# Livestock Breeding and Genomics - Solution 8

Peter von Rohr

2023-11-03

## Problem 1: Regression Model

Use the dataset on weaning weight and fit a regression model of weaning weight on breast circumference. The following tasks are to be completed.

- Use matrix-vector notation to specify the model and fill the information from the dataset into the model.
- Compute the solution for the estimated regression coefficient using a least squares approach.
- Use R to verify your result.

The data set is available from

```
## https://charlotte-ngs.github.io/lbgfs2023/data/beef_data_bc.csv
```

### Solution

The regression model in matrix-vector notation is given by the following formula

$$y = X\beta + e$$

where $y$ is the vector of known responses. For our example this is the vector of weaning weights. The vector $\beta$ contains the intercept and the regression coefficient. The vector $e$ contains the random residuals. The matrix $X$ contains a column of all ones and a second column with breast circumference.

The data are imported to a tibble using

```r
tbl_ww <- readr::read_delim(file = s_data_path, delim = ",",
                            col_types = readr::cols(
                              Animal = readr::col_integer(),
                              Sire = readr::col_integer(),
                              Dam = readr::col_integer(),
                              Herd = readr::col_factor(),
                              `Weaning Weight` = readr::col_double(),
                              `Breast Circumference` = readr::col_double()
                            ))
head(tbl_ww)
```

```
## # A tibble: 6 x 6
##   Animal  Sire   Dam Herd  `Weaning Weight` `Breast Circumference`
##    <int> <int> <int> <fct>            <dbl>                  <dbl>
```

```
## 1      12      1      4 1                       2.61                      1.62
## 2      13      1      4 1                       2.31                      1.96
## 3      14      1      5 1                       2.44                      1.48
## 4      15      1      5 1                       2.41                      1.47
## 5      16      1      6 2                       2.51                      1.5
## 6      17      1      6 2                       2.55                      1.47
```

The information for the known vector $y$ and for matrix $X$ are taken from the dataset. The vector $y$ contains all weaning weight values

```
vec_y <- tbl_ww$`Weaning Weight`
```

The matrix $X$ is a matrix with two columns and as many rows as there are observations.

```
n_nr_obs <- nrow(tbl_ww)
mat_X <- matrix(c(rep(1,n_nr_obs), tbl_ww$`Breast Circumference`), nrow = n_nr_obs, ncol = 2)
```

The least squares estimate is

$$\hat{\beta} = (X^TX)^{-1}X^Ty$$

Inserting the numbers into the formula leads to

```
mat_xtx <- crossprod(mat_X)
mat_xty <- crossprod(mat_X, vec_y)
vec_hat_beta <- solve(mat_xtx, mat_xty)
vec_hat_beta
```

```
##          [,1]
## [1,] 1.638243
## [2,] 0.561967
```

The standard deviation of the residuals is used as an estimate of the spread of the error terms.

```
vec_r <- vec_y - crossprod(t(mat_X), vec_hat_beta)
n_sd_res <- sqrt(crossprod(vec_r) / (n_nr_obs-2))
n_sd_res
```

```
##          [,1]
## [1,] 0.315731
```

Verification of the results using R

```
lm_ww <- lm(`Weaning Weight` ~ `Breast Circumference`, data = tbl_ww)
summary(lm_ww)
```

```
##
## Call:
## lm(formula = `Weaning Weight` ~ `Breast Circumference`, data = tbl_ww)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45921 -0.15326  0.01814  0.10909  0.61699
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.6382     0.7775   2.107   0.0536 .
## `Breast Circumference`     0.5620     0.5019   1.120   0.2817
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3157 on 14 degrees of freedom
## Multiple R-squared:  0.0822, Adjusted R-squared:  0.01664
## F-statistic: 1.254 on 1 and 14 DF,  p-value: 0.2817
```

## Problem 2: Fixed Linear Effects Model

Use the same dataset as in Problem 1 and fit a fixed linear effects model using breast circumference and herd as fixed effects in a model. Use the same path to get to the solution as in Problem 1 and complete the same set of tasks.

### Solution

The model in matrix vector notation is the same as in Problem 1

$$y = X\beta + e$$

But the definition of the vector $\beta$ and the matrix $X$ are different. The vectors $y$ and $e$ are defined the same way.

The check with R

```r
lm_ww_bc_herd <- lm(`Weaning Weight` ~ `Breast Circumference` + Herd, data = tbl_ww)
summary(lm_ww_bc_herd)
```

```
##
## Call:
## lm(formula = `Weaning Weight` ~ `Breast Circumference` + Herd,
##     data = tbl_ww)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50908 -0.12643  0.00342  0.10719  0.57491
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.48141    0.88203   1.680    0.117
## `Breast Circumference`    0.63990    0.54870   1.166    0.264
## Herd2                     0.07344    0.17260   0.426    0.677
##
## Residual standard error: 0.3254 on 13 degrees of freedom
## Multiple R-squared:  0.0948, Adjusted R-squared:  -0.04446
## F-statistic: 0.6808 on 2 and 13 DF,  p-value: 0.5234
```