Linear Regression

Peter von Rohr

2023-02-20

Goal

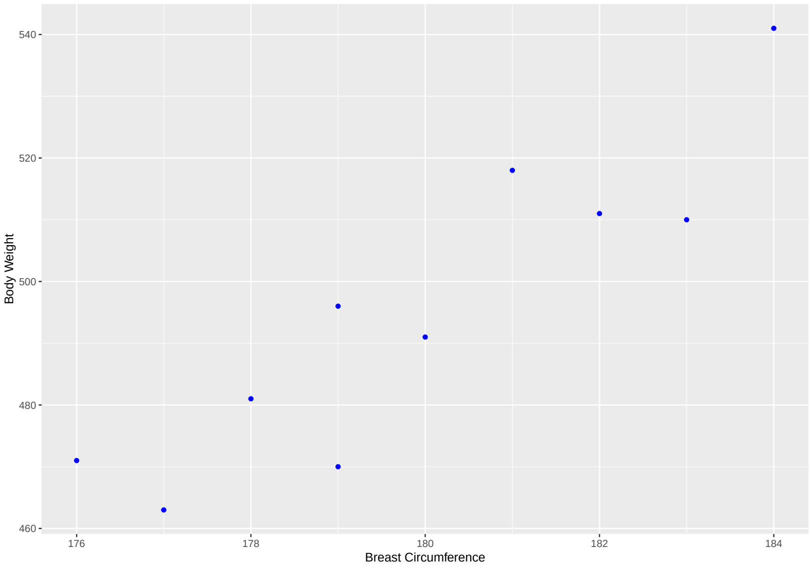
Assessment of relationship between

- 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



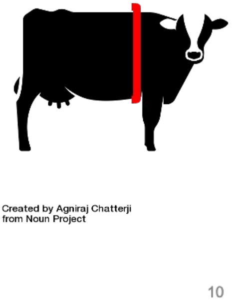
Body Weight
540
520
500
480
460
176
178
180
182
184
Breast Circumference

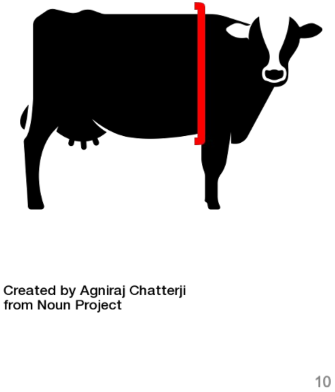
Observations

- 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

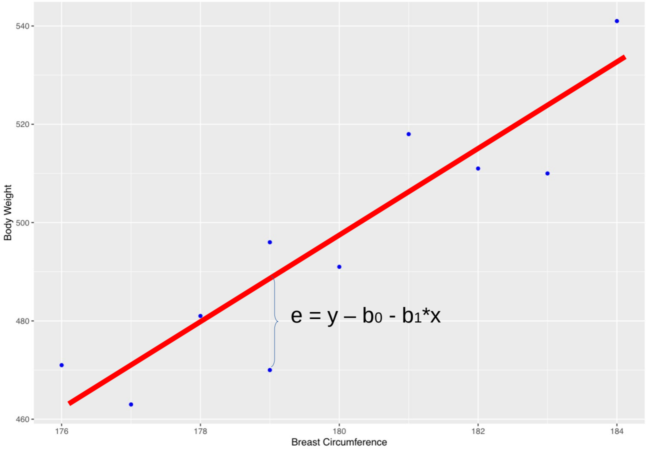
$$E(y) = b_0 + b_1 * x$$

- 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

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

is minimal

Minimization

$$
\begin{aligned}
\frac{\partial \mathbf{e}^T\mathbf{e}}{\partial b_0} &= -2\sum_{i=1}^{N}\left[y_i - b_0 - b_1 x_i\right] \\
&= -2\left[\sum_{i=1}^{N} y_i - Nb_0 - b_1\sum_{i=1}^{N} x_i\right]
\end{aligned}
$$

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

Notation

$$x_{\cdot} = \sum_{i=1}^{N} x_i$$

$$y_{\cdot} = \sum_{i=1}^{N} y_i$$

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

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

$$\bar{x}_{\cdot} = \frac{x_{\cdot}}{N}$$

$$\bar{y}_{\cdot} = \frac{y_{\cdot}}{N}$$

Solutions

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

and

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