0.262
17  ## 2 no        time          2    22    5.90 9.00e- 3 *         0.181
18  ## 3 no        diet:time     2    22    2.91 7.60e- 2 ""        0.09
19  ## 4 yes       diet          1    11    8.60 1.40e- 2 *         0.066
20  ## 5 yes       time          2    22  148.   1.73e-13 *         0.746
21  ## 6 yes       diet:time     2    22    7.81 3.00e- 3 *         0.216
```

There was a statistically significant simple two-way interaction between diet and time for exercises:yes group, F(2, 22) = 7.81, p = 0.0027, but not for exercises:no group, F(2, 22) = 2.91, p = 0.075.

Note that, statistical significance of a simple two-way interaction was accepted at a Bonferroni-adjusted alpha level of 0.025. This corresponds to the current level you declare statistical significance at (i.e., p < 0.05) divided by the number of simple two-way interaction you are computing (i.e., 2).

Compute simple simple main effect

A statistically significant simple two-way interaction can be followed up with **simple simple main effects**.

In our example, you could therefore investigate the effect of `time` on the weight loss score at every level of `diet` and/or investigate the effect of `diet` at every level of `time`.

Note that, you will only need to do this for the simple two-way interaction for "exercises:yes" group, as this was the only simple two-way interaction that was statistically significant.

Group the data by `exercises` and `diet`, and analyze the simple main effect of `time`:

```
 1   time.effect <- weightloss %>%
 2     group_by(exercises, diet) %>%
 3     anova_test(dv = score, wid = id, within = time) %>%
 4     get_anova_table()
 5   time.effect %>% filter(exercises == "yes")
 6   ## # A tibble: 2 x 9
 7   ##    diet  exercises Effect   DFn   DFd     F        p `p<.05`   ges
 8   ##    <fct> <fct>     <chr>  <dbl> <dbl> <dbl>    <dbl> <chr>    <dbl>
 9   ## 1 no    yes       time       2    22  78.8 9.30e-11 *        0.801
10   ## 2 yes   yes       time       2    22  30.9 4.06e- 7 *        0.655
```

In the table above, we only need the results for `exercises = yes`. Statistical significance of a simple simple main effect was accepted at a Bonferroni-adjusted alpha level of 0.025, that is 0.05 divided y the number of simple simple main effects you are computing (i.e., 2).

The simple simple main effect of time on weight loss score was statistically significant under exercises condition for both diet:no (F(2,22) = 78.81, p < 0.0001) and diet:yes (F(2, 22) = 30.92, p < 0.0001) groups.

Compute simple simple comparisons

A statistically significant simple simple main effect can be followed up by **multiple pairwise comparisons** to determine which group means are different.

Recall that, you will only need to concentrate on the pairwise comparison results for exercises:yes.

Group the data by `exercises` and `diet`, and perform pairwise comparisons between `time` points with Bonferroni adjustment. Paired t-test is used:

```
 1   # compute pairwise comparisons
 2   pwc <- weightloss %>%
 3     group_by(exercises, diet) %>%
 4     pairwise_t_test(
 5       score ~ time, paired = TRUE,
 6       p.adjust.method = "bonferroni"
 7       ) %>%
 8     select(-statistic, -df) # Remove details
 9   # Focus on the results of exercises:yes group
10   pwc %>% filter(exercises == "yes") %>%
11     select(-p)    # Remove p column
12   ## # A tibble: 6 x 9
```

```
13  ##   diet  exercises .y.   group1 group2   n1   n2      p.adj
    p.adj.signif
14  ##   <fct> <fct>      <chr> <chr>  <chr>  <int> <int>       <dbl> <chr>

15  ## 1 no    yes        score t1     t2        12   12 0.000741     ***

16  ## 2 no    yes        score t1     t3        12   12 0.0000000121 ****

17  ## 3 no    yes        score t2     t3        12   12 0.000257     ***

18  ## 4 yes   yes        score t1     t2        12   12 0.01         **

19  ## 5 yes   yes        score t1     t3        12   12 0.00000124   ****

20  ## 6 yes   yes        score t2     t3        12   12 0.02         *
```

All simple simple pairwise comparisons were run between the different time points under exercises condition (i.e., exercises:yes) for diet:no and diet:yes trials. A Bonferroni adjustment was applied.

The mean weight loss score was significantly different in all time point comparisons when exercises are performed (p < 0.05).

Report

A three-way mixed ANOVA was performed to evaluate the effects of diet, exercises and time on weight loss.

There were no extreme outliers, as assessed by box plot method. The data was normally distributed, as assessed by Shapiro-Wilk's test of normality (p > 0.05). There was homogeneity of variances (p > 0.05) as assessed by Levene's test of homogeneity of variances. For the three-way interaction effect, Mauchly's test of sphericity indicated that the assumption of sphericity was met (p > 0.05).

There was a statistically significant three-way interaction between exercises, diet and time F(2, 44) = 9.96, p < 0.001.

For the simple two-way interactions and simple simple main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the p < 0.025 level.

There was a statistically significant simple two-way interaction between diet and time for exercises:yes group, F(2, 22) = 7.81, p = 0.0027, but not for exercises:no group, F(2, 22) = 2.91, p = 0.075.

The simple simple main effect of time on weight loss score was statistically significant under exercises condition for both diet:no (F(2,22) = 78.81, p < 0.0001) and diet:yes (F(2, 22) = 30.92, p < 0.0001) groups.

All simple simple pairwise comparisons were run between the different time points under exercises condition (i.e., exercises:yes) for diet:no and diet:yes trials. A Bonferroni adjustment was applied. The mean weight loss score was significantly different in all time point comparisons when exercises are performed (p < 0.05).

```
1   # Visualization: box plots with p-values
2   pwc <- pwc %>% add_xy_position(x = "exercises")
3   pwc.filtered <- pwc %>% filter(exercises == "yes")
4   bxp +
5     stat_pvalue_manual(pwc.filtered, tip.length = 0, hide.ns = TRUE) +
6     labs(
7       subtitle = get_test_label(res.aov, detailed = TRUE),
8       caption = get_pwc_label(pwc)
9     )
```

Anova, $F(2,44) = 9.97$, $p = 0.00027$, $\eta^2_g = 0.15$

pwc: **T test**; p.adjust: **Bonferroni**
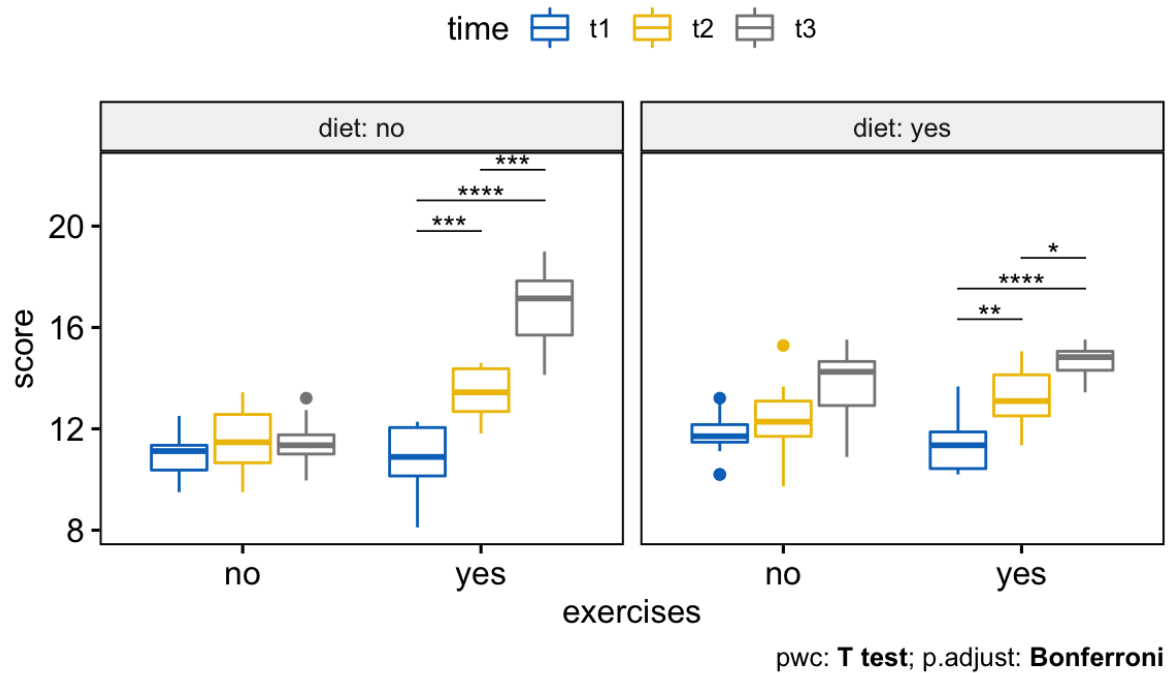
Summary

This article describes how to compute and interpret mixed ANOVA in R. We also explain the assumptions made by mixed ANOVA tests and provide practical examples of R codes to check whether the test assumptions are met.