Applied Statistical Methods - Solution 7

AUTHOR
Peter von Rohr

PUBLISHED
April 22, 2024

WEBR STATUS
Ready!

Problem 1: Sum Contrasts

Use the following dataset on Body.Weight and `Breed` of beef cattle animals. The data is available from

```
[1] "https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_breed.csv"
```

Fit a fixed linear model with Body.Weight as response and `Breed` as predictor variable. Use the `sum` contrasts for reporting the different effects in the model. Validate the estimates by computing the estimates based on a solution of the least squares normal equations.

Tasks

- Read the data

Run Code

```
s_ex07p01_data_path <- "https://charlotte-ngs.github.io/asmasss2024/
tbl_bw_br <- read.table(s_ex07p01_data_path,
                                header = T, sep = ",")
tbl_bw_br
```

```
   Animal Body.Weight     Breed
1       1         471     Angus
2       2         463     Angus
3       4         470     Angus
4       7         518  Limousin
5       8         511  Limousin
6       9         510  Limousin
7      10         541  Limousin
8       3         481 Simmental
9       5         496 Simmental
10      6         491 Simmental
```

- Change contrasts and fit linear model The type of contrasts can directly be specified when fitting the linear model. For more information see the help function of `contrasts`

Run Code

```
lm_bw_br_con_sum <- lm(Body.Weight ~ Breed,
                       data = tbl_bw_br,
                       contrasts = list(Breed = "contr.sum"))
(smry_lm_bw_br_con_sum <- summary(lm_bw_br_con_sum))
```

```
Call:
lm(formula = Body.Weight ~ Breed, data = tbl_bw_br, contrasts = list(Breed = "contr.sum"))

Residuals:
     Min       1Q   Median       3Q      Max
-10.0000  -7.5000  -0.1667   2.7500  21.0000

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  492.444      3.370 146.113 1.86e-13 ***
Breed1       -24.444      4.873  -5.016 0.001538 **
Breed2        27.556      4.545   6.063 0.000509 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.56 on 7 degrees of freedom
Multiple R-squared:  0.8597,    Adjusted R-squared:  0.8196
F-statistic: 21.44 on 2 and 7 DF,  p-value: 0.001035
```

- Solutions of Least Squares Normal Equations

Run Code

```
mat_X <- model.matrix(Body.Weight ~ 0 + Breed, tbl_bw_br)
attr(mat_X, "assign") <- NULL
attr(mat_X, "contrasts") <- NULL
colnames(mat_X) <- NULL
mat_X <- cbind(matrix(rep(1,nrow(mat_X)), ncol = 1), mat_X)
mat_xtx <- crossprod(mat_X)
mat_xtx_ginv <- MASS::ginv(mat_xtx)
mat_xty <- crossprod(mat_X, tbl_bw_br$Body.Weight)
mat_b_sol <- crossprod(mat_xtx_ginv, mat_xty)
mat_b_sol
```

```
        [,1]
[1,] 369.33333
[2,]  98.66667
[3,] 150.66667
[4,] 120.00000
```

- Contrasts Matrix for Sum Contrasts From the contrasts matrix, we get the matrix of estimable functions.

Run Code

```
fac_breed <- as.factor(tbl_bw_br$Breed)
contr_mat_breed_sum <- contrasts(C(fac_breed, sum))
contr_mat_breed_sum <- cbind(matrix(rep(1,nrow(contr_mat_breed_sum)))
est_mat_breed_sum <- solve(contr_mat_breed_sum)
est_mat_breed_sum
```

```
          Angus   Limousin  Simmental
[1,]  0.3333333  0.3333333  0.3333333
[2,]  0.6666667 -0.3333333 -0.3333333
[3,] -0.3333333  0.6666667 -0.3333333
```

The first row of the above matrix `est_mat_breed_sum` shows how the intercept estimate is computed from the observation means. This means that with the sum contrasts, the intercept is the weighted mean of the mean observation for all breeds. Hence, we get

Run Code

```
tbl_bw_br_an <- dplyr::filter(tbl_bw_br, Breed == "Angus")
tbl_bw_br_li <- dplyr::filter(tbl_bw_br, Breed == "Limousin")
tbl_bw_br_si <- dplyr::filter(tbl_bw_br, Breed == "Simmental")
sum(c(mean(tbl_bw_br_an$Body.Weight),
      mean(tbl_bw_br_li$Body.Weight),
```

```
    6         mean(tbl_bw_br_si$Body.Weight)))/3
```

```
[1] 492.4444
```

Comparing that to the result of `lm()` from above, we see that they are equal.

Run Code

```
    1   smry_lm_bw_br_con_sum$coefficients["(Intercept)","Estimate"]
```

```
[1] 492.4444
```

For the effects estimates, we are looking at the second and the third row of the matrix `est_mat_breed_sum`. We are prepending a column of zeroes to the second and the third row of `est_mat_breed_sum`.

Run Code

```
    1   mat_q_efun <- cbind(matrix(rep(0, (nrow(est_mat_breed_sum)-1)), ncol
    2   crossprod(t(mat_q_efun), mat_b_sol)
```

```
          [,1]
[1,] -24.44444
[2,]  27.55556
```

These values correspond to the effect estimates from `lm()`

Run Code

```
    1   smry_lm_bw_br_con_sum$coefficients[2:3,1]
```

```
   Breed1    Breed2
-24.44444  27.55556
```

Problem 2: Custom Contrasts

Use the dataset from Problem 1 and use your own contrasts. Your new contrasts should compute the intercept estimate as is done in the `sum` contrasts. The `Breed` effects should be computed the same way as is done in the `treatment` contrast.

- Read the dataset

Run Code

```
    1   s_ex07p02_data_path <- "https://charlotte-ngs.github.io/asmasss2024/
    2   tbl_bw_br <- read.table(s_ex07p02_data_path,
    3                           header = T, sep = ",")
```

- Matrix of Estimable Functions The matrix of estimable functions is a combination of the matrices from the sum contrasts and from the treatment contrasts.

Run Code

```
    1   fact_breed <- as.factor(tbl_bw_br$Breed)
    2   # treatment
    3   mat_cont_treat <- contrasts(C(fact_breed, treatment))
    4   mat_cont_treat <- cbind(matrix(rep(1, nrow(mat_cont_treat)), ncol =
    5   mat_estf_treat <- solve(mat_cont_treat)
    6   # sum
    7   mat_cont_sum <- contrasts(C(fact_breed, sum))
    8   mat_cont_sum <- cbind(matrix(rep(1, nrow(mat_cont_sum)), ncol = 1),
    9   mat_estf_sum <- solve(mat_cont_sum)
```

```
# custom
mat_estf_cust <- rbind(mat_estf_sum[1,], mat_estf_treat[2:3,])
mat_cont_cust <- solve(mat_estf_cust)
mat_cont_cust <- mat_cont_cust[,2:3]
mat_cont_cust
```

```
           Limousin  Simmental
Angus     -0.3333333 -0.3333333
Limousin   0.6666667 -0.3333333
Simmental -0.3333333  0.6666667
```

Using that contrasts matrix in lm leads to

Run Code

```
lm_bw_br_con_cust <- lm(Body.Weight ~ Breed,
                        data = tbl_bw_br,
                        contrasts = list(Breed = mat_cont_cust))
(smry_lm_bw_br_con_cust <- summary(lm_bw_br_con_cust))
```

```
Call:
lm(formula = Body.Weight ~ Breed, data = tbl_bw_br, contrasts = list(Breed = mat_cont_cust))

Residuals:
     Min       1Q   Median       3Q      Max
-10.0000  -7.5000  -0.1667   2.7500  21.0000

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     492.444      3.370 146.113 1.86e-13 ***
BreedLimousin    52.000      8.066   6.447 0.000351 ***
BreedSimmental   21.333      8.623   2.474 0.042575 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.56 on 7 degrees of freedom
Multiple R-squared:  0.8597,    Adjusted R-squared:  0.8196
F-statistic: 21.44 on 2 and 7 DF,  p-value: 0.001035
```