

contrast coding

sum contrasts (deviation coding)

別名

- sum coding
- deviation coding
- effect coding

options()

- R の各種オプションを設定するコマンド
- contrasts の設定がどうなっているかを確認する
- contrasts のオプションは、unordered と ordered の二種類ある
- このオプションを変えてしまうこともできるが、変えてしまうと、その後もその設定がデフォルトになってしまうので、普通のデフォルトにまた戻さないと、困ったことになる。

```
contrasts(flevels) <- NULL
```

- モデル式のオプションとして、指定する場合は、その場合だけそのオプションになるので、安全

contr.sum

- 以下の「Understanding Sum Contrasts for Regression Models: A Demonstration」の例

```
lm(RT Gender*Strategy, data = df, contrasts = list(Gender = "contr.sum", Strategy = "contr.sum"))
```

- 多重共線性を避けるため、原則、最後のレベルを投入しない。省略する。
 - omit オプションを使うと省略するレベルを明示的に指定できる。

```
omit= "除くもの"
```

具体的手順

1. コントラストがどうなっているかの確認
 - 一番最初の P1 が全部 0 になっている。
2. レベルの表示
3. このレベルに対して、contr.sum() をかける
 - 一番最後の P8 がすべて -1 になっている。
 - この状態で、各カラムは P1 から P7 まで 1 なので、P8 と合わせたら、 ± 0 となる。
4. もとものの contrasts のベクトルを sum contrasts のベクトルに置き換える
5. 置き換わったかどうか確認

tab_model(モデル , show.stat=T)

- Predictors で、各要因の後ろに、レベル名でなく数字が表示される

Gender1 5 (Estimate)

- この 1 はその要因（性別）の複数のレベルの最初のもの（女性）が、参照レベル（sum coding の場合は平均）からどれだけ高いかを示している。
- もう一つのレベル（男性）は、逆にその分だけ参照レベル（平均）から低い数値となる

男女の平均が 50 とすると、女は 55、男は 45 という状況

分析結果解釈上の注意

- 各変数に数字がついたものは
 - 各変数内のレベルが初めのものから順に 1 からつく

レベルが二つの場合

- 変数 1 のみ表示される
- Estimate は、その変数のレベルと「平均」との差
- もう一方のレベルは、平均からその分を引けばよいことになる。
 - 平均が 0 になるようにするのが sum coding なので、二つのレベルは、平均を挟んで、逆方向の等距離にある

レベルが三つの場合

- 変数 1 と変数 2 が表示される。（変数 3 は表示されない）
 - それぞれのレベルと「平均」との差がそれぞれの Estimate
- 変数 3 の Estimate は自分で計算する
 - 三つのレベルの平均が 0 になるようにしてあるので
 - 変数 1 と変数 2 の数字を平均から「引く」と求められる

平均が 100 で、
 変数 1 の Estimate が 10 なら、変数 1 の値は、110
 変数 2 の Estimate が 5 なら、変数 2 の値は、105
 三つの平均が 0（具体的には 100）になるようにするには、
 変数 3 は、85（Estimate は -15）となる。
 $(110 + 105 + 85) / 3 = 100$
 $10 + 5 + (-15) = 0$

交互作用

- 変数一つ目 1: 変数二つ目 1
 - 変数一つ目の最初の水準と変数二つ目の最初の水準を組み合わせた場合、と読む
 - この組み合わせが、有意であるとは、

References

- [Understanding Sum Contrasts for Regression Models: A Demonstration](https://www.rcps.jp/doku.php?id=%E3%83%A1%E3%83%A2:%E3%83%9E%E3%83%AB%E3%83%81%E3%83%AC%E3%83%99%E3%83%AB%E3%83%A2%E3%83%87%E3%83%AB%E4%B8%80%E8%88%AC%E7%B7%9A%E5%BD%A2%E3%83%A2%E3%83%87%E3%83%AB)
- <https://www.rcps.jp/doku.php?id=%E3%83%A1%E3%83%A2:%E3%83%9E%E3%83%AB%E3%83%81%E3%83%AC%E3%83%99%E3%83%AB%E3%83%A2%E3%83%87%E3%83%AB%E4%B8%80%E8%88%AC%E7%B7%9A%E5%BD%A2%E3%83%A2%E3%83%87%E3%83%AB>
- <https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/#DEVIATION>

- <https://cran.r-project.org/web/packages/faux/vignettes/contrasts.html>

Brehm and Alday (2022) Contrast coding in a decade of mixed models

Journal of Memory and Language 125: 1-13

<https://osf.io/jkpxt/>

AMLaP 2020

3 レベル以上の sum coding の場合

- 切片は、grand mean (変数のすべてのレベルの平均)
- 各固定効果は、そのレベルが grand mean と比べ有意に差があるかを示す
- 複数のレベルのうち一つは reference level として表示されない
- grand mean に近いレベルを reference level とする
 - そのレベルは grand mean に一番近いので、差があるかどうかという観点からすれば、一番重要度が低いものとなる (一番差がない)
- こうすることで、個々のレベルが全体の平均からどれだけ離れているかが分かる

指針

- カテゴリー変数を使う場合は、どのコントラストを使ったかを明示的に書くこと。
- デフォルトの treatment coding の場合もその旨書くこと。
 - 例: "Factor A (magenta, green) was treatment coded or the three levels of Factor B, coffee, tea, and cocoa, were coded with two contrasts: (.25,.25, -.5) and (.5, -.5.0)"
- そのうえで、何と何を対比して比較したかをわかりやすく言い換えて説明すること。
 - 例 1: "The model intercept therefore reflects the reference level of factor A, magenta"
 - 例 2: "The first contrast tests caffeinated versus non-caffienated beverages, and the second tests coffee versus tea"
- コントラストが適切に行われていれば、事後テストをしなくてもよい。
 - 必要な場合のみ、emmeans で行えばよい。
 - その場合はその旨明記しておく。例: An additional set of pairwise comparisons was performed to directly compare tea versus cocoa using the R package emmeans.

わかりやすい解説

<https://learnb4ss.github.io/learnB4SS/articles/contrasts.html#sum-contrasts>