Coefficient covariance for permutet data with weights equal 1

Peter Hettegger

2020-11-28

This document contains R-code for analysing the coefficient covariance for two features with sample weights equal 1. We compare coefficient covariance of non-permuted data and permuted data. Therefore we assume a sample design with `age` and `group` (case, control) as covariates. `group` is the hypothesis coefficient.

```
set.seed(1)

library(randRotation)
library(ggplot2)
library(knitr)
library(heatmap3)

# Sample info
samp.inf <- data.frame(age = as.integer(runif(20)*30),
                       group = as.factor(rep(c("case", "control"), c(10,10))))
n <- nrow(samp.inf)
kable(head(samp.inf))
```

age	group
7	case
11	case
17	case
27	case
6	case
26	case

The correlation coefficient of the error terms between feature 1 and feature 2 is assumed as 0.8. However, as it is a constant factor in the covariance of the coefficients, the value of the correlation coefficient solely changes the scales of the plots below, but not the pattern.

```
X <- model.matrix(~ 1 + age + group, samp.inf)

rho12 <- 0.8
# assume sigma1 = sigma2 = 1
sigma12 <- rho12


# define random uniformly distributed weights
W <- matrix(1, nrow = 2, ncol = n, byrow = TRUE)
W
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
## [1,]    1    1    1    1    1    1    1    1    1     1     1     1     1     1
```

```
## [2,]    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##       [,15] [,16] [,17] [,18] [,19] [,20]
## [1,]     1     1     1     1     1     1
## [2,]     1     1     1     1     1     1
```

We now calculate the coefficient covariance for non-permuted data:

```
# whitening of X
X1 <- sqrt(W[1,]) * X
X2 <- sqrt(W[2,]) * X

cov.beta <- sigma12 * solve(t(X1) %*% X1) %*% t(X1) %*% X2 %*% solve(t(X2) %*% X2)
```

The coefficient covariance for permuted data is calculated by simulation (2000 permutations):

```
# group as "hypothesis coefficient"
# intercept and sex as "determined coefficients"
coef.h <- 3

tmp1 <- replicate(2000, {

    # Permute only hypothesis coefficient
    i <- sample(nrow(X))
    X1p <- X
    X2p <- X
    X1p[,coef.h] <- X1p[i,coef.h]
    X2p[,coef.h] <- X2p[i,coef.h]
    X1p <- sqrt(W[1,]) * X1p
    X2p <- sqrt(W[2,]) * X2p

    sigma12 * solve(t(X1p) %*% X1p) %*% t(X1p) %*% X2p %*% solve(t(X2p) %*% X2p)
})


E.cov.beta.r <- apply(tmp1, 1:2, mean)
sd.cov.beta.r <- apply(tmp1, 1:2, function(i)sqrt(var(i)))
```

Coefficient covariance for non-permuted data:

```
kable(cov.beta)
```

	(Intercept)	age	groupcontrol
(Intercept)	0.2271	-0.0092	-0.0772
age	-0.0092	0.0006	-0.0002
groupcontrol	-0.0772	-0.0002	0.1601

Expected coefficient covariance for permuted data:

```
kable(E.cov.beta.r)
```

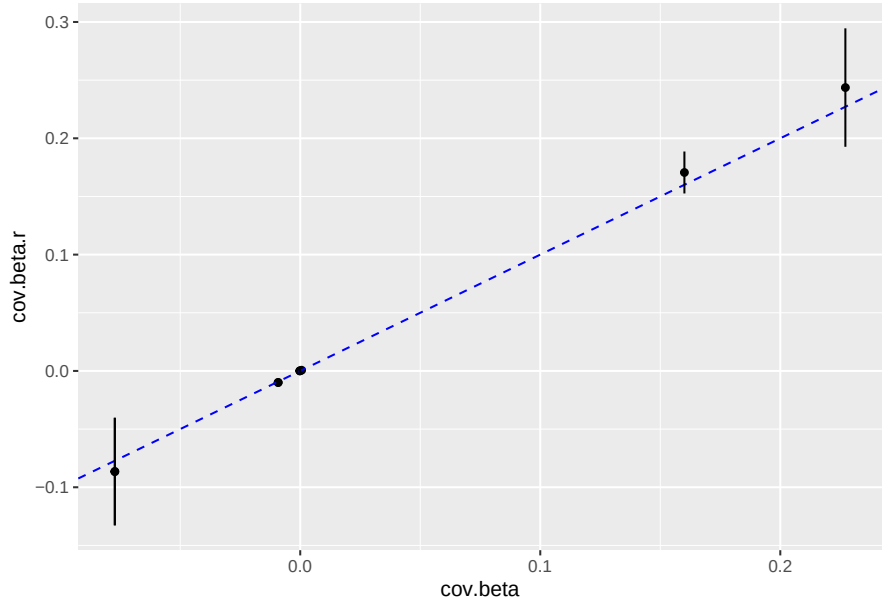
	(Intercept)	age	groupcontrol
(Intercept)	0.2436	-0.0099	-0.0865
age	-0.0099	0.0006	0.0001
groupcontrol	-0.0865	0.0001	0.1706

Standard deviation of coefficient covariance for permuted data:

```
kable(sd.cov.beta.r)
```

	(Intercept)	age	groupcontrol
(Intercept)	0.0509	0.0018	0.0462
age	0.0018	0.0001	0.0028
groupcontrol	0.0462	0.0028	0.0180

The graphical representation of these tables is:



The given example shows, that even for weights equal 1, the dependence structure of coefficient estimates is not exactly retained for the assumed experimental design.

Session Info

```
sessionInfo()
```

```
## R Under development (unstable) (2020-11-14 r79432)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Austria.1252  LC_CTYPE=German_Austria.1252
## [3] LC_MONETARY=German_Austria.1252 LC_NUMERIC=C
## [5] LC_TIME=German_Austria.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] heatmap3_1.1.7     knitr_1.30         ggplot2_3.3.2      randRotation_1.3.4
```

```
##
## loaded via a namespace (and not attached):
##  [1] xml2_1.3.2         magrittr_1.5       munsell_0.5.0      colorspace_2.0-0
##  [5] R6_2.5.0           rlang_0.4.8        highr_0.8          fastcluster_1.1.25
##  [9] stringr_1.4.0      tools_4.1.0        rbibutils_1.4      grid_4.1.0
## [13] gtable_0.3.0       xfun_0.19          withr_2.3.0        ellipsis_0.3.1
## [17] htmltools_0.5.0    yaml_2.2.1         digest_0.6.27      tibble_3.0.4
## [21] lifecycle_0.2.0    crayon_1.3.4       farver_2.0.3       vctrs_0.3.4
## [25] Rdpack_2.1         gbRd_0.4-11        glue_1.4.2         evaluate_0.14
## [29] rmarkdown_2.5      labeling_0.4.2     stringi_1.5.3      pillar_1.4.6
## [33] compiler_4.1.0     scales_1.1.1       pkgconfig_2.0.3
```