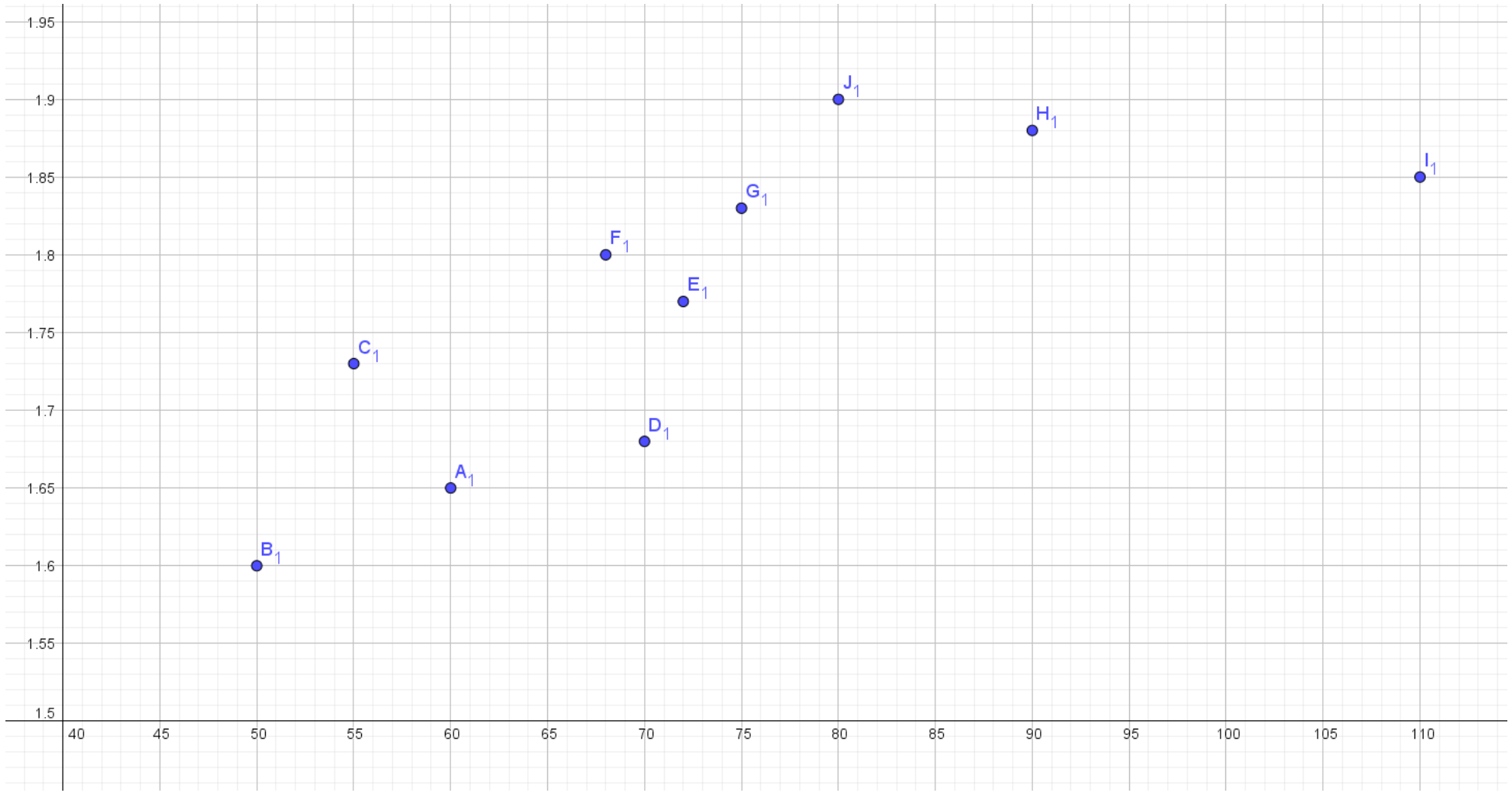
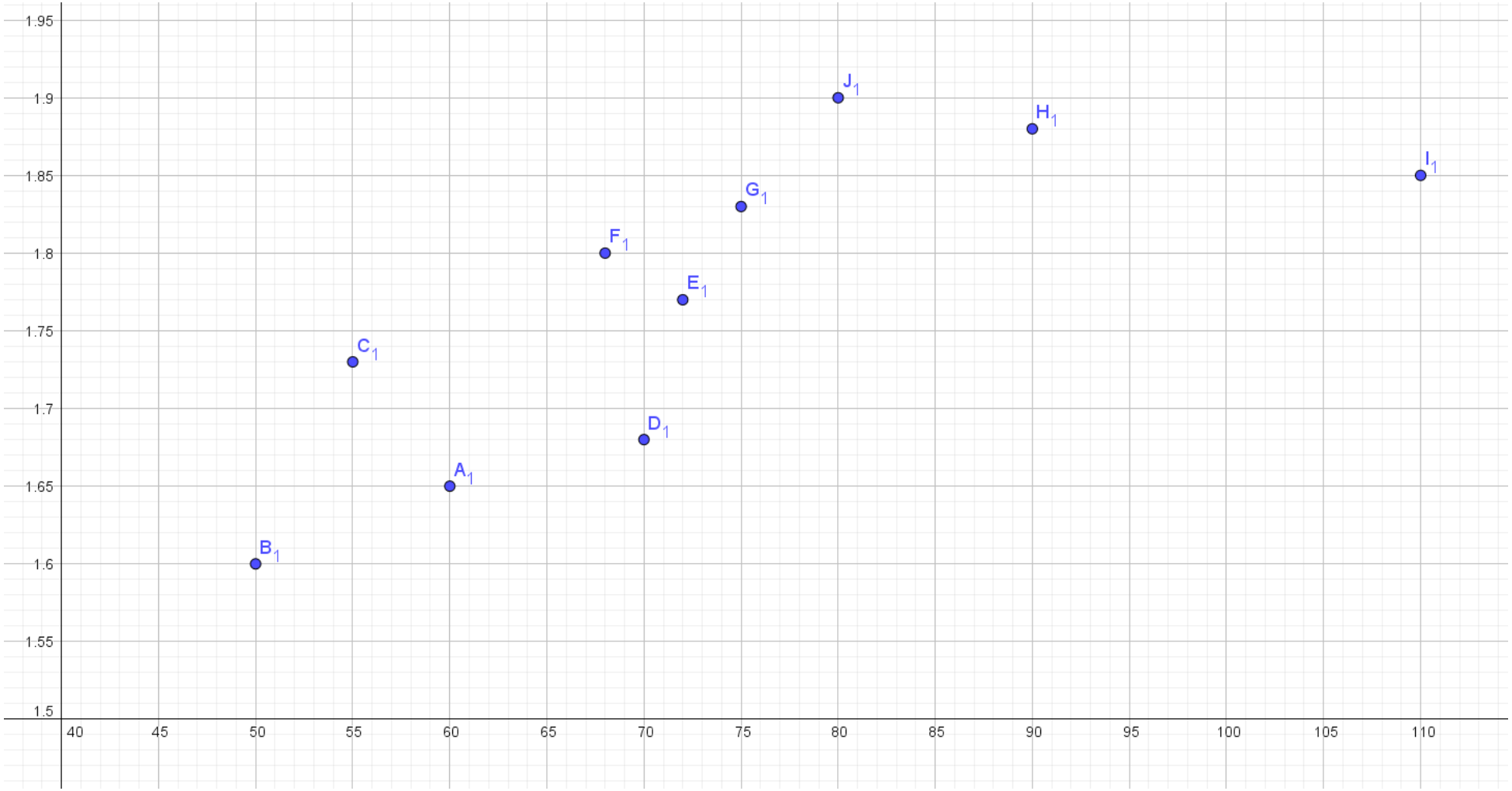
Lineare Methoden zur Regression

Philipp Engels

03.05.2019

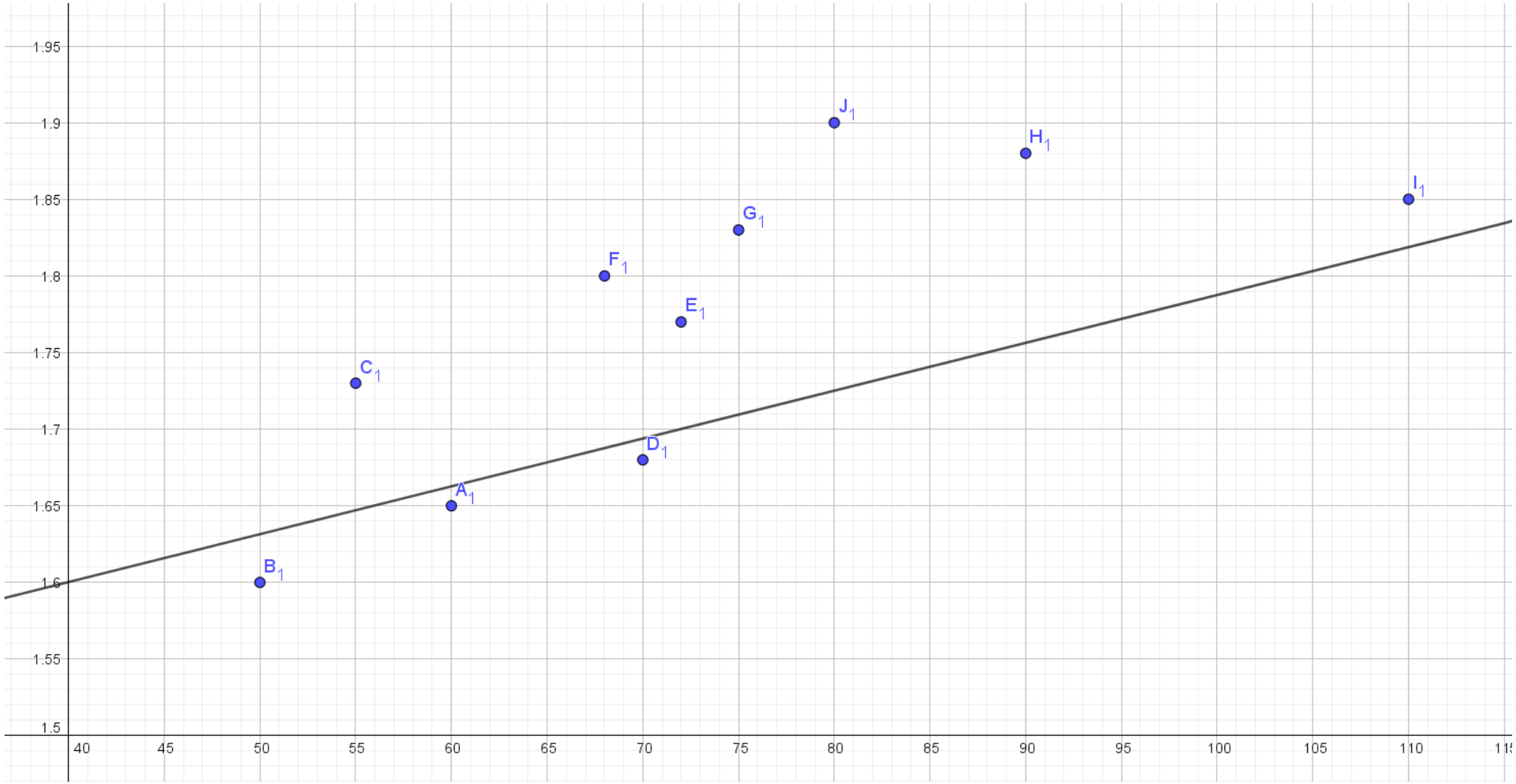
Größe	1,6	1,73	1,65	1,8	1,68	1,77	1,83	1,9	1,88	1,85
Gewicht	50	55	60	68	70	72	75	80	90	110

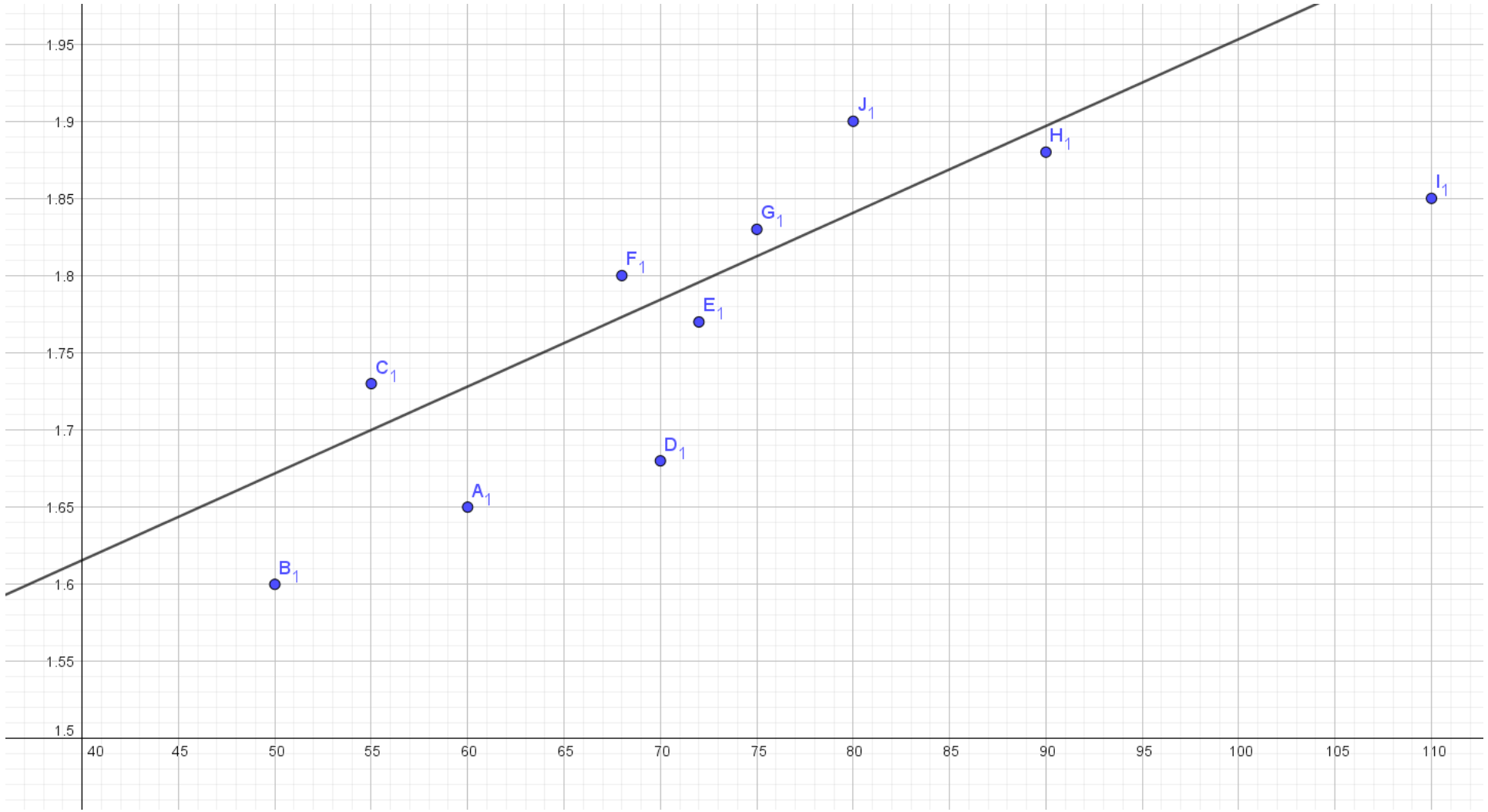




Wir wollen eine Gerade durch unsere Daten legen, die sie möglichst gut abbildet

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \longrightarrow y = \beta_0 + \beta_1 * \text{ Gewicht}$$

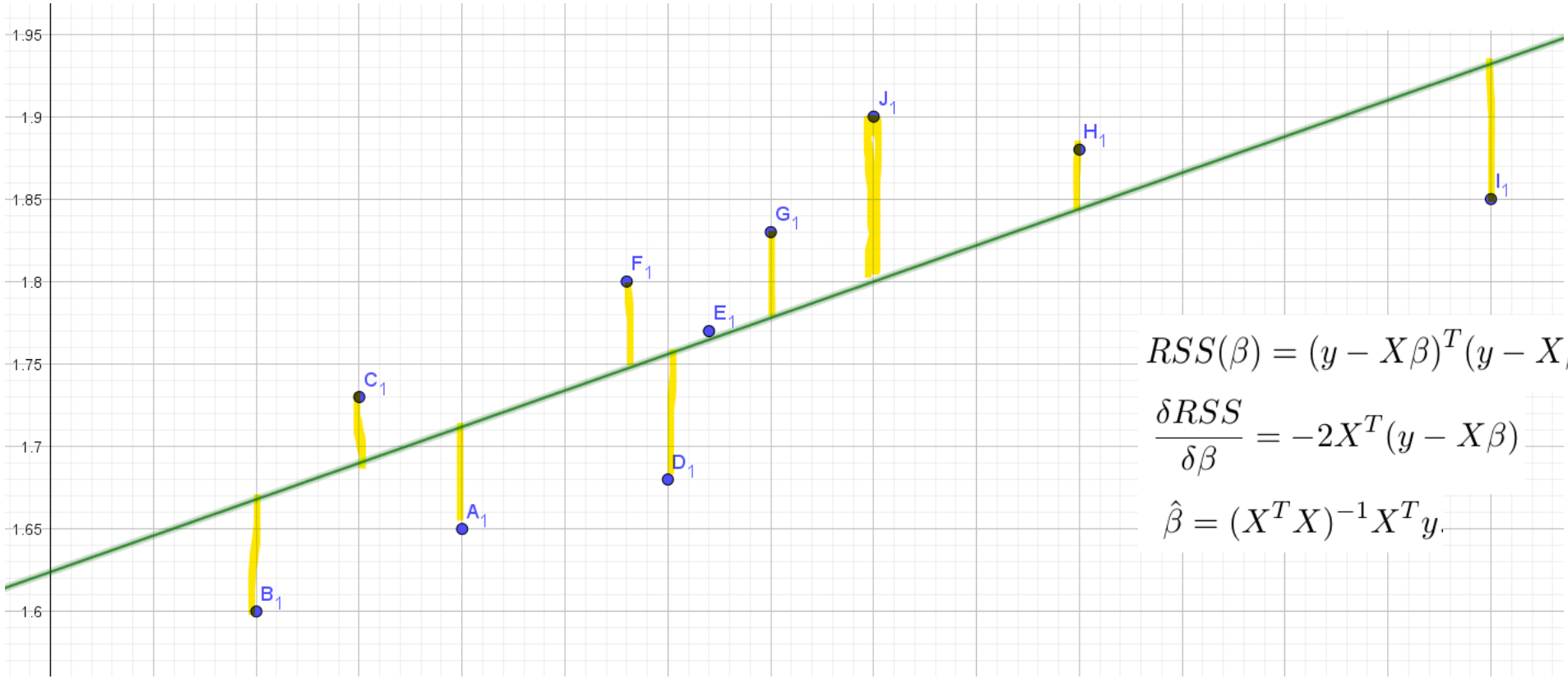

1.95
1.9
1.85
1.8
1.75
1.7
1.65
1.6
1.55
1.5
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
J1
H1
I1
G1
F1
E1
C1
D1
A1
B1


1.95
1.9
1.85
1.8
1.75
1.7
1.65
1.6
1.55
1.5
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
J1
H1
I1
G1
F1
E1
C1
D1
A1
B1

Methode der kleinsten Quadrate

$$RSS(\beta) = \sum_{i=1}^{N}(y_i - f(x_i))^2 = \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$$

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j\beta_j$$

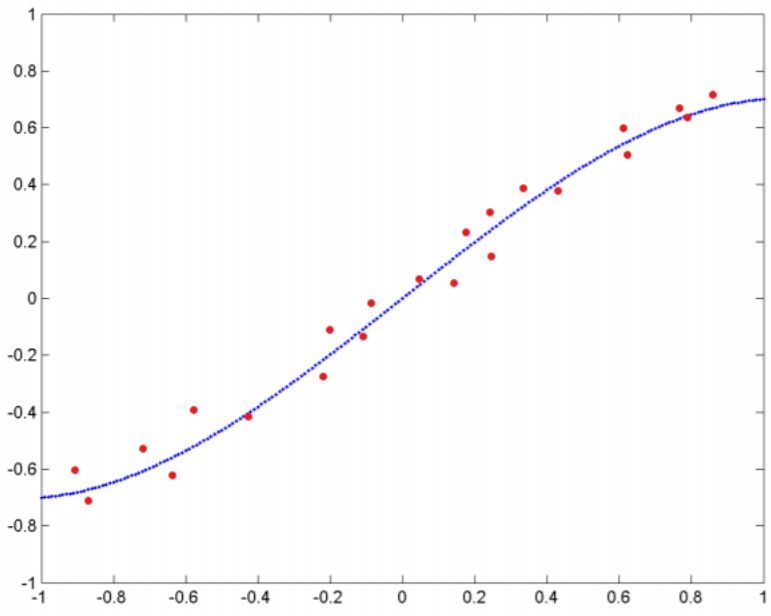


$$RSS(\beta) = (y - X\beta)^T(y - X\beta)$$

$$\frac{\delta RSS}{\delta \beta} = -2X^T(y - X\beta)$$

$$\hat{\beta} = (X^T X)^{-1}X^T y.$$

Keine Grade? – Kein Problem!



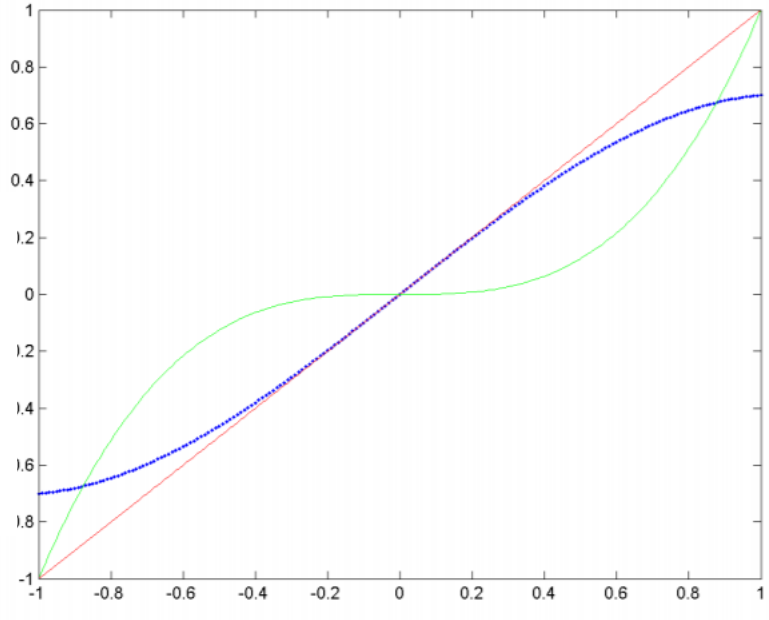
$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$$

$$f(x) = \sum_{j=1}^{4} \beta_j \phi_j(x)$$

$$\phi_1(x) = 1 \quad \phi_2(x) = x \quad \phi_3(x) = x^2 \quad \phi_4(x) = x^3$$

$$\beta = (0, 1, 0, -0.3)^T$$

$$\Rightarrow f(x) = x - 0.3x^2$$



Singulärwertzerlegung

$X = UDV^T$

$U \in \mathbb{R}^{N \times p}, V \in \mathbb{R}^{p \times p}$ orthogonale Matrizen

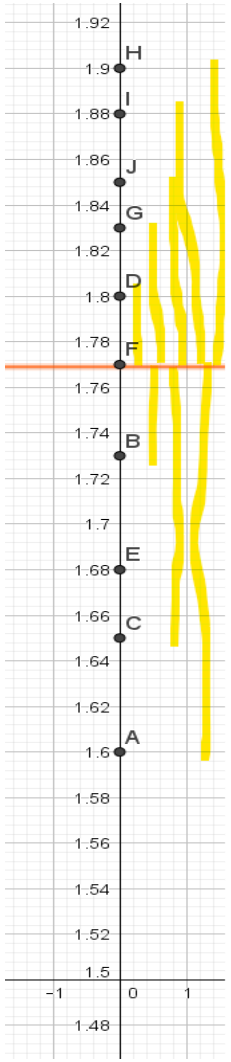
$D \in \mathbb{R}^{p \times p}$ Diagonalmatrix mit Diagonaleinträgen $d_1 \geq d_2 \geq ... \geq d_p \geq 0$

$$A_1 = \begin{pmatrix} 4 & 12 \\ 12 & 11 \end{pmatrix} = \begin{pmatrix} 3/5 & 4/5 \\ 4/5 & -3/5 \end{pmatrix} \begin{pmatrix} 20 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix}$$

$$RSS(\beta) = (y - X\beta)^T(y - X\beta) = ||y - X\beta||^2 = ||X\beta - y||^2$$

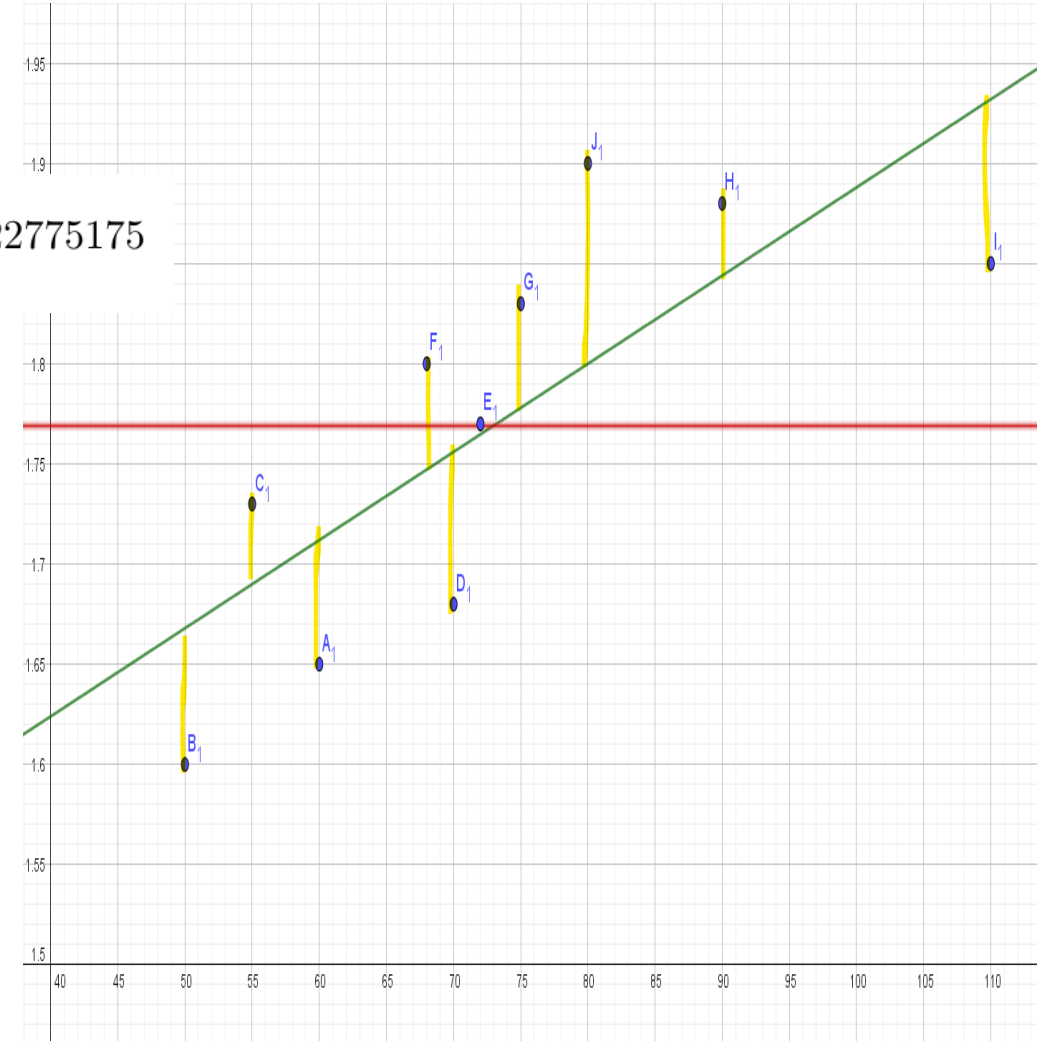
$$\begin{aligned} ||X\beta - y||^2 &= ||UDV^T\beta - y||^2 \\ &= ||UDb - y||^2 \qquad (b = V^T\beta) \\ &= ||U^TUDb - U^Ty||^2 \\ &= ||Db - U^Ty||^2 \\ &= ||Db - u||^2 \qquad (z = U^Ty) \\ &= (d_1b_1 - z_1)^2 + ... + (d_rb_r - z_r)^2 + z_{r+1}^2 + ... + z_N^2 \end{aligned}$$

F-Statistik – Signifikanztest für Koeffizienten



$$F = \frac{(RSS_0 - RSS_1)(p_1 - p_0)}{RSS_1/(N - p_1 - 1)}$$

$$F = \frac{(0,09289 - 0,03938288)(2-1)}{0,03938288/(10-2-1)} = 12,22775175$$



Das Gauß-Markov Theorem

Satz 1. *Wenn wir einen anderen linearen Schätzer* $\tilde{\theta} = c^T y$, *der erwartungstreu für* $a^T\beta$ *ist, also* $E(c^T y) = a^T\beta$, *dann gilt*

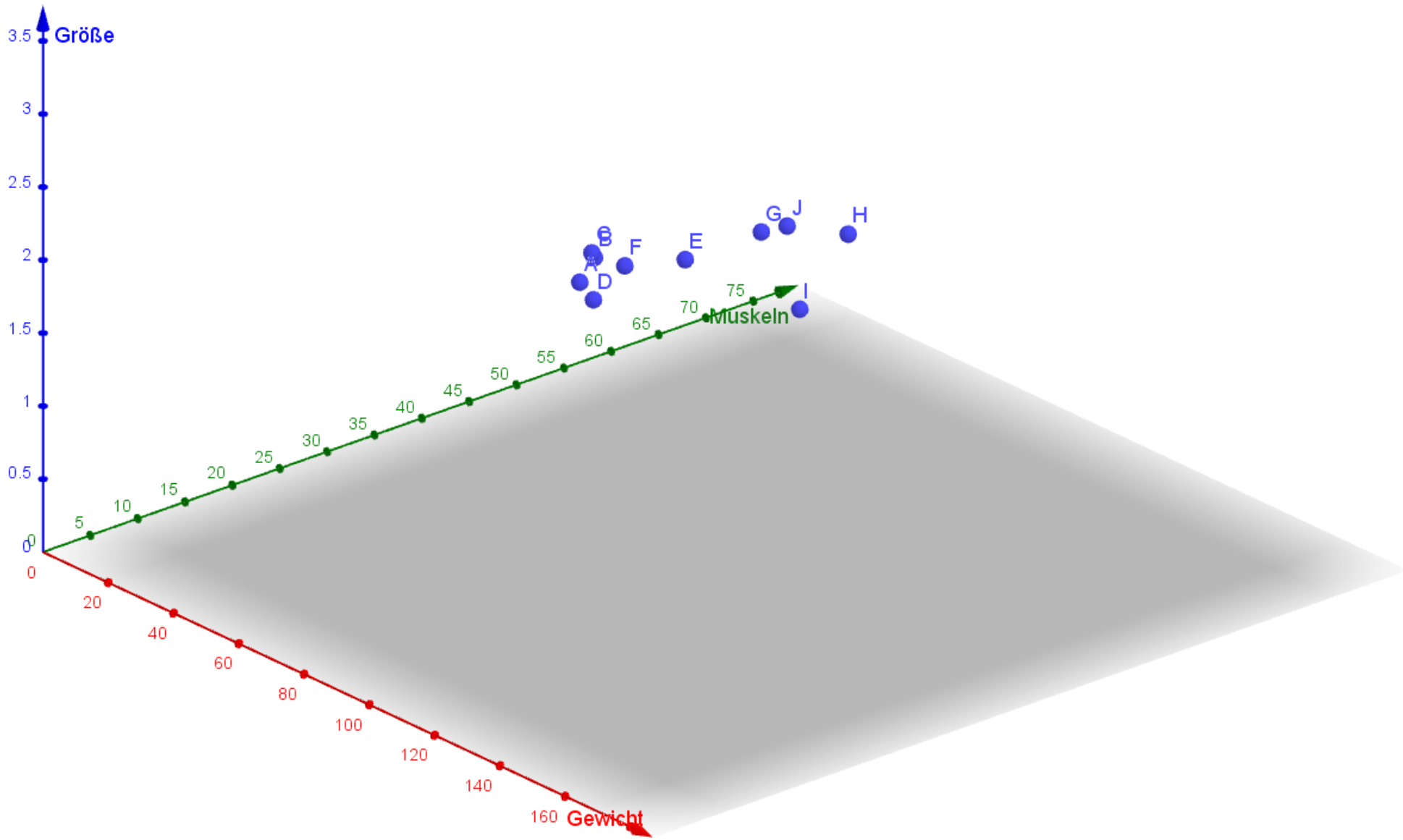
$$Var(a^T\hat{\beta}) \leq Var(c^T y)$$

Durchschnittliche quadratische Fehler eines Schätzers $\tilde{\theta}$:

$$\text{MSE}(\tilde{\theta}) = E(\tilde{\theta} - \theta)^2 = Var(\tilde{\theta}) + [E(\tilde{\theta}) - \theta]^2$$

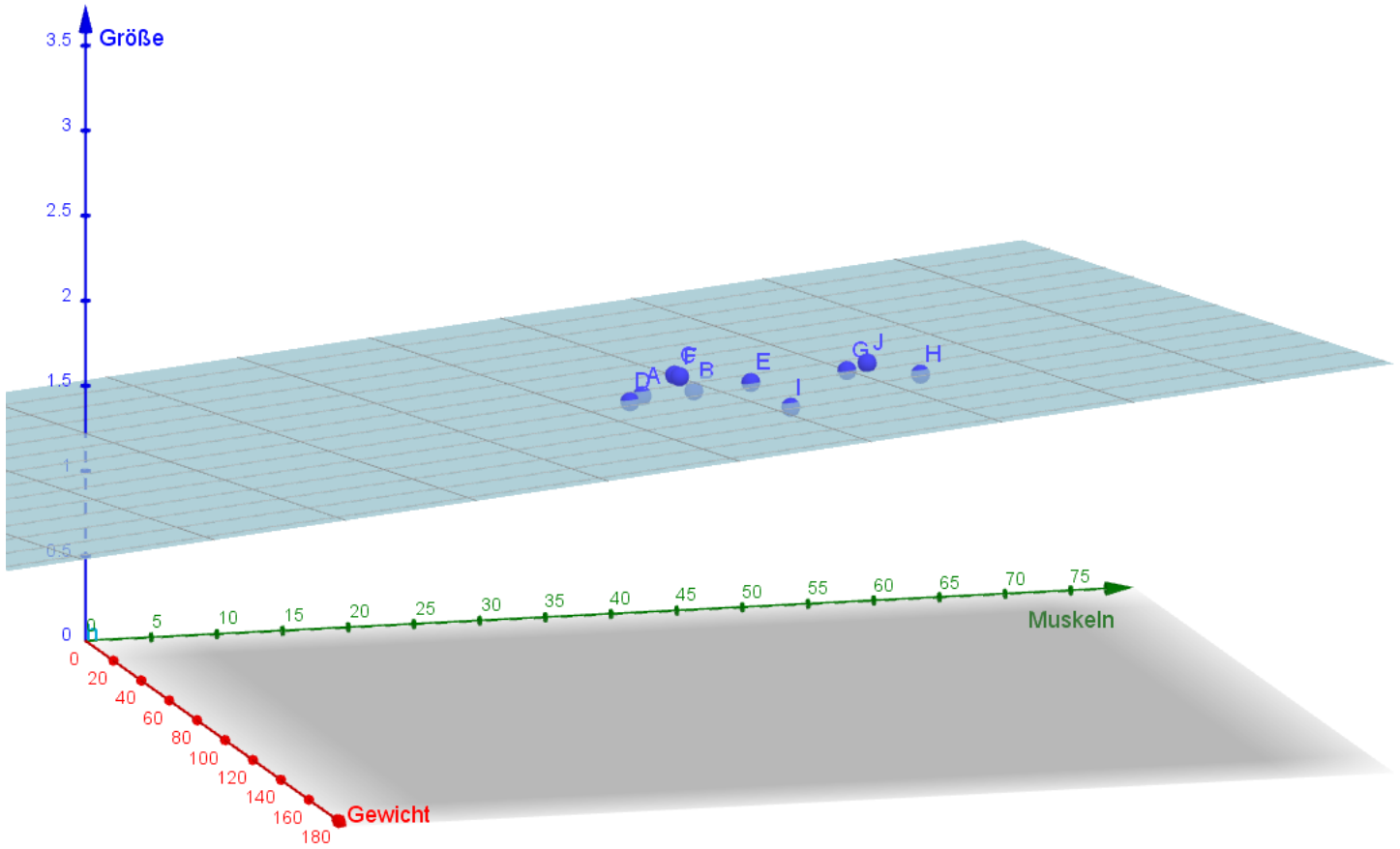
Multiple Regression

Größe	1,6	1,73	1,65	1,8	1,68	1,77	1,83	1,9	1,88	1,85
Gewicht	50	55	60	68	70	72	75	80	90	110
Muskelanteil	41	39	36	38	34	43	50	51	54	42


Größe
3.5
3
2.5
2
1.5
1
0.5
0
C
B
F
E
G
J
H
A
D
I
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Muskeln
0
20
40
60
80
100
120
140
160
Gewicht

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \longrightarrow y = \beta_0 + \beta_1 * \text{Gewicht} + \beta_2 * \text{Muskelanteil}$$

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{i=1}^{p} x_{ij}\beta_j)^2 \longrightarrow \hat{\beta} = (X^T X)^{-1} X^T y$$



Orthogonalisiere die Vektoren x:

$$1. z_0 = x_0 = 1$$

$$2. \text{ Für } j = 1, 2, ..., p$$

$$\hat{\gamma}_{lj} = \frac{\langle z_l, x_j \rangle}{\langle z_l, z_l \rangle}$$

$$z_j = x_j - \sum_{k=0}^{j-1} \hat{\gamma}_{kj} z_k$$

$$3. \hat{\beta}_p = \frac{\langle z_p, y \rangle}{\langle z_p, z_p \rangle}$$

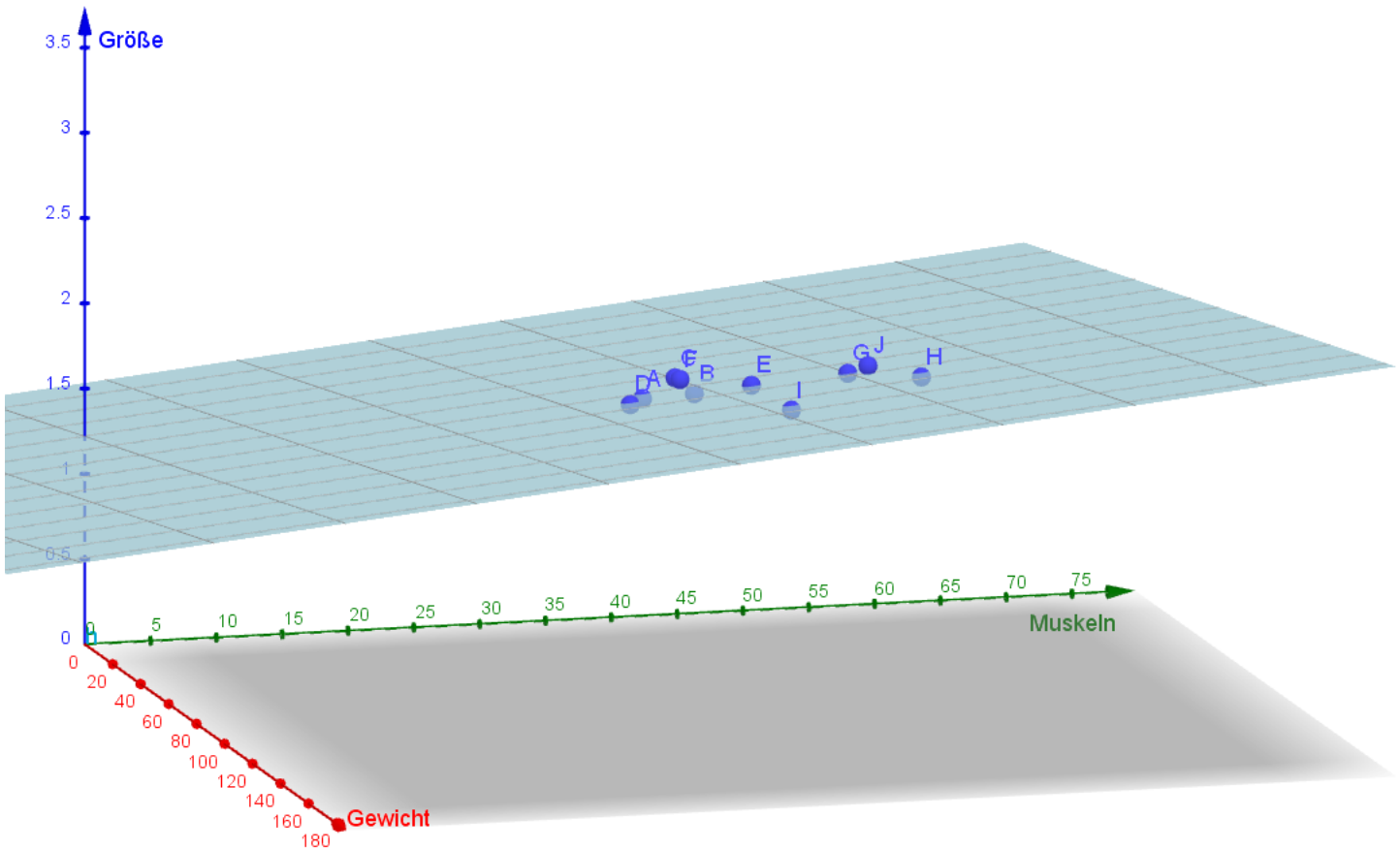
repräsentiert den zusätzlichen Beitrag von x_j auf y, nachdem x_j für $x_0, x_1, .., x_{j-1}, x_{j+1}, ..., x_p$ reguliert wurde.

$$Var(\hat{\beta}_p) = \frac{\sigma^2}{\langle z_p, z_p \rangle} = \frac{\sigma^2}{||z_p||^2}.$$

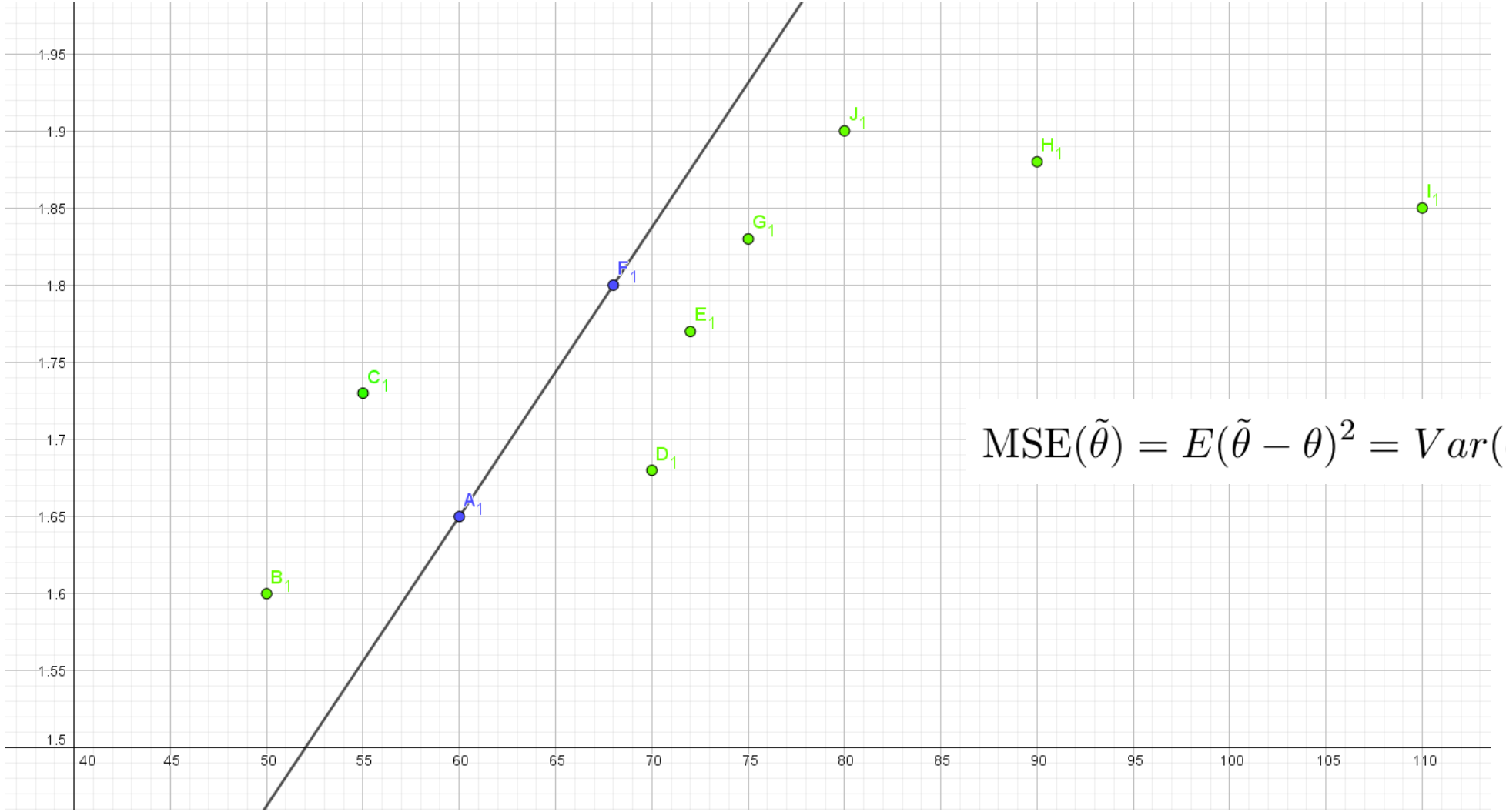
Das heißt die Genauigkeit mit der wir den Schätzer $\hat{\beta}_p$ berechnen können hängt von der Länge von z_p ab, die uns sagt, wie sehr x_p von den anderen x_k nicht erklärt wird.

In unserem Beispiel: $\hat{\beta}_1 = 0,0044$ und $\hat{\beta}_2 = 0,0076$

Tauschen wir Gewicht und Muskelanteil erhalten wir: $\hat{\beta}_1 = 0,0112$ und $\hat{\beta}_2 = 0,0031$



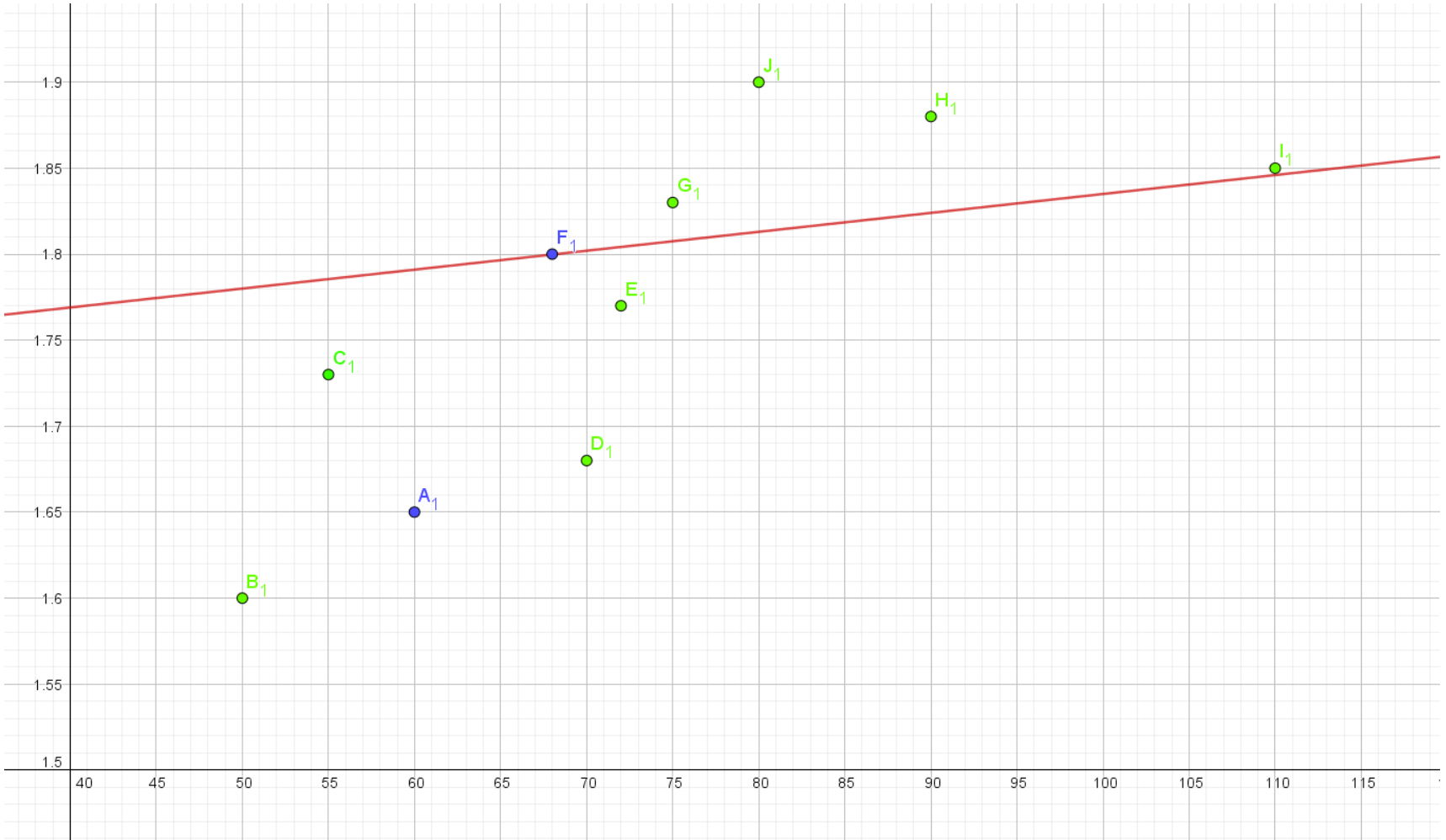
Ridge Regression



$$\mathrm{MSE}(\tilde{\theta}) = E(\tilde{\theta} - \theta)^2 = Var(\tilde{\theta}) + [E(\tilde{\theta}) - \theta]^2$$

$$\hat{\beta} = \min_{\beta}\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_i\beta_j)^2\} \longrightarrow \hat{\beta}^{ridge} = \min_{\beta}\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_i\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\},$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty \longrightarrow \hat{\beta}^{ridge} = (X^TX + \lambda I)^{-1}X^Ty$$



$\lambda = 500$

Lasso Regression

$$\hat{\beta}^{ridge} = \min_{\beta}\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_i\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\},$$

$$\hat{\beta}^{lasso} = \min_{\beta}\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\}.$$

Vorteil:
Größe $= \beta_0 + \beta_1 *$ Gewicht $+\beta_2 *$ Muskelanteil $+\beta_3 *$ Geschwindigkeit eine Schwalbe $+\beta_4 *$ Sonnenstand $+...$

Nachteil:
Keine geschlossene Lösungsformel

Quellen

“The Elements of Statistical Learning", S.42-69 T. Hastie, R. Tibshirani, J. Friedman

“Machine Learning Refined", S.45-72 J. Watt, R. Borhami, A. Katsaggelos