Schad et al (2020)に学ぶ名義尺度のコーディング

Schad et al (2020)日本語注解

小川雅貴

東京大学

2021/12/17 (updated: 2021-12-15)

このHTMLスライドをご覧になる際は，Chromeまたは Firefoxをお使いになることをお勧めいたします。また，HTMLスライドの中身が正しく読み込まれない場合には，ブラウザの更新ボタンを押して，再度スライドを読み込んでください。なお，本スライドのPDF版は こちら にございます。

必要なパッケージ・関数の読み込み

```r
library(tidyverse)
```

次のパッケージ群を呼び出す

- **ggplot2**: データ可視化・図生成
- **dplyr**: データ操作
- **tidyr**: データ整然化
- **readr**: データ読み込み
- **purrr**: 関数型プログラミング
- **tibble**: 現代的データフレームtibbles
- **stringr**: 文字列処理
- **forcats**: カテゴリ変数の操作

```r
library(tidyverse)
library(broom)
```

統計モデルの解析結果をtibbleで表示

```r
library(tidyverse)
library(broom)
library(lme4)
```

一般化線形混合モデルの構築

```r
library(tidyverse)
library(broom)
library(lme4)
library(MASS)
```

対比行列の構築

```r
library(tidyverse)
library(broom)
library(lme4)
library(MASS)
library(hypr)
```

「仮説」行列の構築

```r
library(tidyverse)
library(broom)
library(lme4)
library(MASS)
library(hypr)
library(knitr)
```

Rでの動的文書生成

```r
library(tidyverse)
library(broom)
library(lme4)
library(MASS)
library(hypr)
library(knitr)

source(
  "http://read.psych.uni-potsdam.de/attach
  local = knitr::knit_global()
)
```

mixedDesign()の読み込み

Rmarkdownをknitする際には，**source()**の項で**local = knitr::knit_global()**を指定する
https://bookdown.org/yihui/rmarkdown-cookbook/source-script.html

```
library(tidyverse)
library(broom)
library(lme4)
library(MASS)
library(hypr)
library(knitr)

source(
  "http://read.psych.uni-potsdam.de/attach
  local = knitr::knit_global()
)

#source("http://read.psych.uni-potsdam.de/
```

Schad et al (2020)の元コードはこちら

```r
library(tidyverse)
library(broom)
library(lme4)
library(MASS)
library(hypr)
library(knitr)

source(
  "http://read.psych.uni-potsdam.de/attach
  local = knitr::knit_global()
)

#source("http://read.psych.uni-potsdam.de/

set.seed(1212)
```

乱数の固定

```r
library(tidyverse)
library(broom)
library(lme4)
library(MASS)
library(hypr)
library(knitr)

source(
  "http://read.psych.uni-potsdam.de/attach
  local = knitr::knit_global()
)

#source("http://read.psych.uni-potsdam.de/

set.seed(1212)

options(
  scipen = 999,
  digits = 2
  )
```

指数表記の回避・有効数字2桁に設定

Conceptual explanation of default contrasts (p.3)

以下，各対比の和訳はR commanderにおける和訳に基づく

Treatment contrast 処理対比

- 実験的介入treatmentの効果を，非介入群との比較で検討
  - 実験群 vs 統制群
- 各実験群を同一の統制群と比べる

Treatment contrast 処理対比

- 実験的介入treatmentの効果を，非介入群との比較で検討
  - 実験群 vs 統制群
- 各実験群を同一の統制群と比べる
- 例：様々なプライミング効果を，同一の統制群と比べる

  1. 綴字プライミング
  2. 音韻的プライミング
  3. プライミングなし（統制群）

Treatment contrast 処理対比

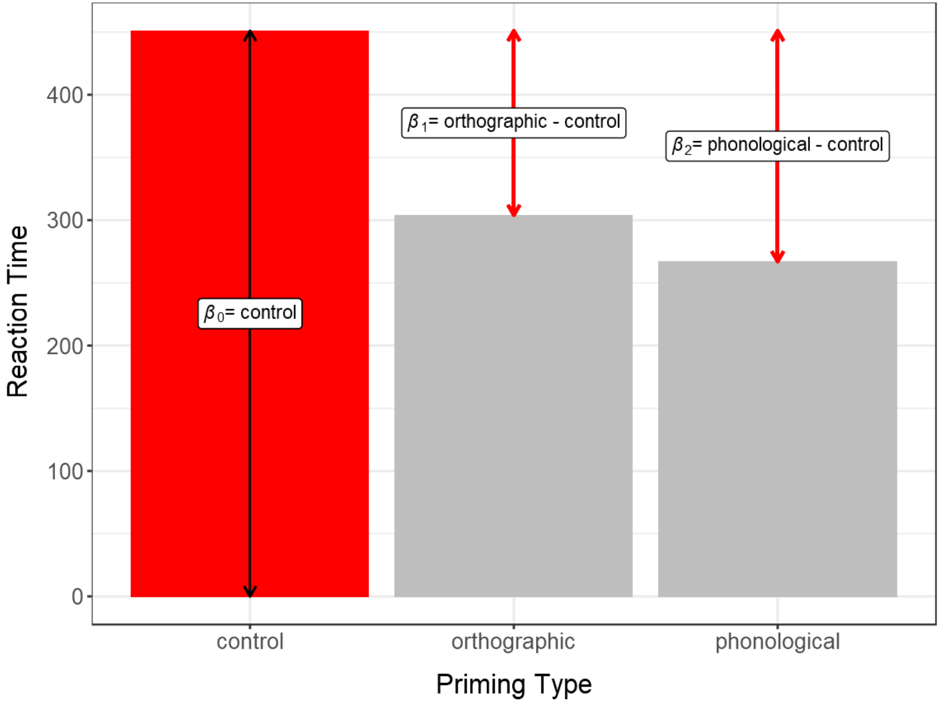
- 実験的介入treatmentの効果を，非介入群との比較で検討
  - 実験群 vs 統制群

- 各実験群を同一の統制群と比べる

- 例：様々なプライミング効果を，同一の統制群と比べる

  1. 綴字プライミング
  2. 音韻的プライミング
  3. プライミングなし（統制群）

- 対比は次の2つ
  1. 綴字プライミング vs プライミングなし（統制群）
  2. 音韻的プライミング vs プライミングなし（統制群）



dummy contrastとも呼ばれる

Sum contrast 零和対比

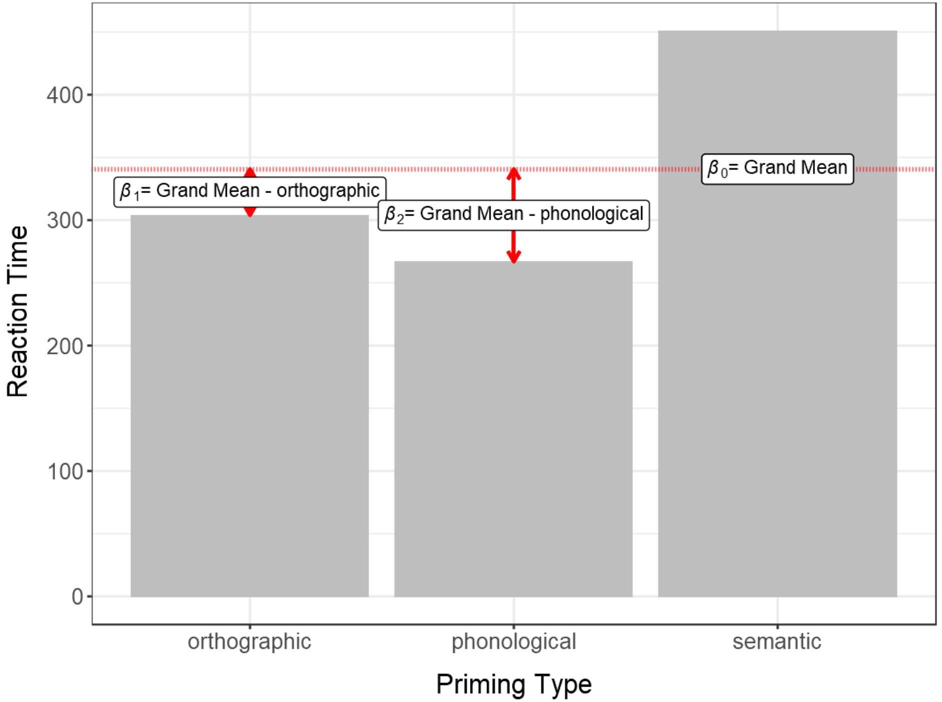
- 実験的介入treatmentの効果を，全ての群の平均との比較で検討
    - 実験群 vs 全体平均

Sum contrast 零和対比

- 実験的介入treatmentの効果を，全ての群の平均との比較で検討
    - 実験群 vs 全体平均
- 例：様々なプライミング効果を，全体平均と比べる
    1. 音韻的プライミング
    2. 綴字プライミング
    3. 意味プライミング

Sum contrast 零和対比

- 実験的介入treatmentの効果を，全ての群の平均との比較で検討
  - 実験群 vs 全体平均
- 例：様々なプライミング効果を，全体平均と比べる
  1. 音韻的プライミング
  2. 綴字プライミング
  3. 意味プライミング
- 対比は次の2つ
  1. 音韻的プライミング vs 全体平均
  2. 綴字プライミング vs 全体平均



Repeated contrast

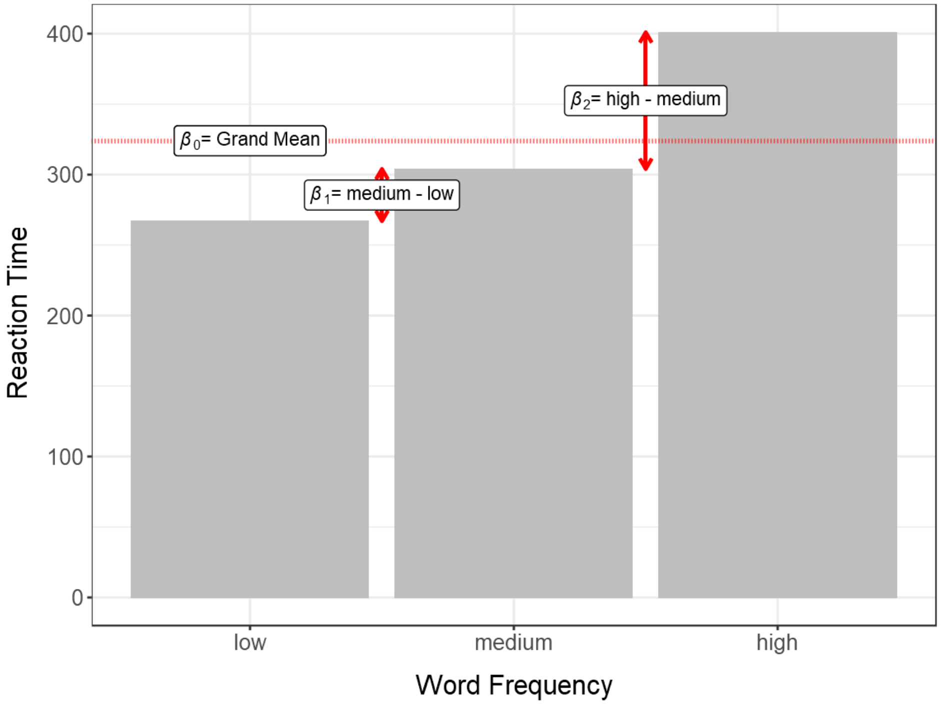
- （順序が決まっている）隣り合う水準同士を比較
  - 実験群1 vs 実験群2
  - 実験群2 vs 実験群3...

Repeated contrast

- （順序が決まっている）隣り合う水準同士を比較
    - 実験群1 vs 実験群2
    - 実験群2 vs 実験群3...
- 例：語の頻度の効果を，隣り合う水準と比べる
    1. 低頻度
    2. 中頻度
    3. 高頻度

Repeated contrast

- （順序が決まっている）隣り合う水準同士を比較
    - 実験群1 vs 実験群2
    - 実験群2 vs 実験群3...
- 例：語の頻度の効果を，隣り合う水準と比べる
    1. 低頻度
    2. 中頻度
    3. 高頻度
- 対比は次の2つ
    1. 中頻度 vs 低頻度
    2. 高頻度 vs 中頻度



Polynomial contrast 多項式対比

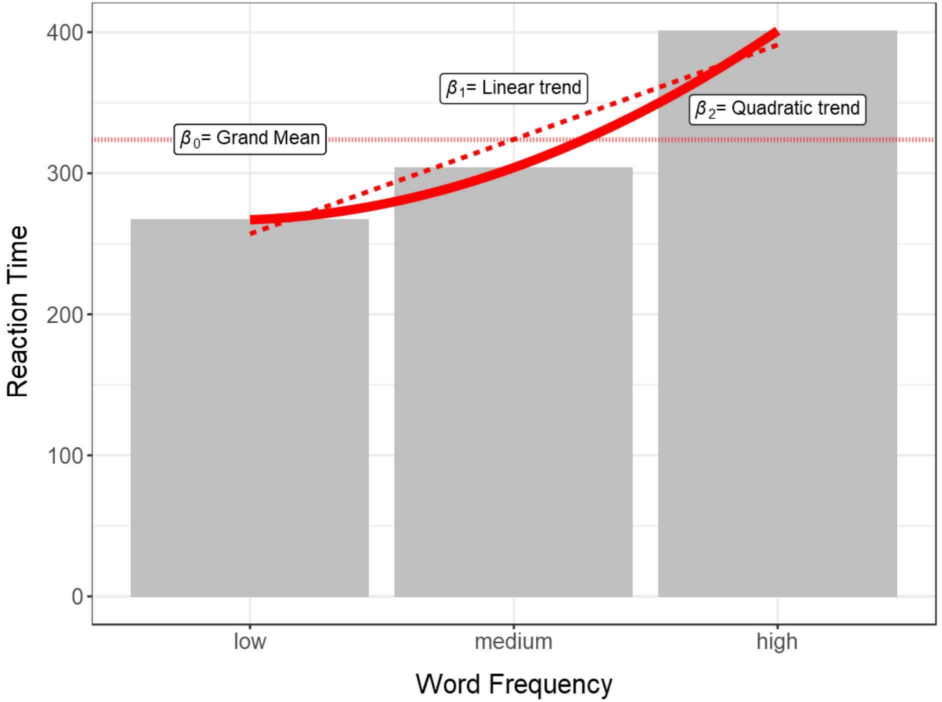
- （順序と間隔が決まっている）隣り合う水準間の増減傾向を比較
  - 水準間で同じペースで増加・減少？
  - 水準間毎に，増加・減少のペースが変わる？

Polynomial contrast 多項式対比

- （順序と間隔が決まっている）隣り合う水準間の増減傾向を比較
    - 水準間で同じペースで増加・減少？
    - 水準間毎に，増加・減少のペースが変わる？
- 例：語の頻度の効果を，隣り合う水準と比べる
    1. 低頻度
    2. 中頻度
    3. 高頻度

Polynomial contrast 多項式対比

- （順序と間隔が決まっている）隣り合う水準間の増減傾向を比較
  - 水準間で同じペースで増加・減少？
  - 水準間毎に，増加・減少のペースが変わる？
- 例：語の頻度の効果を，隣り合う水準と比べる
  1. 低頻度
  2. 中頻度
  3. 高頻度
- 対比は次の2つ
  1. 1次関数ペース（Linear; どの水準間でも一定ペースで増加・減少？）
  2. 2次関数ペース（Quadratic; 次の水準間に移るほど増加・減少が激しい・穏やか？）



This type of coding system should be used only with an ordinal variable in which the levels are equally spaced

Helmert contrast ヘルマート対比

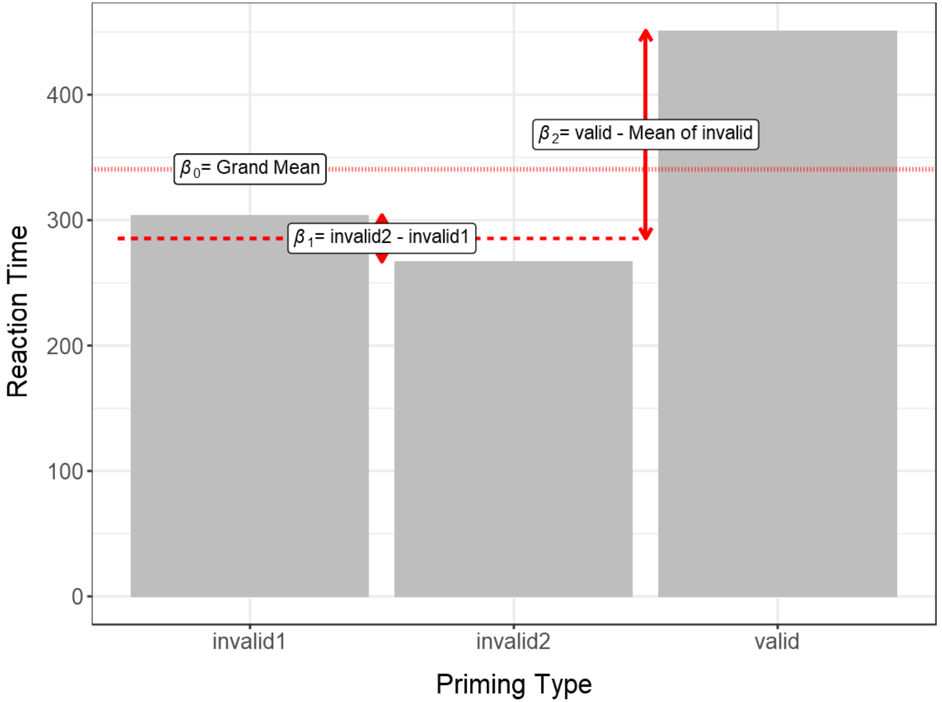
- 「1つの水準を残りの水準の平均と比較する」ことを繰り返す
    - 水準1 vs 残りの水準（水準2と水準3の平均）
    - 水準2 vs 残りの水準（水準3）

Helmert contrast ヘルマート対比

- 「1つの水準を残りの水準の平均と比較する」ことを繰り返す
    - 水準1 vs 残りの水準（水準2と水準3の平均）
    - 水準2 vs 残りの水準（水準3）
- 例：様々なプライミング効果を，種類ごとに比べる
    1. 無効なプライミング1
    2. 無効なプライミング2
    3. 有効なプライミング

Helmert contrast ヘルマート対比

- 「1つの水準を残りの水準の平均と比較する」ことを繰り返す
  - 水準1 vs 残りの水準（水準2と水準3の平均）
  - 水準2 vs 残りの水準（水準3）
- 例：様々なプライミング効果を，種類ごとに比べる
  1. 無効なプライミング1
  2. 無効なプライミング2
  3. 有効なプライミング
- 対比は次の2つ
  1. 無効なプライミング1 vs 無効なプライミング2
  2. 無効なプライミング vs 有効なプライミング



Basic concepts illustrated using a two-level factor (pp.3--8)

1要因2水準の実験で，応答変数の平均値を比較

1. シミュレーション用データを作成
2. シミュレーションの結果を図や表で確認
3. シミュレーション用データに対し線形モデルを構築
    - とりあえず線形モデルを作るとどうなるか試す
4. Default contrast coding: treatment contrasts
5. Defining hypotheses
6. Sum contrast
7. Cell means parameterization

シミュレーション用データを作成

- `mixedDesign()`を使う
- 1要因2水準
    1. F1: 0.8秒
    2. F2: 0.4秒
- 実験参加者数10人
    - 被験者間計画 (between-subject design)
    - 実験参加者は，F1の条件だけ，またはF2の条件だけに接する
    1. F1: 5人
    2. F2: 5人

シミュレーション用データを作成

```
set.seed(1212); simdat <- mixedDesign(
  B = 2,
  W = NULL,
  n = 5,
  M = matrix(
    c(0.8, 0.4),
    nrow = 2,
    ncol = 1,
    byrow = FALSE
  ),
  SD = 0.20,
  long = TRUE
)
```

データフレーム作成

- **B**：被験者間計画での水準数，
- **W**：被験者内計画での水準数，
- **n**：被験者間計画での1水準あたりの実験参加者数，
    - 実験参加者総数は，**B * n**
- **M**：水準ごとの応答変数の平均を示した行列
    - 各行に水準ごとの応答変数の平均を入れる
- **SD**：要因の標準偏差
- **long**：「1試行1データ」の整然データでデータフレームを作成

シミュレーション用データを作成

```
set.seed(1212); simdat <- mixedDesign(
  B = 2,
  W = NULL,
  n = 5,
  M = matrix(
    c(0.8, 0.4),
    nrow = 2,
    ncol = 1,
    byrow = FALSE
  ),
  SD = 0.20,
  long = TRUE
) |>
  rename(F = B_A)
```

要因名をB_AからFに変更

シミュレーション用データを作成

```
set.seed(1212); simdat <- mixedDesign(
  B = 2,
  W = NULL,
  n = 5,
  M = matrix(
    c(0.8, 0.4),
    nrow = 2,
    ncol = 1,
    byrow = FALSE
  ),
  SD = 0.20,
  long = TRUE
) |>
  rename(F = B_A) |>
  mutate(
    F = fct_recode(
      F,
      F1 = "A1",
      F2 = "A2"
    )
  )
```

水準名をA*からF*に変更

シミュレーション用データを作成

```
set.seed(1212); simdat <- mixedDesign(
  B = 2,
  W = NULL,
  n = 5,
  M = matrix(
    c(0.8, 0.4),
    nrow = 2,
    ncol = 1,
    byrow = FALSE
  ),
  SD = 0.20,
  long = TRUE
) |>
  rename(F = B_A) |>
  mutate(
    F = fct_recode(
      F,
      F1 = "A1",
      F2 = "A2"
    )
  )

str(simdat)
```

```
'data.frame':	10 obs. of  3 variables:
 $ F : Factor w/ 2 levels "F1","F2": 1 1 1
 $ id: Factor w/ 10 levels "1","2","3","4"
 $ DV: num  0.737 0.554 0.963 1.044 0.701
```

データフレームの中身の詳細を表示

シミュレーションの結果を図や表で確認

- 1要因2水準
  1. F1: 0.8秒（標準偏差0.2秒）
  2. F2: 0.4秒（標準偏差0.2秒）

シミュレーションの結果を図や表で確認

```r
simdat
```

```
    F id   DV
1  F1  1 0.74
2  F1  2 0.55
3  F1  3 0.96
4  F1  4 1.04
5  F1  5 0.70
6  F2  6 0.26
7  F2  7 0.39
8  F2  8 0.17
9  F2  9 0.48
10 F2 10 0.69
```

データセット

シミュレーションの結果を図や表で確認

```
simdat |>
  group_by(F)
```

```
# A tibble: 10 x 3
# Groups:   F [2]
   F     id       DV
   <fct> <fct> <dbl>
 1 F1    1     0.737
 2 F1    2     0.554
 3 F1    3     0.963
 4 F1    4     1.04
 5 F1    5     0.701
 6 F2    6     0.262
 7 F2    7     0.395
 8 F2    8     0.173
 9 F2    9     0.483
10 F2    10    0.687
```

説明変数の水準毎にデータをまとめ上げ

シミュレーションの結果を図や表で確認

```r
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  )
```

```
# A tibble: 2 x 5
  F         N     M    SD     SE
  <fct> <int> <dbl> <dbl>  <dbl>
1 F1        5   0.8   0.2 0.0894
2 F2        5   0.4   0.2 0.0894
```

水準毎に，

- データの個数**N**，
- 応答変数の平均値**M**，
- 応答変数の平均値の標準偏差**SD**，
- 応答変数の平均値の標準偏差**SE**

を計算

シミュレーションの結果を図や表で確認

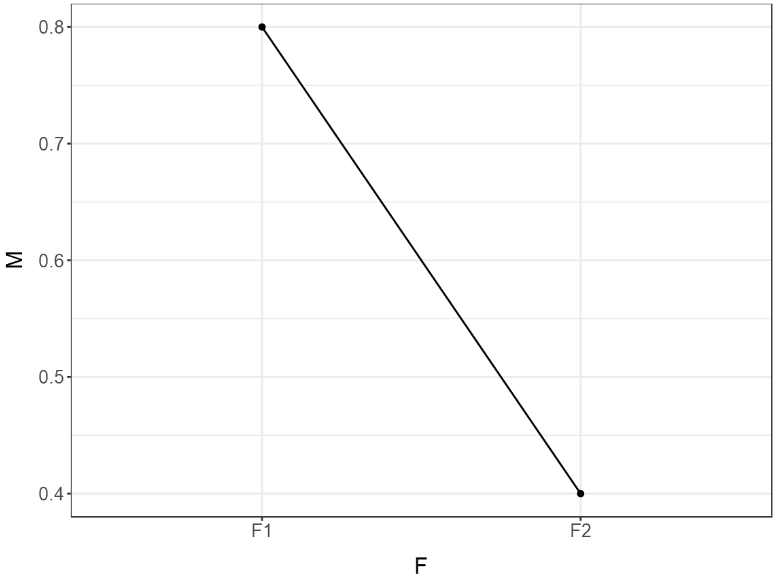
```
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup()
```

```
# A tibble: 2 x 5
  F         N     M    SD     SE
  <fct> <int> <dbl> <dbl>  <dbl>
1 F1        5   0.8   0.2 0.0894
2 F2        5   0.4   0.2 0.0894
```

まとめ上げを解除

シミュレーションの結果を図や表で確認

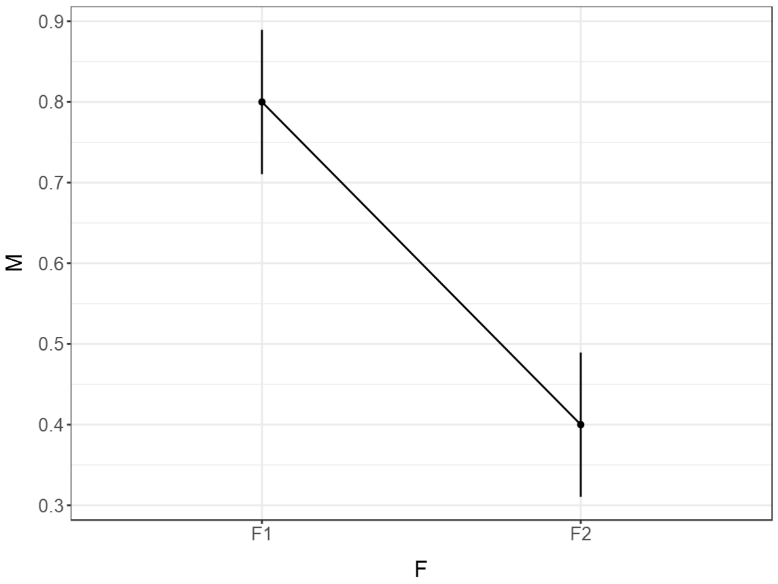
```
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup() |>
  (
    \(d){
      qplot(
        x = F, y = M,
        group = 1,
        data = d,
        geom = c("point", "line")
      )
    }
  )()
```



線グラフを描画

シミュレーションの結果を図や表で確認

```
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup() |>
  (
    \(d){
      qplot(
        x = F, y = M,
        group = 1,
        data = d,
        geom = c("point", "line")
      )
    }
  )() +
  geom_errorbar(
    aes(
      max = M + SE,
      min = M - SE
    ),
    width = 0
  )
```



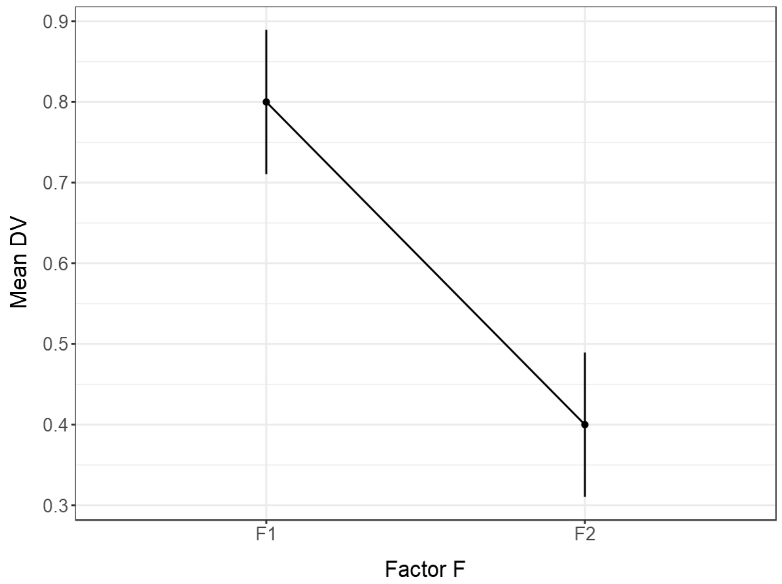
誤差範囲 ($\pm 1SE$) を描画

※ $\pm 1SE$ 自体は95%信頼区間ではない（68%信頼区間）

※信頼区間にする場合，$\pm 1.96SE$

シミュレーションの結果を図や表で確認

```r
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup() |>
  (
    \(d){
      qplot(
        x = F, y = M,
        group = 1,
        data = d,
        geom = c("point", "line")
      )
    }
  )() +
  geom_errorbar(
    aes(
      max = M + SE,
      min = M - SE
    ),
    width = 0
  ) +
  labs(
    y = "Mean DV",
    x = "Factor F"
```



x軸・y軸の軸名を変更

シミュレーションの結果を図や表で確認

```
simdat
```

```
    F id   DV
1  F1  1 0.74
2  F1  2 0.55
3  F1  3 0.96
4  F1  4 1.04
5  F1  5 0.70
6  F2  6 0.26
7  F2  7 0.39
8  F2  8 0.17
9  F2  9 0.48
10 F2 10 0.69
```

データセット

シミュレーションの結果を図や表で確認

```
simdat |>
  group_by(F)
```

```
# A tibble: 10 x 3
# Groups:   F [2]
   F     id       DV
   <fct> <fct> <dbl>
 1 F1    1     0.737
 2 F1    2     0.554
 3 F1    3     0.963
 4 F1    4     1.04
 5 F1    5     0.701
 6 F2    6     0.262
 7 F2    7     0.395
 8 F2    8     0.173
 9 F2    9     0.483
10 F2    10    0.687
```

説明変数の水準毎にデータをまとめ上げ

シミュレーションの結果を図や表で確認

```r
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  )
```

```
# A tibble: 2 x 5
  F         N     M    SD     SE
  <fct> <int> <dbl> <dbl>  <dbl>
1 F1        5   0.8   0.2 0.0894
2 F2        5   0.4   0.2 0.0894
```

水準毎に，

- データの個数**N**，
- 応答変数の平均値**M**，
- 応答変数の平均値の標準偏差**SD**，
- 応答変数の平均値の標準偏差**SE**

を計算

シミュレーションの結果を図や表で確認

```
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup()
```

```
# A tibble: 2 x 5
  F         N     M    SD     SE
  <fct> <int> <dbl> <dbl>  <dbl>
1 F1        5   0.8   0.2 0.0894
2 F2        5   0.4   0.2 0.0894
```

まとめ上げを解除

シミュレーションの結果を図や表で確認

```r
simdat |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup() |>
  kable(
    digits = 2,
    col.names = c(
      'Levels of Factor',
      'N. of data points',
      'Mean RT',
      'Std. Dev.',
      'Std. Err.'
    )
  )
```

Levels of Factor	N. of data points	Mean RT	Std. Dev.	Std. Err.
F1	5	0.8	0.2	0.09
F2	5	0.4	0.2	0.09

表の出力

- **digits**：有効数字の桁数を指定
- **col.names**：表内での各列の名前を変更

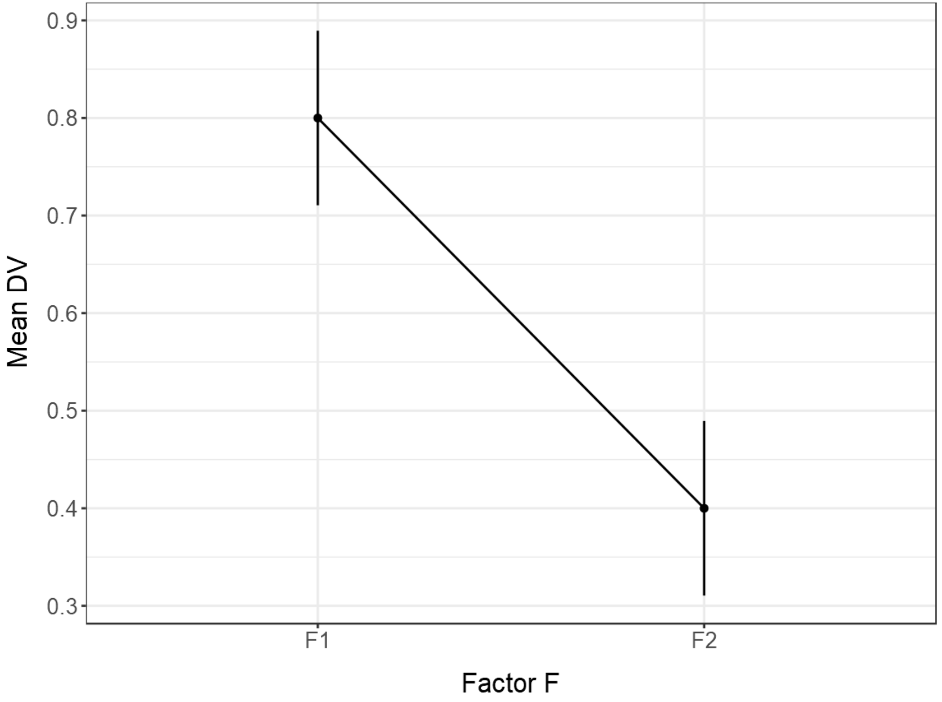
シミュレーション用データに対し線形モデルを構築

- 水準間で応答変数の平均が有意に異なるか，線形モデルで確認

```
m_F <- lm(DV ~ F, simdat)
```

- 切片 (intercept, 0.8)
  - F1の下での応答変数の平均 $\widehat{\mu_{\mathrm{F1}}}$
- 傾き (intercept, -0.4)
  - F2の下での応答変数の平均とF1の下での応答変数の平均の差 $\widehat{\mu_{\mathrm{F2}}} - \widehat{\mu_{\mathrm{F1}}}$

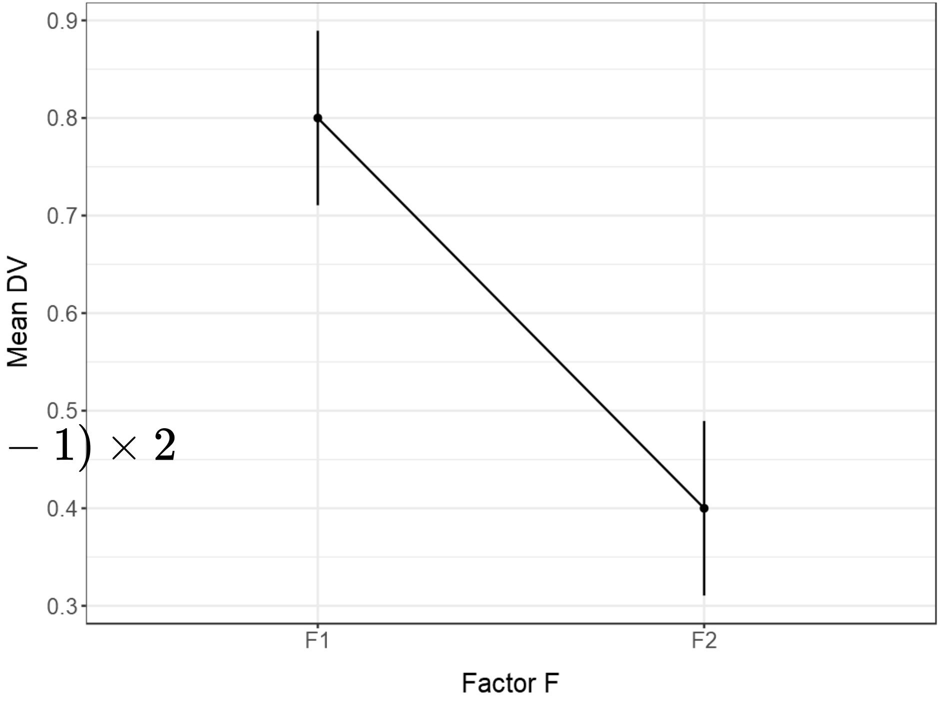
Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	0.8	[0.594, 1.006]	8.9	0.00
FF2	-0.4	[-0.692, -0.108]	-3.2	0.01



シミュレーション用データに対し線形モデルを構築

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	0.8	[0.594, 1.006]	8.9	0.00
FF2	-0.4	[-0.692, -0.108]	-3.2	0.01

- 95%信頼区間 (95% Confidence Interval)
  - 平均値 $\pm 1.96 \times$ 標準誤差 $(\mathrm{Standanrd\ Error})$
  - 「100回実験して95回は真値を含むような区間」
- t 値
  - この t 値は，自由度が（（その水準から得られたデータポイント数） $-1) \times 2$ の t 分布に従う
  - この t 分布の両側2.5%の範囲にあれば，有意
  - $(5-1) \times 2 = 8$ なので，今回の t 値は，$t(8)$ に従う
  - $t(8)$ の両側2.5%の範囲は，$t \leq 2.31, 2.31 \leq t$
  - t 値が上記区間に入っていれば，有意



```
m_F
```

```
Call:
lm(formula = DV ~ F, data = simdat)

Coefficients:
(Intercept)          FF2
        0.8         -0.4
```

```r
m_F |>
  tidy(
    conf.int = TRUE,
    conf.level = 0.95
  )
```

```
# A tibble: 2 x 7
  term        estimate std.error statistic   p.value conf.low co
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)      0.8    0.0894      8.94 0.0000194    0.594
2 FF2             -0.4    0.126      -3.16 0.0133      -0.692
```

```r
m_F |>
  tidy(
    conf.int = TRUE,
    conf.level = 0.95
  ) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(3),
      ", ",
      conf.high |> round(3),
      "]"
    )
  )
```

.panel2-tab-lm-simdat-auto[

```
# A tibble: 2 x 8
  term        estimate std.error statistic   p.value conf.low con
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)      0.8    0.0894      8.94 0.0000194    0.594
2 FF2             -0.4    0.126      -3.16 0.0133      -0.692
```

]

```r
m_F |>
  tidy(
    conf.int = TRUE,
    conf.level = 0.95
  ) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(3),
      ", ",
      conf.high |> round(3),
      "]"
    )
  ) |>
  relocate(
    `95% CI`, .after = estimate
  )
```

.panel2-tab-lm-simdat-auto[

```
# A tibble: 2 x 8
  term        estimate `95% CI`    std.error statistic p.value con
  <chr>          <dbl> <chr>           <dbl>     <dbl>   <dbl>
1 (Intercept)      0.8 [0.594, 1~     0.0894      8.94 1.94e-5
2 FF2             -0.4 [-0.692, ~     0.126      -3.16 1.33e-2    -
```

]

```r
m_F |>
  tidy(
    conf.int = TRUE,
    conf.level = 0.95
  ) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(3),
      ", ",
      conf.high |> round(3),
      "]"
    )
  ) |>
  relocate(
    `95% CI`, .after = estimate
  ) |>
  dplyr::select(
    -c(
      std.error,
      conf.low,
      conf.high
    )
  )
```

```
# A tibble: 2 x 5
  term        estimate `95% CI`         statistic   p.value
  <chr>          <dbl> <chr>                <dbl>     <dbl>
1 (Intercept)      0.8 [0.594, 1.006]        8.94 0.0000194
2 FF2             -0.4 [-0.692, -0.108]     -3.16 0.0133
```

```r
m_F |>
  tidy(
    conf.int = TRUE,
    conf.level = 0.95
  ) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(3),
      ", ",
      conf.high |> round(3),
      "]"
    )
  ) |>
  relocate(
    `95% CI`, .after = estimate
  ) |>
  dplyr::select(
    -c(
      std.error,
      conf.low,
      conf.high
    )
  ) |>
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	0.8	[0.594, 1.006]	8.9	0.00
FF2	-0.4	[-0.692, -0.108]	-3.2	0.01

Default contrast coding: treatment contrasts

- 切片 (intercept, 0.8)
  - F1の下での応答変数の平均 $\widehat{\mu_{\mathrm{F1}}}$
- 傾き (intercept, -0.4)
  - F2の下での応答変数の平均とF1の下での応答変数の平均の差 $\widehat{\mu_{\mathrm{F2}}} - \widehat{\mu_{\mathrm{F1}}}$

なぜ，切片と傾きがこの値になるのか？？

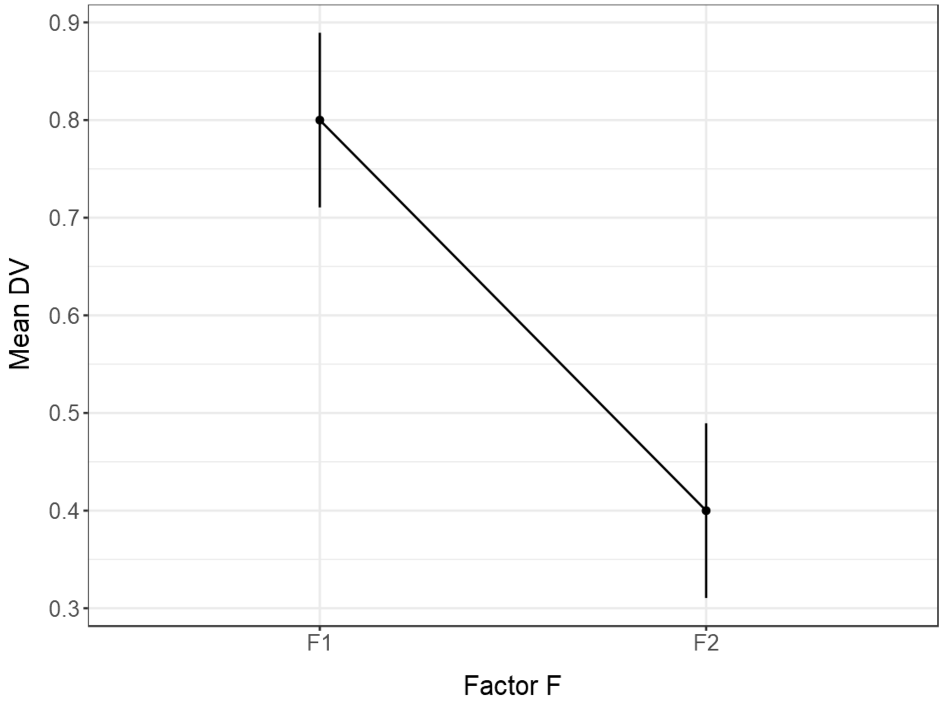
Default contrast coding: treatment contrasts

- 切片 (intercept, 0.8)
  - F1の下での応答変数の平均 $\widehat{\mu_{\mathrm{F1}}}$
- 傾き (intercept, -0.4)
  - F2の下での応答変数の平均とF1の下での応答変数の平均の差 $\widehat{\mu_{\mathrm{F2}}} - \widehat{\mu_{\mathrm{F1}}}$

なぜ，切片と傾きがこの値になるのか？？

∵ デフォルトでは，Rは，名義尺度にtreatment contrastを適用するため

```
m_F <- lm(DV ~ F, simdat)
```



Treatment contrastsとその意味

```r
contrasts(simdat$F)
```

```
   F2
F1  0
F2  1
```

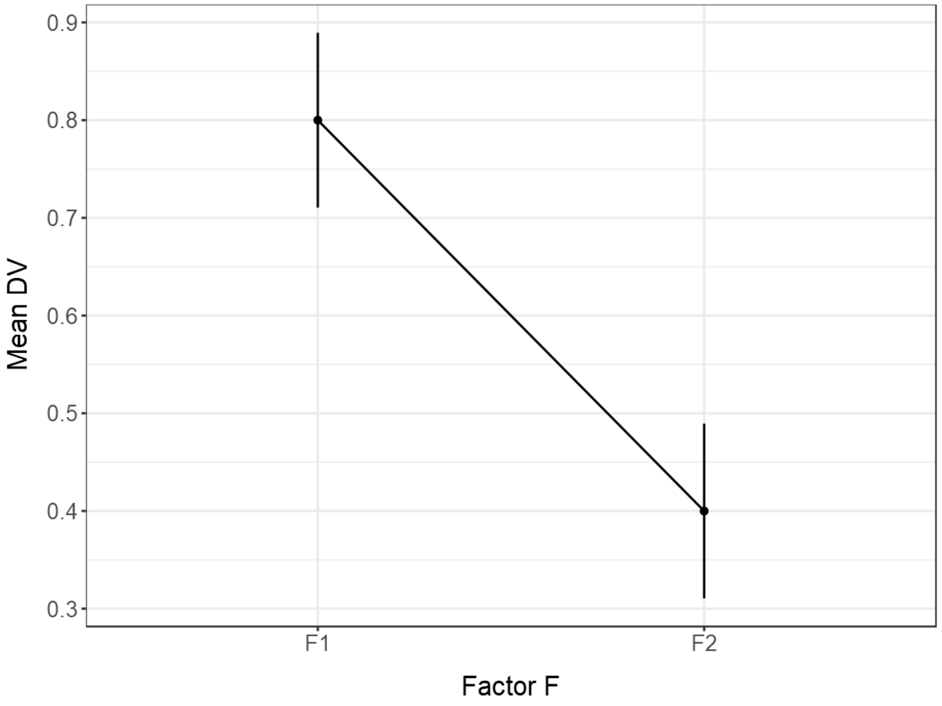
$$x = \begin{cases} 0 & \text{F1 のとき} \\ 1 & \text{F2 のとき} \end{cases}$$

Default contrast coding: treatment contrasts

Rがデフォルトで名義尺度に適用する対比

$$x = \begin{cases} 0 & \text{F1 のとき} \\ 1 & \text{F2 のとき} \end{cases}$$

$$\begin{aligned}
\overbrace{y}^{\text{応答変数}} &= \overbrace{\beta_0}^{\text{切片}} + \overbrace{\beta_1}^{\text{傾き}} x \\
&= \begin{cases} \beta_0 + \beta_1 \times 0 & \text{F1 のとき} \\ \beta_0 + \beta_1 \times 1 & \text{F2 のとき} \end{cases} \\
\therefore \quad &= \begin{cases} \beta_0 & \text{F1 のとき} \\ \beta_0 + \beta_1 & \text{F2 のとき} \end{cases} \\
&= \begin{cases} 0.8 & \text{F1 のとき} \\ 0.8 + (-0.4) & \text{F2 のとき} \end{cases} \\
&= \begin{cases} 0.8 & \text{F1 のとき} \\ 0.4 & \text{F2 のとき} \end{cases}
\end{aligned}$$



Defining hypotheses

Treatment contrastの下では，

- 傾き β_1
  - 「実験条件の応答変数の平均値」と「統制条件での応答変数の平均値」の差
  - ($\mu_{\mathrm{F2}} - \mu_{\mathrm{F1}}$)
- 切片 β_0
  - 統制条件での応答変数の平均値
  - (μ_{F1})

Defining hypotheses

Treatment contrastの下では，

- 傾き β_1
  - 「実験条件の応答変数の平均値」と「統制条件での応答変数の平均値」の差
  - ($\mu_{\mathrm{F2}} - \mu_{\mathrm{F1}}$)
- 切片 β_0
  - 統制条件での応答変数の平均値
  - (μ_{F1})

↑切片や傾きの意味を言語化しただけ

Defining hypotheses

Treatment contrastの下では，

- 傾き β_1
  - 「実験条件の応答変数の平均値」と「統制条件での応答変数の平均値」の差
  - ($\mu_{\mathrm{F2}} - \mu_{\mathrm{F1}}$)
- 切片 β_0
  - 統制条件での応答変数の平均値
  - (μ_{F1})

↑切片や傾きの意味を言語化しただけ

切片や傾きを通じて，どのような帰無仮説を棄却しようとしているのか？

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

$$H_0 : \mu_{\mathrm{F1}} = 0$$

統制条件とする水準を変えるには

- Rは，水準をアルファベット順に読み込み，アルファベット順に数値（0,1）を付与する
- アルファベット順に関係なく，特定の水準を統制条件としたい場合には

統制条件とする水準を変えるには

```r
simdat <- simdat

simdat
```

```
    F id   DV
1   F1  1 0.74
2   F1  2 0.55
3   F1  3 0.96
4   F1  4 1.04
5   F1  5 0.70
6   F2  6 0.26
7   F2  7 0.39
8   F2  8 0.17
9   F2  9 0.48
10  F2 10 0.69
```

統制条件とする水準を変えるには

```r
simdat <- simdat |>
  mutate(
    Fb = fct_relevel(
      F,
      "F2",
      "F1"
    )
  )

simdat
```

```
    F id   DV Fb
1  F1  1 0.74 F1
2  F1  2 0.55 F1
3  F1  3 0.96 F1
4  F1  4 1.04 F1
5  F1  5 0.70 F1
6  F2  6 0.26 F2
7  F2  7 0.39 F2
8  F2  8 0.17 F2
9  F2  9 0.48 F2
10 F2 10 0.69 F2
```

統制条件とする水準を変えるには

```r
simdat <- simdat |>
  mutate(
    Fb = fct_relevel(
      F,
      "F2",
      "F1"
    )
  )

contrasts(simdat$Fb)

simdat
```

```
   F1
F2  0
F1  1
```

```
    F id   DV Fb
1  F1  1 0.74 F1
2  F1  2 0.55 F1
3  F1  3 0.96 F1
4  F1  4 1.04 F1
5  F1  5 0.70 F1
6  F2  6 0.26 F2
7  F2  7 0.39 F2
8  F2  8 0.17 F2
9  F2  9 0.48 F2
10 F2 10 0.69 F2
```

統制条件とする水準を変えるには

```r
simdat <- simdat |>
  mutate(
    Fb = fct_relevel(
      F,
      "F2",
      "F1"
    )
  )

contrasts(simdat$Fb)

contrasts(simdat$F)

simdat
```

```
   F1
F2  0
F1  1
```

```
   F2
F1  0
F2  1
```

```
    F id   DV Fb
1  F1  1 0.74 F1
2  F1  2 0.55 F1
3  F1  3 0.96 F1
4  F1  4 1.04 F1
5  F1  5 0.70 F1
6  F2  6 0.26 F2
7  F2  7 0.39 F2
8  F2  8 0.17 F2
9  F2  9 0.48 F2
10 F2 10 0.69 F2
```

F2を統制条件にした場合

```
m1_mr <- lm(DV ~ Fb, simdat)
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	0.4	[0.194, 0.606]	4.5	0.00
FbF1	0.4	[0.108, 0.692]	3.2	0.01

$$x = \begin{cases} 0 & \text{F2 のとき} \\ 1 & \text{F1 のとき} \end{cases}$$

$$\begin{aligned} y &= \beta_0 + \beta_1 x \\ &= \begin{cases} \beta_0 + \beta_1 \times 0 & \text{F2 のとき} \\ \beta_0 + \beta_1 \times 1 & \text{F1 のとき} \end{cases} \\ \therefore \quad &= \begin{cases} 0.4 & \text{F2 のとき} \\ 0.4 + (0.4) & \text{F1 のとき} \end{cases} \\ &= \begin{cases} 0.4 & \text{F2 のとき} \\ 0.8 & \text{F1 のとき} \end{cases} \end{aligned}$$

F1が統制条件である場合

```
m_F <- lm(DV ~ F, simdat)
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	0.8	[0.594, 1.006]	8.9	0.00
FF2	-0.4	[-0.692, -0.108]	-3.2	0.01

$$x = \begin{cases} 0 & \text{F1 のとき} \\ 1 & \text{F2 のとき} \end{cases}$$

$$\begin{aligned} y &= \beta_0 + \beta_1 x \\ &= \begin{cases} \beta_0 + \beta_1 \times 0 & \text{F1 のとき} \\ \beta_0 + \beta_1 \times 1 & \text{F2 のとき} \end{cases} \\ \therefore \quad &= \begin{cases} 0.8 & \text{F1 のとき} \\ 0.8 + (-0.4) & \text{F2 のとき} \end{cases} \\ &= \begin{cases} 0.8 & \text{F1 のとき} \\ 0.4 & \text{F2 のとき} \end{cases} \end{aligned}$$

F2を統制条件にした場合

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

F2を統制条件にした場合

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

F2を統制条件にした場合

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F1}} - \mu_{\mathrm{F2}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

F2を統制条件にした場合

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F1}} - \mu_{\mathrm{F2}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

F2を統制条件にした場合

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F1}} - \mu_{\mathrm{F2}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

$$H_0 : \mu_{\mathrm{F2}} = 0$$

（参考）F1が統制条件である場合

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

$$H_0 : \mu_{\mathrm{F1}} = 0$$

Sum contrast

- （2水準の場合）一方の水準に-1を，他方の水準に1を充てる
- 効果全体を全体平均に「中心化」する
  - 全体平均 (Grand Mean)：各水準の平均を平均した値
- scaled sum contrasts; effect coding
  - （2水準の場合）一方の水準に-0.5を，他方の水準に0.5を充てる
  - 傾きが，2水準の差と同じ値になる

Scaled sum contrastsとその意味

```r
contrasts(simdat$F) <- c(-0.5,

contrasts(simdat$F)
```

```
   [,1]
F1 -0.5
F2  0.5
```

$$x = \begin{cases} -0.5 & \text{F1 のとき} \\ 0.5 & \text{F2 のとき} \end{cases}$$

Sum contrast

- （2水準の場合）一方の水準に-1を，他方の水準に1を充てる
- 効果全体を全体平均に「中心化」する
  - 全体平均 (Grand Mean)：各水準の平均を平均した値
- scaled sum contrasts; effect coding
  - （2水準の場合）一方の水準に-0.5を，他方の水準に0.5を充てる
  - 傾きが，2水準の差と同じ値になる

```
m1_mr <- lm(DV ~ F, simdat)
```

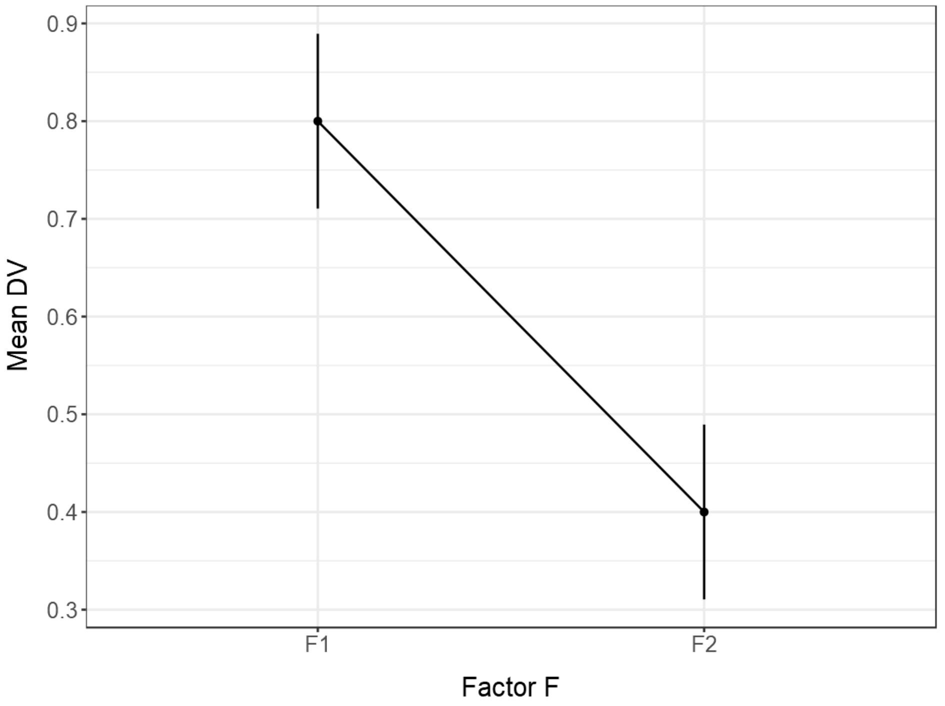
Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	0.6	[0.454, 0.746]	9.5	0.00
F1	-0.4	[-0.692, -0.108]	-3.2	0.01

(Scaled) Sum contrasts

$$x = \begin{cases} -0.5 & \text{F1 のとき} \\ 0.5 & \text{F2 のとき} \end{cases}$$

$$\begin{aligned}
\overbrace{y}^{\text{応答変数}} &= \overbrace{\beta_0}^{\text{切片}} + \overbrace{\beta_1}^{\text{傾き}} x \\
&= \begin{cases} \beta_0 + \beta_1 \times (-0.5) & \text{F1 のとき} \\ \beta_0 + \beta_1 \times 0.5 & \text{F2 のとき} \end{cases} \\
\therefore \quad &= \begin{cases} \beta_0 - 0.5\beta_1 & \text{F1 のとき} \\ \beta_0 + 0.5\beta_1 & \text{F2 のとき} \end{cases} \\
&= \begin{cases} 0.6 - 0.5 \times (-0.4) & \text{F1 のとき} \\ 0.6 + 0.5 \times (-0.4) & \text{F2 のとき} \end{cases} \\
&= \begin{cases} 0.8 & \text{F1 のとき} \\ 0.4 & \text{F2 のとき} \end{cases}
\end{aligned}$$

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	0.6	[0.454, 0.746]	9.5	0.00
F1	-0.4	[-0.692, -0.108]	-3.2	0.01



(Scaled) Sum contrasts

(Scaled) Sum contrastsの下では，

- 傾き β_1
  - 「一方の条件の応答変数の平均値」と「他方の条件での応答変数の平均値」の差
  - ($\mu_{\mathrm{F2}} - \mu_{\mathrm{F1}}$)
- 切片 β_0
  - 「一方の条件の応答変数の平均値」と「他方の条件での応答変数の平均値」の平均値
  - ($\frac{\mu_{\mathrm{F1}} + \mu_{\mathrm{F2}}}{2}$)

(Scaled) Sum contrastsの帰無仮説

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

(Scaled) Sum contrastsの帰無仮説

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

(Scaled) Sum contrastsの帰無仮説

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

(Scaled) Sum contrastsの帰無仮説

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

(Scaled) Sum contrastsの帰無仮説

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

$$H_0 : \frac{\mu_{\mathrm{F1}} + \mu_{\mathrm{F2}}}{2} = 0$$

（参考）F1を統制条件とするTreatment contrastの場合

傾き β_1 に関する帰無仮説 H_0

帰無仮説：傾きが0

$$H_0 : \beta_1 = 0$$

$$H_0 : \mu_{\mathrm{F2}} - \mu_{\mathrm{F1}} = 0$$

切片 β_0 に関する帰無仮説 H_0

帰無仮説：切片が0

$$H_0 : \beta_0 = 0$$

$$H_0 : \mu_{\mathrm{F1}} = 0$$

(Un)balanced dataでのGrand Meanの意味

```
simdat
```

```
    F id   DV Fb
1  F1  1 0.74 F1
2  F1  2 0.55 F1
3  F1  3 0.96 F1
4  F1  4 1.04 F1
5  F1  5 0.70 F1
6  F2  6 0.26 F2
7  F2  7 0.39 F2
8  F2  8 0.17 F2
9  F2  9 0.48 F2
10 F2 10 0.69 F2
```

データセット

(Un)balanced dataでのGrand Meanの意味

```
simdat |>
  mutate(
    gross.mean = mean(DV)
  )
```

```
    F id   DV Fb gross.mean
1   F1  1 0.74 F1        0.6
2   F1  2 0.55 F1        0.6
3   F1  3 0.96 F1        0.6
4   F1  4 1.04 F1        0.6
5   F1  5 0.70 F1        0.6
6   F2  6 0.26 F2        0.6
7   F2  7 0.39 F2        0.6
8   F2  8 0.17 F2        0.6
9   F2  9 0.48 F2        0.6
10  F2 10 0.69 F2        0.6
```

応答変数全ての平均を計算

(Un)balanced dataでのGrand Meanの意味

```
simdat |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F)
```

```
# A tibble: 10 x 5
# Groups:   F [2]
   F     id       DV Fb    gross.mean
   <fct> <fct> <dbl> <fct>      <dbl>
 1 F1    1     0.737 F1           0.6
 2 F1    2     0.554 F1           0.6
 3 F1    3     0.963 F1           0.6
 4 F1    4     1.04  F1           0.6
 5 F1    5     0.701 F1           0.6
 6 F2    6     0.262 F2           0.6
 7 F2    7     0.395 F2           0.6
 8 F2    8     0.173 F2           0.6
 9 F2    9     0.483 F2           0.6
10 F2    10    0.687 F2           0.6
```

説明変数の水準毎にデータをまとめ上げ

(Un)balanced dataでのGrand Meanの意味

```
simdat |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  )
```

```
# A tibble: 10 x 6
# Groups:   F [2]
   F     id       DV Fb    gross.mean lev.
   <fct> <fct> <dbl> <fct>      <dbl>
 1 F1    1     0.737 F1           0.6
 2 F1    2     0.554 F1           0.6
 3 F1    3     0.963 F1           0.6
 4 F1    4     1.04  F1           0.6
 5 F1    5     0.701 F1           0.6
 6 F2    6     0.262 F2           0.6
 7 F2    7     0.395 F2           0.6
 8 F2    8     0.173 F2           0.6
 9 F2    9     0.483 F2           0.6
10 F2    10    0.687 F2           0.6
```

水準毎に応答変数の平均値を計算

(Un)balanced dataでのGrand Meanの意味

```r
simdat |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup()
```

```
# A tibble: 10 x 6
   F     id      DV Fb    gross.mean lev.
   <fct> <fct> <dbl> <fct>      <dbl>
 1 F1    1     0.737 F1           0.6
 2 F1    2     0.554 F1           0.6
 3 F1    3     0.963 F1           0.6
 4 F1    4     1.04  F1           0.6
 5 F1    5     0.701 F1           0.6
 6 F2    6     0.262 F2           0.6
 7 F2    7     0.395 F2           0.6
 8 F2    8     0.173 F2           0.6
 9 F2    9     0.483 F2           0.6
10 F2    10    0.687 F2           0.6
```

まとめ上げを解除

(Un)balanced dataでのGrand Meanの意味

```
simdat |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  )
```

```
# A tibble: 2 x 3
  F     gross.mean lev.wise.mean
  <fct>      <dbl>         <dbl>
1 F1           0.6           0.8
2 F2           0.6           0.4
```

応答変数全ての平均と水準毎の平均を抽出

(Un)balanced dataでのGrand Meanの意味

```
simdat |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  )
```

```
# A tibble: 2 x 2
  gross.mean mean.of.lev.wise.mean
       <dbl>                 <dbl>
1        0.6                   0.6
2        0.6                   0.6
```

水準毎の平均の平均を計算

(Un)balanced dataでのGrand Meanの意味

```
simdat |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  ) |>
  distinct(
    gross.mean, mean.of.lev.wis
  )
```

```
# A tibble: 1 x 2
  gross.mean mean.of.lev.wise.mean
       <dbl>                 <dbl>
1        0.6                   0.6
```

Grand Mean：「各水準で得られた応答変数の平均値」の平均値

Unbalanced dataでのGrand Meanの意味

```r
set.seed(1)
```

乱数固定

Unbalanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat
```

```
    F id   DV Fb
1  F1  1 0.74 F1
2  F1  2 0.55 F1
3  F1  3 0.96 F1
4  F1  4 1.04 F1
5  F1  5 0.70 F1
6  F2  6 0.26 F2
7  F2  7 0.39 F2
8  F2  8 0.17 F2
9  F2  9 0.48 F2
10 F2 10 0.69 F2
```

データセット

Unbalanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F)
```

```
<list_of<
  tbl_df<
    F : factor<1ffb8>
    id: factor<49edf>
    DV: double
    Fb: factor<e285c>
  >
>[2]>
[[1]]
# A tibble: 5 x 4
  F     id       DV Fb
  <fct> <fct> <dbl> <fct>
1 F1    1     0.737 F1
2 F1    2     0.554 F1
3 F1    3     0.963 F1
4 F1    4     1.04  F1
5 F1    5     0.701 F1

[[2]]
# A tibble: 5 x 4
  F     id       DV Fb
  <fct> <fct> <dbl> <fct>
1 F2    6     0.262 F2
2 F2    7     0.395 F2
3 F2    8     0.173 F2
4 F2    9     0.483 F2
5 F2    10    0.687 F2
```

データを水準ごとに分割

Unbalanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  )
```

```
# A tibble: 3 x 4
  F     id        DV Fb
  <fct> <fct> <dbl> <fct>
1 F1    1     0.737 F1
2 F1    4     1.04  F1
3 F2    6     0.262 F2
```

水準毎のデータから，指定した分だけ無作為抽出（**F1とF2でデータ数が異なるように抽出**）

Unbalanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  )
```

```
# A tibble: 3 x 5
  F     id       DV Fb    gross.mean
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.681
2 F1    4     1.04  F1         0.681
3 F2    6     0.262 F2         0.681
```

無作為抽出後のデータについて，応答変数全ての平均を計算

Unbalanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F)
```

```
# A tibble: 3 x 5
# Groups:   F [2]
  F     id       DV Fb    gross.mean
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.681
2 F1    4     1.04  F1         0.681
3 F2    6     0.262 F2         0.681
```

説明変数の水準毎にデータをまとめ上げ

Unbalanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  )
```

```
# A tibble: 3 x 6
# Groups:   F [2]
  F     id       DV Fb    gross.mean lev.w
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.681
2 F1    4     1.04  F1         0.681
3 F2    6     0.262 F2         0.681
```

水準毎に応答変数の平均値を計算

Unbalanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup()
```

```
# A tibble: 3 x 6
  F     id        DV Fb    gross.mean lev.w
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.681
2 F1    4     1.04  F1         0.681
3 F2    6     0.262 F2         0.681
```

まとめ上げを解除

Unbalanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  )
```

```
# A tibble: 2 x 3
  F     gross.mean lev.wise.mean
  <fct>      <dbl>         <dbl>
1 F1         0.681         0.890
2 F2         0.681         0.262
```

応答変数全ての平均と水準毎の平均を抽出

Unbalanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  )
```

```
# A tibble: 2 x 2
  gross.mean mean.of.lev.wise.mean
       <dbl>                 <dbl>
1      0.681                 0.576
2      0.681                 0.576
```

水準毎の平均の平均を計算

Unbalanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  ) |>
  distinct(
    gross.mean, mean.of.lev.wis
  )
```

```
# A tibble: 1 x 2
  gross.mean mean.of.lev.wise.mean
       <dbl>                 <dbl>
1      0.681                 0.576
```

「各水準で得られた応答変数の平均値」の平均値と，全データの平均が**一致しない**

Unbalanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1から2つ，F2から1つデータを取
    .y = c(2, 1),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  ) |>
  distinct(
    gross.mean, mean.of.lev.wis
  ) |>
  kable(
    digit = 2,
```

応答変数全ての平均	「各水準の平均」の平均
0.68	0.58

「各水準で得られた応答変数の平均値」の平均値と，全データの平均が**一致しない**

Balanced dataでのGrand Meanの意味

```
set.seed(1)
```

乱数固定

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat
```

```
    F id   DV Fb
1  F1  1 0.74 F1
2  F1  2 0.55 F1
3  F1  3 0.96 F1
4  F1  4 1.04 F1
5  F1  5 0.70 F1
6  F2  6 0.26 F2
7  F2  7 0.39 F2
8  F2  8 0.17 F2
9  F2  9 0.48 F2
10 F2 10 0.69 F2
```

データセット

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F)
```

```
<list_of<
  tbl_df<
    F : factor<1ffb8>
    id: factor<49edf>
    DV: double
    Fb: factor<e285c>
  >
>[2]>
[[1]]
# A tibble: 5 x 4
  F     id       DV Fb
  <fct> <fct> <dbl> <fct>
1 F1    1     0.737 F1
2 F1    2     0.554 F1
3 F1    3     0.963 F1
4 F1    4     1.04  F1
5 F1    5     0.701 F1

[[2]]
# A tibble: 5 x 4
  F     id       DV Fb
  <fct> <fct> <dbl> <fct>
1 F2    6     0.262 F2
2 F2    7     0.395 F2
3 F2    8     0.173 F2
4 F2    9     0.483 F2
5 F2    10    0.687 F2
```

データを水準ごとに分割

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを取
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  )
```

```
# A tibble: 4 x 4
  F     id       DV Fb
  <fct> <fct> <dbl> <fct>
1 F1    1     0.737 F1
2 F1    4     1.04  F1
3 F2    6     0.262 F2
4 F2    7     0.395 F2
```

水準毎のデータから，指定した分だけ無作為抽出（**F1とF2でデータ数が一致するように抽出**）

Balanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  )
```

```
# A tibble: 4 x 5
  F     id        DV Fb    gross.mean
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.609
2 F1    4     1.04  F1         0.609
3 F2    6     0.262 F2         0.609
4 F2    7     0.395 F2         0.609
```

無作為抽出後のデータについて，応答変数全ての平均を計算

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを耳
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F)
```

```
# A tibble: 4 x 5
# Groups:   F [2]
  F     id       DV Fb    gross.mean
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.609
2 F1    4     1.04  F1         0.609
3 F2    6     0.262 F2         0.609
4 F2    7     0.395 F2         0.609
```

説明変数の水準毎にデータをまとめ上げ

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを耳
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  )
```

```
# A tibble: 4 x 6
# Groups:   F [2]
  F     id       DV Fb    gross.mean lev.w
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.609
2 F1    4     1.04  F1         0.609
3 F2    6     0.262 F2         0.609
4 F2    7     0.395 F2         0.609
```

水準毎に応答変数の平均値を計算

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを耳
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup()
```

```
# A tibble: 4 x 6
  F     id       DV Fb    gross.mean lev.w
  <fct> <fct> <dbl> <fct>      <dbl>
1 F1    1     0.737 F1         0.609
2 F1    4     1.04  F1         0.609
3 F2    6     0.262 F2         0.609
4 F2    7     0.395 F2         0.609
```

まとめ上げを解除

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを耳
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  )
```

```
# A tibble: 2 x 3
  F     gross.mean lev.wise.mean
  <fct>      <dbl>         <dbl>
1 F1         0.609         0.890
2 F2         0.609         0.328
```

応答変数全ての平均と水準毎の平均を抽出

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを耳
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  )
```

```
# A tibble: 2 x 2
  gross.mean mean.of.lev.wise.mean
       <dbl>                 <dbl>
1      0.609                 0.609
2      0.609                 0.609
```

水準毎の平均の平均を計算

Balanced dataでのGrand Meanの意味

```
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  ) |>
  distinct(
    gross.mean, mean.of.lev.wis
  )
```

```
# A tibble: 1 x 2
  gross.mean mean.of.lev.wise.mean
       <dbl>                 <dbl>
1      0.609                 0.609
```

「各水準で得られた応答変数の平均値」の平均値と，全データの平均が一致する

Balanced dataでのGrand Meanの意味

```r
set.seed(1)

simdat |>
  group_split(F) |>
  map2_dfr(
#### F1・F2ともに，2つずつデータを取
    .y = c(2, 2),
    ~ slice_sample(.x, n = .y)
  ) |>
  mutate(
    gross.mean = mean(DV)
  ) |>
  group_by(F) |>
  mutate(
    lev.wise.mean = mean(DV)
  ) |>
  ungroup() |>
  distinct(
    F, gross.mean, lev.wise.mea
  ) |>
  summarise(
    gross.mean = gross.mean,
    mean.of.lev.wise.mean = mea
  ) |>
  distinct(
    gross.mean, mean.of.lev.wis
  ) |>
  kable(
    digit = 2,
```

応答変数全ての平均	「各水準の平均」の平均
0.61	0.61

「各水準で得られた応答変数の平均値」の平均値と，全データの平均が一致する

Balanced dataでのGrand Mean

各水準のデータ数が等しいデータセット（一切欠損値がないデータセット）では，

Grand Meanは，応答変数全体の平均に一致

応答変数全ての平均	「各水準の平均」の平均
0.61	0.61

Unbalanced dataでのGrand Mean

Grand Meanは，**応答変数全体の平均に一致しない**

応答変数全ての平均	「各水準の平均」の平均
0.68	0.58

(*)Cell means parameterization

- 傾き（水準間の差）や切片を推定しない
- 各水準の下で得られる応答変数の平均が0か否かを検定する
  - （2水準の場合）帰無仮説は2つ
    1. $H_0 : \mu_1 = 0$
    2. $H_0 : \mu_2 = 0$

```
m2_mr <- lm(DV ~ -1 + F, simdat)
```

Predictor	Estimate	95% CI	t-value	p-value
FF1	0.8	[0.594, 1.006]	8.9	0
FF2	0.4	[0.194, 0.606]	4.5	0

Examples of different default contrast types (pp.8--9)

さまざまな対比行列

1. Contrast matrices
2. Treatment contrast
3. Sum contrast
4. Repeated contrast
5. Polynomial contrast
6. Helmert contrast
7. 任意の対比行列を充てる方法

Contrast matrices

- 対比を表す係数値（傾き・切片）は，係数値の行列で表現される
- 係数値の行列は，対比行列で実装される

Treatment contrast 処理対比の対比行列

- 例：様々なプライミング効果を，同一の統制群と比べる
  1. プライミングなし（統制群）
  2. 音韻的プライミング（実験群1）
  3. 綴字プライミング（実験群2）

```r
#contr.treatment(3)
contr.treatment(demoDataTreatment$priming)
```

```
              orthographic phonological
control                  0            0
orthographic             1            0
phonological             0            1
```

Treatment contrast 処理対比の対比行列

- 例：様々なプライミング効果を，同一の統制群と比べる
    1. プライミングなし（統制群）
    2. 音韻的プライミング（実験群1）
    3. 綴字プライミング（実験群2）

```
#contr.treatment(3)
contr.treatment(demoDataTreatment$priming)
```

```
              orthographic phonological
control                  0            0
orthographic             1            0
phonological             0            1
```

- 各行：**水準** 上から
    1. **統制群**：プライミングなし
    2. **実験群1**：綴字プライミング
    3. **実験群2**：音韻的プライミング

Treatment contrast 処理対比の対比行列

- 例：様々なプライミング効果を，同一の統制群と比べる
    1. プライミングなし（統制群）
    2. 音韻的プライミング（実験群1）
    3. 綴字プライミング（実験群2）

```
#contr.treatment(3)
contr.treatment(demoDataTreatment$priming)
```

```
             orthographic phonological
control                 0            0
orthographic            1            0
phonological            0            1
```

- 各行：**水準** 上から
    1. **統制群**：プライミングなし
    2. **実験群1**：綴字プライミング
    3. **実験群2**：音韻的プライミング

- 各列：**対比する水準の組合せ** 左から
    1. **実験群1 vs 統制群**
       （綴字プライミング vs プライミングなし）
    2. **実験群2 vs 統制群**
       （音韻的プライミング vs プライミングなし）

Sum contrast 零和対比の対比行列

- 例：様々なプライミング効果を，全体平均と比べる
  1. 音韻的プライミング（実験群1）
  2. 綴字プライミング（実験群2）
  3. 意味プライミング

```
#contr.sum(3)
contr.sum(demoDataSum$priming)
```

```
              [,1] [,2]
orthographic     1    0
phonological     0    1
semantic        -1   -1
```

Sum contrast 零和対比の対比行列

- 例：様々なプライミング効果を，全体平均と比べる
    1. 音韻的プライミング（実験群1）
    2. 綴字プライミング（実験群2）
    3. 意味プライミング

```
#contr.sum(3)
contr.sum(demoDataSum$priming)
```

```
              [,1] [,2]
orthographic     1    0
phonological     0    1
semantic        -1   -1
```

- 各行：**水準** 上から
    1. 音韻的プライミング
    2. 綴字プライミング
    3. 意味プライミング

Sum contrast 零和対比の対比行列

- 例：様々なプライミング効果を，全体平均と比べる
  1. 音韻的プライミング（実験群1）
  2. 綴字プライミング（実験群2）
  3. 意味プライミング

```
#contr.sum(3)
contr.sum(demoDataSum$priming)
```

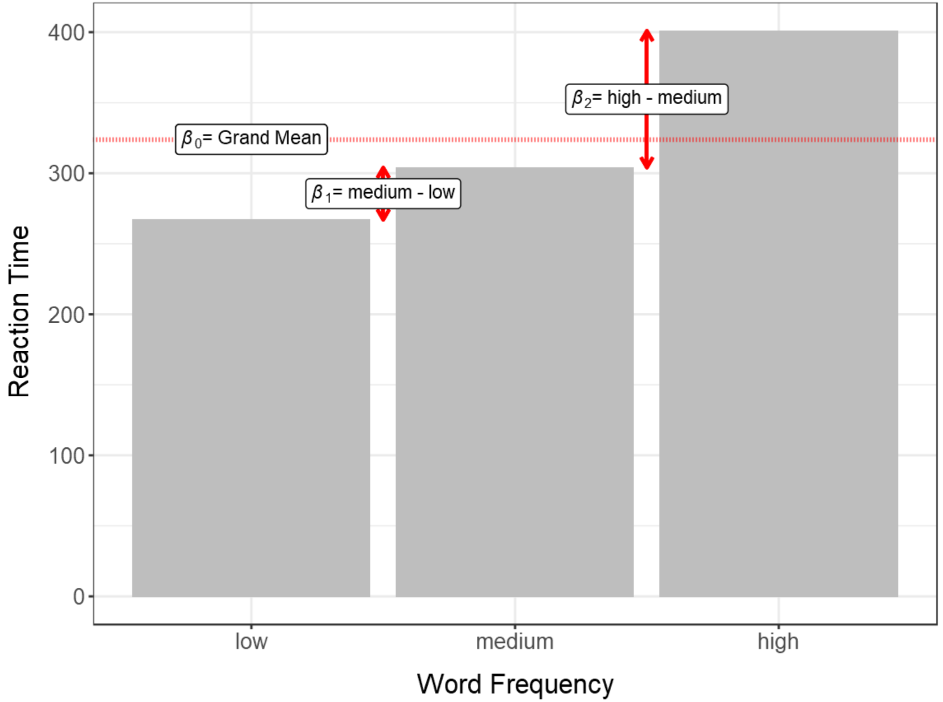
```
              [,1] [,2]
orthographic     1    0
phonological     0    1
semantic        -1   -1
```

- 各行：**水準** 上から
  1. 音韻的プライミング
  2. 綴字プライミング
  3. 意味プライミング

- 各列：**対比する水準の組合せ** 左から
  1. **実験群1 vs 全体平均**
     （音韻的プライミング vs 全体平均）
  2. **実験群2 vs 全体平均**
     （綴字プライミング vs 全体平均）

Repeated contrastの対比行列

- 例：語の頻度の効果を，隣り合う水準と比べる
  1. 低頻度
  2. 中頻度
  3. 高頻度

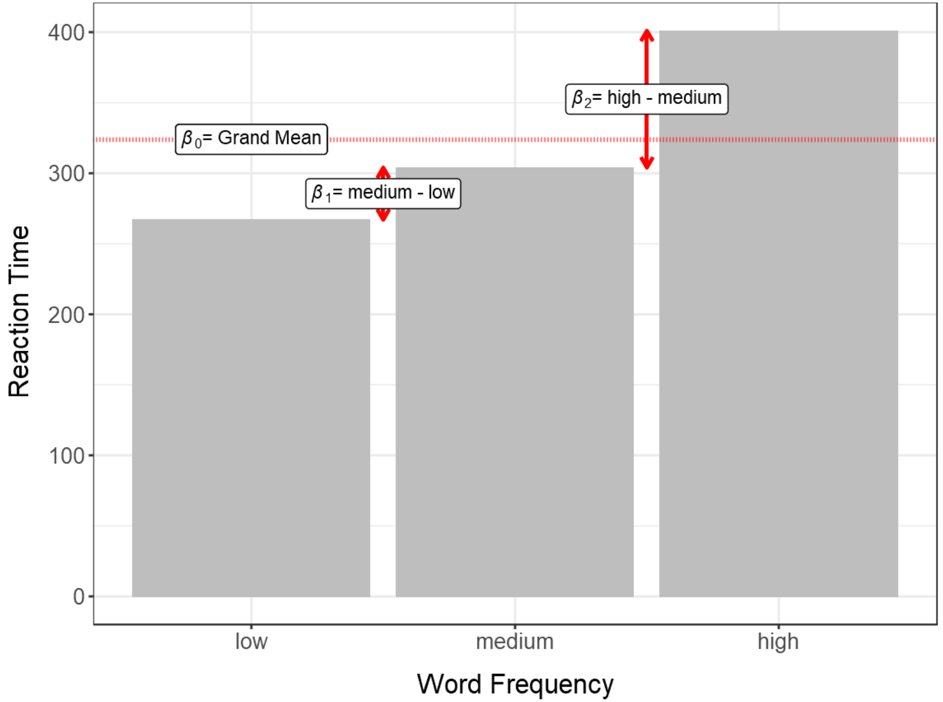



```
#contr.sdif(3)
contr.sdif(demoDataFreq$frequency)
```

```
       medium-low high-medium
low         -0.67       -0.33
medium       0.33       -0.33
high         0.33        0.67
```

Repeated contrastの対比行列

- 例：語の頻度の効果を，隣り合う水準と比べる
  1. 低頻度
  2. 中頻度
  3. 高頻度



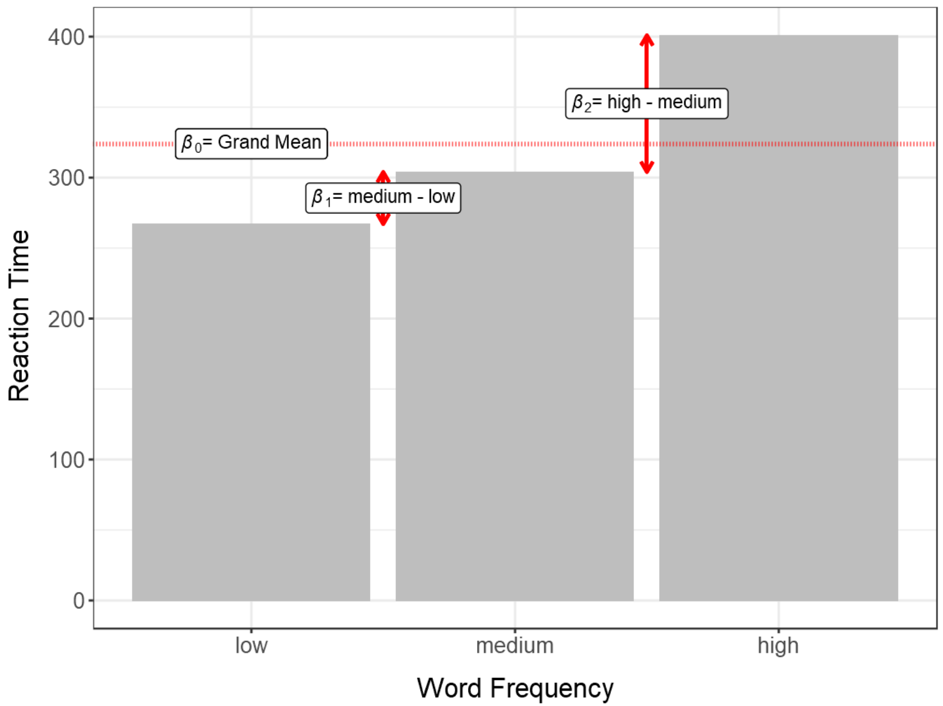

```
#contr.sdif(3)
contr.sdif(demoDataFreq$frequency)
```

```
       medium-low high-medium
low         -0.67       -0.33
medium       0.33       -0.33
high         0.33        0.67
```

- 各行：**水準** 上から
  1. 低頻度
  2. 中頻度
  3. 高頻度

Repeated contrastの対比行列

- 例：語の頻度の効果を，隣り合う水準と比べる
  1. 低頻度
  2. 中頻度
  3. 高頻度




```
#contr.sdif(3)
contr.sdif(demoDataFreq$frequency)
```

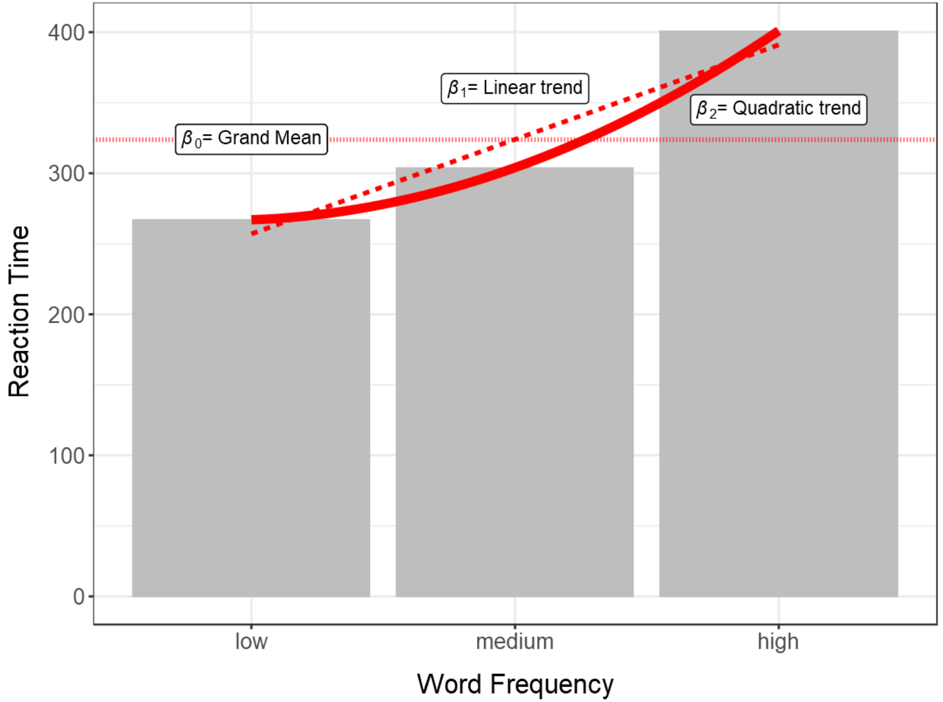
```
       medium-low high-medium
low         -0.67       -0.33
medium       0.33       -0.33
high         0.33        0.67
```

- 各行：**水準** 上から
  1. 低頻度
  2. 中頻度
  3. 高頻度

- 各列：**対比する水準の組合せ** 左から
  1. **第2水準 vs 第1水準**
     （中頻度 vs 低頻度）
  2. **第3水準 vs 第2水準**
     （高頻度 vs 中頻度）

Polynomial contrast 多項式対比の対比行列

- 例：語の頻度の効果を，隣り合う水準と比べる
    1. 低頻度
    2. 中頻度
    3. 高頻度

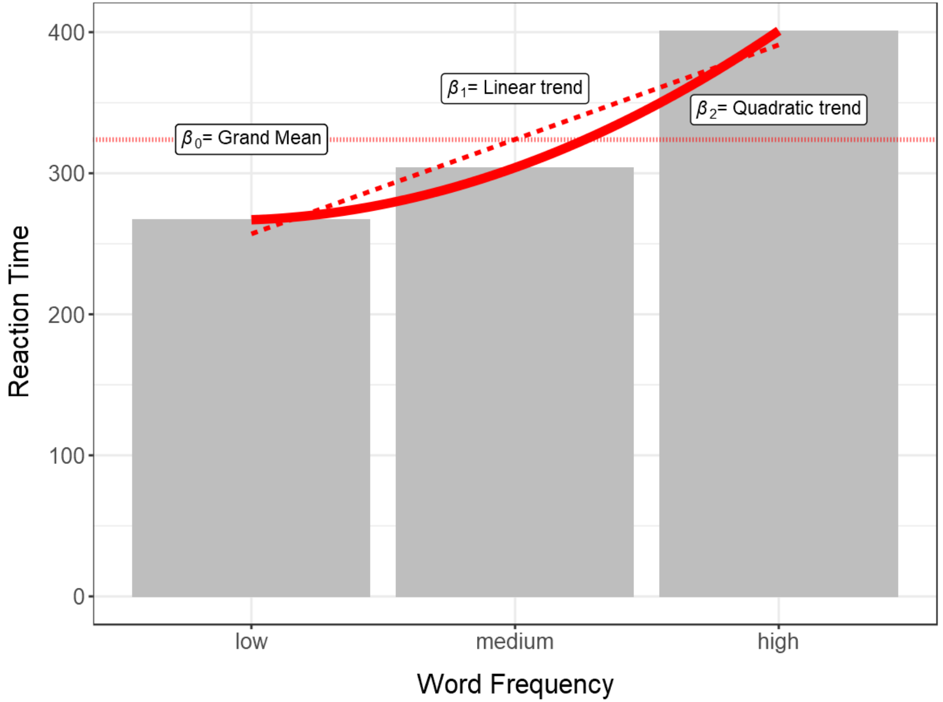



```r
#contr.poly(3)
contr.poly(demoDataFreq$frequency)
```

```
                         .L    .Q
[1,] -0.70710678118654757  0.41
[2,] -0.00000000000000079 -0.82
[3,]  0.70710678118654746  0.41
```

Polynomial contrast 多項式対比の対比行列

- 例：語の頻度の効果を，隣り合う水準と比べる
    1. 低頻度
    2. 中頻度
    3. 高頻度



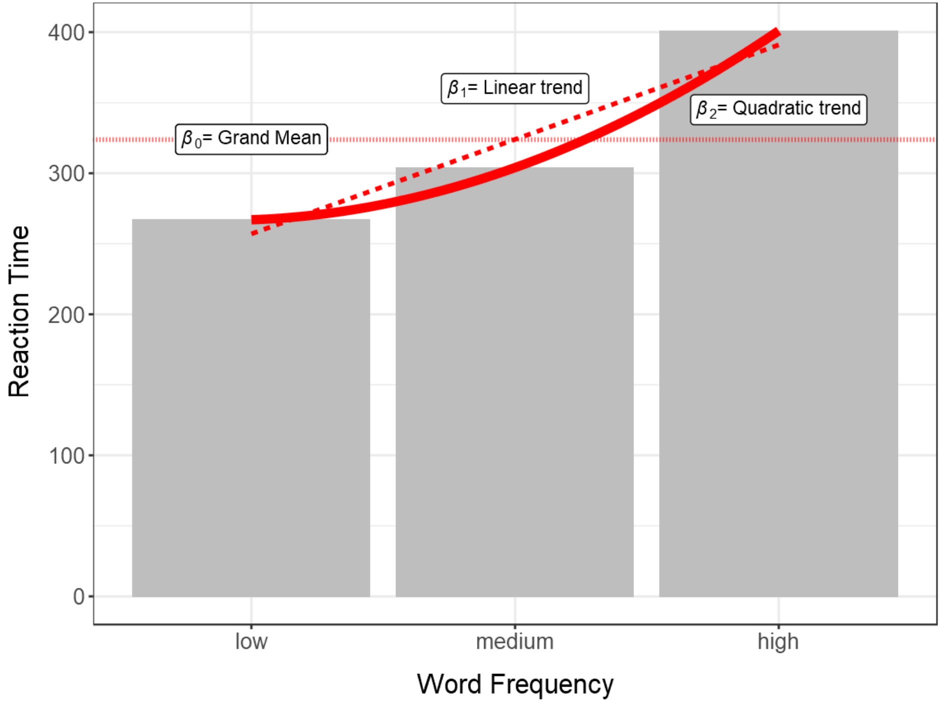

```
#contr.poly(3)
contr.poly(demoDataFreq$frequency)
```

```
                         .L    .Q
[1,] -0.70710678118654757  0.41
[2,] -0.00000000000000079 -0.82
[3,]  0.70710678118654746  0.41
```

- 各行：**水準** 上から
    1. 低頻度
    2. 中頻度
    3. 高頻度

Polynomial contrast 多項式対比の対比行列

- 例：語の頻度の効果を，隣り合う水準と比べる
    1. 低頻度
    2. 中頻度
    3. 高頻度




```
#contr.poly(3)
contr.poly(demoDataFreq$frequency)
```

```
                        .L    .Q
[1,] -0.7071067811865473  0.41
[2,] -0.0000000000000000079 -0.82
[3,]  0.7071067811865462  0.41
```

- 各行：**水準** 上から
    1. 低頻度
    2. 中頻度
    3. 高頻度

- 各列：**対比の組合せ** 左から
    1. **1次関数ペース**（Linear; どの水準間でも一定ペースで増加・減少？）
    2. **2次関数ペース**（Quadratic; 次の水準間ほどに増加・減少が激しい・穏やか？）

Helmert contrast ヘルマート対比の対比行列

- 例：様々なプライミング効果を，種類ごとに比べる
    1. 無効なプライミング1
    2. 無効なプライミング2
    3. 有効なプライミング

```
#contr.helmert(3)
contr.helmert(demoDataHelmert$priming)
```

```
         [,1] [,2]
invalid1   -1   -1
invalid2    1   -1
valid       0    2
```

Helmert contrast ヘルマート対比の対比行列

- 例：様々なプライミング効果を，種類ごとに比べる
    1. 無効なプライミング1
    2. 無効なプライミング2
    3. 有効なプライミング

```
#contr.helmert(3)
contr.helmert(demoDataHelmert$priming)
```

```
         [,1] [,2]
invalid1   -1   -1
invalid2    1   -1
valid       0    2
```

- 各行：**水準** 上から
    1. 無効なプライミング1
    2. 無効なプライミング2
    3. 有効なプライミング

Helmert contrast ヘルマート対比の対比行列

- 例：様々なプライミング効果を，種類ごとに比べる
  1. 無効なプライミング1
  2. 無効なプライミング2
  3. 有効なプライミング

```
#contr.helmert(3)
contr.helmert(demoDataHelmert$priming)
```

```
         [,1] [,2]
invalid1   -1   -1
invalid2    1   -1
valid       0    2
```

- 各行：**水準** 上から
  1. 無効なプライミング1
  2. 無効なプライミング2
  3. 有効なプライミング
- 各列：**対比する水準の組合せ** 左から
  1. **小グループ1 vs 小グループ2**
     （無効なプライミング1 vs 無効なプライミング2）
  2. **小グループ連合 vs 別グループ**
     （無効なプライミング vs 有効なプライミン

任意の対比行列を充てる方法

まずは，シミュレーション用データ`simdat2`を作成

- **応答変数**：語彙判断課題の反応時間
- **説明変数**：判断する単語の頻度
  - 頻度が低いほど，反応時間が長い
- 被験者間計画での水準数
  - 3水準
- 被験者内計画での水準数
  - なし
- 被験者間計画での1水準あたりの実験参加者数
  - 4人
  - 計 $4 \times 3 = 12$ 人
- 各水準での応答変数の平均値
  - 500, 450, 400
- 要因の標準偏差
  - 20

```r
set.seed(1212); simdat2 <- mixedDesign(
  B = 3,
  W = NULL,
  n = 4,
  M = matrix(
    c(500, 450, 400),
    nrow = 3,
    ncol = 1,
    byrow = FALSE
  ),
  SD = 20,
  long = TRUE
) |>
  rename(F = B_A) |>
  mutate(
    F = fct_recode(
      F,
      low    = "A1",
      medium = "A2",
      high   = "A3"
    )
  )
```

任意の対比行列を充てる方法

```r
contrasts(simdat2$F) <- contr.treatment(3)

contrasts(simdat2$F)

lm(DV ~ F, data = simdat2) |>
  summary()
```

```
       2 3
low    0 0
medium 1 0
high   0 1
```

```
Call:
lm(formula = DV ~ F, data = simdat2)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.875 -13.011   0.419  14.211  24.685 

Coefficients:
            Estimate Std. Error t value
(Intercept)    500.0       10.0   50.00 0.
F2             -50.0       14.1   -3.54
F3            -100.0       14.1   -7.07 0.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*
Residual standard error: 20 on 9 degrees o
Multiple R-squared:  0.847,    Adjusted R-
F-statistic:    25 on 2 and 9 DF,  p-value:
```

Treatment contrastsを適用

任意の対比行列を充てる方法

```r
contrasts(simdat2$F) <- contr.sum(3)

contrasts(simdat2$F)

lm(DV ~ F, data = simdat2) |>
  summary()
```

```
       [,1] [,2]
low       1    0
medium    0    1
high     -1   -1
```

```
Call:
lm(formula = DV ~ F, data = simdat2)

Residuals:
    Min      1Q  Median      3Q     Max
-25.875 -13.011   0.419  14.211  24.685

Coefficients:
                        Estimate
(Intercept) 450.0000000000001705   5.77350
F1           50.0000000000001066   8.16496
F2           -0.0000000000000508   8.16496

(Intercept) ***
F1          ***
F2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*

Residual standard error: 20 on 9 degrees o
Multiple R-squared:  0.847,    Adjusted R-
```

Sum contrastsを適用

任意の対比行列を充てる方法

```r
contrasts(simdat2$F) <- contr.sdif(3)

contrasts(simdat2$F)

lm(DV ~ F, data = simdat2) |>
  summary()
```

```
        2-1   3-2
low    -0.67 -0.33
medium  0.33 -0.33
high    0.33  0.67
```

```

Call:
lm(formula = DV ~ F, data = simdat2)

Residuals:
    Min      1Q  Median      3Q     Max
-25.875 -13.011   0.419  14.211  24.685

Coefficients:
            Estimate Std. Error t value
(Intercept)   450.00       5.77   77.94 0.
F2-1          -50.00      14.14   -3.54
F3-2          -50.00      14.14   -3.54
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*

Residual standard error: 20 on 9 degrees c
Multiple R-squared:  0.847,    Adjusted R-
F-statistic:    25 on 2 and 9 DF,  p-value:
```

Repeated contrastsを適用

任意の対比行列を充てる方法

```r
contrasts(simdat2$F) <- contr.poly(3)

contrasts(simdat2$F)

lm(DV ~ F, data = simdat2) |>
  summary()
```

```
                        .L    .Q
low    -0.7071067811865475730  0.41
medium -0.0000000000000000079 -0.82
high    0.7071067811865474620  0.41
```

```
Call:
lm(formula = DV ~ F, data = simdat2)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.875 -13.011   0.419  14.211  24.685 

Coefficients:
                          Estimate
(Intercept) 450.0000000000001705    5.77350
F.L         -70.7106781186548830   10.00000
F.Q           0.0000000000000631   10.00000

(Intercept) ***
F.L         ***
F.Q
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*

Residual standard error: 20 on 9 degrees o
Multiple R-squared:  0.847,    Adjusted R-
```

Polynomial contrastsを適用

任意の対比行列を充てる方法

```r
contrasts(simdat2$F) <- contr.helmert(3)

contrasts(simdat2$F)

lm(DV ~ F, data = simdat2) |>
  summary()
```

```
       [,1] [,2]
low      -1   -1
medium    1   -1
high      0    2
```

```

Call:
lm(formula = DV ~ F, data = simdat2)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.875 -13.011   0.419  14.211  24.685 

Coefficients:
            Estimate Std. Error t value 
(Intercept)   450.00       5.77   77.94 0.
F1            -25.00       7.07   -3.54 
F2            -25.00       4.08   -6.12 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*

Residual standard error: 20 on 9 degrees o
Multiple R-squared:  0.847,	Adjusted R-
F-statistic:   25 on 2 and 9 DF,  p-value:
```

Helmert contrastsを適用

The hypothesis matrix illustrated with a three-level factor (pp.9--12)

さまざまな対比行列

1. シミュレーション用データについて
2. Sum contrasts
3. The hypothesis matrix
    1. 検証したい仮説を用意する
    2. 重みづけを抽出し，「仮説行列」 *hypothesis matrix* に書く
    3. 一般化逆行列 *generalised matrix inverse* を使い，仮説行列を対比行列に変換
    4. 対比行列を要因に当てはめ，線形モデルを走らせる

シミュレーション用データについて

- データ：`simdat2`
  - 本資料では，前のセクションで作成済
- **応答変数**：語彙判断課題の反応時間
- **説明変数**：判断する単語の頻度
  - 頻度が低いほど，反応時間が長い
- 被験者間計画での水準数
  - 3水準（低頻度・中頻度・高頻度）
- 被験者内計画での水準数
  - なし
- 被験者間計画での1水準あたりの実験参加者数
  - 4人
  - 計 $4 \times 3 = 12$ 人
- 各水準での応答変数の平均値
  - 500, 450, 400
- 要因の標準偏差
  - 20

検証する仮説

1. 低頻度語への反応時間 μ_1 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - 全体平均より長い？短い？
2. 中頻度語への反応時間 μ_2 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - 全体平均より長い？短い？

$$H_{01}: \overbrace{\mu_1}^{\text{低頻度語への反応時間}} = \overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}$$

Sum contrast

μ_1, μ_2, μ_3 には重みづけがある

$$H_{01}: \overbrace{\mu_1}^{\text{低頻度語への反応時間}} = \overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}$$

$$\Leftrightarrow \mu_1 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3$$

$$\Leftrightarrow \frac{2}{3}\mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 = 0$$

$$(\underbrace{\frac{2}{3}}_{\mu_1\text{への重み}}\mu_1) + (\underbrace{-\frac{1}{3}}_{\mu_2\text{への重み}}\mu_2) + (\underbrace{-\frac{1}{3}}_{\mu_3\text{への重み}}\mu_3) = 0$$

$$H_{02}: \overbrace{\mu_2}^{\text{中頻度語への反応時間}} = \overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}$$

$$\Leftrightarrow \mu_2 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3$$

$$\Leftrightarrow -\frac{1}{3}\mu_1 + \frac{2}{3}\mu_2 - \frac{1}{3}\mu_3 = 0$$

$$(\underbrace{-\frac{1}{3}}_{\mu_1\text{への重み}}\mu_1) + (\underbrace{\frac{2}{3}}_{\mu_2\text{への重み}}\mu_2) + (\underbrace{-\frac{1}{3}}_{\mu_3\text{への重み}}\mu_3) = 0$$

Hypotheis matrix

- ある対比を使って本当に仮説を正しく検証できるか確認する
  - 例：
  - 「低頻度語への反応時間 μ_1 が，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？」を調べるのに，$H_{01}: \mu_1 = \frac{\mu_1+\mu_2+\mu_3}{3}$ という対比を使って問題ないか
- 検証したい帰無仮説や使いたい実験デザインによっては，Rの`contr.*()`関数で作れる対比行列を当てはめられないことがある
  - **どのように対比行列が作られるのか理解する意義がある**

対比行列を自分で作る

1. 検証したい帰無仮説を用意する
2. 重みづけ（例：sum contrastであればこの重みづけ）を抽出し，「仮説行列」 *hypothesis matrix* に書く
3. 一般化逆行列 *generalised matrix inverse* を使い，仮説行列を対比行列に変換
4. 対比行列を要因に当てはめ，線形モデルを実行

仮説行列?? →

1	0
0	1
-1	-1

検証したい帰無仮説を用意する

- **応答変数**：語彙判断課題の反応時間
- **説明変数**：判断する単語の頻度
  - 頻度が低いほど，反応時間が長い
- 被験者間計画での水準数
  - 3水準（低頻度・中頻度・高頻度）

$$H_{0_1}: \overbrace{\mu_1}^{\text{低頻度語への反応時間}} = \overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}$$

$$H_{0_2}: \overbrace{\mu_2}^{\text{中頻度語への反応時間}} = \overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}$$

$$H_{0_0}: \overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}} = 0$$

検証する仮説

1. 低頻度語への反応時間 μ_1 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - 全体平均より長い？短い？
2. 中頻度語への反応時間 μ_2 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - 全体平均より長い？短い？
3. 全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ は0秒と同じ？

検証したい帰無仮説を用意する

研究上の問い

- **応答変数**：語彙判断課題の反応時間
- **説明変数**：判断する単語の頻度
  - 頻度が低いほど，反応時間が長い
- 被験者間計画での水準数
  - 3水準（低頻度・中頻度・高頻度）

検証したい帰無仮説を用意する

検証する仮説

1. 低頻度語への反応時間 μ_1 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
    - 全体平均より長い？短い？
2. 中頻度語への反応時間 μ_2 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
    - 全体平均より長い？短い？
3. 全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ は0秒と同じ？

$$H_{0_1}:\overbrace{\mu_1}^{\text{低頻度語への反応時間}}=\overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}$$

$$H_{0_2}:\overbrace{\mu_2}^{\text{中頻度語への反応時間}}=\overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}$$

$$H_{0_0}:\overbrace{\frac{\mu_1+\mu_2+\mu_3}{3}}^{\text{全体平均}}=0$$

検証したい帰無仮説を用意する

検証する仮説

1. 低頻度語への反応時間 μ_1 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - 全体平均より長い？短い？
2. 中頻度語への反応時間 μ_2 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - 全体平均より長い？短い？
3. 全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ は0秒と同じ？

$$H_{01} : (\underbrace{\frac{2}{3}}_{\mu_1\text{への重み}} \mu_1) + (\underbrace{-\frac{1}{3}}_{\mu_2\text{への重み}} \mu_2) + (\underbrace{-\frac{1}{3}}_{\mu_3\text{への重み}} \mu_3) = 0$$

$$H_{02} : (\underbrace{-\frac{1}{3}}_{\mu_1\text{への重み}} \mu_1) + (\underbrace{\frac{2}{3}}_{\mu_2\text{への重み}} \mu_2) + (\underbrace{-\frac{1}{3}}_{\mu_3\text{への重み}} \mu_3) = 0$$

$$H_{00} : (\underbrace{\frac{1}{3}}_{\mu_1\text{への重み}} \mu_1) + (\underbrace{\frac{1}{3}}_{\mu_2\text{への重み}} \mu_2) + (\underbrace{\frac{1}{3}}_{\mu_3\text{への重み}} \mu_3) = 0$$

重みづけを抽出し，「仮説行列」に書く

μ_1, μ_2, μ_3 には重みづけがある

$$H_{01} : (\underbrace{\frac{2}{3}}_{\mu_1\text{への重み}}\mu_1) + (\underbrace{-\frac{1}{3}}_{\mu_2\text{への重み}}\mu_2) + (\underbrace{-\frac{1}{3}}_{\mu_3\text{への重み}}\mu_3) = 0$$

$$H_{02} : (\underbrace{-\frac{1}{3}}_{\mu_1\text{への重み}}\mu_1) + (\underbrace{\frac{2}{3}}_{\mu_2\text{への重み}}\mu_2) + (\underbrace{-\frac{1}{3}}_{\mu_3\text{への重み}}\mu_3) = 0$$

$$H_{00} : (\underbrace{\frac{1}{3}}_{\mu_1\text{への重み}}\mu_1) + (\underbrace{\frac{1}{3}}_{\mu_2\text{への重み}}\mu_2) + (\underbrace{\frac{1}{3}}_{\mu_3\text{への重み}}\mu_3) = 0$$

- 各行：**帰無仮説**
  - H_{00}: `cH00`
  - H_{01}: `cH01`
  - H_{02}: `cH02`
- 各列：各水準への**重み**

```
HcSum <- rbind(
  cH00 = c(low =  1/3, med =  1/3, hi =  1/3),
  cH01 = c(low =  2/3, med = -1/3, hi = -1/3),
  cH02 = c(low = -1/3, med =  2/3, hi = -1/3)
)
```

重みづけを抽出し，「仮説行列」に書く

```
HcSum
```

```
       low   med    hi
cH00  0.33  0.33  0.33
cH01  0.67 -0.33 -0.33
cH02 -0.33  0.67 -0.33
```

仮説行列

重みづけを抽出し，「仮説行列」に書く

```
HcSum |>
  fractions()
```

```
      low  med  hi
cH00  1/3  1/3  1/3
cH01  2/3 -1/3 -1/3
cH02 -1/3  2/3 -1/3
```

小数表記を分数表記に改める

重みづけを抽出し，「仮説行列」に書く

```
HcSum |>
  fractions()

HcSum
```

```
     low  med  hi
cH00  1/3  1/3  1/3
cH01  2/3 -1/3 -1/3
cH02 -1/3  2/3 -1/3
```

```
       low   med    hi
cH00  0.33  0.33  0.33
cH01  0.67 -0.33 -0.33
cH02 -0.33  0.67 -0.33
```

仮説行列

重みづけを抽出し，「仮説行列」に書く

```
HcSum |>
  fractions()

HcSum |>
  t()
```

```
      low  med  hi
cH00  1/3  1/3  1/3
cH01  2/3 -1/3 -1/3
cH02 -1/3  2/3 -1/3
```

```
    cH00  cH01  cH02
low 0.33  0.67 -0.33
med 0.33 -0.33  0.67
hi  0.33 -0.33 -0.33
```

仮説行列を転置 *transpose*
今度は各行が水準，各列が仮説に変わる

重みづけを抽出し，「仮説行列」に書く

```
HcSum |>
  fractions()

HcSum |>
  t() |>
  fractions()
```

```
      low  med  hi
cH00  1/3  1/3  1/3
cH01  2/3 -1/3 -1/3
cH02 -1/3  2/3 -1/3
```

```
     cH00 cH01 cH02
low  1/3  2/3 -1/3
med  1/3 -1/3  2/3
hi   1/3 -1/3 -1/3
```

小数表記を分数表記に改める

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcSum <- HcSum

XcSum
```

```
        low   med    hi
cH00  0.33  0.33  0.33
cH01  0.67 -0.33 -0.33
cH02 -0.33  0.67 -0.33
```

仮説行列に対し…

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcSum <- HcSum |>
  hypr::ginv2()

XcSum
```

```
    cH00 cH01 cH02
low  1    1    0
med  1    0    1
hi   1   -1   -1
```

一般化逆行列を適用，仮説行列を対比行列に変換

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcSum <- HcSum |>
  hypr::ginv2()

XcSum <- HcSum

XcSum
```

```
        low   med    hi
cH00   0.33  0.33  0.33
cH01   0.67 -0.33 -0.33
cH02  -0.33  0.67 -0.33
```

仮説行列に対し…

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcSum <- HcSum |>
  hypr::ginv2()

XcSum <- HcSum |>
  ginv()

XcSum
```

```
     [,1] [,2]                 [,3]
[1,]    1    1 -0.00000000000000022
[2,]    1    0  0.99999999999999956
[3,]    1   -1 -1.00000000000000044
```

一般化逆行列を適用，仮説行列を対比行列に変換

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcSum <- HcSum |>
  hypr::ginv2()

XcSum <- HcSum |>
  ginv() |>
  provideDimnames(base = dimnam

XcSum
```

```
     cH00 cH01                cH02
low     1    1 -0.000000000000000022
med     1    0  0.999999999999999956
hi      1   -1 -1.000000000000000044
```

行名・列名を再付与

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcSum <- HcSum |>
  hypr::ginv2()

XcSum <- HcSum |>
  ginv() |>
  provideDimnames(base = dimnam
  fractions()

XcSum
```

```
    cH00 cH01 cH02
low  1    1    0
med  1    0    1
hi   1   -1   -1
```

分数表記に

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcSum
```

```
    cH00 cH01 cH02
low  1    1    0
med  1    0    1
hi   1   -1   -1
```

仮説行列に一般化逆行列を適用したもの

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcSum

contr.sum(3)
```

```
    cH00 cH01 cH02
low  1    1    0
med  1    0    1
hi   1   -1   -1
```

```
  [,1] [,2]
1    1    0
2    0    1
3   -1   -1
```

3水準の時のsum contrast

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcSum

contr.sum(3)

1
```

```
    cH00 cH01 cH02
low    1    1    0
med    1    0    1
hi     1   -1   -1
```

```
  [,1] [,2]
1    1    0
2    0    1
3   -1   -1
```

```
[1] 1
```

1を…

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcSum

contr.sum(3)

1 |>
  cbind(contr.sum(3))
```

```
    cH00 cH01 cH02
low    1    1    0
med    1    0    1
hi     1   -1   -1
```

```
  [,1] [,2]
1    1    0
2    0    1
3   -1   -1
```

```
  [,1] [,2] [,3]
1    1    1    0
2    1    0    1
3    1   -1   -1
```

3水準の時のsum contrastの行列の1列目に付与

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcSum

contr.sum(3)

1 |>
  cbind(contr.sum(3)) |>
  fractions()
```

```
    cH00 cH01 cH02
low  1    1    0
med  1    0    1
hi   1   -1   -1
```

```
  [,1] [,2]
1    1    0
2    0    1
3   -1   -1
```

```
  [,1] [,2] [,3]
1  1    1    0
2  1    0    1
3  1   -1   -1
```

分数表記に

対比行列を要因に当てはめ，線形モデルを実行

```r
contrasts(simdat2$F) <- XcSum[, colnames(XcSum) != "

contrasts(simdat2$F)

lm(DV ~ F, data = simdat2) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(2),
      ", ",
      conf.high |> round(2),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

```
       cH01 cH02
low       1    0
medium    0    1
high     -1   -1
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	450	[436.94, 463.06]	77.9	0
FcH01	50	[31.53, 68.47]	6.1	0
FcH02	0	[-18.47, 18.47]	0.0	1

仮説行列から作った
Sum contrastsを適用

対比行列を要因に当てはめ，線形モデルを実行

```r
contrasts(simdat2$F) <- contr.sum(3)

contrasts(simdat2$F)

lm(DV ~ F, data = simdat2) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(2),
      ", ",
      conf.high |> round(2),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

```
       [,1] [,2]
low       1    0
medium    0    1
high     -1   -1
```

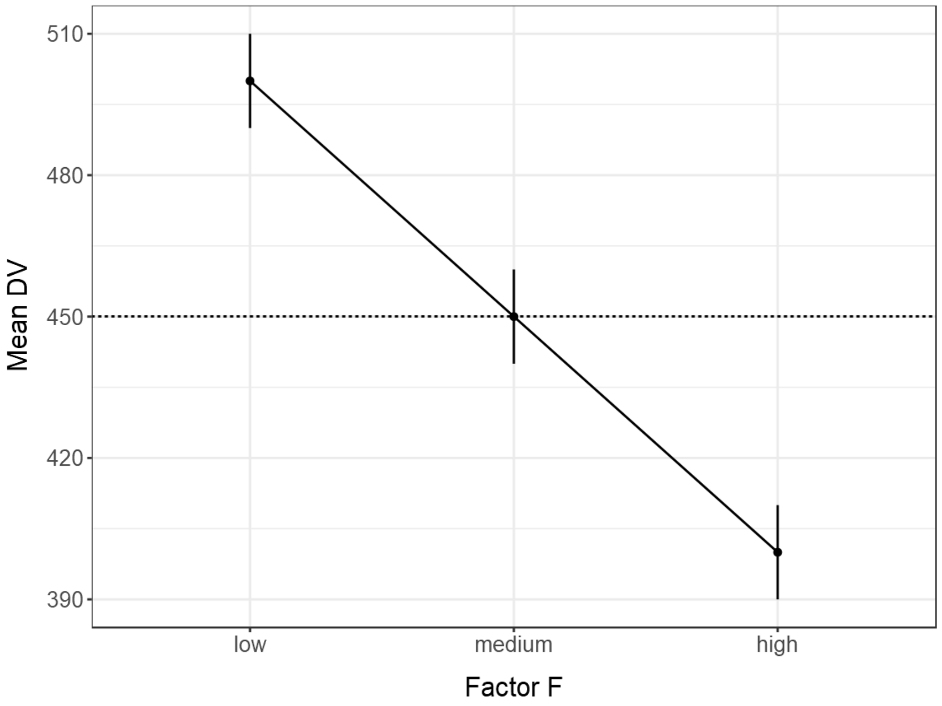
Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	450	[436.94, 463.06]	77.9	0
F1	50	[31.53, 68.47]	6.1	0
F2	0	[-18.47, 18.47]	0.0	1

contr.sum(3)で作った
Sum contrastsを適用

統計モデルから仮説を検証

1. 低頻度語への反応時間 μ_1 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - `FcH01`より， $\mu_1 - \frac{\mu_1+\mu_2+\mu_3}{3} = 50$
   - t 値が $t(9)$ の両側2.5% $t \leq -2.26, 2.26 \leq t$ にあるため，低頻度語への反応時間は全体平均より有意に長い
2. 中頻度語への反応時間 μ_2 は，全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ と同じか？
   - `FcH02`より， $\mu_2 - \frac{\mu_1+\mu_2+\mu_3}{3} = 0$
   - t 値が $t(9)$ の両側2.5%にないため，低頻度語への反応時間と全体平均の差は有意ではない
3. 全体平均 $\frac{\mu_1+\mu_2+\mu_3}{3}$ は0秒と同じ？
   - `(Intercept)`より， $\frac{\mu_1+\mu_2+\mu_3}{3} = 450$

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	450	[437, 463]	77.9	0
FcH01	50	[32, 68]	6.1	0
FcH02	0	[-18, 18]	0.0	1



Further examples of contrasts illustrated with a factor with four levels (pp.12--16)

1要因4水準のデータに，さまざまな対比行列を当てはめる

1. Repeated contrast
2. Contrasts in linear regression analysis: The design or model matrix
3. Polynomial contrasts
4. Custom contrasts

シミュレーション用データについて

- データ：`simdat3`
  - Heister, Würzner, & Kliegl (2012)の実験結果に基づくシミュレーション
- **応答変数**：fixation duration
- **説明変数**：単語の頻度
- 被験者間計画での水準数
  - 4水準
- 被験者内計画での水準数
  - なし
- 被験者間計画での1水準あたりの実験参加者数
  - 5人
  - 計 $4 \times 5 = 20$ 人
- 各水準での応答変数の平均値
  - 10, 20, 10, 40
- 要因の標準偏差
  - 10

```
set.seed(1212); simdat3 <- mixedDesign(
  B = 4,
  W = NULL,
  n = 5,
  M = matrix(
    c(10, 20, 10, 40),
    nrow = 4,
    ncol = 1,
    byrow = FALSE
  ),
  SD = 10,
  long = TRUE
) |>
  rename(F = B_A) |>
  mutate(
    F = fct_recode(
      F,
      low             = "A1",
      `medium-low`    = "A2",
      `medium-high`   = "A3",
      high            = "A4"
    )
  )
```

シミュレーションの結果を図や表で確認

```
simdat3
```

```
             F id   DV
1          low  1  3.6
2          low  2 11.3
3          low  3  7.3
4          low  4  1.2
5          low  5 26.5
6   medium-low  6  9.2
7   medium-low  7 36.5
8   medium-low  8 17.9
9   medium-low  9 18.0
10  medium-low 10 18.4
11 medium-high 11 -3.7
12 medium-high 12 13.7
13 medium-high 13 22.6
14 medium-high 14  4.4
15 medium-high 15 13.0
16        high 16 51.2
17        high 17 34.0
18        high 18 35.9
19        high 19 28.9
20        high 20 50.0
```

データセット

シミュレーションの結果を図や表で確認

```r
simdat3 |>
  group_by(F)
```

```
# A tibble: 20 x 3
# Groups:   F [4]
   F           id       DV
   <fct>       <fct> <dbl>
 1 low         1      3.57
 2 low         2     11.3 
 3 low         3      7.34
 4 low         4      1.22
 5 low         5     26.5 
 6 medium-low  6      9.24
 7 medium-low  7     36.5 
 8 medium-low  8     17.9 
 9 medium-low  9     18.0 
10 medium-low  10    18.4 
11 medium-high 11    -3.71
12 medium-high 12    13.7 
13 medium-high 13    22.6 
14 medium-high 14     4.44
15 medium-high 15    13.0 
16 high        16    51.2 
17 high        17    34.0 
18 high        18    35.9 
19 high        19    28.9 
20 high        20    50.0 
```

説明変数の水準毎にデータをまとめ上げ

シミュレーションの結果を図や表で確認

```
simdat3 |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  )
```

```
# A tibble: 4 x 5
  F               N     M    SD    SE
  <fct>       <int> <dbl> <dbl> <dbl>
1 low             5    10    10  4.47
2 medium-low      5    20    10  4.47
3 medium-high     5    10    10  4.47
4 high            5    40    10  4.47
```

水準毎に，

- データの個数**N**，
- 応答変数の平均値**M**，
- 応答変数の平均値の標準偏差**SD**，
- 応答変数の平均値の標準偏差**SE**

を計算

シミュレーションの結果を図や表で確認

```
simdat3 |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup()
```

```
# A tibble: 4 x 5
  F                N     M    SD    SE
  <fct>        <int> <dbl> <dbl> <dbl>
1 low              5    10    10  4.47
2 medium-low       5    20    10  4.47
3 medium-high      5    10    10  4.47
4 high             5    40    10  4.47
```

まとめ上げを解除

シミュレーションの結果を図や表で確認

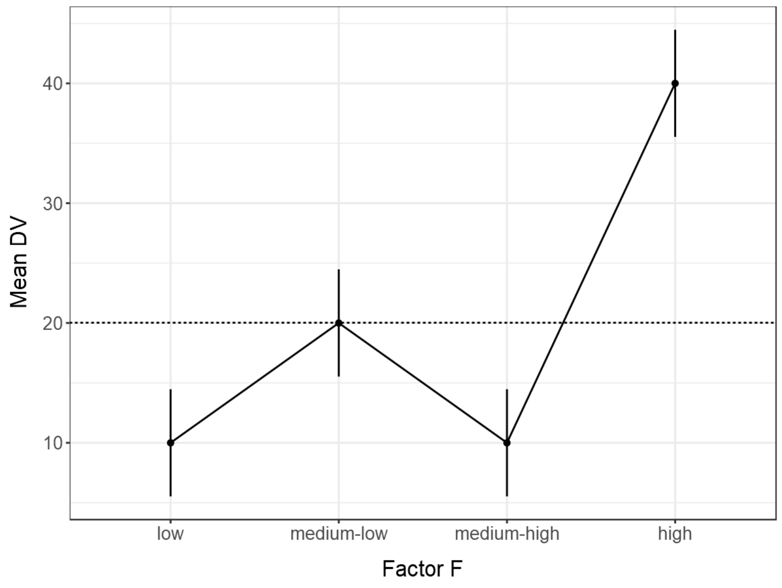
```
simdat3 |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup() |>
  mutate(
    GM = mean(M)
  )
```

```
# A tibble: 4 x 6
  F               N     M    SD    SE     G
  <fct>       <int> <dbl> <dbl> <dbl> <dbl
1 low             5    10    10  4.47     2
2 medium-low      5    20    10  4.47     2
3 medium-high     5    10    10  4.47     2
4 high            5    40    10  4.47     2
```

線グラフを描画

シミュレーションの結果を図や表で確認

```
simdat3 |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup() |>
  mutate(
    GM = mean(M)
  ) |>
  (
    \(d){
      qplot(
        x = F, y = M,
        group = 1,
        data = d,
        geom = c("point", "line")
      ) +
      geom_errorbar(
        aes(
          max = M + SE,
          min = M - SE
        ),
        width = 0
      ) +
      labs(
        y = "Mean DV",
```



誤差範囲 ($\pm 1SE$) を描画

※ $\pm 1SE$ 自体は95%信頼区間ではない（68%信頼区間）

※信頼区間にする場合，$\pm 1.96SE$

シミュレーションの結果を図や表で確認

```
simdat3
```

```
             F id   DV
1          low  1  3.6
2          low  2 11.3
3          low  3  7.3
4          low  4  1.2
5          low  5 26.5
6   medium-low  6  9.2
7   medium-low  7 36.5
8   medium-low  8 17.9
9   medium-low  9 18.0
10  medium-low 10 18.4
11 medium-high 11 -3.7
12 medium-high 12 13.7
13 medium-high 13 22.6
14 medium-high 14  4.4
15 medium-high 15 13.0
16        high 16 51.2
17        high 17 34.0
18        high 18 35.9
19        high 19 28.9
20        high 20 50.0
```

データセット

シミュレーションの結果を図や表で確認

```
simdat3 |>
  group_by(F)
```

```
# A tibble: 20 x 3
# Groups:   F [4]
   F           id       DV
   <fct>       <fct> <dbl>
 1 low         1      3.57
 2 low         2     11.3
 3 low         3      7.34
 4 low         4      1.22
 5 low         5     26.5
 6 medium-low  6      9.24
 7 medium-low  7     36.5
 8 medium-low  8     17.9
 9 medium-low  9     18.0
10 medium-low  10    18.4
11 medium-high 11    -3.71
12 medium-high 12    13.7
13 medium-high 13    22.6
14 medium-high 14     4.44
15 medium-high 15    13.0
16 high        16    51.2
17 high        17    34.0
18 high        18    35.9
19 high        19    28.9
20 high        20    50.0
```

説明変数の水準毎にデータをまとめ上げ

シミュレーションの結果を図や表で確認

```
simdat3 |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  )
```

```
# A tibble: 4 x 5
  F               N     M    SD    SE
  <fct>       <int> <dbl> <dbl> <dbl>
1 low             5    10    10  4.47
2 medium-low      5    20    10  4.47
3 medium-high     5    10    10  4.47
4 high            5    40    10  4.47
```

水準毎に，

- データの個数**N**，
- 応答変数の平均値**M**，
- 応答変数の平均値の標準偏差**SD**，
- 応答変数の平均値の標準偏差**SE**

を計算

シミュレーションの結果を図や表で確認

```
simdat3 |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup()
```

```
# A tibble: 4 x 5
  F                N     M    SD    SE
  <fct>        <int> <dbl> <dbl> <dbl>
1 low              5    10    10  4.47
2 medium-low       5    20    10  4.47
3 medium-high      5    10    10  4.47
4 high             5    40    10  4.47
```

まとめ上げを解除

シミュレーションの結果を図や表で確認

```
simdat3 |>
  group_by(F) |>
  summarise(
    N = length(DV),
    M = mean(DV),
    SD = sd(DV),
    SE = SD / sqrt(N)
  ) |>
  ungroup() |>
  kable(
    digits = 2,
    col.names = c(
      'Levels of Factor',
      'N. of data points',
      'Mean RT',
      'Std. Dev.',
      'Std. Err.'
    )
  )
```

Levels of Factor	N. of data points	Mean RT	Std. Dev.	Std. Err.
low	5	10	10	4.5
medium-low	5	20	10	4.5
medium-high	5	10	10	4.5
high	5	40	10	4.5

表の出力

- **digits**：有効数字の桁数を指定
- **col.names**：表内での各列の名前を変更

Repeated contrasts

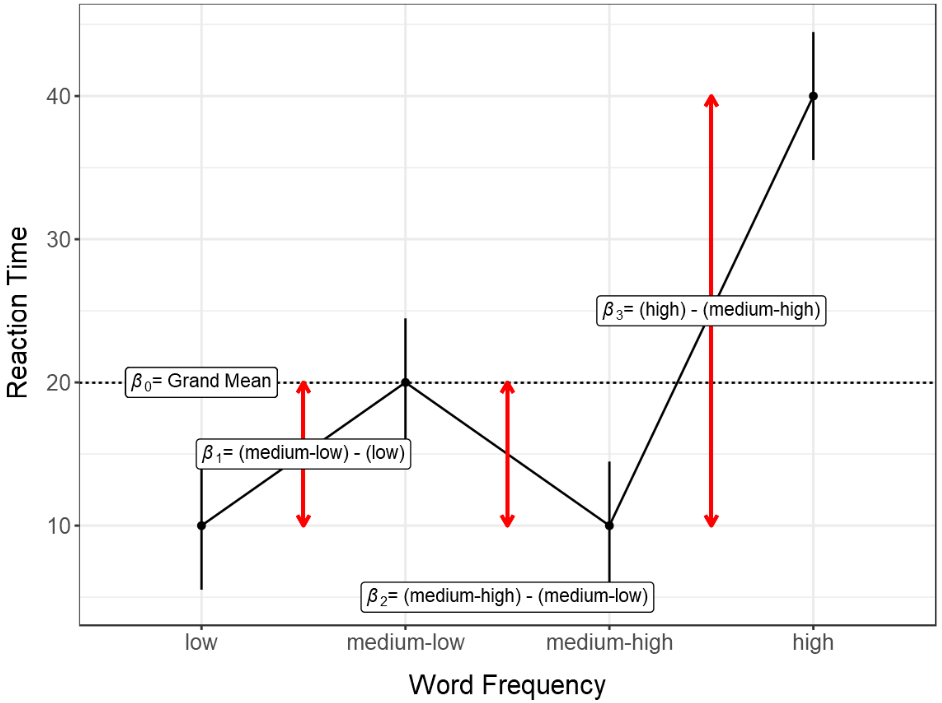
- （順序が決まっている）隣り合う水準同士を比較
  - 実験群1 vs 実験群2
  - 実験群2 vs 実験群3
  - 実験群3 vs 実験群4...

Repeated contrasts

- （順序が決まっている）隣り合う水準同士を比較
  - 実験群1 vs 実験群2
  - 実験群2 vs 実験群3
  - 実験群3 vs 実験群4...
- 例：語の頻度の効果を，隣り合う水準と比べる
  1. 低頻度
  2. 中頻度・低
  3. 中頻度・高
  4. 高頻度

Repeated contrasts

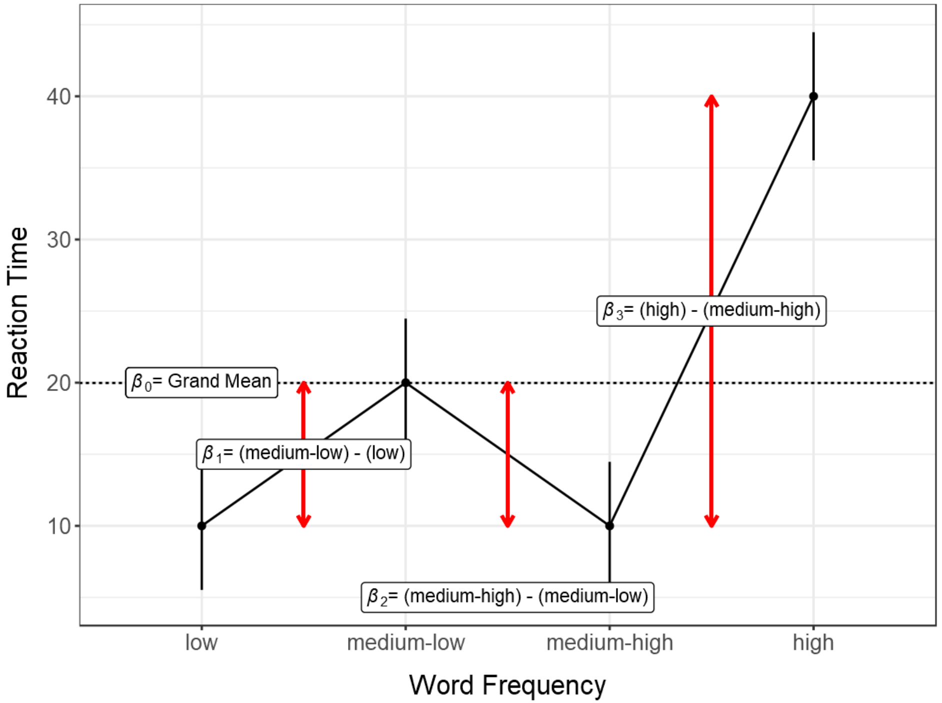
- （順序が決まっている）隣り合う水準同士を比較
    - 実験群1 vs 実験群2
    - 実験群2 vs 実験群3
    - 実験群3 vs 実験群4...
- 例：語の頻度の効果を，隣り合う水準と比べる
    1. 低頻度
    2. 中頻度・低
    3. 中頻度・高
    4. 高頻度
- 主な対比は次の3つ
    1. 中頻度・低のRT < 低頻度のRT
    2. 中頻度・高のRT < 中頻度・低のRT
    3. 高頻度のRT < 中頻度・高のRT



Repeated contrasts

検証する帰無仮説

1. 中頻度・低のRT μ_2 は，低頻度語へのRT μ_1 と同じか？
2. 中頻度・高のRT μ_3 は，中頻度・低のRT μ_2 と同じか？
3. 高頻度のRT μ_4 は， 中頻度・高のRT μ_3 と同じか？

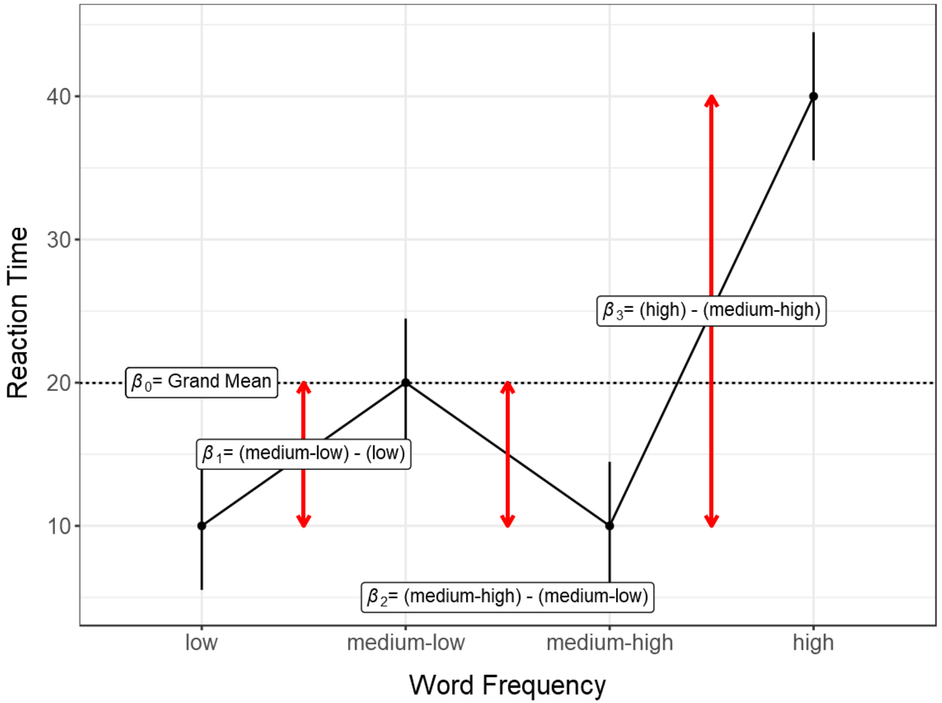


Repeated contrasts

検証する帰無仮説

言葉を数式に変換

$$H_{0_{2-1}}: \mu_2 = \mu_1$$
$$H_{0_{3-2}}: \mu_3 = \mu_2$$
$$H_{0_{4-3}}: \mu_4 = \mu_3$$

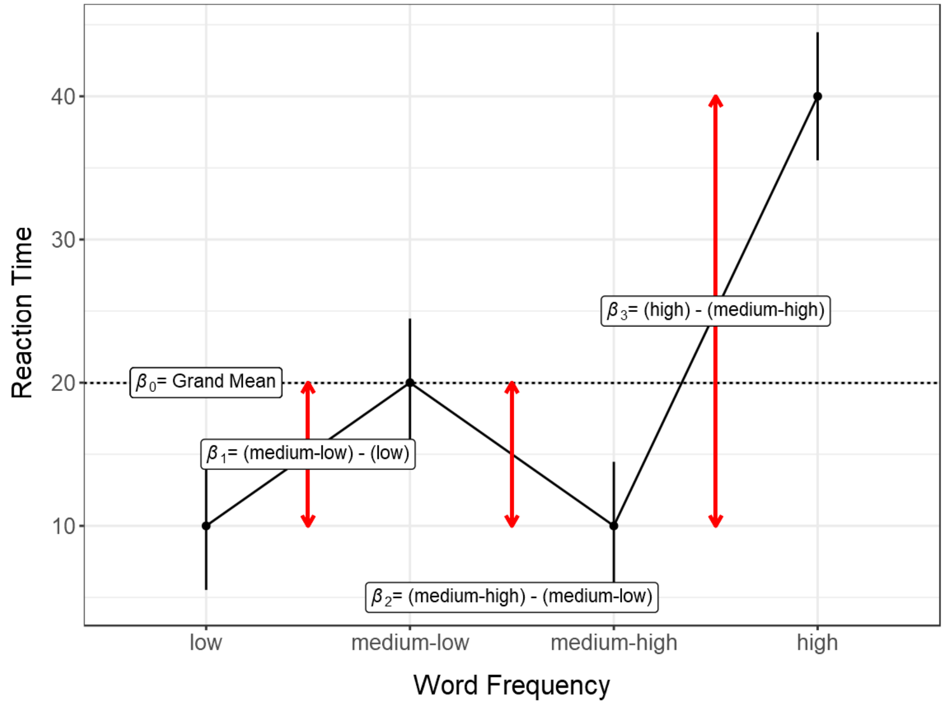


Repeated contrasts

検証する帰無仮説

両辺から右辺部分を引く

$$H_{0_{2-1}}: \mu_2 - \mu_1 = 0$$
$$H_{0_{3-2}}: \mu_3 - \mu_2 = 0$$
$$H_{0_{4-3}}: \mu_4 - \mu_3 = 0$$

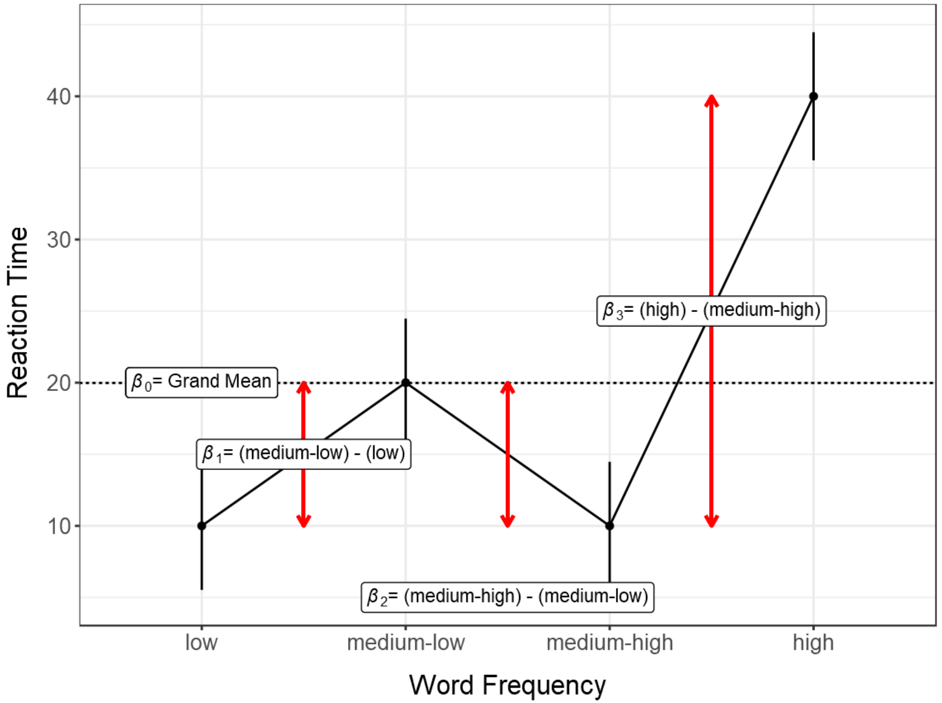


Repeated contrasts

検証する帰無仮説

- $\mu_1, \ldots, \mu_4$ の内，各式に存在しないものを書き加える
- 書き加えるときには，追加するものに0を掛けてから

$$
\begin{aligned}
H_{0 2-1} &: \mu_2 - \mu_1 + 0 \times \mu_3 + 0 \times \mu_4 = 0 \\
H_{0 3-2} &: \mu_3 - \mu_2 + 0 \times \mu_1 + 0 \times \mu_4 = 0 \\
H_{0 4-3} &: \mu_4 - \mu_3 + 0 \times \mu_1 + 0 \times \mu_2 = 0
\end{aligned}
$$

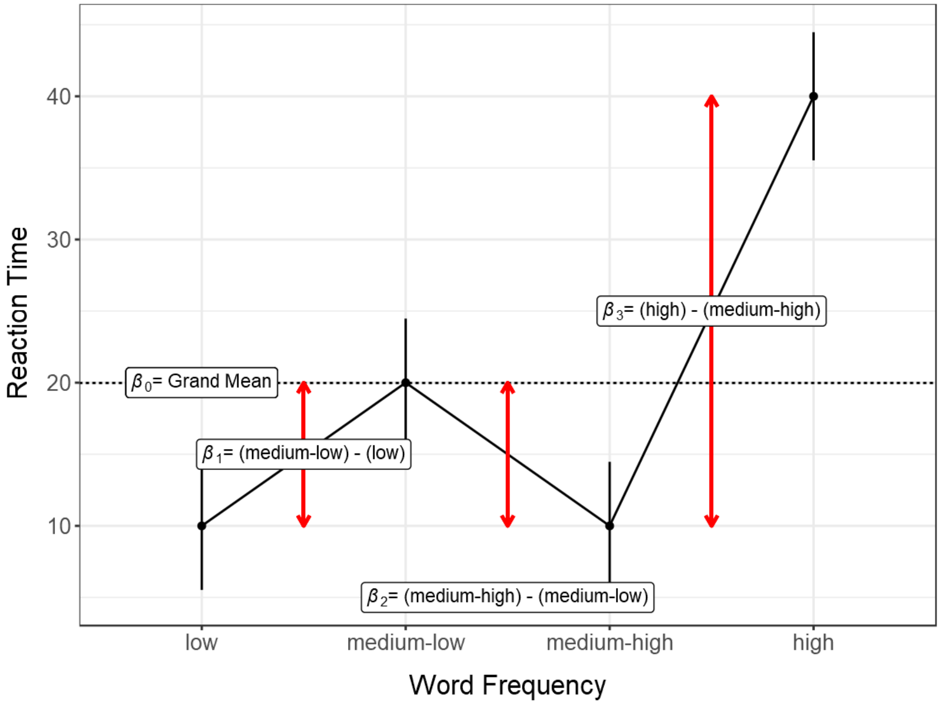


Repeated contrasts

検証する帰無仮説

- $\mu_1, \ldots, \mu_4$ の順番を整理
- $\mu_1, \ldots, \mu_4$に掛かっている数が，それぞれに対する重みづけ

$$
\begin{aligned}
H_{0 2-1} &: (-1) \times \mu_1 + 1 \times \mu_2 + 0 \times \mu_3 + 0 \times \mu_4 = 0 \\
H_{0 3-2} &: 0 \times \mu_1 + (-1) \times \mu_2 + 1 \times \mu_3 + 0 \times \mu_4 = 0 \\
H_{0 4-3} &: 0 \times \mu_1 + 0 \times \mu_2 + (-1) \times \mu_3 + 1 \times \mu_4 = 0
\end{aligned}
$$



重みづけを抽出し，「仮説行列」に書く

$\mu_1, \mu_2, \mu_3, \mu_4$ には重みづけがある

$$
\begin{aligned}
H_{02-1} &:(-1) \times \mu_1 + 1 \times \mu_2 + 0 \times \mu_3 + 0 \times \mu_4 = 0 \\
H_{03-2} &:0 \times \mu_1 + (-1) \times \mu_2 + 1 \times \mu_3 + 0 \times \mu_4 = 0 \\
H_{04-3} &:0 \times \mu_1 + 0 \times \mu_2 + (-1) \times \mu_3 + 1 \times \mu_4 = 0
\end{aligned}
$$

- 各行：**帰無仮説**
  - H_{02-1}: `c2vs1`
  - H_{03-2}: `c3vs2`
  - H_{04-3}: `c4vs3`
- 各列：各水準への**重み**

```
HcRE <- rbind(
  c2vs1 = c(low = -1, `med-low` =  1, `med-hi` =  0
  c3vs2 = c(low =  0, `med-low` = -1, `med-hi` =  1
  c4vs3 = c(low =  0, `med-low` =  0, `med-hi` = -1
)
```

重みづけを抽出し，「仮説行列」に書く

```
HcRE
```

```
      low med-low med-hi hi
c2vs1  -1       1      0  0
c3vs2   0      -1      1  0
c4vs3   0       0     -1  1
```

仮説行列

重みづけを抽出し，「仮説行列」に書く

```r
HcRE |>
  t()
```

```
         c2vs1 c3vs2 c4vs3
low         -1     0     0
med-low      1    -1     0
med-hi       0     1    -1
hi           0     0     1
```

仮説行列を転置 *transpose*
今度は各行が水準，各列が仮説に変わる

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcRE <- HcRE

XcRE
```

```
       low med-low med-hi hi
c2vs1  -1        1      0  0
c3vs2   0       -1      1  0
c4vs3   0        0     -1  1
```

仮説行列に対し…

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcRE <- HcRE |>
  hypr::ginv2()

XcRE
```

```
         c2vs1 c3vs2 c4vs3
low       -3/4  -1/2  -1/4
med-low    1/4  -1/2  -1/4
med-hi     1/4   1/2  -1/4
hi         1/4   1/2   3/4
```

一般化逆行列を適用，仮説行列を対比行列に変換

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcRE <- HcRE |>
  hypr::ginv2()

XcRE <- HcRE

XcRE
```

```
      low med-low med-hi hi
c2vs1  -1       1      0  0
c3vs2   0      -1      1  0
c4vs3   0       0     -1  1
```

仮説行列に対し…

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcRE <- HcRE |>
  hypr::ginv2()

XcRE <- HcRE |>
  ginv()

XcRE
```

```
      [,1] [,2]  [,3]
[1,] -0.75 -0.5 -0.25
[2,]  0.25 -0.5 -0.25
[3,]  0.25  0.5 -0.25
[4,]  0.25  0.5  0.75
```

一般化逆行列を適用，仮説行列を対比行列に変換

一般化逆行列を使い，仮説行列を対比行列に変換

```r
XcRE <- HcRE |>
  hypr::ginv2()

XcRE <- HcRE |>
  ginv() |>
  provideDimnames(
    base = dimnames(HcRE)[2:1]
  )

XcRE
```

```
         c2vs1 c3vs2 c4vs3
low      -0.75  -0.5 -0.25
med-low   0.25  -0.5 -0.25
med-hi    0.25   0.5 -0.25
hi        0.25   0.5  0.75
```

行名・列名を再付与

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcRE <- HcRE |>
  hypr::ginv2()

XcRE <- HcRE |>
  ginv() |>
  provideDimnames(
    base = dimnames(HcRE)[2:1]
  ) |>
  fractions()

XcRE
```

```
        c2vs1 c3vs2 c4vs3
low     -3/4  -1/2  -1/4
med-low  1/4  -1/2  -1/4
med-hi   1/4   1/2  -1/4
hi       1/4   1/2   3/4
```

分数表記に

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcRE
```

```
         c2vs1 c3vs2 c4vs3
low      -3/4  -1/2  -1/4
med-low   1/4  -1/2  -1/4
med-hi    1/4   1/2  -1/4
hi        1/4   1/2   3/4
```

仮説行列に一般化逆行列を適用したもの

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcRE

contr.sdif(4)
```

```
         c2vs1 c3vs2 c4vs3
low      -3/4  -1/2  -1/4
med-low   1/4  -1/2  -1/4
med-hi    1/4   1/2  -1/4
hi        1/4   1/2   3/4
```

```
    2-1  3-2   4-3
1 -0.75 -0.5 -0.25
2  0.25 -0.5 -0.25
3  0.25  0.5 -0.25
4  0.25  0.5  0.75
```

4水準の時のrepeated contrast

一般化逆行列を使い，仮説行列を対比行列に変換

```
XcRE

contr.sdif(4) |>
  fractions()
```

```
         c2vs1 c3vs2 c4vs3
low      -3/4  -1/2  -1/4
med-low   1/4  -1/2  -1/4
med-hi    1/4   1/2  -1/4
hi        1/4   1/2   3/4
```

```
  2-1  3-2  4-3
1 -3/4 -1/2 -1/4
2  1/4 -1/2 -1/4
3  1/4  1/2 -1/4
4  1/4  1/2  3/4
```

分数表記に

対比行列を要因に当てはめ，線形モデルを実行

```r
contrasts(simdat3$F) <- contr.sdif(4) |> fractions()

contrasts(simdat3$F)

lm(DV ~ F, data = simdat3) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(2),
      ", ",
      conf.high |> round(2),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

```
            2-1  3-2  4-3
low         -3/4 -1/2 -1/4
medium-low   1/4 -1/2 -1/4
medium-high  1/4  1/2 -1/4
high         1/4  1/2  3/4
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	20	[15.26, 24.74]	8.9	0.00
F2-1	10	[-3.41, 23.41]	1.6	0.13
F3-2	-10	[-23.41, 3.41]	-1.6	0.13
F4-3	30	[16.59, 43.41]	4.7	0.00

contr.sdif(3)で作ったrepeated contrastsを適用

対比行列を要因に当てはめ，線形モデルを実行

```r
contrasts(simdat3$F) <- XcRE

contrasts(simdat3$F)

lm(DV ~ F, data = simdat3) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(2),
      ", ",
      conf.high |> round(2),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

```
            c2vs1 c3vs2 c4vs3
low          -3/4  -1/2  -1/4
medium-low    1/4  -1/2  -1/4
medium-high   1/4   1/2  -1/4
high          1/4   1/2   3/4
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	20	[15.26, 24.74]	8.9	0.00
Fc2vs1	10	[-3.41, 23.41]	1.6	0.13
Fc3vs2	-10	[-23.41, 3.41]	-1.6	0.13
Fc4vs3	30	[16.59, 43.41]	4.7	0.00

仮説行列から作ったrepeated contrastsを適用

統計モデルから仮説を検証

1. 中頻度・低のRT μ_2 は，低頻度語へのRT μ_1 と同じか？
   - `Fc2vs1`より， $\mu_2 - \mu_1 = 10$
   - t 値が $t(16)$ の両側2.5% $t \leq -2.12, 2.12 \leq t$にないため，中頻度・低へのRTは低頻度語へのRTより有意に長いとは言えない
2. 中頻度・高へのRT μ_3 は，中頻度・低へのRT μ_2 と同じか？
   - `Fc3vs2`より， $\mu_3 - \mu_2 = -10$
   - t 値が $t(16)$ の両側2.5%にないため，中頻度・高へのRTは中頻度・低へのRTより有意に長いとは言えない
3. 高頻度へのRT μ_4 は， 中頻度・高のRT μ_3 と同じか？
   - `Fc4vs3`より， $\mu_4 - \mu_3 = 30$

```
            c2vs1 c3vs2 c4vs3
low         -3/4  -1/2  -1/4
medium-low   1/4  -1/2  -1/4
medium-high  1/4   1/2  -1/4
high         1/4   1/2   3/4
```

Predictor	**Estimate**	**95% CI**	t**-value**	p**-value**
(Intercept)	20	[15.26, 24.74]	8.9	0.00
Fc2vs1	10	[-3.41, 23.41]	1.6	0.13
Fc3vs2	-10	[-23.41, 3.41]	-1.6	0.13
Fc4vs3	30	[16.59, 43.41]	4.7	0.00

Contrasts in linear regression analysis: The design or model matrix

- 質的変数の対比により，離散的な水準を数値の変数にする
  - 離散的：とびとびの値を取る
- どの水準間の差について検定しようとしているか，数値・数式として符号する

Contrast matrix 対比行列

それぞれの水準をどのようにコーディングしたか，1行ずつ記述したもの

```
XcRE
```

```
         c2vs1 c3vs2 c4vs3
low      -3/4  -1/2  -1/4
med-low   1/4  -1/2  -1/4
med-hi    1/4   1/2  -1/4
hi        1/4   1/2   3/4
```

Design matrix 配置行列

各データポイントに対し，具体的にどの対比が使われたか記述したもの

```
covars <- model.matrix(
  ~ 1 + F, data = simdat3
)
```

配置行列のオブジェクトを作る

Design matrix 配置行列

各データポイントに対し，具体的にどの対比が使われたか記述したもの

```
covars <- model.matrix(
  ~ 1 + F, data = simdat3
) |>
  as_tibble()
```

tibbleデータフレームにする

Design matrix 配置行列

各データポイントに対し，具体的にどの対比が使われたか記述したもの

```
covars <- model.matrix(
  ~ 1 + F, data = simdat3
) |>
  as_tibble()

covars
```

```
# A tibble: 20 x 4
   `(Intercept)` Fc2vs1 Fc3vs2 Fc4vs3
           <dbl>  <dbl>  <dbl>  <dbl>
 1             1 -0.75  -0.5    -0.25
 2             1 -0.75  -0.5    -0.25
 3             1 -0.75  -0.5    -0.25
 4             1 -0.75  -0.5    -0.25
 5             1 -0.75  -0.5    -0.25
 6             1  0.250 -0.5    -0.25
 7             1  0.250 -0.5    -0.25
 8             1  0.250 -0.5    -0.25
 9             1  0.250 -0.5    -0.25
10             1  0.250 -0.5    -0.25
11             1  0.250  0.500  -0.25
12             1  0.250  0.500  -0.25
13             1  0.250  0.500  -0.25
14             1  0.250  0.500  -0.25
15             1  0.250  0.500  -0.25
16             1  0.25   0.5     0.75
17             1  0.25   0.5     0.75
18             1  0.25   0.5     0.75
19             1  0.25   0.5     0.75
20             1  0.25   0.5     0.75
```

Contrast matrix 対比行列

```
        c2vs1 c3vs2 c4vs3
low     -3/4  -1/2  -1/4
med-low  1/4  -1/2  -1/4
med-hi   1/4   1/2  -1/4
hi       1/4   1/2   3/4
```

Design matrix 配置行列

- 対比行列を各データポイントに拡張したもの
- どのデータポイントに，どの水準（コーディング）が**配置**されているか示す
- `Fc2vs1`のような列は，「質的変数の各水準を，1つずつ量的な説明変数に替えたもの」

```
   (Intercept)           F Fc2vs1 Fc3vs2 Fc4vs3
1            1         low  -0.75   -0.5  -0.25
2            1         low  -0.75   -0.5  -0.25
3            1         low  -0.75   -0.5  -0.25
4            1         low  -0.75   -0.5  -0.25
5            1         low  -0.75   -0.5  -0.25
6            1  medium-low   0.25   -0.5  -0.25
7            1  medium-low   0.25   -0.5  -0.25
8            1  medium-low   0.25   -0.5  -0.25
9            1  medium-low   0.25   -0.5  -0.25
10           1  medium-low   0.25   -0.5  -0.25
11           1 medium-high   0.25    0.5  -0.25
12           1 medium-high   0.25    0.5  -0.25
13           1 medium-high   0.25    0.5  -0.25
14           1 medium-high   0.25    0.5  -0.25
15           1 medium-high   0.25    0.5  -0.25
```

Design matrixを線形モデルの要因として使う

Design matrixにおける`Fc2vs1`のような列は，「質的変数の各水準を，1つずつの量的な説明変数に替えたもの」

- `Fc2vs1`などを「『質的な説明変数』の水準」の1つではなく，「量的な説明変数」の1つとして捉える
  - `Fc2vs1`は，質的な説明変数`F`の水準の1つ
  - `Fc2vs1`は，量的な説明変数`Fc2vs1`であり，$-\frac{3}{4}$ または $\frac{1}{4}$ を取る

Design matrixにおける`Fc2vs1`などを説明変数として利用できる

Design matrixを線形モデルの要因として使う

Design matrixをsimdat3に結合

```r
simdat3_DM <- covars

simdat3_DM
```

```
# A tibble: 20 x 4
   `(Intercept)` Fc2vs1 Fc3vs2 Fc4vs3
           <dbl>  <dbl>  <dbl>  <dbl>
 1             1 -0.75  -0.5    -0.25
 2             1 -0.75  -0.5    -0.25
 3             1 -0.75  -0.5    -0.25
 4             1 -0.75  -0.5    -0.25
 5             1 -0.75  -0.5    -0.25
 6             1  0.250 -0.5    -0.25
 7             1  0.250 -0.5    -0.25
 8             1  0.250 -0.5    -0.25
 9             1  0.250 -0.5    -0.25
10             1  0.250 -0.5    -0.25
11             1  0.250  0.500  -0.25
12             1  0.250  0.500  -0.25
13             1  0.250  0.500  -0.25
14             1  0.250  0.500  -0.25
15             1  0.250  0.500  -0.25
16             1  0.25   0.5     0.75
17             1  0.25   0.5     0.75
18             1  0.25   0.5     0.75
19             1  0.25   0.5     0.75
20             1  0.25   0.5     0.75
```

配置行列のオブジェクト

Design matrixを線形モデルの要因として使う

Design matrixをsimdat3に結合

```r
simdat3_DM <- covars |>
  dplyr::select(-`(Intercept)`)

simdat3_DM
```

```
# A tibble: 20 x 3
   Fc2vs1 Fc3vs2 Fc4vs3
    <dbl>  <dbl>  <dbl>
 1 -0.75  -0.5    -0.25
 2 -0.75  -0.5    -0.25
 3 -0.75  -0.5    -0.25
 4 -0.75  -0.5    -0.25
 5 -0.75  -0.5    -0.25
 6  0.250 -0.5    -0.25
 7  0.250 -0.5    -0.25
 8  0.250 -0.5    -0.25
 9  0.250 -0.5    -0.25
10  0.250 -0.5    -0.25
11  0.250  0.500  -0.25
12  0.250  0.500  -0.25
13  0.250  0.500  -0.25
14  0.250  0.500  -0.25
15  0.250  0.500  -0.25
16  0.25   0.5     0.75
17  0.25   0.5     0.75
18  0.25   0.5     0.75
19  0.25   0.5     0.75
20  0.25   0.5     0.75
```

(Intercept)列を除去

Design matrixを線形モデルの要因として使う

Design matrixをsimdat3に結合

```
simdat3_DM <- covars |>
  dplyr::select(-`(Intercept)`)
  bind_cols(simdat3)

simdat3_DM
```

```
# A tibble: 20 x 6
   Fc2vs1 Fc3vs2 Fc4vs3 F           id
    <dbl>  <dbl>  <dbl> <fct>       <fct>
 1 -0.75  -0.5    -0.25 low         1
 2 -0.75  -0.5    -0.25 low         2
 3 -0.75  -0.5    -0.25 low         3
 4 -0.75  -0.5    -0.25 low         4
 5 -0.75  -0.5    -0.25 low         5
 6  0.250 -0.5    -0.25 medium-low  6
 7  0.250 -0.5    -0.25 medium-low  7
 8  0.250 -0.5    -0.25 medium-low  8
 9  0.250 -0.5    -0.25 medium-low  9
10  0.250 -0.5    -0.25 medium-low  10
11  0.250  0.500  -0.25 medium-high 11
12  0.250  0.500  -0.25 medium-high 12
13  0.250  0.500  -0.25 medium-high 13
14  0.250  0.500  -0.25 medium-high 14
15  0.250  0.500  -0.25 medium-high 15
16  0.25   0.5     0.75 high        16
17  0.25   0.5     0.75 high        17
18  0.25   0.5     0.75 high        18
19  0.25   0.5     0.75 high        19
20  0.25   0.5     0.75 high        20
```

(Intercept)列を除去した配置行列を，**simdat3**に結合

Design matrixを線形モデルの要因として使う

Design matrixをsimdat3に結合

```
simdat3_DM <- covars |>
  dplyr::select(-`(Intercept)`)
  bind_cols(simdat3) |>
  relocate(F)

simdat3_DM
```

```
# A tibble: 20 x 6
   F           Fc2vs1 Fc3vs2 Fc4vs3 id
   <fct>        <dbl>  <dbl>  <dbl> <fct>
 1 low         -0.75  -0.5    -0.25 1
 2 low         -0.75  -0.5    -0.25 2
 3 low         -0.75  -0.5    -0.25 3
 4 low         -0.75  -0.5    -0.25 4
 5 low         -0.75  -0.5    -0.25 5
 6 medium-low   0.250 -0.5    -0.25 6
 7 medium-low   0.250 -0.5    -0.25 7
 8 medium-low   0.250 -0.5    -0.25 8
 9 medium-low   0.250 -0.5    -0.25 9
10 medium-low   0.250 -0.5    -0.25 10
11 medium-high  0.250  0.500  -0.25 11
12 medium-high  0.250  0.500  -0.25 12
13 medium-high  0.250  0.500  -0.25 13
14 medium-high  0.250  0.500  -0.25 14
15 medium-high  0.250  0.500  -0.25 15
16 high         0.25   0.5     0.75 16
17 high         0.25   0.5     0.75 17
18 high         0.25   0.5     0.75 18
19 high         0.25   0.5     0.75 19
20 high         0.25   0.5     0.75 20
```

F列が最先頭列になるよう列を並べ替え

Design matrixを線形モデルの要因として使う

```r
m3_mr <- lm(DV ~ F, data = simc

m3_mr |>
  tidy(conf.int = TRUE, conf.le
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(2),
      ", ",
      conf.high |> round(2),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = e
  dplyr::select(-c(std.error, c
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	20	[15.26, 24.74]	8.9	0.00
Fc2vs1	10	[-3.41, 23.41]	1.6	0.13
Fc3vs2	-10	[-23.41, 3.41]	-1.6	0.13
Fc4vs3	30	[16.59, 43.41]	4.7	0.00

Repeated contrastを適用した**F**を使いモデルを構築した場合

Design matrixを線形モデルの要因として使う

```r
m3_mr <- lm(DV ~ Fc2vs1 + Fc3vs

m3_mr |>
  tidy(conf.int = TRUE, conf.le
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(2),
      ", ",
      conf.high |> round(2),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = e
  dplyr::select(-c(std.error, c
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

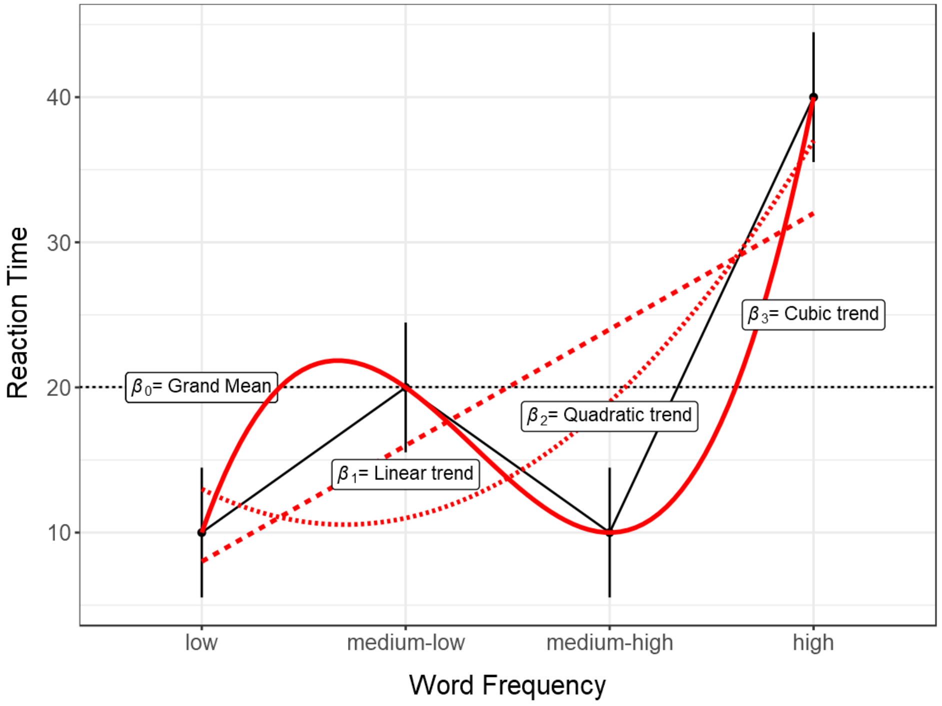
Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	20	[15.26, 24.74]	8.9	0.00
Fc2vs1	10	[-3.41, 23.41]	1.6	0.13
Fc3vs2	-10	[-23.41, 3.41]	-1.6	0.13
Fc4vs3	30	[16.59, 43.41]	4.7	0.00

Repeated contrastに基づくdesign matrixを使いモデルを構築した場合

Repeated contrastを適用したFを使いモデルを構築した場合と結果が完全に一致！

Polynomial contrasts

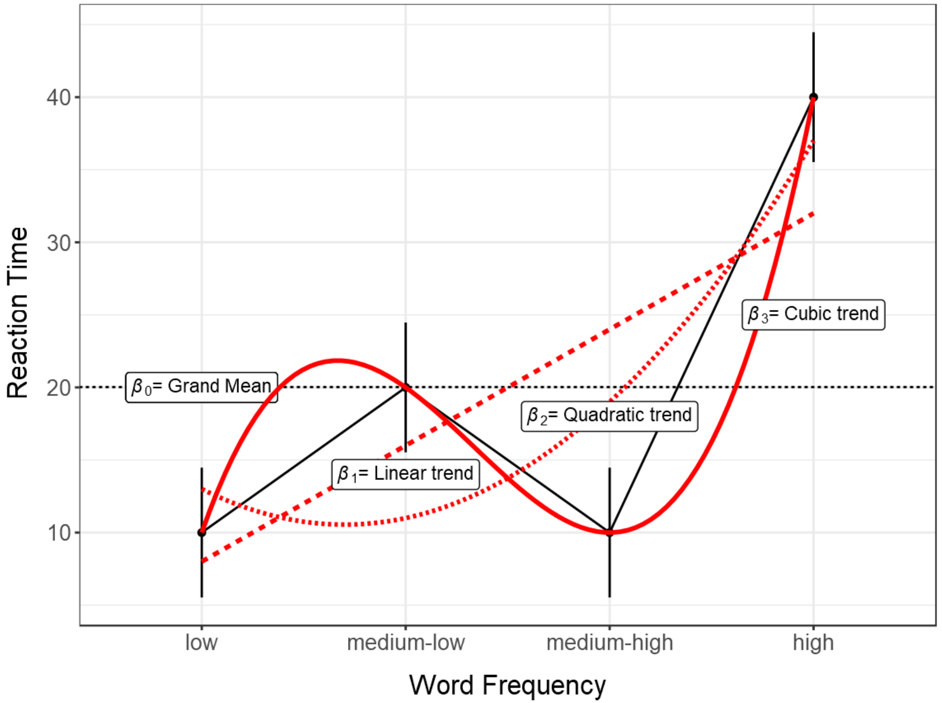
- （順序が決まっている）隣り合う水準同士を比較
  - 水準間で同じペースで増加・減少？
  - 水準間毎に，増加・減少のペースが変わる？
- 例：語の頻度の効果を，隣り合う水準と比べる
  1. 低頻度
  2. 中頻度・低
  3. 中頻度・高
  4. 高頻度
- 主な対比は次の3つ
  1. 1次関数ペース（Linear; どの水準間でも一定ペースで増加・減少？）
  2. 2次関数ペース（Quadratic; 次の水準間に移るほど増加・減少が激しい・穏やか？）
  3. 3次関数ペース（Cubic; 次の水準間に移るほど増加・減少が入れ替わる？）



Polynomial contrasts

検証する帰無仮説

1. 1次関数ペース（Linear; どの水準間でも一定ペースで増加・減少？）
2. 2次関数ペース（Quadratic; 次の水準間に移るほど増加・減少が激しい・穏やか？）
3. 3次関数ペース（Cubic; 次の水準間に移るほど増加・減少が入れ替わる？）



Polynomial contrastの対比行列

```
contrasts(simdat3$F) <- contr.p
```

4水準の時のPolynomial contrastの対比行列

Polynomial contrastの対比行列

```
contrasts(simdat3$F) <- contr.p
contrasts(simdat3$F)
```

```
                 .L   .Q    .C
low          -0.67  0.5 -0.22
medium-low   -0.22 -0.5  0.67
medium-high   0.22 -0.5 -0.67
high          0.67  0.5  0.22
```

希望するcontrastsになったか確認

対比行列を要因に当てはめ，線形モデルを実行

```r
lm(DV ~ F, data = simdat3)
```

```
Call:
lm(formula = DV ~ F, data = simdat3

Coefficients:
(Intercept)          F.L            F
       20.0         17.9           1(
```

統計モデル構築

対比行列を要因に当てはめ，線形モデルを実行

```
lm(DV ~ F, data = simdat3) |>
  tidy(conf.int = TRUE, conf.level = 0.95)
```

```
# A tibble: 4 x 7
  term        estimate std.error st
  <chr>          <dbl>     <dbl>
1 (Intercept)     20        2.24
2 F.L             17.9      4.47
3 F.Q             10        4.47
4 F.C             13.4      4.47
```

結果表示（各係数の95%信頼区間も算出）

対比行列を要因に当てはめ，線形モデルを実行

```
lm(DV ~ F, data = simdat3) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  )
```

```
# A tibble: 4 x 8
  term        estimate std.error st
  <chr>          <dbl>     <dbl>
1 (Intercept)     20        2.24
2 F.L             17.9      4.47
3 F.Q             10        4.47
4 F.C             13.4      4.47
```

各係数の95%信頼区間を1列にまとめる

対比行列を要因に当てはめ，線形モデルを実行

```r
lm(DV ~ F, data = simdat3) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate)
```

```
# A tibble: 4 x 8
  term        estimate `95% CI` std
  <chr>          <dbl> <chr>
1 (Intercept)     20   [15, 25]
2 F.L             17.9 [8, 27]
3 F.Q             10   [1, 19]
4 F.C             13.4 [4, 23]
```

各係数の95%信頼区間をまとめた列を，係数の列の後ろに配置

対比行列を要因に当てはめ，線形モデルを実行

```r
lm(DV ~ F, data = simdat3) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
```

```
# A tibble: 4 x 5
  term        estimate `95% CI` sta
  <chr>          <dbl> <chr>
1 (Intercept)     20   [15, 25]
2 F.L             17.9 [8, 27]
3 F.Q             10   [1, 19]
4 F.C             13.4 [4, 23]
```

標準誤差，95%信頼区間の上・下限の列を削除

対比行列を要因に当てはめ，線形モデルを実行

```r
lm(DV ~ F, data = simdat3) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	20	[15, 25]	8.9	0.00
F.L	18	[8, 27]	4.0	0.00
F.Q	10	[1, 19]	2.2	0.04
F.C	13	[4, 23]	3.0	0.01

表にまとめる

統計モデルから仮説を検証

1. 1次関数ペース（Linear; どの水準間でも一定ペースで増加・減少？）
    - `F_L`より，linear trend 有意
2. 2次関数ペース（Quadratic; 次の水準間に移るほど増加・減少が激しい・穏やか？）
    - `F_Q`より，quadratic trend 有意
3. 3次関数ペース（Cubic; 次の水準間に移るほど増加・減少が入れ替わる？）
    - `F_C`より，cubic trend 有意

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	20	[15, 25]	8.9	0.00
F.L	18	[8, 27]	4.0	0.00
F.Q	10	[1, 19]	2.2	0.04
F.C	13	[4, 23]	3.0	0.01

Polynomial contrastの長所

- Polynomial contrastでlinear trendを指定することで，全体的な増加傾向をlinear trendの係数として表現することができる
    - 全体的な増加傾向の検出力も上がる
    - （データに応じ，全体的な減少傾向もlinear trendの係数として表現可能なはず）
- 要因の水準が増えれば増えるほど，他の対比方法（コーディング方法）やオムニバスF検定よりも，polynomial contrastを使う時の方が，少ない変数で応答変数を説明できる
    - オムニバスF検定：全ての水準から2つ選んでできる水準の組み合わせ全てについて1回ずつ統計を掛ける

Custom contrasts

- 先行研究の結果や推論に基づいて，特定の応答変数（の平均値）のパターンが予測できることがある
- そうしたパターンを検定するための対比

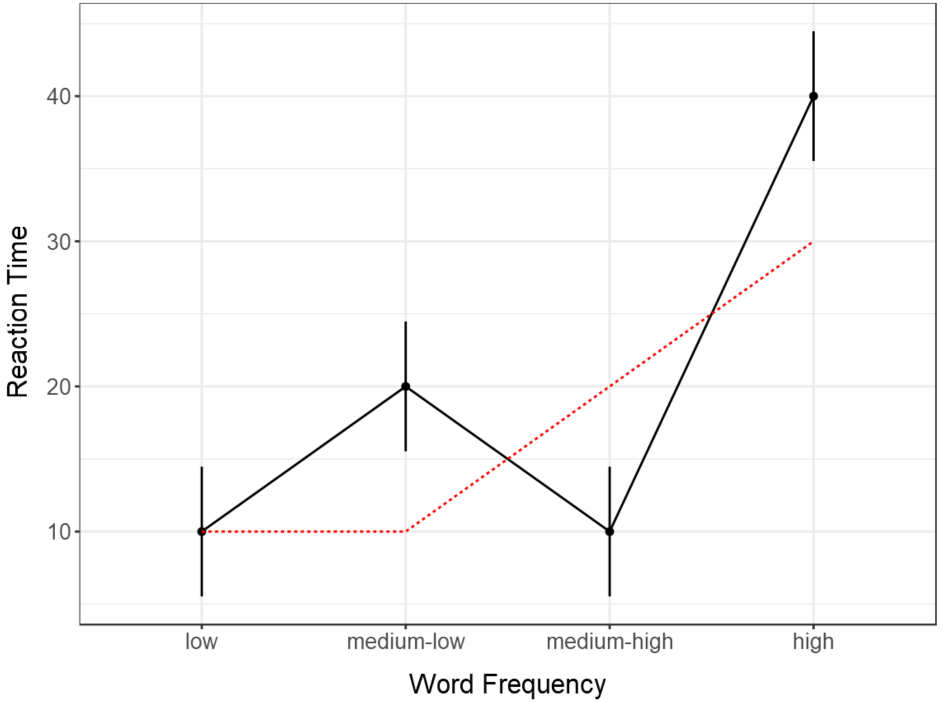
Custom contrasts

- 例：語の頻度の効果について
  1. 低頻度へのRTと中頻度・低へのRTは，同じ
  2. 中頻度・高へのRTから高頻度へのRTにかけて，一定ペースで増加する

1. ↑の予測に基づき，応答変数の平均値のパターンを考える
2. 応答変数の平均値を中心化
   - 各水準へのRTが，全体平均からどれだけ離れているか
3. 中心化した値の比を同じ数で割り，最小の整数値から成る比にする

他にも，例えば このような方法 がある

	低頻度へのRT	中頻度・低へのRT	中頻度・高へのRT	高頻度へのRT
1	10	10	20	30
2	-7.5	-7.5	2.5	12.5
3	-3	-3	1	5



Custom contrastの使い方

- Schad et al. (2020), p.16 では，custom contrastの対比行列と，それに基づく配置行列を作っている
- ここでは，custom contrastを新しい説明変数として直接データに組み込む方法を紹介

```r
simdat3_CC <- simdat3

simdat3_CC
```

```
             F id   DV
1          low  1  3.6
2          low  2 11.3
3          low  3  7.3
4          low  4  1.2
5          low  5 26.5
6   medium-low  6  9.2
7   medium-low  7 36.5
8   medium-low  8 17.9
9   medium-low  9 18.0
10  medium-low 10 18.4
11 medium-high 11 -3.7
12 medium-high 12 13.7
13 medium-high 13 22.6
14 medium-high 14  4.4
15 medium-high 15 13.0
16        high 16 51.2
17        high 17 34.0
18        high 18 35.9
19        high 19 28.9
20        high 20 50.0
```

データ

Custom contrastの使い方

- Schad et al. (2020), p.16 では，custom contrastの対比行列と，それに基づく配置行列を作っている
- ここでは，custom contrastを新しい説明変数として直接データに組み込む方法を紹介

```
simdat3_CC <- simdat3 |>
  mutate(
    Fcust = case_when(
      F == "low" | F == "mediun
      F == "medium-high" ~ 1,
      F == "high" ~ 5,
      TRUE ~ NaN
    )
  )

simdat3_CC
```

```
              F id   DV Fcust
1           low  1  3.6    -3
2           low  2 11.3    -3
3           low  3  7.3    -3
4           low  4  1.2    -3
5           low  5 26.5    -3
6    medium-low  6  9.2    -3
7    medium-low  7 36.5    -3
8    medium-low  8 17.9    -3
9    medium-low  9 18.0    -3
10   medium-low 10 18.4    -3
11 medium-high 11 -3.7     1
12 medium-high 12 13.7     1
13 medium-high 13 22.6     1
14 medium-high 14  4.4     1
15 medium-high 15 13.0     1
16         high 16 51.2     5
17         high 17 34.0     5
18         high 18 35.9     5
19         high 19 28.9     5
20         high 20 50.0     5
```

Fの水準に応じて異なる値を取る，新しい説明変数**Fcust**を作る

Custom contrastで，線形モデルを実行

```r
lm(DV ~ Fcust, data = simdat3_CC)
```

```

Call:
lm(formula = DV ~ Fcust, data = sim

Coefficients:
(Intercept)        Fcust
      20.00         2.73
```

統計モデル構築

Custom contrastで，線形モデルを実行

```r
lm(DV ~ Fcust, data = simdat3_CC) |>
  tidy(conf.int = TRUE, conf.level = 0.95)
```

```
# A tibble: 2 x 7
  term        estimate std.error st
  <chr>          <dbl>     <dbl>
1 (Intercept)    20        2.87
2 Fcust           2.73     0.865
```

結果表示（各係数の95%信頼区間も算出）

Custom contrastで，線形モデルを実行

```r
lm(DV ~ Fcust, data = simdat3_CC) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  )
```

```
# A tibble: 2 x 8
  term        estimate std.error st
  <chr>          <dbl>     <dbl>
1 (Intercept)    20        2.87
2 Fcust           2.73     0.865
```

各係数の95%信頼区間を1列にまとめる

Custom contrastで，線形モデルを実行

```
lm(DV ~ Fcust, data = simdat3_CC) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate)
```

```
# A tibble: 2 x 8
  term        estimate `95% CI` std
  <chr>          <dbl> <chr>
1 (Intercept)    20    [14, 26]
2 Fcust           2.73 [1, 5]
```

各係数の95%信頼区間をまとめた列を，係数の列の後ろに配置

Custom contrastで，線形モデルを実行

```r
lm(DV ~ Fcust, data = simdat3_CC) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
```

```
# A tibble: 2 x 5
  term        estimate `95% CI` sta
  <chr>          <dbl> <chr>
1 (Intercept)    20    [14, 26]
2 Fcust           2.73 [1, 5]
```

標準誤差，95%信頼区間の上・下限の列を削除

Custom contrastで，線形モデルを実行

```r
lm(DV ~ Fcust, data = simdat3_CC) |>
  tidy(conf.int = TRUE, conf.level = 0.95) |>
  mutate(
    `95% CI` = paste0(
      "[",
      conf.low |> round(0),
      ", ",
      conf.high |> round(0),
      "]"
    )
  ) |>
  relocate(`95% CI`, .after = estimate) |>
  dplyr::select(-c(std.error, conf.low, conf.high))
  kable(
    digits = 2,
    escape = FALSE,
    col.names = c(
      'Predictor',
      'Estimate',
      '95% CI',
      '$t$-value',
      '$p$-value'
    )
  )
```

Predictor	Estimate	95% CI	t-value	p-value
(Intercept)	20.0	[14, 26]	7.0	0.00
Fcust	2.7	[1, 5]	3.1	0.01

表にまとめる