Linear Regression

Peter von Rohr

2024-02-26

Goal

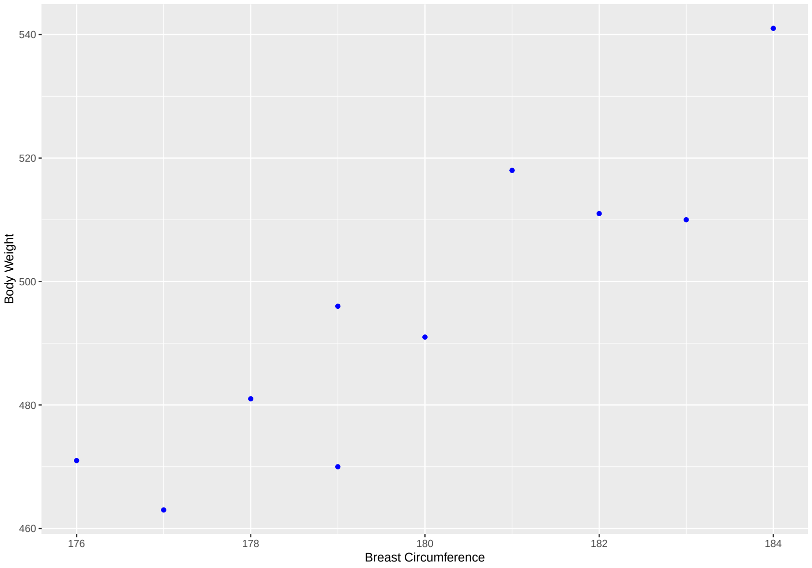
Assessment of relationship between

- a given variable (response) and
- other measurements or observations (predictors) on the same animal

Example

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

Diagram



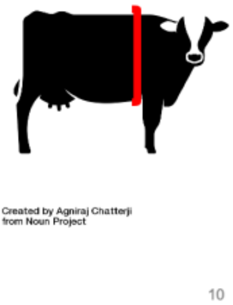
540
520
500
480
460
Body Weight
176
178
180
182
184
Breast Circumference

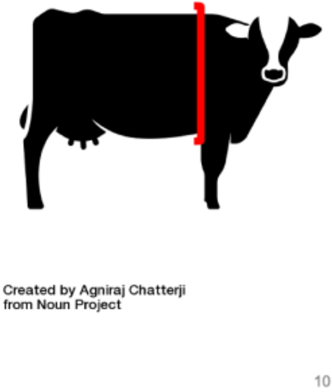
Observations

- relationship between breast circumference and body weight: heavier animals tend to have larger values for breast circumference
- same relationship across whole range $\rightarrow$ **linear** relationship

Regression Model

- quantify relationship between body weight and breast circumference
- practical application: measure band for animals





Model Building

- expected body weight ($E(y)$ in kg) based on an observed value of x cm for breast circumference

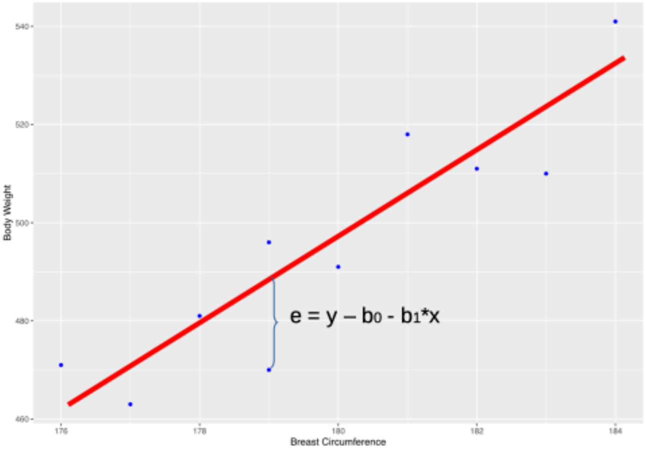
$$E(y) = b_0 + b_1 * x$$

- b_0 and b_1 are unknown parameters of the model
- model is linear function of parameters $\rightarrow$ linear model

Parameter Estimation

- How to find values for b_0 and b_1
- several techniques available: start with Least Squares

Least Squares



Estimators

Find values $\hat{b}_0$ and $\hat{b}_1$ such that

$$\mathbf{e}^T\mathbf{e} = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} \left[y_i - E(y_i)\right]^2 = \sum_{i=1}^{N} \left[y_i - b_0 - b_1 * x_i\right]^2$$

is minimal

Minimization

$$\frac{\partial \mathbf{e}^T\mathbf{e}}{\partial b_0} = -2\sum_{i=1}^{N}\left[y_i - b_0 - b_1 x_i\right]$$
$$= -2\left[\sum_{i=1}^{N} y_i - N b_0 - b_1 \sum_{i=1}^{N} x_i\right]$$

$$\frac{\partial \mathbf{e}^T\mathbf{e}}{\partial b_1} = -2\sum_{i=1}^{N} x_i\left[y_i - b_0 - b_1 x_i\right]$$
$$= -2\left[\sum_{i=1}^{N} x_i y_i - b_0 \sum_{i=1}^{N} x_i - b_1 \sum_{i=1}^{N} x_i^2\right]$$

Minimization II

- Expressions $\frac{\partial \mathbf{e}^T\mathbf{e}}{\partial b_0}$ and $\frac{\partial \mathbf{e}^T\mathbf{e}}{\partial b_1}$ both set to 0
- Solutions obtained will be called $\widehat{b_0}$ and $\widehat{b_1}$
- First introduce simplifying notation

Notation

$$x. = \sum_{i=1}^{N} x_i \quad \text{and} \quad \bar{x}. = \frac{x.}{N}$$

$$y. = \sum_{i=1}^{N} y_i \quad \text{and} \quad \bar{y}. = \frac{y.}{N}$$

$$(x^2). = \sum_{i=1}^{N} x_i^2$$

$$(xy). = \sum_{i=1}^{N} x_i y_i$$

Normal Equations

$$N\widehat{b_0} + \widehat{b_1}x. = y.$$

$$\widehat{b_0}x. + \widehat{b_1}(x^2). = (xy).$$

Solutions

$$\hat{b}_0 = \bar{y}. - \hat{b}_1 \bar{x}.$$

$$\hat{b}_1 = \frac{(xy). - N\bar{x}.\bar{y}.}{(x^2). - N\bar{x}.^2}$$

Example Dataset

$$N = 10,\ \bar{x}. = 179.9,\ \bar{y}. = 495.2$$

$$(xy). = 8.91393 \times 10^5,\ (x^2). = 3.23701 \times 10^5$$

$$\hat{b}_1 = \frac{8.91393 \times 10^5 - 10 * 179.9 * 495.2}{3.23701 \times 10^5 - 10 * 179.9^2} = 8.673$$

$$\hat{b}_0 = 495.2 - 8.6732348 * 179.9 = -1065.115$$

Estimates in R

```
lm_bw_bc <- lm(`Body Weight` ~ `Breast Circumference`, data = tbl_reg)
summary(lm_bw_bc)
```

```
Call:
lm(formula = `Body Weight` ~ `Breast Circumference`, data = tbl_reg)

Residuals:
     Min       1Q   Median       3Q      Max 
-17.3941  -6.5525  -0.0673   9.3707  13.2594 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)            -1065.115    255.483  -4.169 0.003126 ** 
`Breast Circumference`     8.673      1.420   6.108 0.000287 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.08 on 8 degrees of freedom
Multiple R-squared:  0.8234,    Adjusted R-squared:  0.8014 
F-statistic: 37.31 on 1 and 8 DF,  p-value: 0.000287
```

General Case

- More x variables ...
- Matrix Vector Notation

$$\mathbf{X} = \begin{bmatrix} x_{10} & x_{11} & x_{12} \\ x_{20} & x_{21} & x_{22} \\ . & . & . \\ . & . & . \\ . & . & . \\ x_{N0} & x_{N1} & x_{N2} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_N \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_N \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$