looking.R

fuxiao

2022-12-05

```
### libraries
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────
──────── tidyverse 1.3.2 ──
## ✔ ggplot2 3.4.0      ✔ purrr   0.3.5
## ✔ tibble  3.1.8      ✔ dplyr   1.0.10
## ✔ tidyr   1.2.1      ✔ stringr 1.4.1
## ✔ readr   2.1.3      ✔ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────────────
── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(Hmisc)
```

```
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
##
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(PerformanceAnalytics)
```

```
## Loading required package: xts
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
##
## Attaching package: 'xts'
##
## The following objects are masked from 'package:dplyr':
##
##     first, last
##
##
## Attaching package: 'PerformanceAnalytics'
##
## The following object is masked from 'package:graphics':
##
##     legend
```

```
library(mboost)
```

```
## Loading required package: parallel
## Loading required package: stabs
##
## Attaching package: 'mboost'
##
## The following object is masked from 'package:magrittr':
##
##     extract
##
## The following object is masked from 'package:tidyr':
##
##     extract
##
## The following object is masked from 'package:ggplot2':
##
##     %+%
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

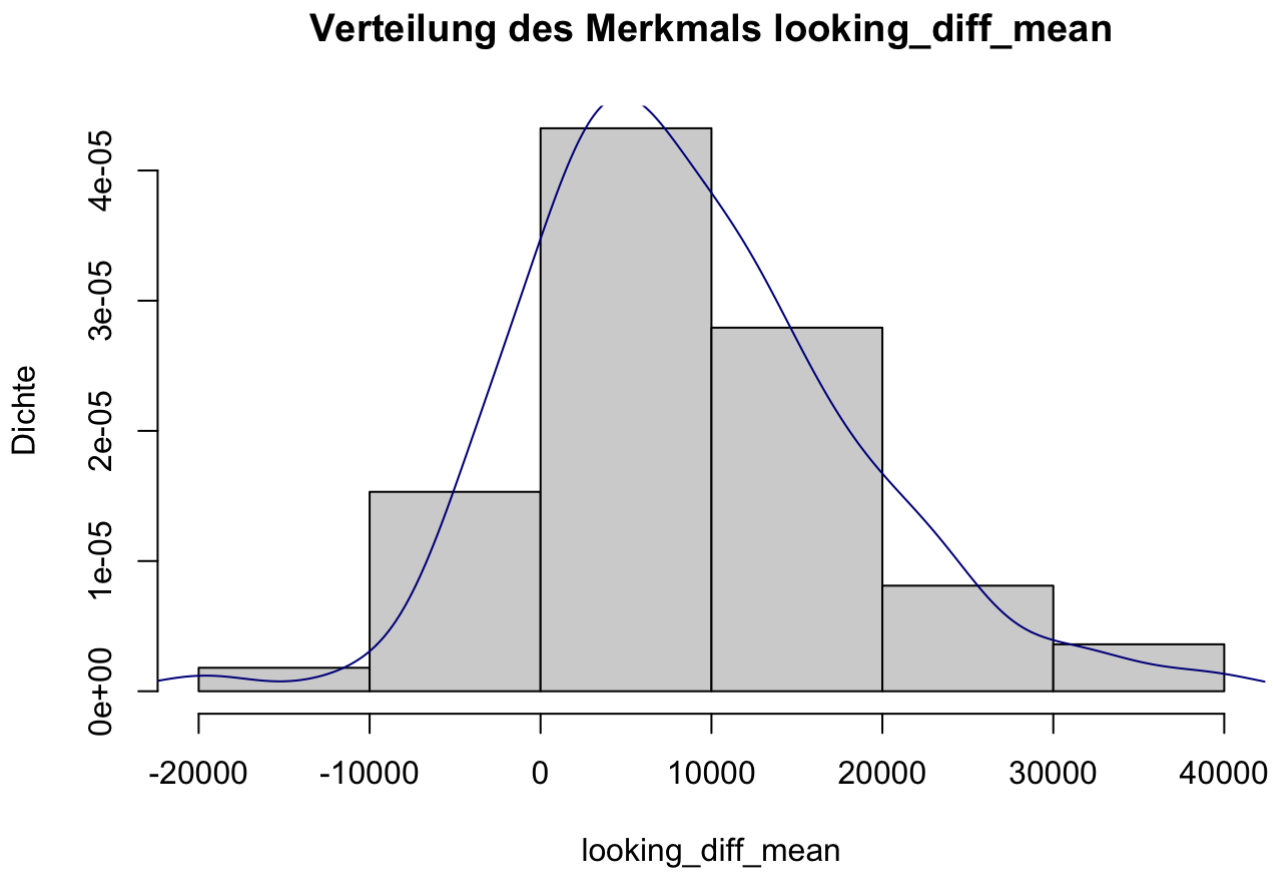
```
library(effects)
```

```
## Use the command
##     lattice::trellis.par.set(effectsTheme())
##   to customize lattice options for effects plots.
## See ?effectsTheme for details.
```

```
### data sets
load('looking.RData')
load('fNIRSData.RData')
load('age_sex.RData')

### create tidydata dataset
fNIRS2 <- pivot_wider(data = fNIRSData,
                      names_from = c(condition, channel),
                      values_from = c(HbO, HbR))
looking2 <- pivot_wider(data = looking,
                        names_from = c(trial, view),
                        values_from = duration)
data.list <- list(fNIRS2, looking2, age_sex)
fNIRS.looking.age_sex <- data.list %>%
  reduce(full_join, by = "id")


####### Frage 3: Korrelation einzelner channels zu looking experiment ---------------
--------
### Spalten für channel_total hinzufügen
fNIRS.looking.age_sex$channel1_total <- (fNIRS.looking.age_sex$HbO_online_1 - fNIRS.l
ooking.age_sex$HbR_online_1) - (fNIRS.looking.age_sex$HbO_delayed_1 - fNIRS.looking.a
ge_sex$HbR_delayed_1)
fNIRS.looking.age_sex$channel2_total <- (fNIRS.looking.age_sex$HbO_online_2 - fNIRS.l
ooking.age_sex$HbR_online_2) - (fNIRS.looking.age_sex$HbO_delayed_2 - fNIRS.looking.a
ge_sex$HbR_delayed_2)
fNIRS.looking.age_sex$channel3_total <- (fNIRS.looking.age_sex$HbO_online_3 - fNIRS.l
ooking.age_sex$HbR_online_3) - (fNIRS.looking.age_sex$HbO_delayed_3 - fNIRS.looking.a
ge_sex$HbR_delayed_3)
fNIRS.looking.age_sex$channel4_total <- (fNIRS.looking.age_sex$HbO_online_4 - fNIRS.l
ooking.age_sex$HbR_online_4) - (fNIRS.looking.age_sex$HbO_delayed_4 - fNIRS.looking.a
ge_sex$HbR_delayed_4)
fNIRS.looking.age_sex$channel5_total <- (fNIRS.looking.age_sex$HbO_online_5 - fNIRS.l
ooking.age_sex$HbR_online_5) - (fNIRS.looking.age_sex$HbO_delayed_5 - fNIRS.looking.a
ge_sex$HbR_delayed_5)
fNIRS.looking.age_sex$channel6_total <- (fNIRS.looking.age_sex$HbO_online_6 - fNIRS.l
ooking.age_sex$HbR_online_6) - (fNIRS.looking.age_sex$HbO_delayed_6 - fNIRS.looking.a
ge_sex$HbR_delayed_6)
fNIRS.looking.age_sex$channel7_total <- (fNIRS.looking.age_sex$HbO_online_7 - fNIRS.l
ooking.age_sex$HbR_online_7) - (fNIRS.looking.age_sex$HbO_delayed_7 - fNIRS.looking.a
ge_sex$HbR_delayed_7)
fNIRS.looking.age_sex$channel8_total <- (fNIRS.looking.age_sex$HbO_online_8 - fNIRS.l
ooking.age_sex$HbR_online_8) - (fNIRS.looking.age_sex$HbO_delayed_8 - fNIRS.looking.a
ge_sex$HbR_delayed_8)
fNIRS.looking.age_sex$channel9_total <- (fNIRS.looking.age_sex$HbO_online_9 - fNIRS.l
ooking.age_sex$HbR_online_9) - (fNIRS.looking.age_sex$HbO_delayed_9 - fNIRS.looking.a
ge_sex$HbR_delayed_9)
fNIRS.looking.age_sex$channel10_total <- (fNIRS.looking.age_sex$HbO_online_10 - fNIR
S.looking.age_sex$HbR_online_10) - (fNIRS.looking.age_sex$HbO_delayed_10 - fNIRS.look
ing.age_sex$HbR_delayed_10)
fNIRS.looking.age_sex$channel11_total <- (fNIRS.looking.age_sex$HbO_online_11 - fNIR
S.looking.age_sex$HbR_online_11) - (fNIRS.looking.age_sex$HbO_delayed_11 - fNIRS.look
ing.age_sex$HbR_delayed_11)
fNIRS.looking.age_sex$channel12_total <- (fNIRS.looking.age_sex$HbO_online_12 - fNIR
S.looking.age_sex$HbR_online_12) - (fNIRS.looking.age_sex$HbO_delayed_12 - fNIRS.look
```

```
ing.age_sex$HbR_delayed_12)
fNIRS.looking.age_sex$channel13_total <- (fNIRS.looking.age_sex$HbO_online_13 - fNIR
S.looking.age_sex$HbR_online_13) - (fNIRS.looking.age_sex$HbO_delayed_13 - fNIRS.look
ing.age_sex$HbR_delayed_13)
fNIRS.looking.age_sex$channel14_total <- (fNIRS.looking.age_sex$HbO_online_14 - fNIR
S.looking.age_sex$HbR_online_14) - (fNIRS.looking.age_sex$HbO_delayed_14 - fNIRS.look
ing.age_sex$HbR_delayed_14)
fNIRS.looking.age_sex$channel15_total <- (fNIRS.looking.age_sex$HbO_online_15 - fNIR
S.looking.age_sex$HbR_online_15) - (fNIRS.looking.age_sex$HbO_delayed_15 - fNIRS.look
ing.age_sex$HbR_delayed_15)
fNIRS.looking.age_sex$channel16_total <- (fNIRS.looking.age_sex$HbO_online_16 - fNIR
S.looking.age_sex$HbR_online_16) - (fNIRS.looking.age_sex$HbO_delayed_16 - fNIRS.look
ing.age_sex$HbR_delayed_16)
fNIRS.looking.age_sex$channel17_total <- (fNIRS.looking.age_sex$HbO_online_17 - fNIR
S.looking.age_sex$HbR_online_17) - (fNIRS.looking.age_sex$HbO_delayed_17 - fNIRS.look
ing.age_sex$HbR_delayed_17)
fNIRS.looking.age_sex$channel18_total <- (fNIRS.looking.age_sex$HbO_online_18 - fNIR
S.looking.age_sex$HbR_online_18) - (fNIRS.looking.age_sex$HbO_delayed_18 - fNIRS.look
ing.age_sex$HbR_delayed_18)
fNIRS.looking.age_sex$channel19_total <- (fNIRS.looking.age_sex$HbO_online_19 - fNIR
S.looking.age_sex$HbR_online_19) - (fNIRS.looking.age_sex$HbO_delayed_19 - fNIRS.look
ing.age_sex$HbR_delayed_19)
fNIRS.looking.age_sex$channel20_total <- (fNIRS.looking.age_sex$HbO_online_20 - fNIR
S.looking.age_sex$HbR_online_20) - (fNIRS.looking.age_sex$HbO_delayed_20 - fNIRS.look
ing.age_sex$HbR_delayed_20)
fNIRS.looking.age_sex$channel21_total <- (fNIRS.looking.age_sex$HbO_online_21 - fNIR
S.looking.age_sex$HbR_online_21) - (fNIRS.looking.age_sex$HbO_delayed_21 - fNIRS.look
ing.age_sex$HbR_delayed_21)
fNIRS.looking.age_sex$channel22_total <- (fNIRS.looking.age_sex$HbO_online_22 - fNIR
S.looking.age_sex$HbR_online_22) - (fNIRS.looking.age_sex$HbO_delayed_22 - fNIRS.look
ing.age_sex$HbR_delayed_22)
fNIRS.looking.age_sex$channel23_total <- (fNIRS.looking.age_sex$HbO_online_23 - fNIR
S.looking.age_sex$HbR_online_23) - (fNIRS.looking.age_sex$HbO_delayed_23 - fNIRS.look
ing.age_sex$HbR_delayed_23)
fNIRS.looking.age_sex$channel24_total <- (fNIRS.looking.age_sex$HbO_online_24 - fNIR
S.looking.age_sex$HbR_online_24) - (fNIRS.looking.age_sex$HbO_delayed_24 - fNIRS.look
ing.age_sex$HbR_delayed_24)
fNIRS.looking.age_sex$channel25_total <- (fNIRS.looking.age_sex$HbO_online_25 - fNIR
S.looking.age_sex$HbR_online_25) - (fNIRS.looking.age_sex$HbO_delayed_25 - fNIRS.look
ing.age_sex$HbR_delayed_25)
fNIRS.looking.age_sex$channel26_total <- (fNIRS.looking.age_sex$HbO_online_26 - fNIR
S.looking.age_sex$HbR_online_26) - (fNIRS.looking.age_sex$HbO_delayed_26 - fNIRS.look
ing.age_sex$HbR_delayed_26)
fNIRS.looking.age_sex$channel27_total <- (fNIRS.looking.age_sex$HbO_online_27 - fNIR
S.looking.age_sex$HbR_online_27) - (fNIRS.looking.age_sex$HbO_delayed_27 - fNIRS.look
ing.age_sex$HbR_delayed_27)
fNIRS.looking.age_sex$channel28_total <- (fNIRS.looking.age_sex$HbO_online_28 - fNIR
S.looking.age_sex$HbR_online_28) - (fNIRS.looking.age_sex$HbO_delayed_28 - fNIRS.look
ing.age_sex$HbR_delayed_28)
fNIRS.looking.age_sex$channel29_total <- (fNIRS.looking.age_sex$HbO_online_29 - fNIR
S.looking.age_sex$HbR_online_29) - (fNIRS.looking.age_sex$HbO_delayed_29 - fNIRS.look
ing.age_sex$HbR_delayed_29)
fNIRS.looking.age_sex$channel30_total <- (fNIRS.looking.age_sex$HbO_online_30 - fNIR
S.looking.age_sex$HbR_online_30) - (fNIRS.looking.age_sex$HbO_delayed_30 - fNIRS.look
ing.age_sex$HbR_delayed_30)
```

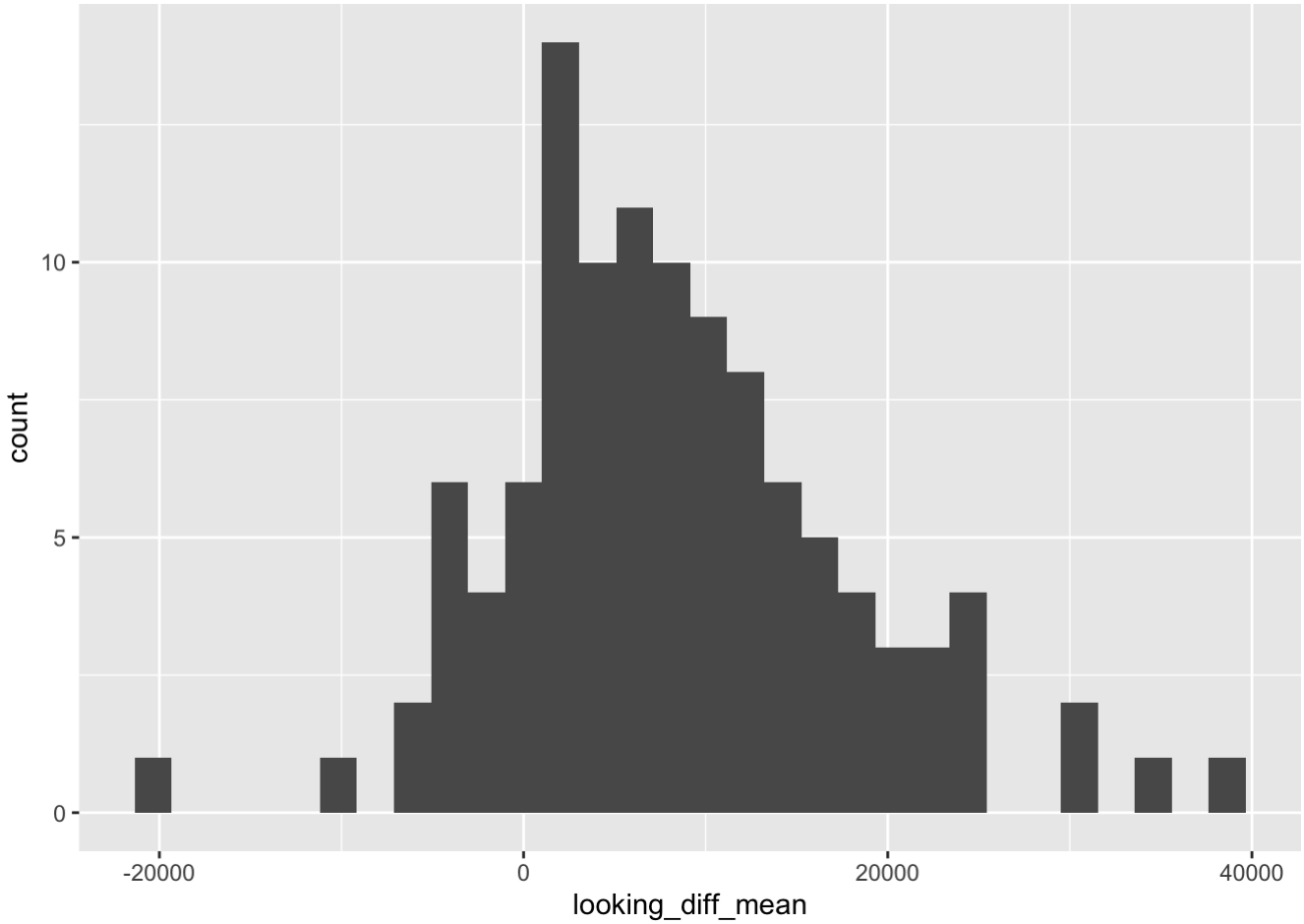
```
### Spalten für ego - mirrored von looking experiment hinzufügen
fNIRS.looking.age_sex$looking_diff_1 <- fNIRS.looking.age_sex$`1_ego` - fNIRS.lookin
g.age_sex$`1_mirrored`
fNIRS.looking.age_sex$looking_diff_2 <- fNIRS.looking.age_sex$`2_ego` - fNIRS.lookin
g.age_sex$`2_mirrored`
fNIRS.looking.age_sex$looking_diff_3 <- fNIRS.looking.age_sex$`3_ego` - fNIRS.lookin
g.age_sex$`3_mirrored`
fNIRS.looking.age_sex$looking_diff_4 <- fNIRS.looking.age_sex$`4_ego` - fNIRS.lookin
g.age_sex$`4_mirrored`
### Spalten für looking_diff_mean hinzufügen
fNIRS.looking.age_sex$looking_diff_mean <- rowMeans(fNIRS.looking.age_sex[163:166], n
a.rm = TRUE)
### subdatensatz für Regression erstellen
mydata <- fNIRS.looking.age_sex[1:53,c(133:162,167)]
## ersetzen Na durch Mittelwert
for(i in 1:ncol(mydata)) {
  mydata[ , i][is.na(mydata[ , i])] <- mean(mydata[[i]], na.rm=TRUE)
}

### plot der Verteilung des Merkmals looking_diff_mean
hist(fNIRS.looking.age_sex$looking_diff_mean, freq =FALSE,
     main = "Verteilung des Merkmals looking_diff_mean",
     ylab = "Dichte", xlab = "looking_diff_mean")
lines(density(fNIRS.looking.age_sex$looking_diff_mean, na.rm = TRUE), col= "darkblue"
)
```



```
ggplot(data = fNIRS.looking.age_sex, aes(looking_diff_mean)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 8 rows containing non-finite values (`stat_bin()`).
```




```
### Variablenselektion per boosting-Verfahren
# Modell mit nu = 0.05:
set.seed(123)
model_0.05 <- glmboost(looking_diff_mean ~ ., data = mydata,
                       control = boost_control(mstop = 2000, nu = 0.05))
cv_0.05 <- cvrisk(object = model_0.05,
                  folds = cv(weights = model.weights(model_0.05),
                             type = "kfold"))
mstop(cv_0.05)
```

```
## [1] 86
```

```
mean(cv_0.05[, mstop(cv_0.05) + 1])
```

```
## [1] 121698225
```

```
# Modell mit nu = 0.05:
set.seed(123)
model_0.1 <- glmboost(looking_diff_mean ~ ., data = mydata,
                      control = boost_control(mstop = 2000, nu = 0.1))
cv_0.1 <- cvrisk(object = model_0.1,
                 folds = cv(weights = model.weights(model_0.1),
                            type = "kfold"))
mstop(cv_0.1)
```

```
## [1] 23
```

```
mean(cv_0.1[, mstop(cv_0.1) + 1])
```

```
## [1] 121461563
```

```
# Modell mit nu = 0.2:
set.seed(123)
model_0.2 <- glmboost(looking_diff_mean ~ ., data = mydata,
                      control = boost_control(mstop = 2000, nu = 0.2))
cv_0.2<- cvrisk(object = model_0.2,
                folds = cv(weights = model.weights(model_0.2),
                           type = "kfold"))
mstop(cv_0.2)
```

```
## [1] 22
```

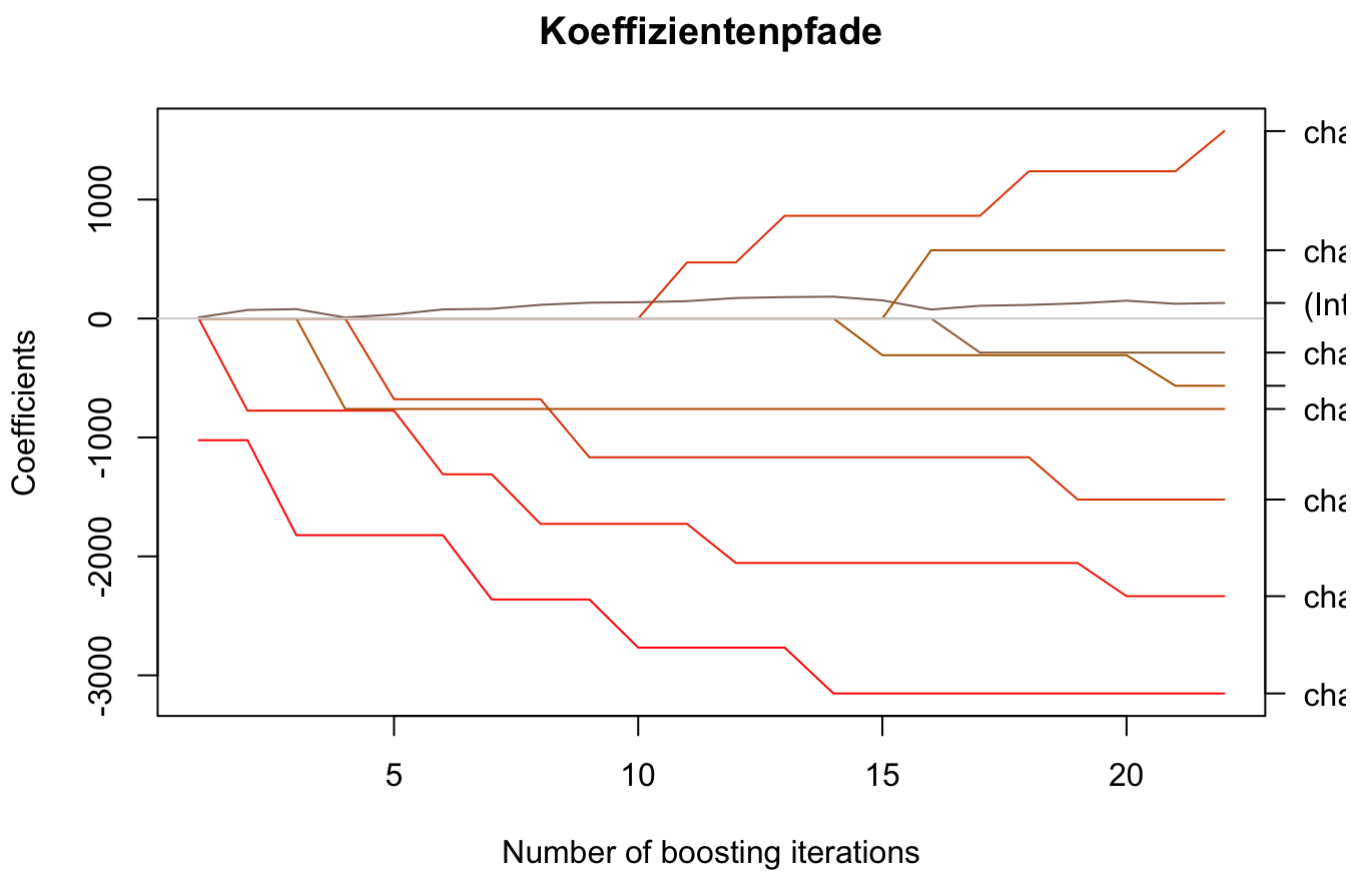
```
mean(cv_0.2[, mstop(cv_0.2) + 1])
```

```
## [1] 121347802
```

```
##  Der niedrigste MSE  wird für das Modell mit den Parametern mstop = 22
##  und v = 0.2 erreicht.

model_boosting_opt <- glmboost(looking_diff_mean ~., data = mydata,
                           control = boost_control(mstop = 22, nu = 0.2))
summary(model_boosting_opt)
```

```
##
##   Generalized Linear Models Fitted via Gradient Boosting
##
## Call:
## glmboost.formula(formula = looking_diff_mean ~ ., data = mydata,     control = boo
st_control(mstop = 22, nu = 0.2))
##
##
##   Squared Error (Regression)
##
## Loss function: (y - f)^2
##
##
## Number of boosting iterations: mstop = 22
## Step size:  0.2
## Offset:  9385.34
##
## Coefficients:
##     (Intercept)  channel1_total  channel2_total channel13_total channel17_total
##        130.6180       1575.7277      -2334.6156      -1522.2023       -286.7144
## channel18_total channel23_total channel24_total channel27_total
##      -3152.5738       -565.7368       -760.6204        574.1845
## attr(,"offset")
## [1] 9385.34
##
## Selection frequencies:
##  channel2_total channel18_total  channel1_total channel13_total channel23_total
##      0.22727273      0.22727273      0.18181818      0.13636364      0.09090909
## channel17_total channel24_total channel27_total
##      0.04545455      0.04545455      0.04545455
```

```
plot(x = model_boosting_opt, main = "Koeffizientenpfade")
```



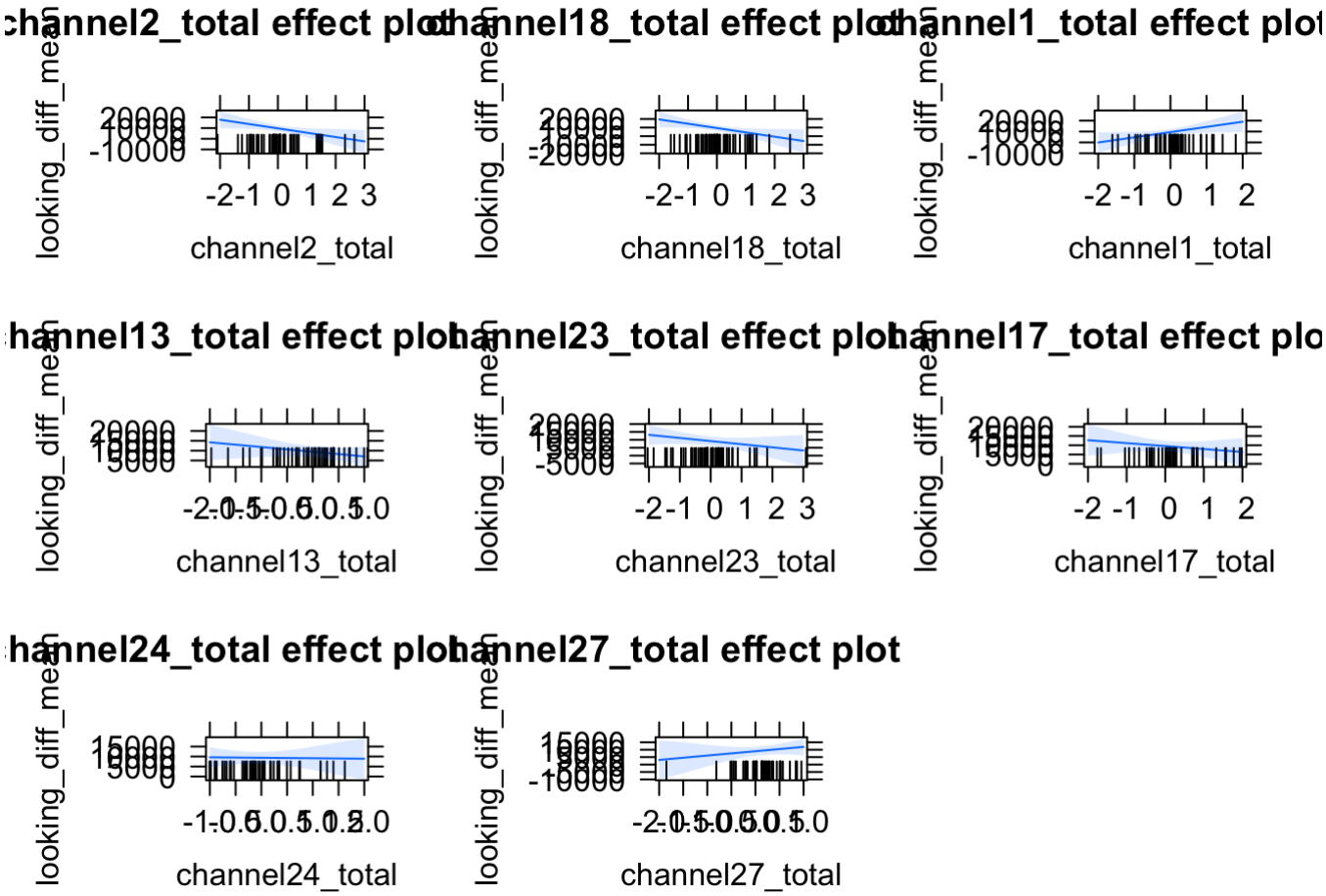

```
## glm
model_all <- glm(looking_diff_mean ~ ., family = gaussian(), data = mydata)
summary(model_all)
```

```
##
## Call:
## glm(formula = looking_diff_mean ~ ., family = gaussian(), data = mydata)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -18915   -5293    1677    4644   18790
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     10664.25    2688.28   3.967 0.000654 ***
## channel1_total   5888.41    4620.38   1.274 0.215804
## channel2_total  -5285.73    3112.36  -1.698 0.103552
## channel3_total   4243.67    4263.76   0.995 0.330419
## channel4_total   1508.55    5599.52   0.269 0.790127
## channel5_total  -4216.16    4666.76  -0.903 0.376076
## channel6_total   -650.70    4588.93  -0.142 0.888530
## channel7_total    -21.41    4081.92  -0.005 0.995862
## channel8_total   4503.21    4438.49   1.015 0.321335
## channel9_total   -561.59    2502.20  -0.224 0.824487
## channel10_total  2187.75    6425.67   0.340 0.736733
## channel11_total -5076.03    6957.03  -0.730 0.473316
## channel12_total  -478.81    5392.26  -0.089 0.930048
## channel13_total -7326.17    5944.21  -1.232 0.230776
## channel14_total  -423.66    3499.42  -0.121 0.904739
## channel15_total -1464.56    5434.16  -0.270 0.790049
## channel16_total  -386.52    5682.73  -0.068 0.946387
## channel17_total   232.72    4227.18   0.055 0.956592
## channel18_total -6234.47    4489.14  -1.389 0.178797
## channel19_total  -284.70    3891.15  -0.073 0.942335
## channel20_total  2095.11    4895.62   0.428 0.672843
## channel21_total  3798.61    5414.21   0.702 0.490285
## channel22_total  2790.98    3445.88   0.810 0.426646
## channel23_total -5615.13    3892.79  -1.442 0.163265
## channel24_total  1724.14    5515.62   0.313 0.757534
## channel25_total -4924.78    6710.11  -0.734 0.470739
## channel26_total   780.55    4814.02   0.162 0.872675
## channel27_total  3870.16    6926.15   0.559 0.581958
## channel28_total  3007.26    5318.38   0.565 0.577492
## channel29_total -1341.44    3593.92  -0.373 0.712534
## channel30_total   815.67    3212.05   0.254 0.801900
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 149230149)
##
##     Null deviance: 6531001509  on 52  degrees of freedom
## Residual deviance: 3283063279  on 22  degrees of freedom
## AIC: 1165.3
##
## Number of Fisher Scoring iterations: 2
```

```
model_boost <- glm(looking_diff_mean ~ channel2_total + channel18_total+
                channel1_total + channel13_total + channel23_total + channel17_total
+
                channel24_total + channel27_total,
              family = gaussian, data = mydata)
summary(model_boost)
```

```
##
## Call:
## glm(formula = looking_diff_mean ~ channel2_total + channel18_total +
##     channel1_total + channel13_total + channel23_total + channel17_total +
##     channel24_total + channel27_total, family = gaussian, data = mydata)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -17746   -6334    2262    5786   19863
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       9498.5     1462.8   6.493 6.34e-08 ***
## channel2_total   -4057.2     1715.1  -2.366   0.0225 *
## channel18_total  -5027.0     2181.8  -2.304   0.0260 *
## channel1_total    4773.3     2212.5   2.157   0.0365 *
## channel13_total  -2382.7     2062.7  -1.155   0.2543
## channel23_total  -1974.7     1502.8  -1.314   0.1957
## channel17_total  -1590.4     1845.0  -0.862   0.3933
## channel24_total   -233.2     2352.3  -0.099   0.9215
## channel27_total   2940.5     2958.1   0.994   0.3256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 94602047)
##
##     Null deviance: 6531001509  on 52  degrees of freedom
## Residual deviance: 4162490077  on 44  degrees of freedom
## AIC: 1133.9
##
## Number of Fisher Scoring iterations: 2
```

```
plot(allEffects(model_boost))
```




```
avPlots(model_boost)
```