Appendix 1: R Functions

Manuscript: Simulation-Based Power-Analysis for Factorial ANOVA Designs

Daniël Lakens & Aaron R. Caldwell

The goal of ANOVApower is to easily simulate ANOVA designs and calculate the observed power. In the manuscript we referenced numerous simulations, and within this appendix we have reproduced these simulations using the functions in the ANOVApower R package.

Installation

First, install the released version of ANOVApower from GitHub with:

```
devtools::install_github("Lakens/ANOVApower")

library(ANOVApower)
```

ANOVA_design function

The ANOVA_design function can create designs up three factors, for both within, between, and mixed designs. It requires the following input: string, n, mu, sd, r, and labelnames.

1. string: string that specifies the design (see below).
2. n: the sample size for each between subject condition.
3. mu: a vector with the means for each condition.
4. sd: the population standard deviation. Assumes homogeneity of variances (only one standard deviation can be provided).
5. r: the correlation for within designs (or 0 for between designs).
6. labelnames: This is a vector of words that indicate factor names and level names (see below).
7. A final optional setting is to specify if you want to output a plot or not (plot = TRUE or FALSE)

ANOVA_power function

The ANOVA_power function is used to perform the simulation. A design must be specified from the ANOVA_design function. It requires the following input: design_result, alpha_level, p_adjust, nsims, seed, and verbose.

1. design_result: output from the ANOVA_design function
2. alpha_level: the alpha level used to determine statistical significance
3. p_adjust: correction for multiple comparisons; set to none is none required or desired
4. nsims: number of simulations to perform
5. seed: set for reproducible results
6. verbose: Set to FALSE to not print results (default = TRUE)

Settings from the manuscript

In order to reproduce the simulations from the manuscript we need to set the number of repetitions in each simulation and the seed number.

```
nsims <- 100000


manuscript_seed = 2019
```

#Simple 2 group between-subjects design

The initial study from the manuscript described a study wherein participants interact with an artificial voice assistant who sounds either cheerful or sad, and enjoyment is measured on scale (-5 to +5). In the ANOVA_design function this a two-level between-participant design.

`string = "2b"`

In this study, expected mean in the cheerful condition is 1 and 0 in the sad condition with a standard deviation of 2.

```
mu = c(1,0)

sd = 2
```

We also assumed we will recruit 80 participants in each group.

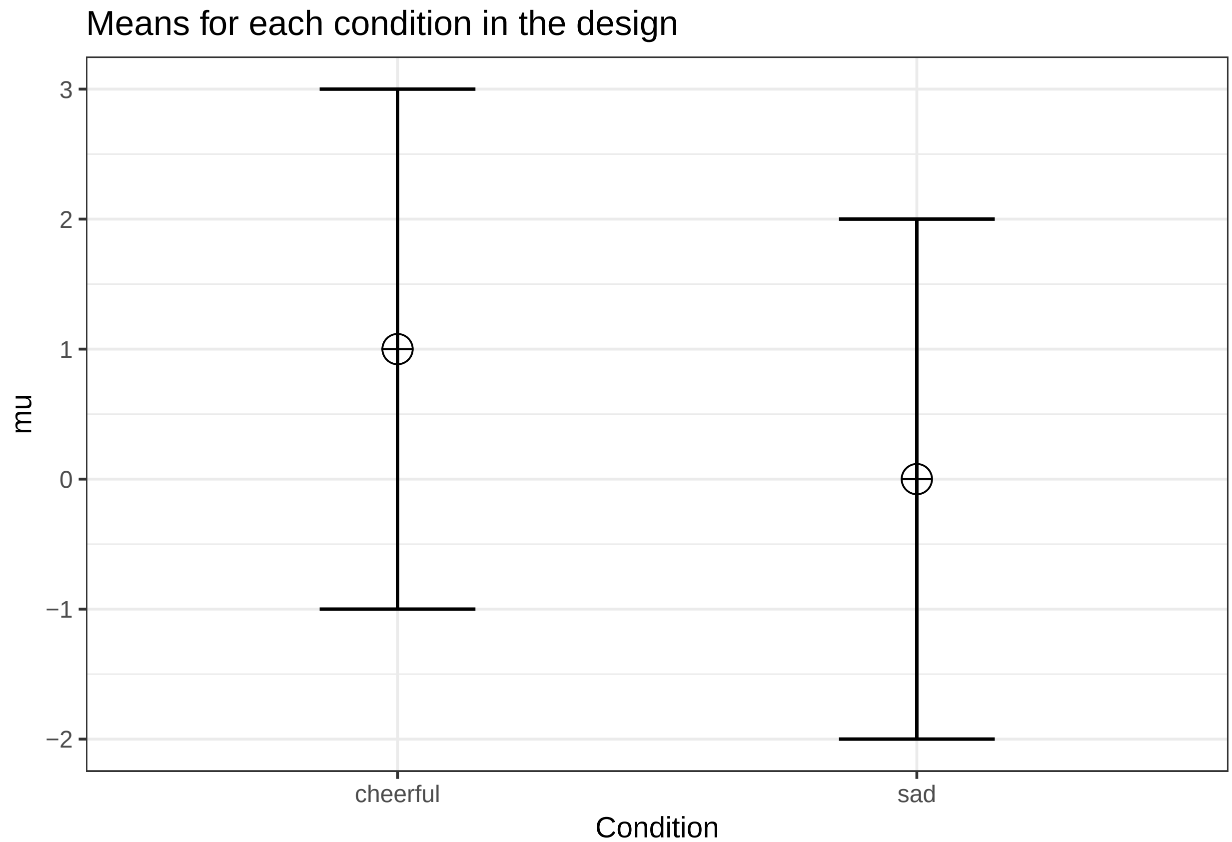
`n = 80`

For ease of interpretation the factors and levels our named in our example, factor = "Condition" with levels = "cheerful" and "sad".

`labelnames = c("Condition", "cheerful", "sad")`

We now can put this all together to into the ANOVA_design function.

```
design_result <- ANOVA_design(string = "2b", n = 80, mu = c(1, 0), sd = 2,
                              labelnames = c("Condition", "cheerful", "sad"))
```



With the design defined, we can run the simulation with the significance level at 0.05 (`alpha_level = 0.05`) and we will run the previously defined number of simulations (`nsims = nsims`) and seed number (`seed = manuscript_seed`).

```
power_result <- ANOVA_power(design_result,
                            alpha_level = 0.05,
                            nsims = nsims,
                            seed = manuscript_seed,
                            verbose = TRUE)
```

```
## Power and Effect sizes for ANOVA tests
##                  power effect_size
## anova_Condition 88.03      0.0595
##
## Power and Effect sizes for contrasts
##                                    power effect_size
## p_Condition_cheerful_Condition_sad 88.03     -0.5012
```

The results of a simulation study will vary each time the simulation is performed, but, in this case the results are reproducible by specifying a `seed`. When 100000 simulations are performed with a seed set to 2019, we see the statistical power (`Power and Effect sizes for ANOVA tests`) is 88.03% and the average parital η^2 (eta-squared) is 0.0595. In this scenario, the ANOVA results are exactly the same as the power from a simple t-test (`Power and Effect sizes for contrasts`), but the effect size, Cohen's d, is -0.5012.

#Extending to 3 conditions

In the next example, we explored what would happen if we extended the design to 3 between-participant conditions.

`string = "3b"`

This was accomplished by adding "neutral" condition.

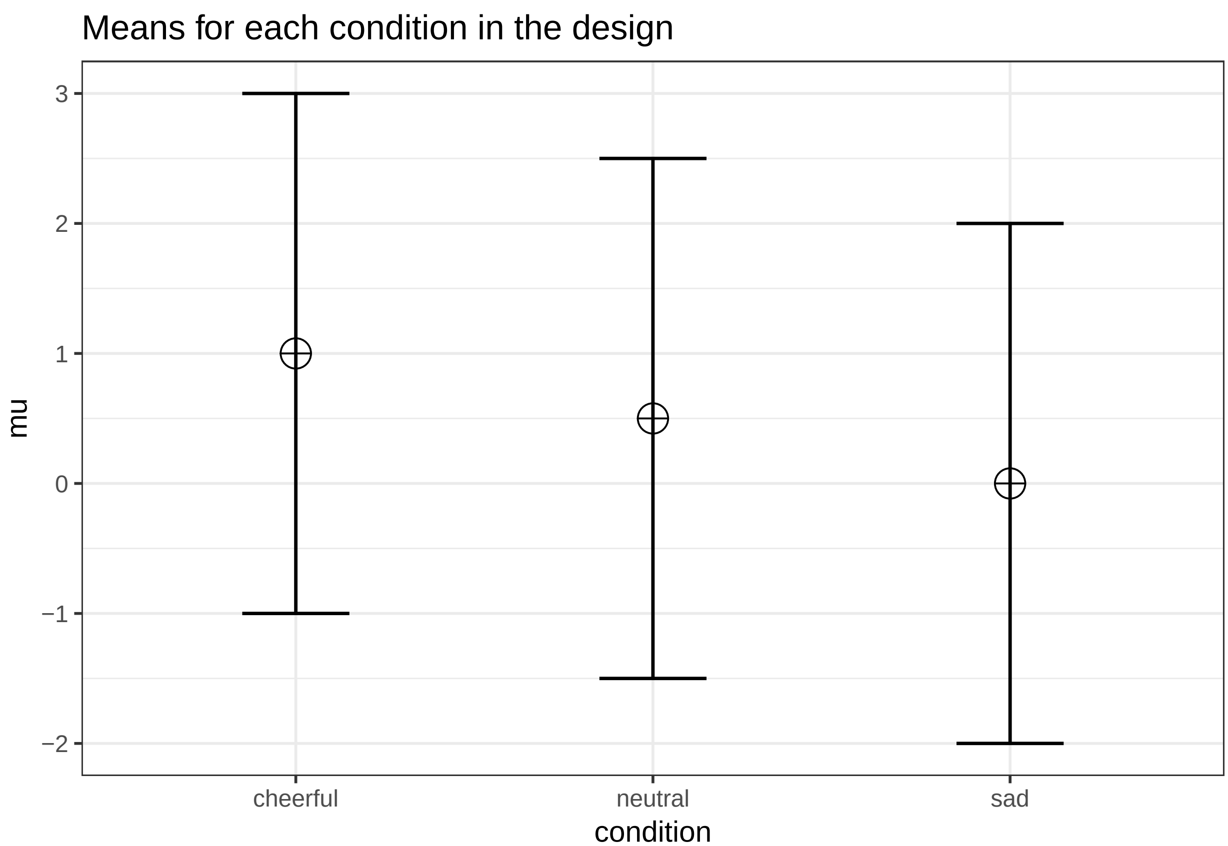
`labelnames = c("condition", "cheerful", "neutral", "sad")`

In one scenario, the expected means were 1 (cheerful), 0.5 (neutral), and 0 (sad).

`mu = c(1, 0.5, 0)`

This full design can be implemented as the following in the `ANOVA_design` function.

```
design_result_1 <- ANOVA_design(string = "3b", n = 80, mu = c(1, 0.5, 0), sd = 2,
                                labelnames = c("condition", "cheerful", "neutral", "sad"))
```



Now we can simulate this design with the `ANOVA_power` function.

```
power_result_1 <- ANOVA_power(design_result_1,
                              alpha_level = 0.05,
                              nsims = nsims,
                              seed = manuscript_seed,
                              verbose = TRUE)
```

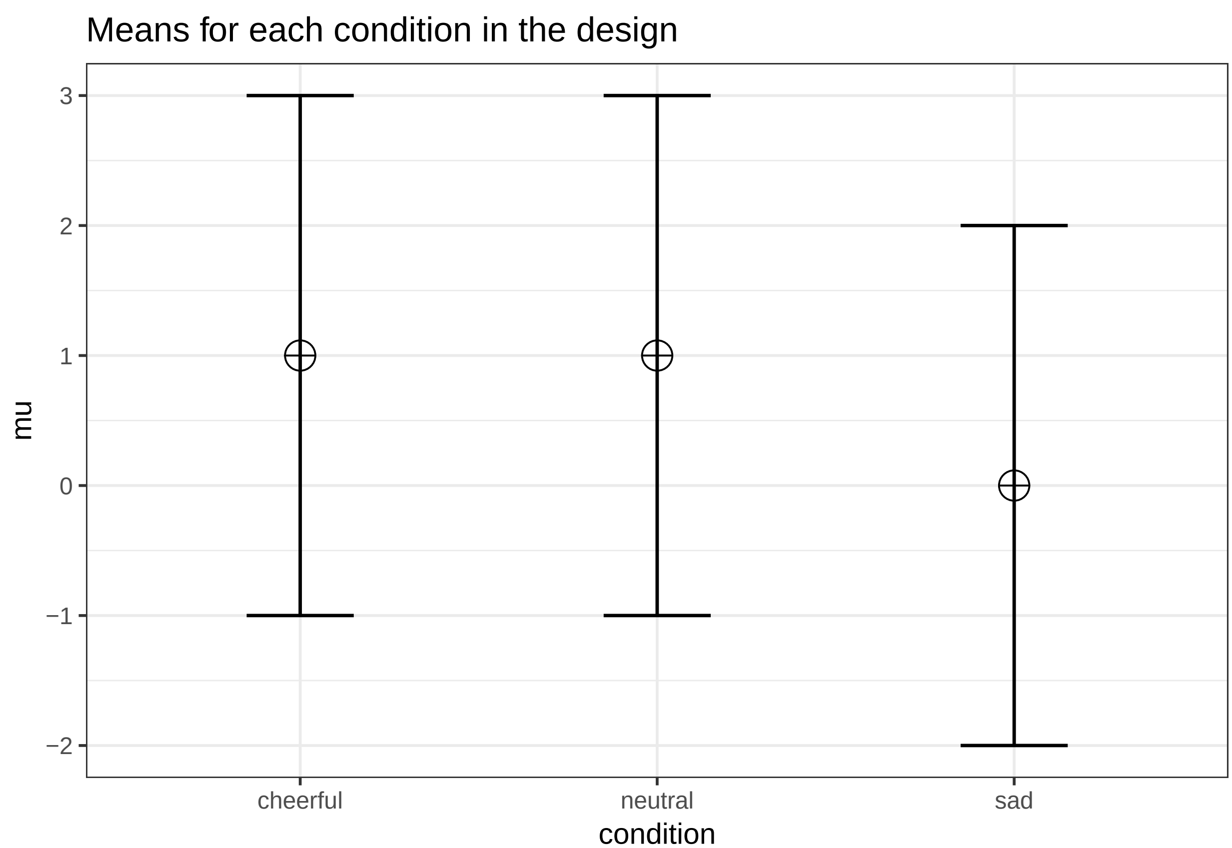
```
## Power and Effect sizes for ANOVA tests
##                 power effect_size
## anova_condition 81.71      0.0451
```

```
##
## Power and Effect sizes for contrasts
##                                           power effect_size
## p_condition_cheerful_condition_neutral 35.59     -0.2525
## p_condition_cheerful_condition_sad     88.64     -0.5051
## p_condition_neutral_condition_sad      35.94     -0.2525
```

In the second scenario, we assume that the enjoyment is the same (mean = 1) in the cheerful and neutral condition.

```
mu = c(1, 1, 0)
```

```
design_result_2 <- ANOVA_design(string = "3b", n = 80, mu = c(1, 1, 0), sd = 2,
                                labelnames = c("condition", "cheerful", "neutral", "sad"))
```



Again, we simulate this new design.

```
power_result_2 <- ANOVA_power(design_result_2,
                              alpha_level = 0.05,
                              nsims = nsims,
                              seed = manuscript_seed,
                              verbose = TRUE)
```

```
## Power and Effect sizes for ANOVA tests
##                 power effect_size
## anova_condition 91.57      0.0576
```

```
##
## Power and Effect sizes for contrasts
##                                         power effect_size
## p_condition_cheerful_condition_neutral  5.32     -0.0012
## p_condition_cheerful_condition_sad     88.64     -0.5051
## p_condition_neutral_condition_sad      88.18     -0.5037
```

As we can see the ANOVA-level power in scenario #1 was 81.71 % compared to 91.57 % in scenario #2. Further, the effect size, parital η^2, increased from 0.0451 to 0.0576. As we can see, changing the design from `2b` to `3b` has an effect on power dependent upon the expected pattern of the means.

#Chaning to a within-subjects design

Now we modify the design by changing it to a within-subjects (i.e., repeated measures) design. Therefore, the first modification we must make in the `ANOVA_design` function is to the design definition.

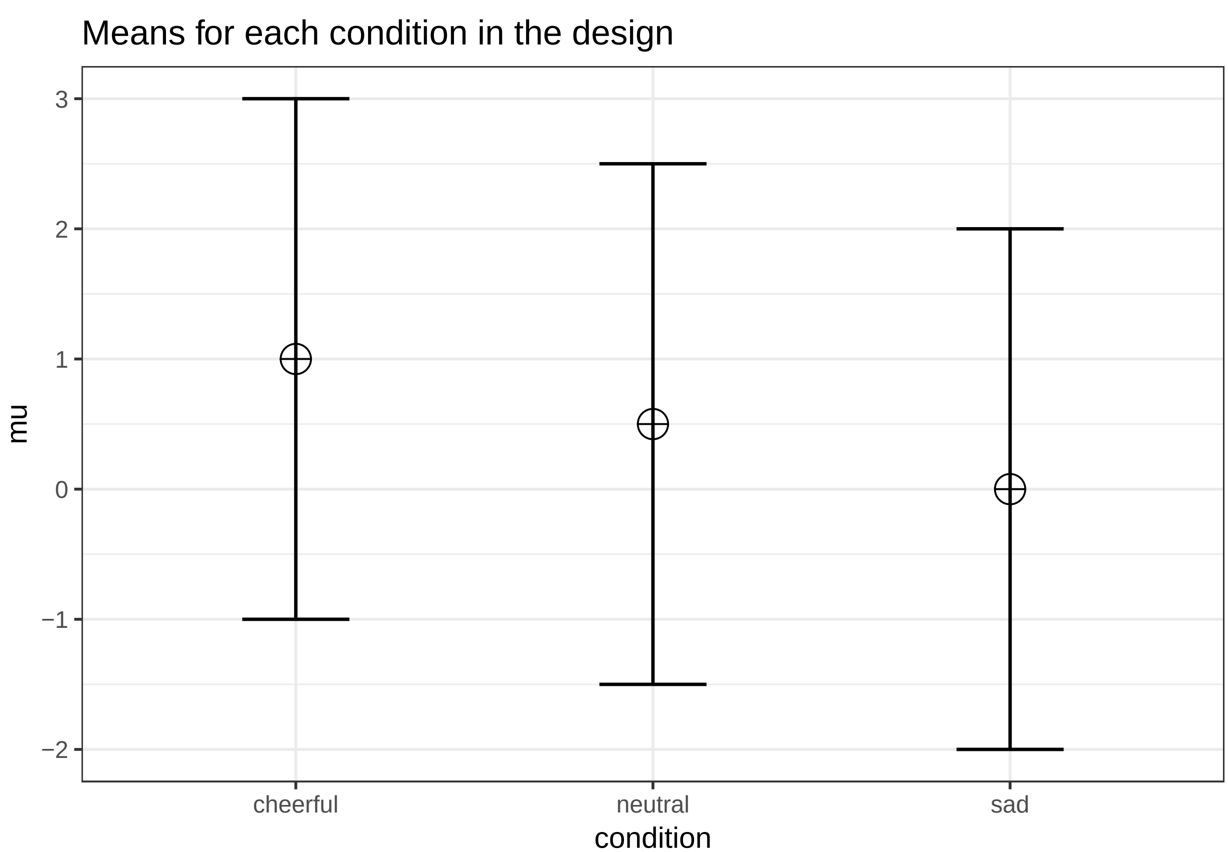
`string = "3w"`

We must also specify the correlation between dependent measures (below).

```
r = 0.5
```

Now we have the information to define the ANOVA design.

```
design_result_within_1 <- ANOVA_design(string = "3w", n = 80, mu = c(1, 0.5, 0), sd = 2, r = 0.5,
                                      labelnames = c("condition", "cheerful", "neutral", "sad"))
```



Now, visually speaking, nothing in the design appears to have changed. However, let's look at the correlation matrix from the `3b` design. Notice (below) that there is no correlation between measurements

```
design_result_1$cor_mat
```

```
##   X1 X2 X3
## 1  1  0  0
## 2  0  1  0
## 3  0  0  1
```

Let's look at the correlation matrix for the `3w` design.

```
design_result_within_1$cor_mat
```

```
##    X1  X2  X3
## 1 1.0 0.5 0.5
## 2 0.5 1.0 0.5
## 3 0.5 0.5 1.0
```

The within-subjects design has been established; now we run the simulation.

```
power_result_within_1 <- ANOVA_power(design_result_within_1,
                                     alpha_level = 0.05,
                                     nsims = nsims,
                                     seed = manuscript_seed,
                                     verbose = TRUE)
```

```
## Power and Effect sizes for ANOVA tests
##                 power effect_size
## anova_condition 98.29      0.1187
##
## Power and Effect sizes for contrasts
##                                      power effect_size
## p_condition_cheerful_condition_neutral 60.12     -0.2545
## p_condition_cheerful_condition_sad     99.24     -0.5062
## p_condition_neutral_condition_sad      59.77     -0.2517
```

Power has now increased to 98.29% from the `3b` design (81.71%). The total number of participants is lower in the within-subjects design (N=240 to N=80), but there are now three observations per participant (80 x 3 = 240).

#Power for Interactions

In addition to simple one-way effects, in the manuscript we demonstrate the power analysis for 2x2 between-subjects design. In this scenario there two factors "condition" and "voice". Again, participants are exposed to a "sad" or "cheerful" condition. However, the voice is either human-like ("human") or robotic ("robot").

The design has changed so we must change the design definition.

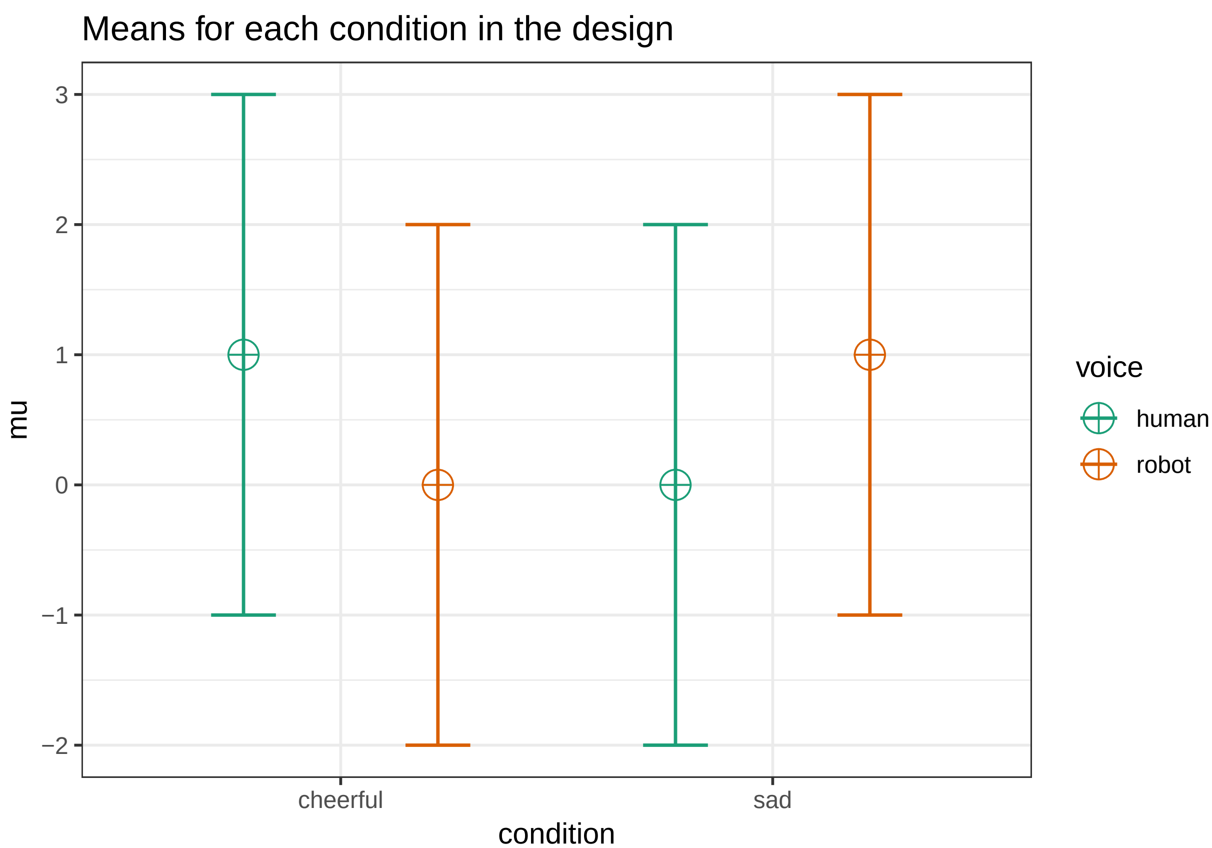
We must also define a new vector of means for the given design. In this case we expect a cross-over interaction wherein the mean response to cheerful-human will be the same as sad-robot (mean = 1) and cheerful-robot is the same as sad-human (mean = 0). Further, the standard deviation in each condition is equal to 2.

```
mu = c(1, 0, 0, 1)
sd = 2
```

Now that the basics are laid out we can set up the design.

```r
design_result_cross_40 <- ANOVA_design(string = "2b*2b", n = 40, mu = c(1, 0, 0, 1), sd = 2,
                                       labelnames = c("condition", "cheerful", "sad",
                                                      "voice", "human", "robot"),
                                       plot = FALSE)

design_result_cross_80 <- ANOVA_design(string = "2b*2b", n = 80, mu = c(1, 0, 0, 1), sd = 2,
                                       labelnames = c("condition", "cheerful", "sad",
                                                      "voice", "human", "robot"))
```



Only one means plot is produced because the vector of means and standard deviation are the same.

Power for a sample size of 40.

```r
power_result_cross_40 <- ANOVA_power(design_result_cross_40,
                                     alpha_level = 0.05,
                                     nsims = nsims,
                                     seed = manuscript_seed,
                                     verbose = TRUE)
```

```
## Power and Effect sizes for ANOVA tests
##                       power effect_size
## anova_condition        5.20      0.0030
## anova_voice            4.47      0.0029
## anova_condition:voice 88.02      0.0605
##
## Power and Effect sizes for contrasts
```

```
##                                                               power
## p_condition_cheerful_voice_human_condition_cheerful_voice_robot 59.99
## p_condition_cheerful_voice_human_condition_sad_voice_human      60.06
## p_condition_cheerful_voice_human_condition_sad_voice_robot       4.76
## p_condition_cheerful_voice_robot_condition_sad_voice_human       4.93
## p_condition_cheerful_voice_robot_condition_sad_voice_robot      58.81
## p_condition_sad_voice_human_condition_sad_voice_robot           59.59
##                                                               effect_size
## p_condition_cheerful_voice_human_condition_cheerful_voice_robot     -0.5044
## p_condition_cheerful_voice_human_condition_sad_voice_human          -0.5067
## p_condition_cheerful_voice_human_condition_sad_voice_robot          -0.0036
## p_condition_cheerful_voice_robot_condition_sad_voice_human          -0.0025
## p_condition_cheerful_voice_robot_condition_sad_voice_robot           0.5012
## p_condition_sad_voice_human_condition_sad_voice_robot                0.5033
```

Power for a sample size of 80.

```r
power_result_cross_80 <- ANOVA_power(design_result_cross_80,
                                     alpha_level = 0.05,
                                     nsims = nsims,
                                     seed = manuscript_seed,
                                     verbose = TRUE)
```

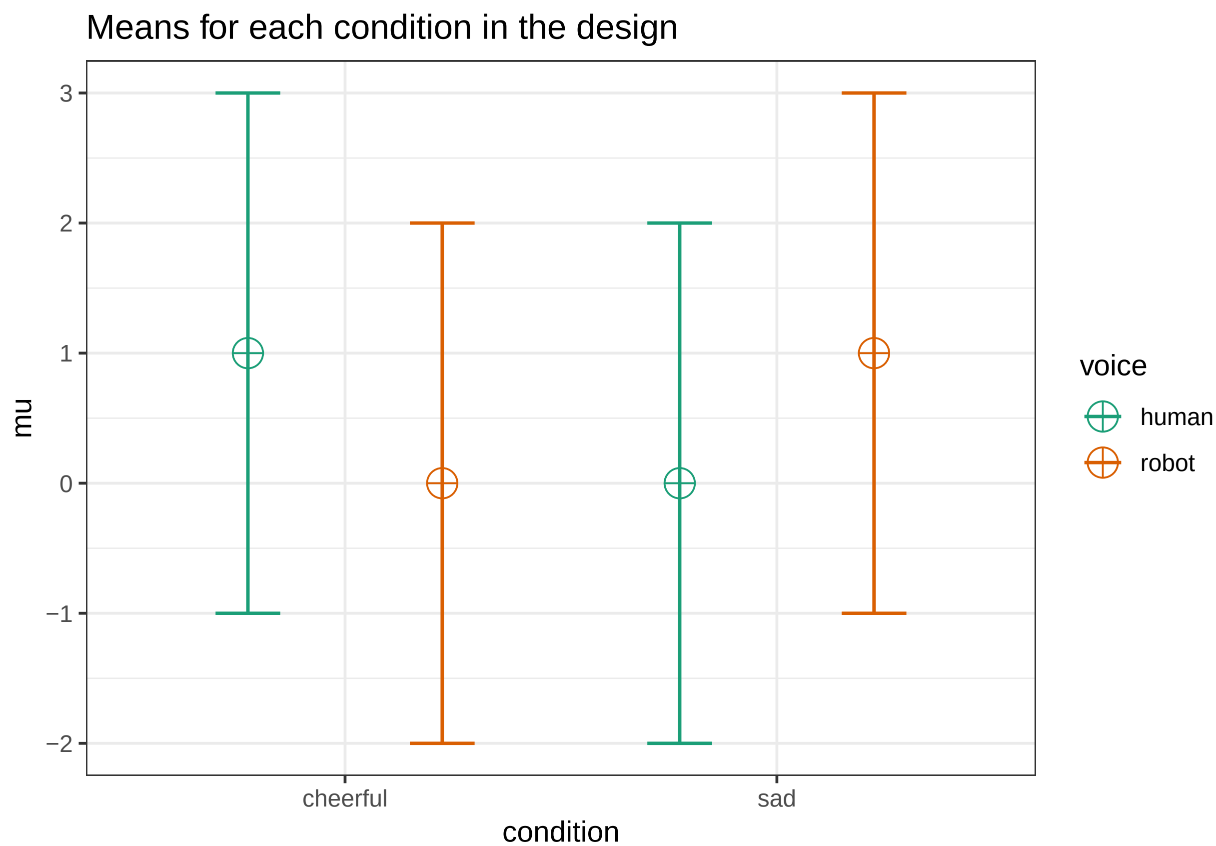
```
## Power and Effect sizes for ANOVA tests
##                         power effect_size
## anova_condition          4.80      0.0014
## anova_voice              4.95      0.0015
## anova_condition:voice   99.47      0.0597
##
## Power and Effect sizes for contrasts
##                                                               power
## p_condition_cheerful_voice_human_condition_cheerful_voice_robot 88.55
## p_condition_cheerful_voice_human_condition_sad_voice_human      88.30
## p_condition_cheerful_voice_human_condition_sad_voice_robot       4.91
## p_condition_cheerful_voice_robot_condition_sad_voice_human       5.00
## p_condition_cheerful_voice_robot_condition_sad_voice_robot      88.06
## p_condition_sad_voice_human_condition_sad_voice_robot           88.22
##                                                               effect_size
## p_condition_cheerful_voice_human_condition_cheerful_voice_robot     -0.5056
## p_condition_cheerful_voice_human_condition_sad_voice_human          -0.5041
## p_condition_cheerful_voice_human_condition_sad_voice_robot          -0.0023
## p_condition_cheerful_voice_robot_condition_sad_voice_human           0.0015
## p_condition_cheerful_voice_robot_condition_sad_voice_robot           0.5037
## p_condition_sad_voice_human_condition_sad_voice_robot                0.5021
```

The simulations show we have 99.47% power when 80 participants are collected per condition (N = 320). If the sample size per cell is reduced from 80 to 40, power to detect the interaction is 88.02%.

#Adjusting for Multiple Comparisons

As mentioned in the manuscript, the number of pairwise comparisons, if left unadjusted, can increase the Type I error rate. So in the manuscript, we revisited the 40 person per group study from the cross-over interaction example. So the design below is the exact same as `design_result_cross_40`.

```r
design_result_holm_correction <- ANOVA_design(string = "2b*2b", n = 40, mu = c(1, 0, 0, 1), sd = 2,
                                              labelnames = c("condition", "cheerful", "sad",
                                                             "voice", "human", "robot"))
```



Now, when we run the simulation we can adjust for multiple comparisons using the `p_adjust` argument. In this case we use the Holm-Bonferroni correction. This feature is derived from `stats` package (base R), and number of different adjustments for multiple comparions can be applied in the simulation (see `?p.adjust` for options).

```r
power_result_holm_correction <- ANOVA_power(design_result_holm_correction,
                                            alpha_level = 0.05,
                                            p_adjust = "holm",
                                            nsims = nsims,
                                            seed = manuscript_seed,
                                            verbose = TRUE)
```

```
## Power and Effect sizes for ANOVA tests
##                          power effect_size
## anova_condition           2.38      0.0030
## anova_voice               2.14      0.0029
## anova_condition:voice    77.07      0.0605
##
## Power and Effect sizes for contrasts
##                                                           power
## p_condition_cheerful_voice_human_condition_cheerful_voice_robot 36.12
```

```
## p_condition_cheerful_voice_human_condition_sad_voice_human      35.98
## p_condition_cheerful_voice_human_condition_sad_voice_robot       1.21
## p_condition_cheerful_voice_robot_condition_sad_voice_human       1.40
## p_condition_cheerful_voice_robot_condition_sad_voice_robot      35.60
## p_condition_sad_voice_human_condition_sad_voice_robot           34.81
##                                                                 effect_size
## p_condition_cheerful_voice_human_condition_cheerful_voice_robot     -0.5044
## p_condition_cheerful_voice_human_condition_sad_voice_human          -0.5067
## p_condition_cheerful_voice_human_condition_sad_voice_robot          -0.0036
## p_condition_cheerful_voice_robot_condition_sad_voice_human          -0.0025
## p_condition_cheerful_voice_robot_condition_sad_voice_robot           0.5012
## p_condition_sad_voice_human_condition_sad_voice_robot                0.5033
```

As we can see from this simulation, power for the first pairwise comparison (cheerful-human vs cheerful-robot) decreases from 59.99% when the Holm-Bonferroni correction is applied to 36.12%.