

Applied Statistical Methods - Solution 4

AUTHOR
Peter von Rohr

PUBLISHED
March 18, 2024

WEBR STATUS
● Ready!

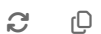
Problem 1: Matrix-Vector Notation

Use the dataset on **Body Weight** and **Breast Circumference** to show that the matrix-vector notation of the regression model yields the same result as the scalar notation.

Tasks

- Read data from https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_bc_reg.csv

▶ Run Code



```
1 # read data
2 s_bw_bc_url <- "https://charlotte-ngs.github.io/asmasss2024/data/asm_
3 df_bw_bc <- read.table(s_bw_bc_url, header = T, sep = ",")
4 df_bw_bc
```

	Animal	Breast.Circumference	Body.Weight
1	1	176	471
2	2	177	463
3	3	178	481
4	4	179	470
5	5	179	496
6	6	180	491
7	7	181	518
8	8	182	511
9	9	183	510
10	10	184	541

- Set up the matrix X and the vector y

▶ Run Code



```
1 # matrix X
2 X <- matrix(data = c(rep(1, nrow(df_bw_bc)),
3                       df_bw_bc$Breast.Circumference),
4             nrow = nrow(df_bw_bc), ncol = 2)
5 X
```

```
      [,1] [,2]
[1,]    1 176
[2,]    1 177
[3,]    1 178
[4,]    1 179
[5,]    1 179
[6,]    1 180
[7,]    1 181
[8,]    1 182
[9,]    1 183
[10,]   1 184
```

the vector y is taken from the dataset directly

▶ Run Code



```
1 # vector y
2 y <- df_bw_bc$Body.Weight
3 y
```

```
[1] 471 463 481 470 496 491 518 511 510 541
```

- Compute the solution for \hat{b}

▶ Run Code



```
1 # compute solution
2 xtx <- crossprod(X)
3 xty <- crossprod(X,y)
4 b_hat <- solve(xtx,xty)
5 b_hat
```

```
      [,1]
```

```
[1,] -1065.114943
```

```
[2,]   8.673235
```

Problem 2: Multiple Linear Regression

Use the dataset on **Body Weight** and multiple other predictor variables to fit a multiple linear regression.

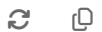
The dataset is available from https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_mult_reg.csv.

Tasks

The same as in Problem 1

- Read data from https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_mult_reg.csv

▶ Run Code



```
1 # read data
2 s_bw_mlr_url <- "https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_mult_reg.csv"
3 df_bw_mlr <- read.table(s_bw_mlr_url, header = T, sep = ",")
4 df_bw_mlr
```

	Animal	Breast.Circumference	Body.Weight	BCS	HEI
1	1	176	471	5.0	161
2	2	177	463	4.2	121
3	3	178	481	4.9	157
4	4	179	470	3.0	165
5	5	179	496	6.8	136
6	6	180	491	4.9	123
7	7	181	518	4.4	163
8	8	182	511	4.4	149
9	9	183	510	3.5	143
10	10	184	541	4.7	130

- Set up the matrix X and the vector y

▶ Run Code



```
1 # matrix X
2 X <- matrix(data = c(rep(1, nrow(df_bw_mlr))),
```

```

3           df_bw_mlr$Breast.Circumference,
4           df_bw_mlr$BCS,
5           df_bw_mlr$HEI),
6           nrow = nrow(df_bw_mlr))
7 X

```

```

      [,1] [,2] [,3] [,4]
[1,]    1  176  5.0  161
[2,]    1  177  4.2  121
[3,]    1  178  4.9  157
[4,]    1  179  3.0  165
[5,]    1  179  6.8  136
[6,]    1  180  4.9  123
[7,]    1  181  4.4  163
[8,]    1  182  4.4  149
[9,]    1  183  3.5  143
[10,]   1  184  4.7  130

```

the vector y is taken from the dataset directly

▶ Run Code



```

1 # vector y
2 # vector y
3 y <- df_bw_mlr$Body.Weight
4 y

```

```
[1] 471 463 481 470 496 491 518 511 510 541
```

- Compute the solution for \hat{b}

▶ Run Code



```

1 # compute solution
2 xtx <- crossprod(X)
3 xty <- crossprod(X,y)
4 b_hat <- solve(xtx,xty)
5 b_hat

```

```

      [,1]
[1,] -1313.0788097
[2,]    9.6492685
[3,]    8.6331873
[4,]    0.2267639

```

- Validation with `lm()` in R

▶ Run Code



```

1 # fit model with lm()
2 lm_bw_mlr <- lm(Body.Weight ~ Breast.Circumference + BCS + HEI,
3                 data = df_bw_mlr)
4 summary(lm_bw_mlr)

```

Call:

```
lm(formula = Body.Weight ~ Breast.Circumference + BCS + HEI,
    data = df_bw_mlr)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-7.686  -5.001  -2.190   5.715   9.613

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1313.0788	209.3310	-6.273	0.000763	***
Breast.Circumference	9.6493	1.0958	8.805	0.000119	***
BCS	8.6332	2.8939	2.983	0.024533	*
HEI	0.2268	0.1736	1.306	0.239335	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.088 on 6 degrees of freedom

Multiple R-squared: 0.9294, Adjusted R-squared: 0.8942

F-statistic: 26.35 on 3 and 6 DF, p-value: 0.0007476