

contrast coding

- ・ カテゴリカル変数をどう対比してコード化するか
- ・ デフォルトで treatment coding になっている
- ・ 明示的にコーディングのことを説明する必要がある
- ・ sum coding のことを、contrast coding と呼ぶ人がいるが、紛らわしいのでやめるべきだと Brehm & Alday は言っている。

二種類のコード化

treatment coding (treatment contrast)

- ・ デフォルトはこちらになっている
 - ・ これを dummy coding と呼ぶこともある
- ・ 参照レベルに 0 を、もう一方に 1 をわりあてる
- ・ 切片は、すべて 0 の状態の値
- ・ 二要因ある場合、効果は、一方が 0 の時の効果を示す
 - ・ 特定のレベル (0) の時の効果なので「」に相当する

sum coding (sum contrasts)

- ・ `contr.Sum()` で設定する
- ・ 一方を -1 に、もう一方を 1 にする
 - ・ 足すと 0 になる。0 が平均となり、それが参照レベルとなる
- ・ 「参照レベル」は、各要因の平均 (なので、いわゆる「」となる)
 - ・ reference level は 0 となるが、0 のカテゴリーがあるわけではなく、平均 0 が reference level となる。
- ・ 切片は、総平均 (grand mean)
- ・ 分析結果表で、変数名に 1 がついているのは、を意味する。
 - ・ 切片に、その推測値を足した値がその変数全体の平均値を意味する。
 - ・ 男女という性別の場合、男の平均値と女の平均値の真ん中の値のこととなる。

注意すべき点

主効果を見たいのに単純主効果を調べて、それを主効果と誤解する点

- ・ デフォルトでは、単純主効果を調べるようになっている
 - ・ アルファベットの早い方を「0」として「参照レベル」(比較の基準)にしている
- ・ 単純主効果では有意でも、主効果は有意ではないことがある。
 - ・ 主効果としては有意ではないのに、単純主効果が有意なのを、主効果が有意だと誤解してしまう恐れがある (偽陽性 Type I error)

単純主効果だとわかってやっていれば、ある意味便利

- ・ 従来の ANOVA だと、主効果の交互作用があった時に、さらに追加で単純主効果を調べるということをしていた
 - ・ 最初から単純主効果が出るので、追加でなくてよい
 - ・ 今で言えば、あとから emmeans とかでペアごとの差を検定しなくてよい

- ・ しかし、逆に、主効果じたいは分かっていない状態

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

- ・ フォークでスープを食べるには時間がかかる
- ・ スプーンでスパゲッティを食べるのにも時間がかかる
- ・ 食べる時間の速さを一般的に考えるときには、ナイフとフォークとスプーンを使った全体的の話でしょ、ということ
- ・ フォークでスープを食べるのに時間がかかるからと言って、食器を使うと食べるのが遅くなる、と判断してしまうのは変でしょ、ということ（部分から全体に及ぼす誤り）

contrasts() でコントラストがどうなっているか表示

- ・ デフォルトは、treatment coding で、0, 1
- ・ アルファベットの若いレベルが 0 で、reference となる
- ・ intercept は、ゼロのほうにセットされる。
 - ・ この場合は、Fork で Salad の場合

参照レベルの変更

relevel(変数 , ref = " 新しい参照レベル ")

- ・ ただし、2 が一番上になったことからわかるように、2 を先頭に持ってくるので、結果のグラフを描く時に、1 , 2 , 3 ではなく、2 , 1 , 3 という順になってしまう。
- ・ そもそも、順序の関係ないカテゴリーであれば問題ない。

基準とするカテゴリーの変更 base= 番号

- ・ 順序は、1 , 2 , 3 で、2 が基準 0 になっている。

References

<https://people.linguistics.mcgill.ca/~morgan/qmld-book/practical-regression-topics-1-multi-level-factors-contrast-coding-interactions.html#contrast-coding>