R Workshop

Multilevel Modelling with R
Contrasts and Pairwise Comparisons

Dr. Daniel Lüdecke
d.luedecke@uke.de
https://github.com/easystats

Universitätsklinikum Hamburg-Eppendorf

Pairwise comparisons and contrasts

- Adjusted predictions / marginal means help understanding the relationship between predictors and outcome.
- Complex models are easier to understand when looking at adjusted predictions instead of a table of regression coefficients.
- Natural next step is to see if there are statistically significant differences. E.g., differences between groups, i.e., between the levels of categorical predictors or whether trends differ significantly from each other.
- The **modelbased** package provides a function, `estimate_contrasts()`, which does exactly this: testing differences of adjusted predictions for statistical significance. This is usually called *contrasts* or *(pairwise) comparisons*.

Pairwise comparisons and contrasts

Remember: whenever we refer to predictions (or marginal means), we are talking about the average expected values of the response variable (at different values of our focal predictors), based on what our model has estimated.

Such estimates (expected values, adjusted predictions, marginal means, ...) can be used to make inferences about relationships between variables.

- To summarize: The expected value of our outcome varies depending on the values of our predictors. E.g., if our outcome measures *functional limitations*, we might expect a higher value of our outcome (i.e., more functional limitations) for an older person compared to a younger person. This is what we call *predictions* or the “adjusted” (marginal) mean.

Simple example: Does our `score` differ between `time` points?

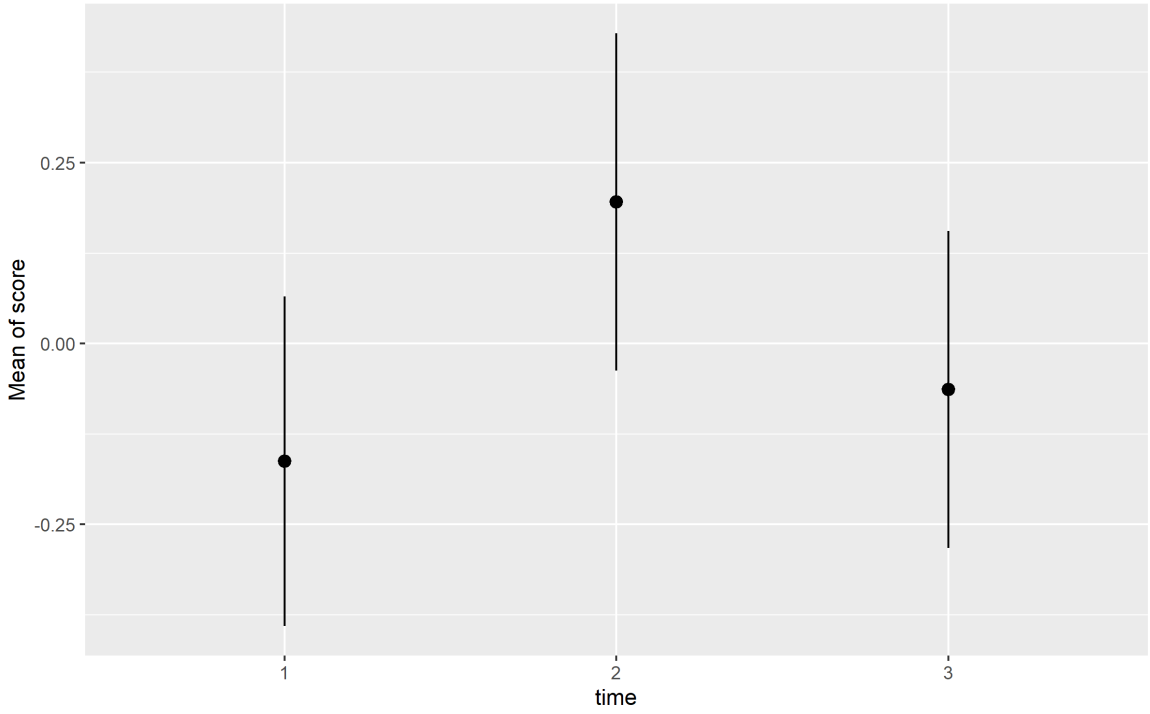
- For this example, we fit a simple linear model and look at the coefficients.

```
set.seed(123)
n <- 200
d <- data.frame(
  score = rnorm(n),
  grp = as.factor(sample(c("treatment", "control"), n, TRUE)),
  time = as.factor(sample(1:3, n, TRUE))
)
model1 <- lm(score ~ grp + time, data = d)
model_parameters(model1)
#> Parameter       | Coefficient |   SE |        95% CI | t(196) |     p
#> ------------------------------------------------------------------------
#> (Intercept)     |       -0.08 | 0.13 | [-0.33, 0.18] |  -0.60 | 0.552
#> grp [treatment] |       -0.17 | 0.13 | [-0.44, 0.09] |  -1.30 | 0.197
#> time [2]        |        0.36 | 0.16 | [ 0.03, 0.68] |   2.18 | 0.031
#> time [3]        |        0.10 | 0.16 | [-0.22, 0.42] |   0.62 | 0.538
```

Simple example: Does our `score` differ between `time` points?

- Let us look at the adjusted predictions:

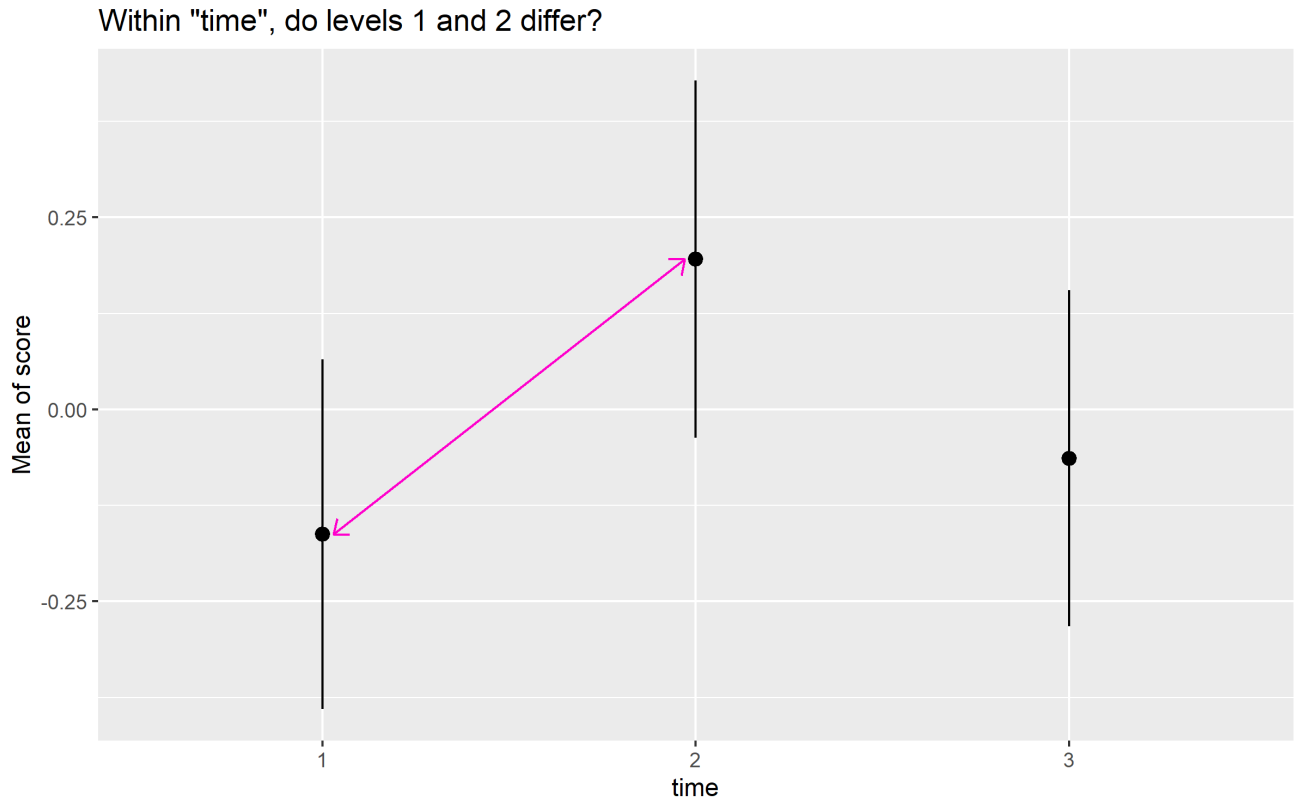
```
predictions <- estimate_means(model1, "time")
predictions
#> Estimated Marginal Means
#>
#> time |  Mean |   SE |         95% CI | t(196)
#> ------------------------------------------------
#> 1    | -0.16 | 0.12 | [-0.39, 0.07] |  -1.41
#> 2    |  0.20 | 0.12 | [-0.04, 0.43] |   1.66
#> 3    | -0.06 | 0.11 | [-0.28, 0.16] |  -0.57
#>
#> Variable predicted: score
#> Predictors modulated: time
#> Predictors averaged: grp
```



- We see that, for instance, the expected average score when `time=2` is 0.20.

Pairwise Comparison

- Next, we could ask whether the predicted outcome (score) for `time=1` is significantly different from the predicted score at `time=2`.



Pairwise Comparison

- We use the `estimate_contrasts()` function to achieve this.
- By default, when all focal terms are categorical, a pairwise comparison is performed.
- You can specify other types of comparisons, including more complex hypothesis testing as well, using the `comparison` argument.

Adjusted predictions

```
estimate_means(model1, "time")[1:5]
#> time |  Mean |   SE |           CI
#> ------------------------------------
#> 1    | -0.16 | 0.12 | [-0.39, 0.07]
#> 2    |  0.20 | 0.12 | [-0.04, 0.43]
#> 3    | -0.06 | 0.11 | [-0.28, 0.16]
```

Pairwise Comparison

```
# we only print some columns
estimate_contrasts(model1, "time")[c(1:3, 9)]
#> Level1 | Level2 | Difference |     p
#> ------------------------------------
#> 1      | 2      |      -0.36 | 0.031
#> 1      | 3      |      -0.10 | 0.538
#> 2      | 3      |       0.26 | 0.112
```

- For our quantity of interest, the contrast between time levels 1 and 2, we see the value -0.36, which is exactly the difference between the predicted score for time = 1 (-0.16) and time = 2 (0.20).
- The related p-value is 0.031, indicating that the difference between the predicted values of our outcome at these two levels of the factor time is indeed statistically significant.

Does same level of time differ between groups?

- The next example includes a pairwise comparison of an interaction between two categorical predictors.

```
model2 <- lm(score ~ grp * time, data = d)
model_parameters(model2)
#> Parameter                   | Coefficient |   SE |         95% CI | t(194) |     p
#> ------------------------------------------------------------------------------------
#> (Intercept)                 |        0.03 | 0.15 | [-0.27, 0.33] |   0.18 | 0.853
#> grp [treatment]             |       -0.42 | 0.23 | [-0.88, 0.04] |  -1.80 | 0.074
#> time [2]                    |        0.20 | 0.22 | [-0.23, 0.63] |   0.94 | 0.350
#> time [3]                    |       -0.07 | 0.22 | [-0.51, 0.37] |  -0.32 | 0.750
#> grp [treatment] × time [2] |        0.36 | 0.33 | [-0.29, 1.02] |   1.09 | 0.277
#> grp [treatment] × time [3] |        0.37 | 0.32 | [-0.27, 1.00] |   1.14 | 0.254
```

Does same level of time differ between groups?

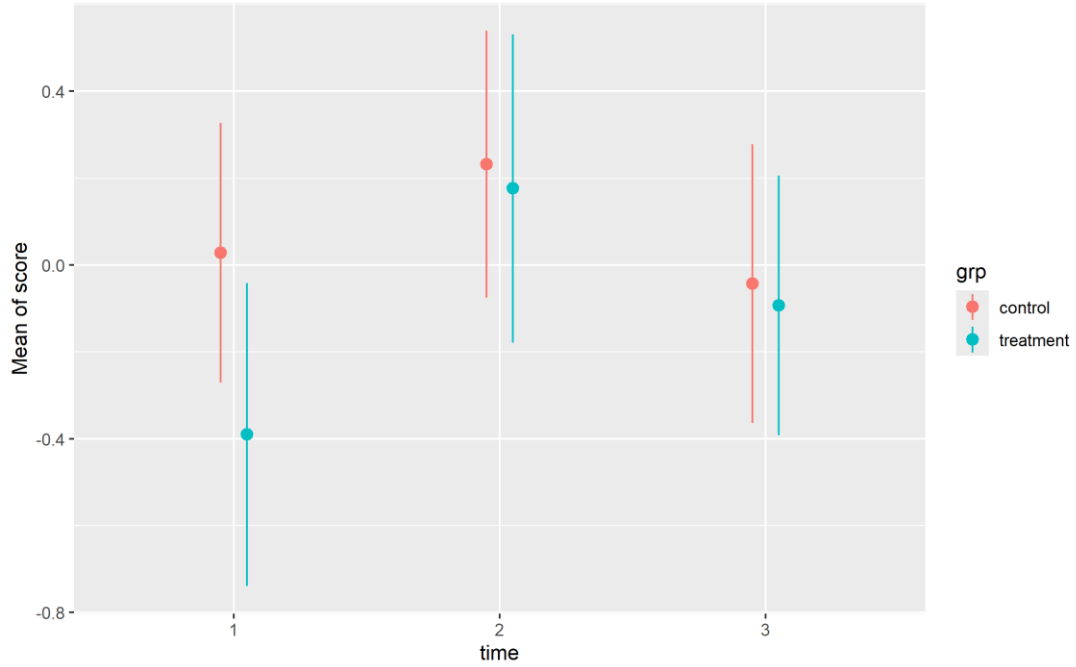
- First, we look at the marginal means of score for all combinations of the involved interaction term.

```
predictions <- estimate_means(model2, c("time", "grp"))
predictions
#> Estimated Marginal Means
#>
#> time | grp       |  Mean |   SE |         95% CI | t(194)
#> ------------------------------------------------------------
#> 1    | control   |  0.03 | 0.15 | [-0.27,  0.33] |   0.18
#> 2    | control   |  0.23 | 0.16 | [-0.08,  0.54] |   1.49
#> 3    | control   | -0.04 | 0.16 | [-0.36,  0.28] |  -0.26
#> 1    | treatment | -0.39 | 0.18 | [-0.74, -0.04] |  -2.21
#> 2    | treatment |  0.18 | 0.18 | [-0.18,  0.53] |   0.98
#> 3    | treatment | -0.09 | 0.15 | [-0.39,  0.21] |  -0.61
#>
#> Variable predicted: score
#> Predictors modulated: time, grp
```

Does same level of time differ between groups?

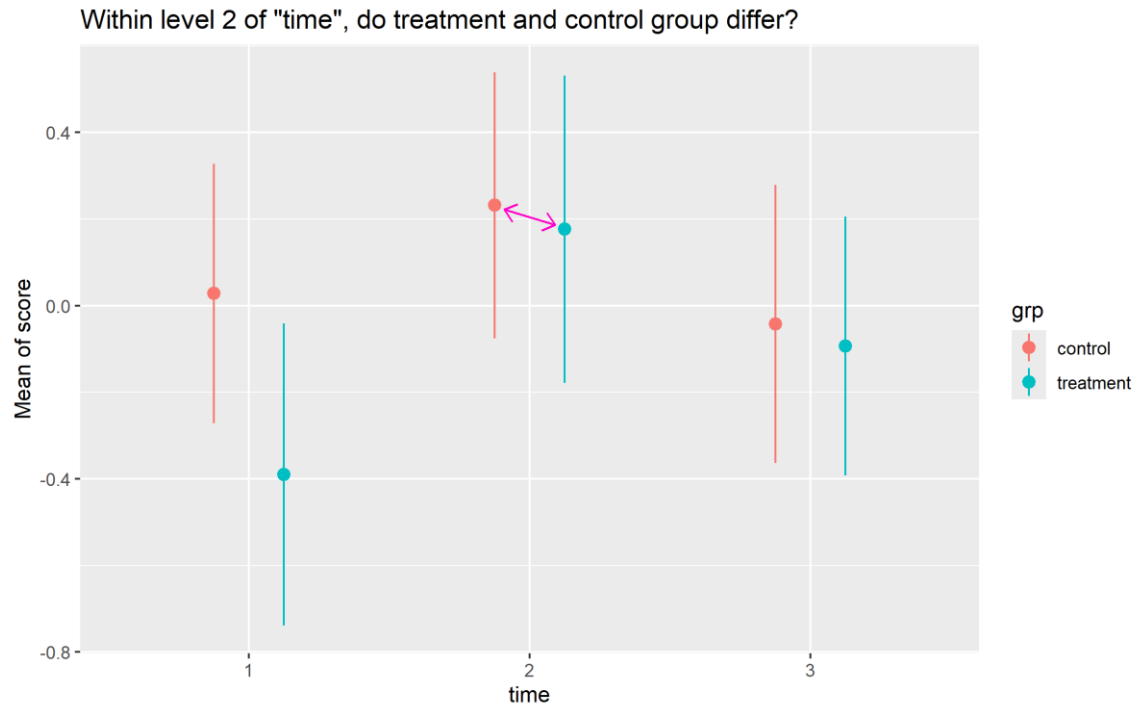
- First, we look at the marginal means of score for all combinations of the involved interaction term.

```
plot(predictions)
```



Does same level of time differ between groups?

- We could ask whether the predicted score for time = 2 is significantly different depending on the level of grp? In other words, do the groups treatment and control differ when time = 2?



Does same level of time differ between groups?

```
# we want to contrast group "2, control" and group "2, treatment"
estimate_contrasts(model2, c("time", "grp"))
#> Marginal Contrasts Analysis
#>
#> Level1       | Level2       | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------------------------
#> 1, treatment | 1, control   |      -0.42 | 0.23 | [-0.88, 0.04] |  -1.80 | 0.074
#> 2, control   | 1, control   |       0.20 | 0.22 | [-0.23, 0.63] |   0.94 | 0.350
#> 2, treatment | 1, control   |       0.15 | 0.24 | [-0.32, 0.61] |   0.63 | 0.529
#> 3, control   | 1, control   |      -0.07 | 0.22 | [-0.51, 0.37] |  -0.32 | 0.750
#> 3, treatment | 1, control   |      -0.12 | 0.21 | [-0.54, 0.30] |  -0.56 | 0.573
#> 2, control   | 1, treatment |       0.62 | 0.24 | [ 0.16, 1.09] |   2.64 | 0.009
#> 2, treatment | 1, treatment |       0.57 | 0.25 | [ 0.07, 1.06] |   2.25 | 0.026
#> 3, control   | 1, treatment |       0.35 | 0.24 | [-0.13, 0.82] |   1.45 | 0.150
#> 3, treatment | 1, treatment |       0.30 | 0.23 | [-0.16, 0.76] |   1.28 | 0.203
#> 2, treatment | 2, control   |      -0.06 | 0.24 | [-0.52, 0.41] |  -0.23 | 0.816
#> 3, control   | 2, control   |      -0.27 | 0.23 | [-0.72, 0.17] |  -1.22 | 0.225
#> 3, treatment | 2, control   |      -0.32 | 0.22 | [-0.75, 0.10] |  -1.49 | 0.137
#> 3, control   | 2, treatment |      -0.22 | 0.24 | [-0.70, 0.26] |  -0.90 | 0.368
#> 3, treatment | 2, treatment |      -0.27 | 0.24 | [-0.73, 0.19] |  -1.14 | 0.254
#> 3, treatment | 3, control   |      -0.05 | 0.22 | [-0.49, 0.39] |  -0.23 | 0.821
```

Does same level of time differ between groups?

- For our quantity of interest, the contrast between groups treatment and control when time = 2 is -0.06.
- We find this comparison in row 10 of the output.

```
#> Marginal Contrasts Analysis
#>
#> Level1       | Level2       | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------------------------
#> 1, treatment | 1, control   |      -0.42 | 0.23 | [-0.88, 0.04] |  -1.80 | 0.074
#> 2, control   | 1, control   |       0.20 | 0.22 | [-0.23, 0.63] |   0.94 | 0.350
#> 2, treatment | 1, control   |       0.15 | 0.24 | [-0.32, 0.61] |   0.63 | 0.529
#> 3, control   | 1, control   |      -0.07 | 0.22 | [-0.51, 0.37] |  -0.32 | 0.750
#> 3, treatment | 1, control   |      -0.12 | 0.21 | [-0.54, 0.30] |  -0.56 | 0.573
#> 2, control   | 1, treatment |       0.62 | 0.24 | [ 0.16, 1.09] |   2.64 | 0.009
#> 2, treatment | 1, treatment |       0.57 | 0.25 | [ 0.07, 1.06] |   2.25 | 0.026
#> 3, control   | 1, treatment |       0.35 | 0.24 | [-0.13, 0.82] |   1.45 | 0.150
#> 3, treatment | 1, treatment |       0.30 | 0.23 | [-0.16, 0.76] |   1.28 | 0.203
#> 2, treatment | 2, control   |      -0.06 | 0.24 | [-0.52, 0.41] |  -0.23 | 0.816
#> 3, control   | 2, control   |      -0.27 | 0.23 | [-0.72, 0.17] |  -1.22 | 0.225
#> 3, treatment | 2, control   |      -0.32 | 0.22 | [-0.75, 0.10] |  -1.49 | 0.137
#> 3, control   | 2, treatment |      -0.22 | 0.24 | [-0.70, 0.26] |  -0.90 | 0.368
```

Finding specific pairwise comparisons

- Instead of searching for the desired comparison in a long output, we can directly request it.
- We therefore must provide the row indices from the terms of interest, using "`b<rownumber>`" (from the marginal means table).

```
estimate_means(model2, c("time", "grp"))
#> Estimated Marginal Means
#>
#> time | grp       |  Mean |   SE |         95% CI | t(194)
#> ------------------------------------------------------------
#> 1    | control   |  0.03 | 0.15 | [-0.27,  0.33] |   0.18
#> 2    | control   |  0.23 | 0.16 | [-0.08,  0.54] |   1.49
#> 3    | control   | -0.04 | 0.16 | [-0.36,  0.28] |  -0.26
#> 1    | treatment | -0.39 | 0.18 | [-0.74, -0.04] |  -2.21
#> 2    | treatment |  0.18 | 0.18 | [-0.18,  0.53] |   0.98
#> 3    | treatment | -0.09 | 0.15 | [-0.39,  0.21] |  -0.61
#>
#> Variable predicted: score
#> Predictors modulated: time, grp
```

Finding specific pairwise comparisons

- In the output, each row is considered as one coefficient of interest.
- Our groups we want to include in our comparison are rows two (`grp=control` and `time=2`) and five (`grp=treatment` and `time=2`).
- Thus, our "quantities of interest" are `b2` and `b5`.

```
#> Estimated Marginal Means
#>
#> time | grp       |  Mean |   SE |         95% CI | t(194)
#> ------------------------------------------------------------
#> 1    | control   |  0.03 | 0.15 | [-0.27,  0.33] |   0.18
#> 2    | control   |  0.23 | 0.16 | [-0.08,  0.54] |   1.49
#> 3    | control   | -0.04 | 0.16 | [-0.36,  0.28] |  -0.26
#> 1    | treatment | -0.39 | 0.18 | [-0.74, -0.04] |  -2.21
#> 2    | treatment |  0.18 | 0.18 | [-0.18,  0.53] |   0.98
#> 3    | treatment | -0.09 | 0.15 | [-0.39,  0.21] |  -0.61
```

Testing specific pairwise comparisons

- Our null hypothesis we want to test is whether both predictions are equal, i.e., `comparison="b5=b2"`. This is how we calculate the desired comparison directly:

```
# compute specific contrast directly
estimate_contrasts(model2, c("time", "grp"), comparison = "b5=b2")
#> Marginal Contrasts Analysis
#>
#> Parameter | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------
#> b5=b2     |      -0.06 | 0.24 | [-0.52, 0.41] |  -0.23 | 0.816
#>
#> Variable predicted: score
#> Predictors contrasted: time, grp
#> p-values are uncorrected.
#> Parameters:
#> b5 = time [2], grp [treatment]
#> b2 = time [2], grp [control]
```

Testing specific pairwise comparisons

- A second option to test comparisons at specific values or levels is to make use of the `contrast` argument, which allows us to specify the levels of the focal terms.
- Although this option is simpler, using `comparison` provides more flexibility. Examples follow on the next slides.

```
# compute specific contrast directly
estimate_contrasts(model2, c("time=2", "grp"))
#> Marginal Contrasts Analysis
#>
#> Level1   | Level2    | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------------------
#> treatment | control |      -0.06 | 0.24 | [-0.52, 0.41] |  -0.23 | 0.816
#>
#> Variable predicted: score
#> Predictors contrasted: time=2, grp
```

Find out more about specifying representative values at https://easystats.github.io/insight/reference/get_datagrid.html

Testing specific pairwise comparisons

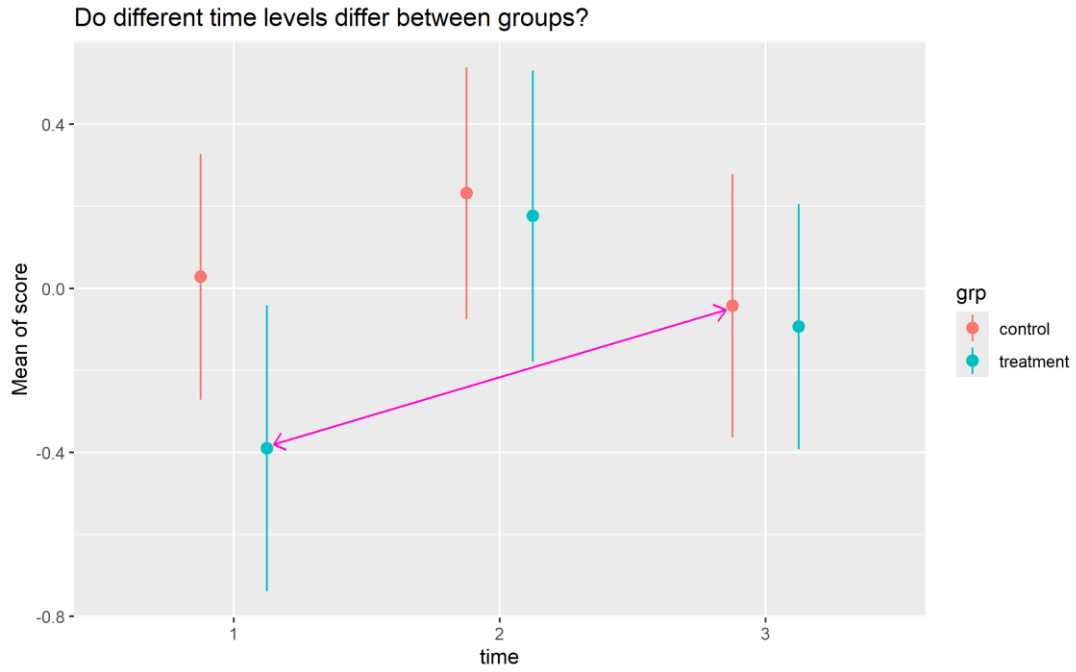
- A third option to test comparisons at specific values or levels is to make use of the `contrast` *and* `by` arguments, which also allows us to specify the levels of the focal terms.

```
# compute specific contrast directly
estimate_contrasts(model2, contrast = "grp", by = "time=2")
#> Marginal Contrasts Analysis
#>
#> Level1  | Level2    | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------------------
#> control | treatment |       0.06 | 0.24 | [-0.41, 0.52] |   0.23 | 0.816
#>
#> Variable predicted: score
#> Predictors contrasted: grp
```

Find out more about specifying representative values at https://easystats.github.io/insight/reference/get_datagrid.html

Do different time levels differ between groups?

- We can repeat the previous steps to test any combination of group levels for differences.
- For instance, is the predicted score for time = 1 in the treatment group significantly different from the predicted score for time = 3 in the control group?



Do different time levels differ between groups?

- The contrast we are interested in is between time = 1 in the treatment group and time = 3 in the control group.
- These are the predicted values in rows three and four, thus we test whether `comparison="b4=b3"`.

```
#> Estimated Marginal Means
#>
#> time | grp       |  Mean |   SE |         95% CI | t(194)
#> ------------------------------------------------------------
#> 1    | control   |  0.03 | 0.15 | [-0.27,  0.33] |   0.18
#> 2    | control   |  0.23 | 0.16 | [-0.08,  0.54] |   1.49
#> 3    | control   | -0.04 | 0.16 | [-0.36,  0.28] |  -0.26
#> 1    | treatment | -0.39 | 0.18 | [-0.74, -0.04] |  -2.21
#> 2    | treatment |  0.18 | 0.18 | [-0.18,  0.53] |   0.98
#> 3    | treatment | -0.09 | 0.15 | [-0.39,  0.21] |  -0.61
```

Do different time levels differ between groups?

- We test whether `comparison="b4=b3"`.

```
# pairwise comparison
estimate_contrasts(model2, c("time", "grp"), comparison = "b4 = b3")
#> Marginal Contrasts Analysis
#>
#> Parameter | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------------
#> b4=b3     |      -0.35 | 0.24 | [-0.82, 0.13] |  -1.45 | 0.150
#>
#> Variable predicted: score
#> Predictors contrasted: time, grp
#> p-values are uncorrected.
#> Parameters:
#> b4 = time [1], grp [treatment]
#> b3 = time [3], grp [control]
```

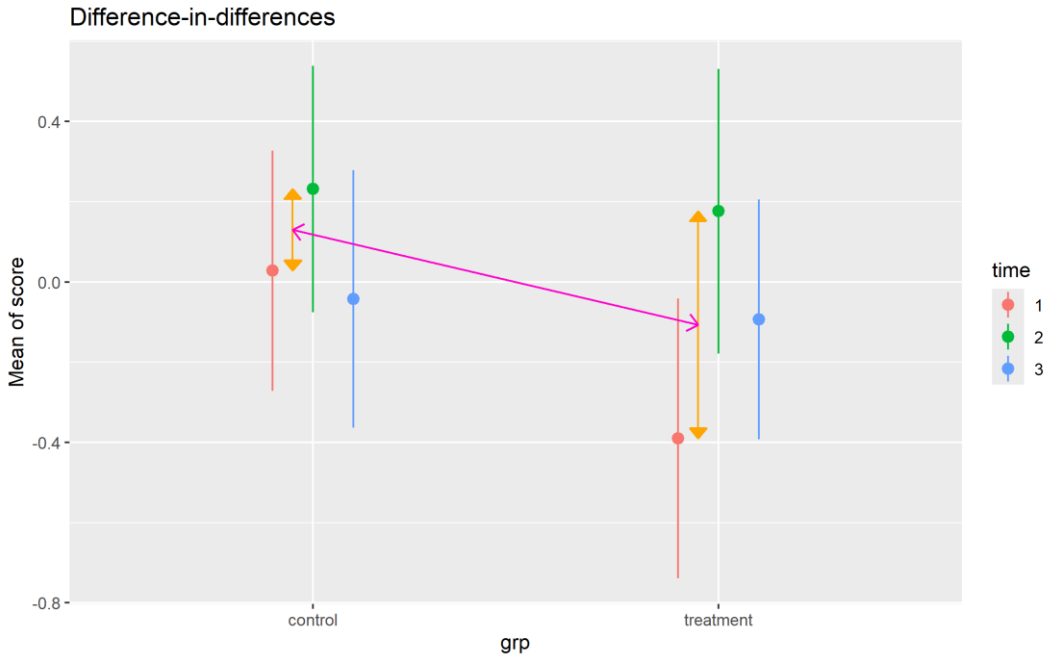
Do different time levels differ *within* groups?

- If we want pairwise comparisons of two levels of one focal term *within* the levels of another focal term, we can simply use the `by` argument to group pairwise comparisons and make the output more readable.

```
# pairwise comparison of "grp" within levels of "time"
estimate_contrasts(model2, "grp", by = "time")
#> Marginal Contrasts Analysis
#>
#> Level1    | Level2  | time | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------------------------
#> treatment | control | 1    |      -0.42 | 0.23 | [-0.88, 0.04] |  -1.80 | 0.074
#> treatment | control | 2    |      -0.06 | 0.24 | [-0.52, 0.41] |  -0.23 | 0.816
#> treatment | control | 3    |      -0.05 | 0.22 | [-0.49, 0.39] |  -0.23 | 0.821
#>
#> Variable predicted: score
#> Predictors contrasted: grp
#> p-values are uncorrected.
```

Does difference between two levels of time in the control group differ from difference of same two levels in the treatment group?

- The `comparison` argument also allows us to compare *difference-in-differences* (also called *interaction contrasts*).
- For example, is the difference between two time levels in one group significantly different from the difference of the same two time levels in the other group?



Does difference between two levels of time in the control group differ from difference of same two levels in the treatment group?

- The difference of time levels 1 and 2 in the control group refers to rows one and two (`b1` and `b2`).
- The difference for the same time levels in the treatment group refers to the difference between rows four and five (`b4` and `b5`).
- Test that these are equal: `comparison="(b1-b2)=(b4-b5)"`.

```
estimate_means(model2, c("time", "grp"))
#> Estimated Marginal Means
#>
#> time | grp       |  Mean |   SE |         95% CI | t(194)
#> ------------------------------------------------------------
#> 1    | control   |  0.03 | 0.15 | [-0.27,  0.33] |   0.18
#> 2    | control   |  0.23 | 0.16 | [-0.08,  0.54] |   1.49
#> 3    | control   | -0.04 | 0.16 | [-0.36,  0.28] |  -0.26
#> 1    | treatment | -0.39 | 0.18 | [-0.74, -0.04] |  -2.21
#> 2    | treatment |  0.18 | 0.18 | [-0.18,  0.53] |   0.98
#> 3    | treatment | -0.09 | 0.15 | [-0.39,  0.21] |  -0.61
```

Does difference between two levels of time in the control group differ from difference of same two levels in the treatment group?

- Test that these are equal: `comparison="(b1-b2)=(b4-b5)"`.

```
estimate_contrasts(model2, c("time", "grp"), comparison = "(b1-b2)=(b4-b5)")
#> Marginal Contrasts Analysis
#>
#> Parameter       | Difference |   SE |        95% CI | t(194) |     p
#> ------------------------------------------------------------------------
#> (b1-b2)=(b4-b5) |       0.36 | 0.33 | [-0.29, 1.02] |   1.09 | 0.277
#>
#> Variable predicted: score
#> Predictors contrasted: time, grp
#> p-values are uncorrected.
#> Parameters:
#> b1 = time [1], grp [control]
#> b2 = time [2], grp [control]
#> b4 = time [1], grp [treatment]
#> b5 = time [2], grp [treatment]
```

Does difference between two levels of time in the control group differ from difference of same two levels in the treatment group?

Let's replicate this step-by-step:

1. Predicted value of score for time = 1 in the control group is 0.03.
2. Predicted value of score for time = 2 in the control group is 0.23.
3. The first difference is -0.20.
4. Predicted value of score for time = 1 in the treatment group is -0.39.
5. Predicted value of score for time = 2 in the treatment group is 0.18.
6. The second difference is -0.57.
7. Our quantity of interest is the difference between these two differences, which is 0.36 *(inaccuracy due to rounding)*. This difference is not statistically significant ($p = 0.277$).