How to get away (legally) with planned comparisons

Umberto Granziol

Department of General Psychology, University of Padova

Psicostat, 14 April 2023

Stories of a life

- Have you ever felt that there is something different from post hoc comparisons?
- Do you ever feel the need to escape from dozens of multiple comparisons?
- If yes, you are not alone!

The framework

- Usually, students and researchers have specific hypotheses on how their results should be.
- To test such hypotheses, sometimes they use (non)linear models (e.g., correlations, regression and structural equation models).
- The interpretation of the (non)statistically significant results (long live the *p* value) usually follows two approaches:

1. The **post-hoc** comparisons approach.
2. The **planned** comparisons approach.

Post-hoc: Pros and cons

- Pros: Such an approach is useful in case of preliminary studies, namely no idea of what to expect.
- Cons 1: Increasing the number comparisons increases the chance of Type I Error.
- Cons 2: Assume you have to test 20 post-hoc comparisons. The chance of getting at least one statistically significant comparison is not negligible. So, you need to adjust the p values to correct such chance. It is not unusual that, as a consequence, the effect that was your target disappears.

Planned comparisons: Pros and cons

- Pros 1: No need to compute dozens of multiple comparisons.
- Pros 2: Reduced chance of getting biased results.
- Cons 1: Expertise is required to code them. They are usually considered as difficult.
- Cons 2: Reviewers usually find in post-hoc comparisons their personal safe-zone.
- Cons 3: Sometimes, we can get confused by our statistical software (in the way we use it).

Short sad stories of sad reviewers

- Comment 1: "Planned comparisons is a niche topic".
- Comment 2: "You used planned comparisons, but I still wonder what if you compare X with Y".
- Comment 3: "Making a paper on the use of planned comparisons is useless..it is an old topic and EVERYBODY uses them".
- Comment 4: "Did you use ANOVA to get the effects?"

When the software is misunderstood

- Suppose you are working on people with a diagnosis of eating disorder and you want to understand if providing a psychotherapy is better than not to apply it.
- Suppose that, in your study, your predictor (e.g. `Treat`) has three conditions: Control, Family Therapy (FT) and Cognitive-Behavioral Therapy (CBT).
- Your Dependent variable is the weight after six month of research.
- Finally, assume that you need to test, in the same model, the hypothesis according to which there is no difference between both therapeutic approaches.

When the software is misunderstood

```
Call:
lm(formula = Postwt ~ Treat, data = MASS::anorexia)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.2941  -3.7299  -0.0021   4.7809  17.9034 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   85.697      1.353  63.321   <2e-16 ***
TreatCont     -4.589      1.968  -2.331   0.0227 *  
TreatFT        4.798      2.226   2.155   0.0347 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.288 on 69 degrees of freedom
Multiple R-squared:  0.2005,	Adjusted R-squared:  0.1773 
F-statistic: 8.651 on 2 and 69 DF,  p-value: 0.0004443
```

```
> pairs(posthoc)
 contrast   estimate   SE df t.ratio p.value
 CBT - Cont     4.59 1.97 69   2.331  0.0581
 CBT - FT      -4.80 2.23 69  -2.155  0.0864
 Cont - FT     -9.39 2.27 69  -4.129  0.0003

P value adjustment: tukey method for comparing a family of 3 estimates
```

Pay attention, to the interpretation!

- `(Intercept) = 85.697` is the average weight when the predictor is set to zero (aka the intercept). Indeed, such an intercept estimates refers to the CBT experimental group, that has been chosen as a baseline.
- `(TreatCont) = -4.589` is the estimated average difference between the control group and the CBT experimental group.
- `(TreatFT) = 4.798` is the estimated average difference between the FT experimental group and the CBT experimental group.
- Problem: it is not what we wanted!

What you see (in your mind) *vs* What your software sees

$$H_0 : \begin{cases} \mu_{intercept} = 0 \\ \frac{1}{2}(\mu_{FT} + \mu_{CBT}) - \mu_{Cont} = 0 \\ \mu_{FT} - \mu_{CBT} = 0 \end{cases}$$

The Contrast Matrix

- Assume a linear model $Y = X\hat{\beta} + \hat{\epsilon}$
- where Y is the response variable
- $\hat{\beta}$ is the set of expected regression coefficients,
- $\hat{\epsilon}$ is the expected error.
- X is the design matrix of the model.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \end{bmatrix}$$

The Contrast Matrix

- The design matrix can be simplified into the Contrast matrix C.
- The first column encodes the intercept.
- The second column encodes the first contrast (or comparison).
- The third column encodes the second comparison.

$$C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Important notes

- Formally, contrasts are weighted linear functions (usually of means; Baguley2012).
- Contrasts help to code and quantify the performance or the comparisons among a set of means. In other words, contrasts make it possible to collapse several comparisons among the levels of a categorical variable into a unique function (as if they were a single effect).
- Given n levels of a variable, it is possible to test (a priori) $n - 1$ contrasts, since each contrast consumes 1 degree of freedom.
- The Contrasts matrix is the result of a series of matrix operation, starting from the so called hypothesis matrix H.

Baguley, T. (2012). Contrasts. In Serious stats: A guide to advanced statistics for the behavioral sciences. Macmillan International Higher Education.

The hypothesis matrix

- An hypothesis matrix contains the hypothesis in rows and the levels of a variable in columns.
- Considering the treatment contrasts (i.e., the example at hand) the related hypothesis matrix will be:

$$H = \begin{bmatrix} CBT & Cont & FT \\ 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

From hypothesis to contrast matrix

- It is possible to compute a contrast matrix starting from its hypothesis one, by using an extract of a formula that we know very well:

$$(H'H)^{-1}H'$$

- Do you remember it? $\hat{\beta} = (H'H)^{-1}H'y$

Treatment contrasts

- In case of treatment contrasts, the aim is to compare the $n - 1$ levels to a level chosen as the reference level (or baseline).

$$H = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- In R, the function to compute treatment contrasts is `contr.treatment()`, setting as argument the number of levels belonging to the variable of interest.

```
> contr.treatment(3)
  2 3
1 0 0
2 1 0
3 0 1
```

In case you need a proof

- By applying $C = (H'H)^{-1}H'$

$$C_{treatment} = \left(\begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 3 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Sum contrasts

- In sum contrasts, the aim is to compare the $n-1$ levels with the grand mean of those levels.
- In this case, one level needs to be selected to represent the grand mean. If $n=3$ the contrast matrix will be:

$$H = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

- In R, the function to compute sum contrasts is `contr.sum()`.

```
>contr.sum(3)
  [,1] [,2]
1    1    0
2    0    1
3   -1   -1
```

Repeated contrasts

- Repeated contrasts (or successive difference contrasts) allow comparing each level's mean with its subsequent one.
- For instance, if $n = 3$, the contrasts will be:

$$H = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} -0.67 & -0.33 \\ 0.33 & -0.33 \\ 0.33 & 0.67 \end{bmatrix}$$

- In R, the function to compute sum contrasts is `contr.sdif()`, from the MASS package in R.

```
> MASS::contr.sdif(3)
         2-1        3-2
1 -0.6666667 -0.3333333
2  0.3333333 -0.3333333
3  0.3333333  0.6666667
```

Centered contrasts

- There is a difference among Sum/Repeated and treatment contrasts: the first ones are centered.
- Taken a set C of contrasts, they are centered if their weights sum to zero.
- In other words, the column sum of each contrast matrix is equal to zero.
- Centering is an advantageous choices since enhances the orthogonality of contrasts.

Polynomial contrasts

- Polynomial contrasts allows testing for possible trends of the predictor's levels (i.e., linear, quadratic, cubic etc.).
- If $n = 3$, the contrast will test the presence of both linear (i.e., the first contrast) and quadratic (i.e., the second contrast) effects:

$$H_{G_{pol}} = \begin{bmatrix} 0.333 & 0.333 & 0.333 \\ -0.707 & 0 & 0.707 \\ 0.408 & -0.816 & 0.408 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & -0.71 & 0.41 \\ 1 & 0 & -0.82 \\ 1 & 0.71 & 0.41 \end{bmatrix}$$

- In R, the function to compute polynomial contrasts is `contr.poly()`.

```
> contr.poly(3)
                .L         .Q
[1,] -7.071068e-01  0.4082483
[2,] -7.850462e-17 -0.8164966
[3,]  7.071068e-01  0.4082483
```

Helmert contrasts

- Helmert contrasts are very useful for variables with $n > 2$.
- If $n = 3$, the first contrast will encode the difference between the first two levels. The second contrast will encode the difference between the third level and the average of the first two conditions. In our example:

$$H = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/2 & 1/2 & 0 \\ -1/6 & -1/6 & 1/3 \end{bmatrix} \qquad C = \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix}$$

- In R, the function to compute helmert contrasts is `contr.helmert()`, setting as argument the number of levels of the variable of interest. In our example:

```
> contr.helmert(3)
  [,1] [,2]
1   -1   -1
2    1   -1
3    0    2
```

Reverse Helmert contrast

- Reverse Helmert contrast follow the same rules of the Helmert ones. The difference is that the contrast weights are reversed.
- In other words, the first contrast tests the difference between the first level and the average of the following two levels. The second contrast tests the difference between the last two levels.

$$H = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & -1/6 & -1/6 \\ 0 & 1/2 & 1/2 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

- In R, a function computing the reverse helmert contrast still does not exist. This problem can be overcome by creating a matrix with the contrasts' weights and assign it to the `contrasts()` function of a variable.

Orthogonal contrasts

- It can be noted that all the contrasts but treatment ones are centered.
- Helmert (and Reverse Helmert) contrast have another characteristic that make them more preferable when testing specific hypothesis: orthogonality.
- Taken two (or more) centered contrasts, they are orthogonal if they are uncorrelated.
- Orthogonality is useful when testing specific hypothesis: since such contrasts uncorrelated, each estimated effect less dependent on the others.
- As a consequence, the interpretation more straightforward.

Interaction: continuous by categorical

- Let'use the following design matrix:

$$X = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 & 0 \\ 1 & 5 & 0 & 0 & 0 & 0 \\ 1 & 7 & 1 & 0 & 7 & 0 \\ 1 & 8 & 1 & 0 & 8 & 0 \\ 1 & 10 & 1 & 0 & 10 & 0 \\ 1 & 2 & 0 & 1 & 0 & 2 \\ 1 & 5 & 0 & 1 & 0 & 5 \\ 1 & 3 & 0 & 1 & 0 & 3 \end{bmatrix}$$

Interaction: continuous by categorical

- The interaction is simply the product between the columns!
- The fifth column is the product between the second and the third columns and represent the difference between the two groups (A vs B) when the continuous variable increases.
- The last column is the product between the second and the fourth columns and represent the difference between the two groups (A vs C) when the continuous variable increases.

Interaction: categorical by categorical

- In case of an interaction between two (or more) categorical predictors, the design matrix follows the same rules.
- For instance, assume a contrast matrix displaying the interaction among two predictors G (three levels: g_1, g_2 and g_3) and A (two levels: a_1 and a_2), coded respectively with helmert and sum contrasts, would be:

$$\begin{bmatrix} g_1 a_1 & -1 & -1 & 1 & -1 & -1 \\ g_1 a_2 & -1 & -1 & -1 & 1 & 1 \\ g_2 a_1 & 1 & -1 & 1 & 1 & -1 \\ g_2 a_2 & 1 & -1 & -1 & -1 & 1 \\ g_3 a_1 & 0 & 2 & 1 & 0 & 2 \\ g_3 a_2 & 0 & 2 & -1 & 0 & -2 \end{bmatrix}$$

Interaction: categorical by categorical

- The first two columns encode the contrasts related to the main effect of the variable G.
- The third column encodes the contrasts related to the main effect of variable A.
- The last two columns encode interaction effect:
- The fourth column encodes the first contrast of the G variable across the levels of the variable A. It is obtained by multiplying the first and the third columns of such contrast matrix.
- The same applies for the fifth column.

Customized contrasts

- Orthogonality is a necessary property that is not sufficient to define a set of contrasts.
- The goodness of a set of contrasts is mainly related to the hypothesis they test.
- Sometimes, the family of defined contrasts might not fit a researcher's questions and may need customization.
- In the case of single variables, it is possible to customize contrasts by creating bundles of means or by extracting the contrast weights directly from the raw means. Let's focus on the first modality.

Customized contrasts

- For example, consider the variable C with five levels (that is, $\{c_0, c_1, c_2, c_3, c_4\}$).
- Assume that the user wants to set the following two comparisons: condition c_0 against (the average of) all the others taken together (i.e., $\{c_0\}$ *vs* $\{c_1, c_2, c_3, c_4\}$);
- Excluding c_0, the (average of) former two conditions with the (average of) latter two (i.e., $\{c_1, c_2\}$ *vs* $\{c_3, c_4\}$).

$$\text{Contrast 1} : c_0 \sim (c_1 + c_2 + c_3 + c_4)/4,$$
$$\text{Contrast 2} : (c_1 + c_2)/2 \sim (c_3 + c_4)/2.$$

Customized contrasts

- There are several ways to code customized contrasts.
- An easy (and lazy) one is by using *appRiori*.
- It is a Shiny app that is programmed through a plug-and-play logic.

appRiori

This Shiny App (it is available on CRAN!) helps users to:

- Understand and learn the logic of planned contrasts.
- Directly plan specific contrasts from an uploaded database: The app can handle all the default contrasts provided by R and allows a researcher to set customized contrasts.
- Set contrasts not only for simple effects but also for two-way and three-way interactions.
- Obtain the corresponding ready-to-use R code to be applied directly before running the analysis: The generated contrasts can be used for analyses of variance as well as linear or generalized (mixed-effects) models, etc.

SHOWTIME!

Take home message and future directions

Conclusions

- In the end, it is possible to plan all the desired comparisons and directly see the estimate from the `summary()` of your model.
- Planned comparisons are not so difficult (even if not so easy).
- If you have problems, use *appRiori*.

Future directions

- It would be great to conduct systematic reviews to understand if it is true that all researchers use planned comparisons in presence of non-exploratory studies
- There are several way to define models providing customized contrasts for interactions.

"So it's time to leave you a preview,
so you too can review what we do.."
(Harder than you think, Publick Enemy; 2007)

Thank you for youR attention!

Some references

- Agathokleous, E., & Yu, L. (2022). Six statistical issues in scientific writing that might lead to rejection of a manuscript. Springer.
- Baguley, T. (2012). Contrasts. In Serious stats: A guide to advanced statistics for the behavioral sciences. Macmillan International Higher Education.
- Buckless, F. A., & Ravenscroft, S. P. (1990). Contrast coding: A refinement of anova in behavioral analysis. Accounting Review, 933–945.
- Chatham, K. (1999). Planned contrasts: An overview of comparison methods. Retrieved from ERIC database.
- Davis, M. J. (2010). Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. Journal of data science, 8 (1),61–73.
- Kuehne, C. C. (1993). The advantages of using planned comparisons over post hoc tests. Retrieved from ERIC database.
- Kwon, M. (1996). The use of planned comparisons in analysis of variance research. Retrieved from ERIC database.
- Rabe, M. M., Vasishth, S., Hohenstein, S., Kliegl, R., & Schad, D. J. (2020). hypr: An r package for hypothesis-driven contrast coding. Journal of Open Source Software, 5 (48), 2134.
- Ruxton, G. D., & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. Behavioral ecology, 19 (3), 690–693.
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. Journal of Memory and Language, 110 , 104038.
- Thompson, B. (1990). Planned versus unplanned and orthogonal versus nonorthogonal contrasts: The neo-classical perspective. Retrieved from ERIC database.
- https://cran.r-project.org/web/packages/appRiori/index.html
-