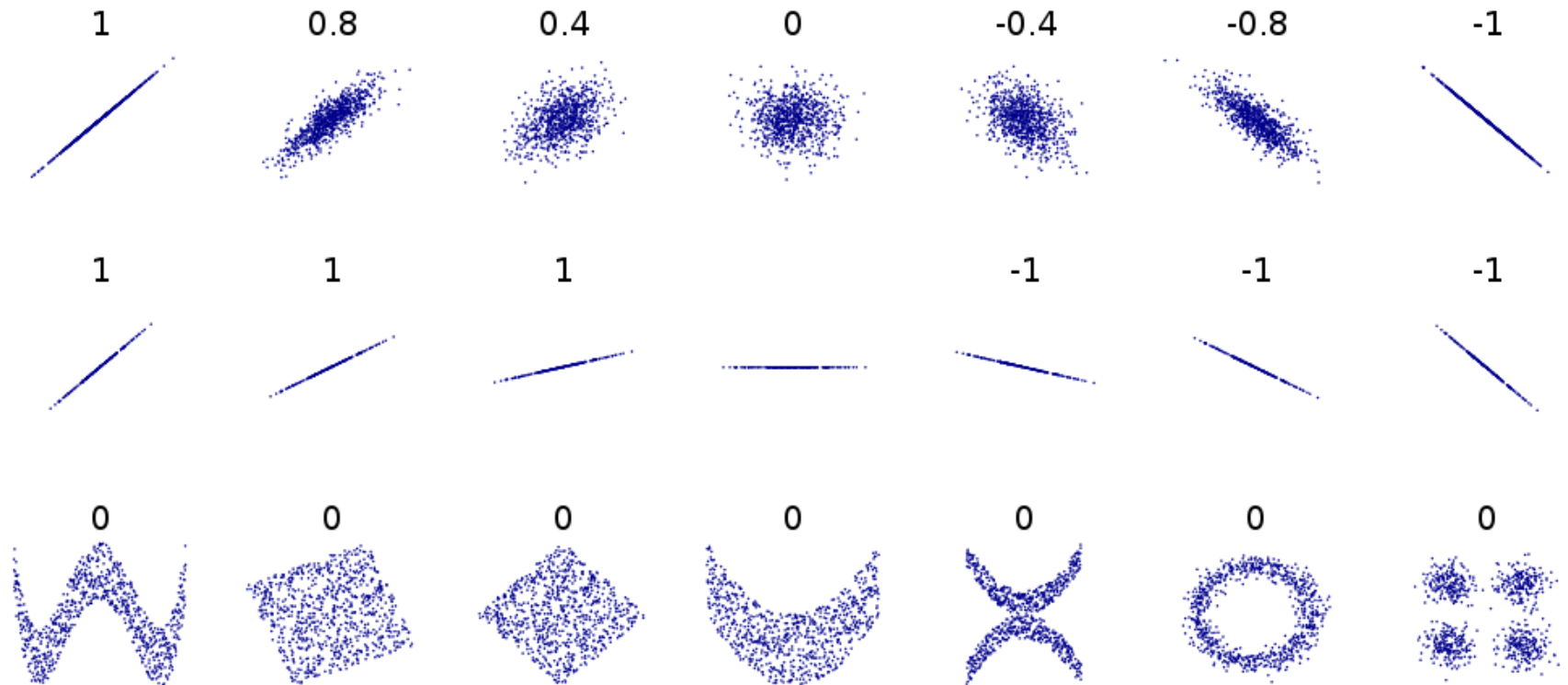


Einfache lineare Regression

Statistik (Biol./Pharm./HST) – FS 2014



Wdh: Korrelation



Big picture: Generalized **Linear** Models (GLMs)

- **Bisher:** Population wird mit einer Verteilung beschrieben
Bsp: Medikament wirkt mit 30% Wa. Wie wa. ist es, dass bei 10 Patienten mindestens 5 gesund werden?

$$X \sim \text{Bin}(10, \pi = 0.3)$$

- **Neu:** **Parameter** dieser Verteilung **hängt von erklärenden Variablen ab.**

Bsp: Wirkwa. hängt von Dosis D ab. Bei welcher Dosis werden im Mittel 90% der Patienten gesund?

$$X \sim \text{Bin}(10, \pi) \text{ und } \pi = f(D)$$

- **Generalized Linear Models:** Zshg zw. erklärenden Variablen (z.B. Dosis) und Parametern einer Verteilung (z.B. Erfolgswa. in Binomialverteilung)

Bsp 1: Wirkung von Medikament

- X: Dosis des Wirkstoffs; n: Patienten, p: Genesungswa.
Y: Anz. gesunder Patienten nach Behandlung

- $Y \sim \text{Bin}(n, p(x))$

- Zshg. zwischen p und x z.B.:

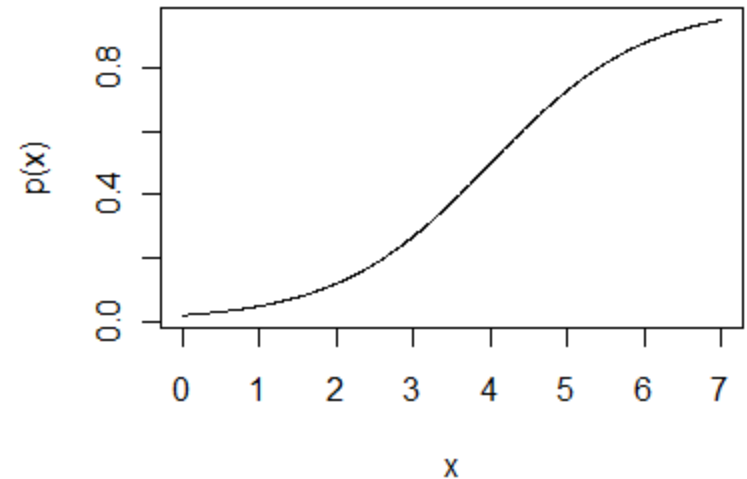
$$p(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

- Kann man umformen zu:

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x$$

Logistische Funktion

Linear in β 's



“Bei welcher Dosis ist die Genesungswa. 80%?”

- “Logistische Regression”, “Binomialregression”

Bsp 2: Anzahl Autounfälle im Winter

- Y: Anz. Autounfälle pro Tag in ZH
X: Temperatur in Celsius

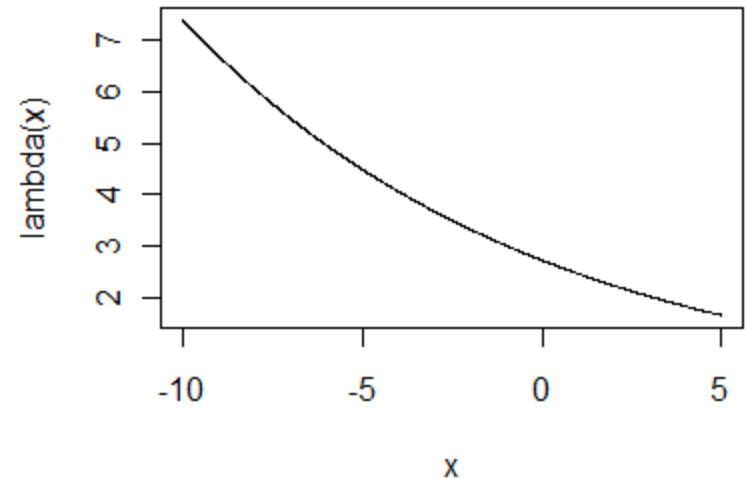
- $Y \sim \text{Pois}(\lambda(x))$

- Zshg. zw. λ und x z.B.:
 $\lambda(x) = \exp(\beta_0 + \beta_1 x)$

- Kann man umformen zu:
 $\log(\lambda(x)) = \beta_0 + \beta_1 x$

Linear in β 's

- “Poissonregression”



Morgen wird es -5 C.

Was ist das 95%-Quantil
der Unfälle morgen?

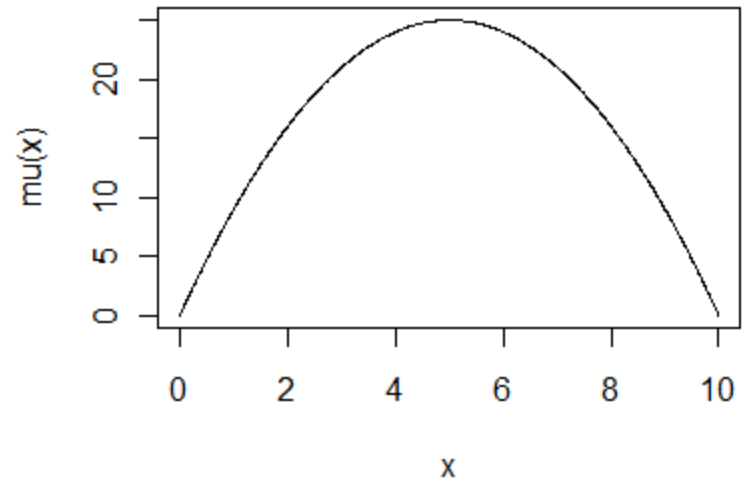
Bsp 3: Kraftzuwachs bei Training

- Y: Kraftzuwachs nach 6 Wochen Training bei Anfängern
X: Trainingszeit pro Woche

- $Y \sim N(\mu(x), \sigma^2)$

- Zshg. zw. μ und x z.B.:
$$\mu(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

Linear in β 's



Welche Trainingsdauer pro Woche
bringt optimalen Kraftzuwachs?

- “**Lineare** Regression”
- Einfache Lineare Regression
 $\mu(x) = \beta_0 + \beta_1 x$ (eine Erklärende)
- Multiple Lineare Regression
 $\mu(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \text{etc}$ (mehrere Erklärende)

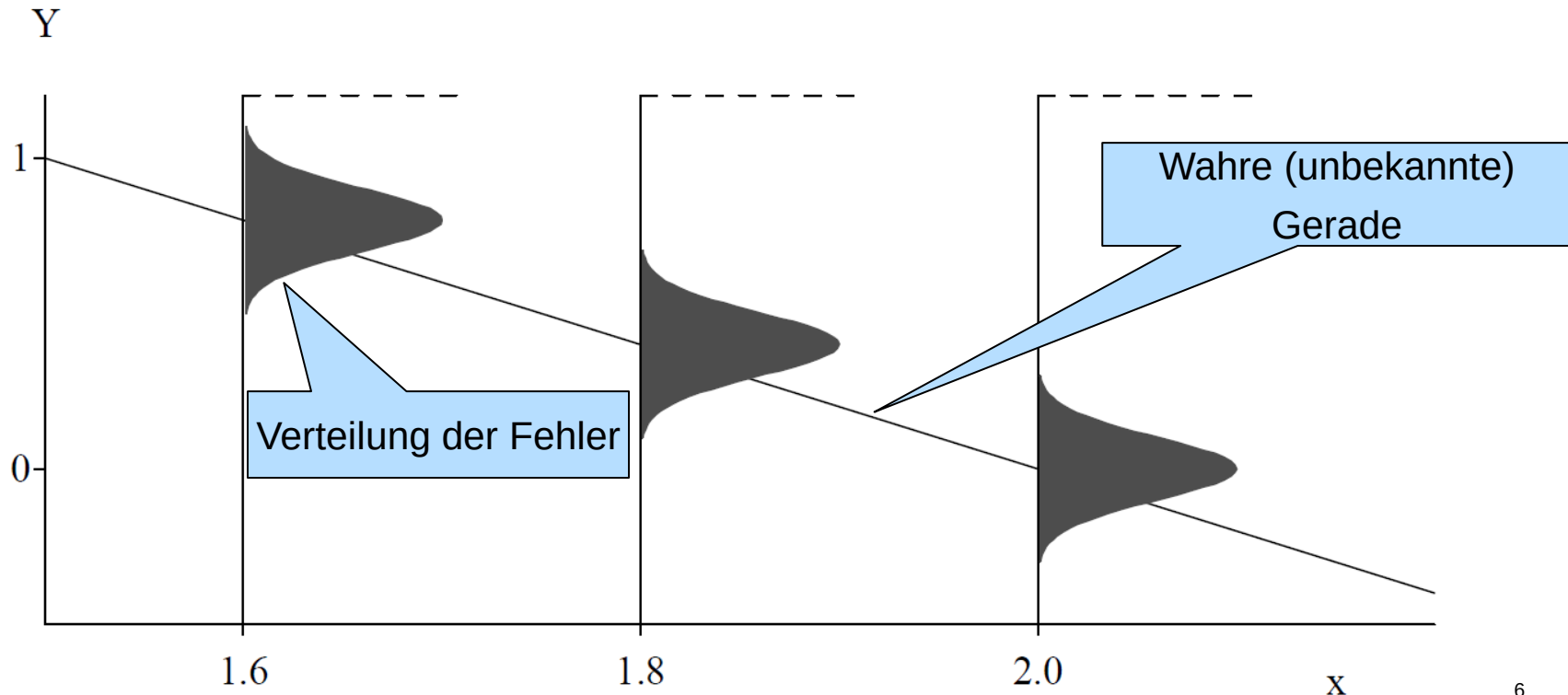
Lineare Regression: Zwei Definitionen

1. $Y \sim N(\mu(x), \sigma^2)$
 $\mu(x) = \beta_0 + \beta_1 x$

Def 1 und Def 2 sind äquivalent

2. $Y = \beta_0 + \beta_1 x + \varepsilon$
 $\varepsilon \sim N(0, \sigma^2)$

$$E(Y) = E(\beta_0 + \beta_1 x + \varepsilon) = \beta_0 + \beta_1 x + E(\varepsilon) = \beta_0 + \beta_1 x$$
$$\text{Var}(Y) = \text{Var}(\beta_0 + \beta_1 x + \varepsilon) = \text{Var}(\varepsilon) = \sigma^2$$





Welche Schlange?

Kasse 1

3

2

3

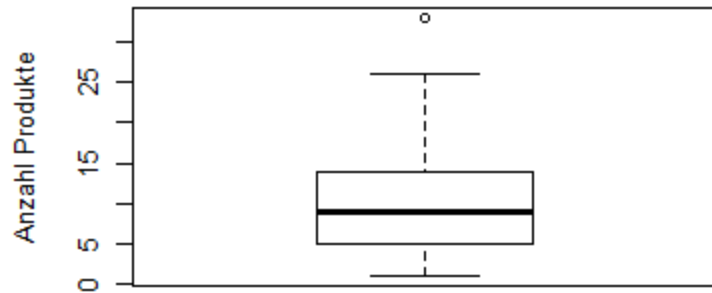
Kasse 2

19

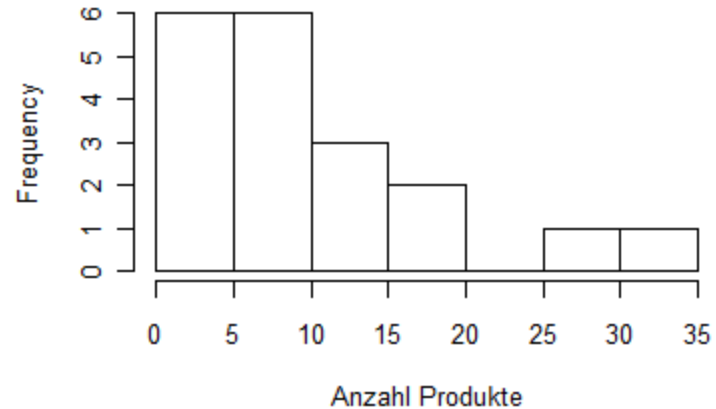
Coop Hauptbahnhof

Di, 22.11.2011, 17:40 – 18:00
(eine Kassierererin)

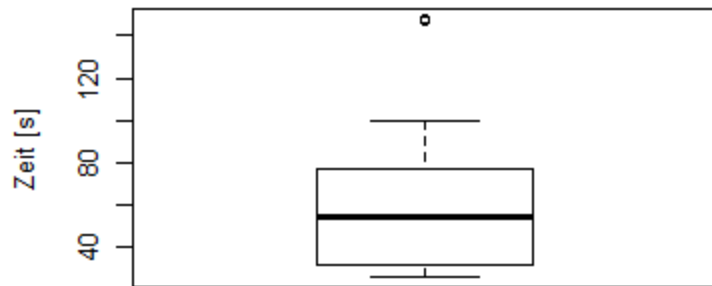
Anzahl Produkte



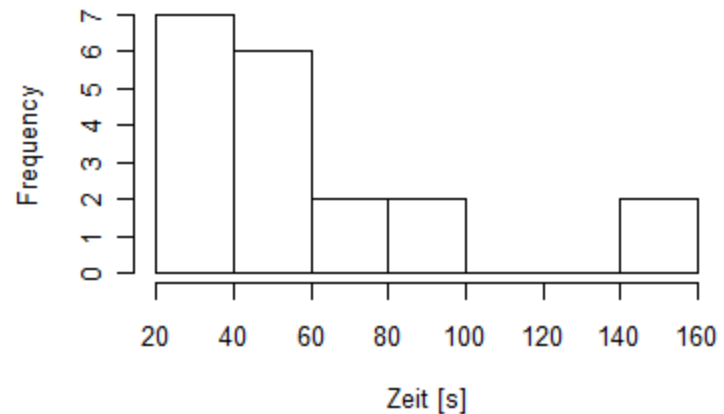
Anzahl Produkte



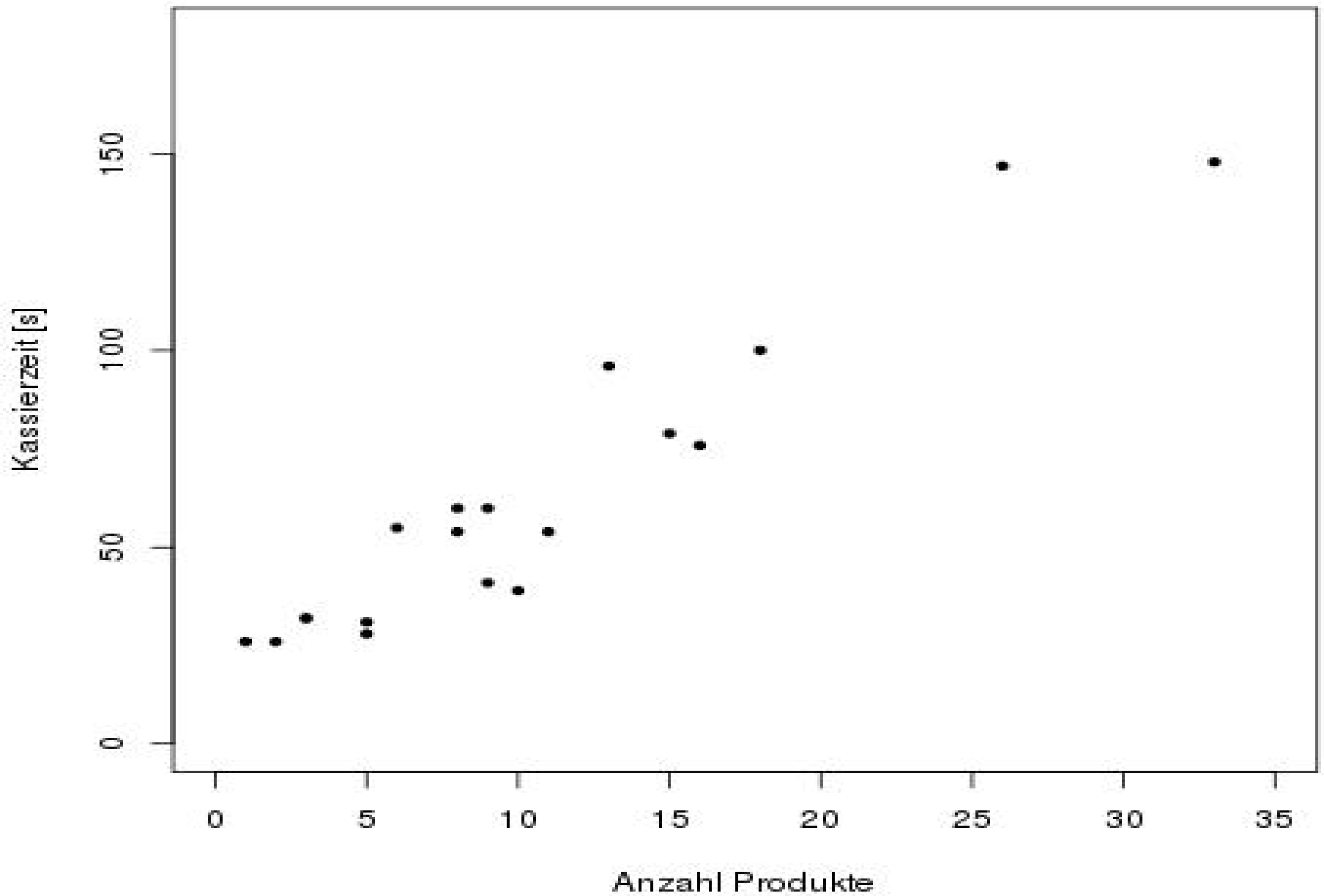
Kassierzeit



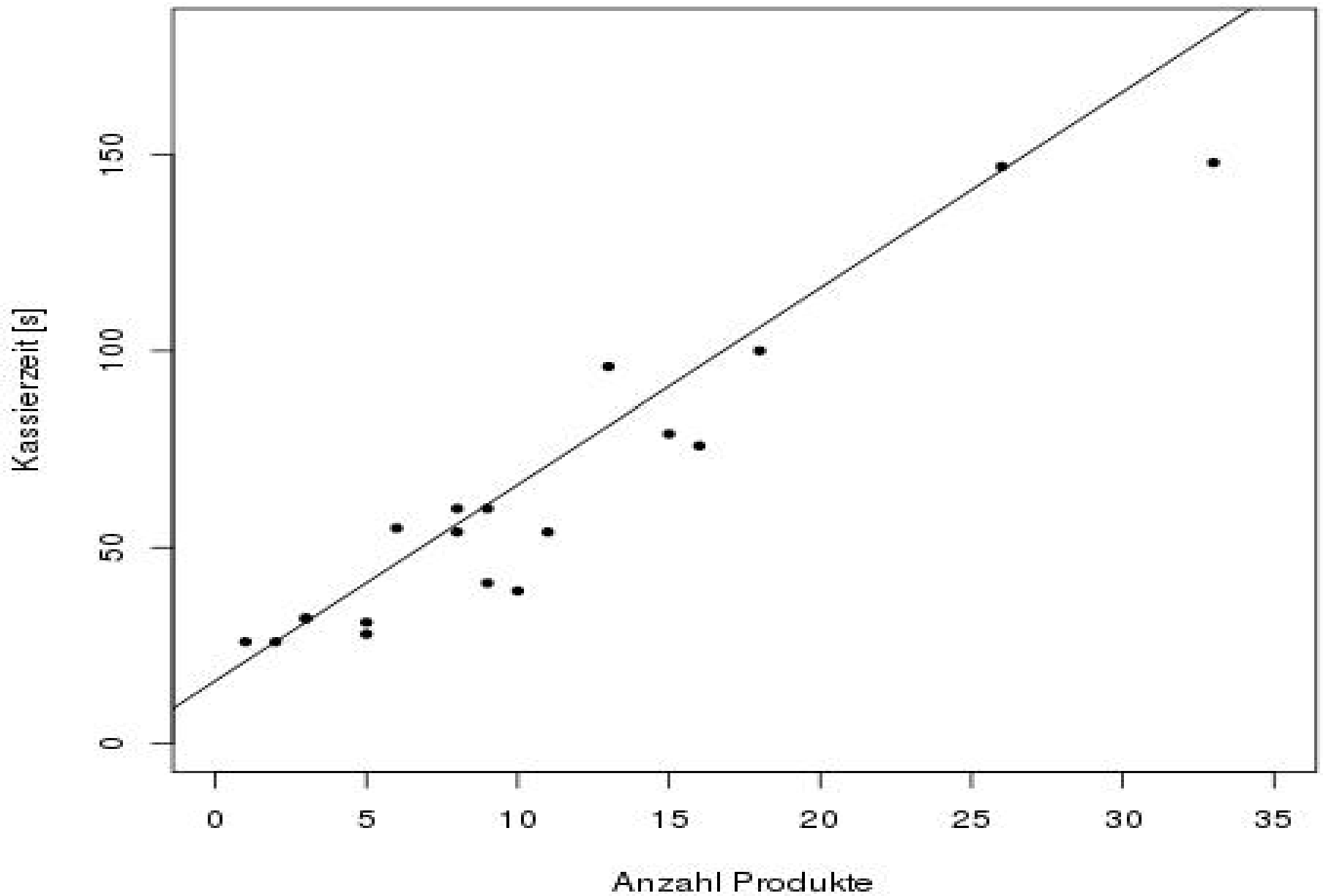
Kassierzeit



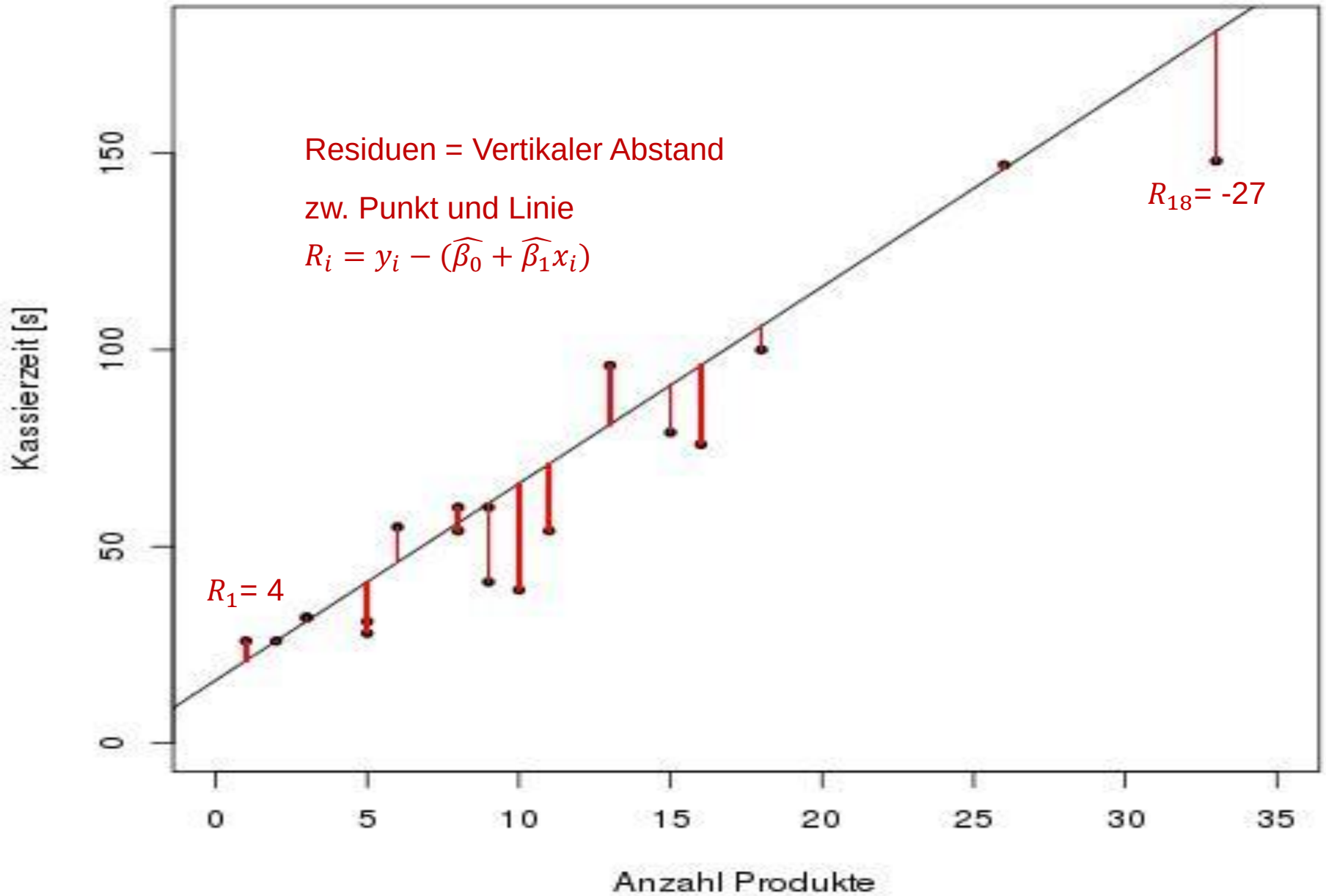
Streudiagramm



Streudiagramm



Streudiagramm



Parameterschätzung – Variante 1:

Methode der kleinsten Quadrate (“Least Squares”, LS)

- Welche Gerade passt am besten zu den Punkten?
- Wähle $\widehat{\beta}_0, \widehat{\beta}_1$ so, dass Summe der quadrierten Residuen minimal ist:

$$\widehat{\beta}_0, \widehat{\beta}_1 \text{ minimieren } \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

- Lösung mit Analysis:

$$\hat{\beta}_1 = \left(\sum_{i=1}^n (y_i - \bar{y}_n)(x_i - \bar{x}_n) / \sum_{i=1}^n (x_i - \bar{x}_n)^2 \right)$$
$$\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$$

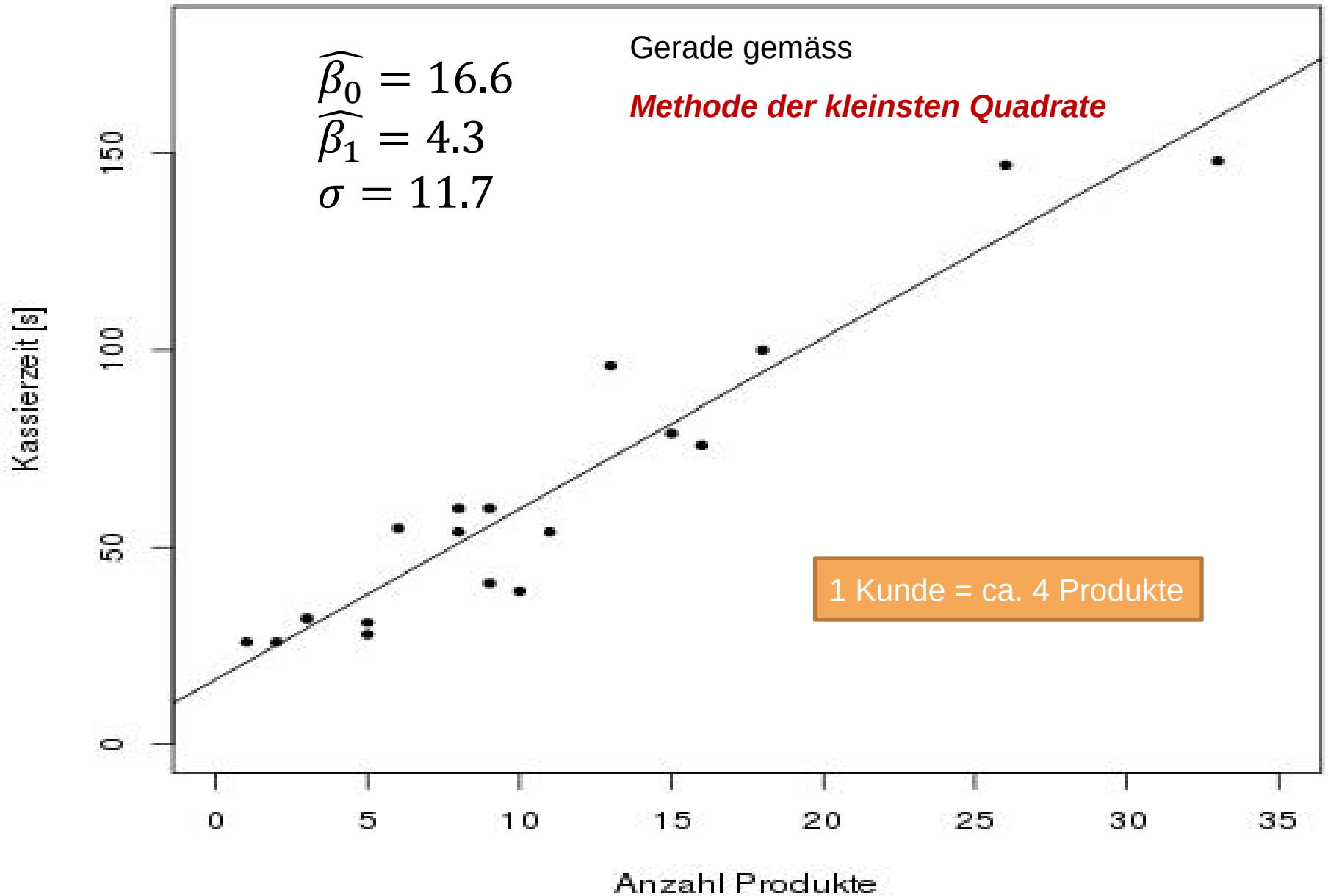
Parameterschätzung: Variante 2

Maximum Likelihood Methode (ML)

- $Y_i \sim N(\mu(x_i), \sigma^2)$ i. i. d.
- Likelihood:
$$L(\beta_0, \beta_1) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{(y_i - \mu(x_i))^2}{\sigma^2}\right)\right) =$$
$$= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^2}\right)\right)$$
- Log-Likelihood:
$$l(\beta_0, \beta_1) = \log(L(\beta_0, \beta_1)) =$$
$$= -n * \log(\sigma\sqrt{2\pi}) - \frac{1}{2} \frac{\left(\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right)}{\sigma^2}$$
- Log-Likelihood ist maximal, wenn $\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$ minimal ist.
- Daher:
Methode ML = Methode LS

Methode der kleinsten Quadrate
und
Maximum Likelihood Methode
sind äquivalent !

Streudiagramm



Welche Schlange?

Kasse 1

3

$$3 + 4 = 7$$

2

$$2 + 4 = 6$$

3

$$3 + 4 = 7$$

20

Kasse 2

19

$$19 + 4 = 23$$

23

Aerobe Leistungsfähigkeit

VO2max: Menge Sauerstoff, die der Körper pro kg maximal pro Minute verwerten kann

- Teuer, aufwändig
- Nicht für breite Masse geeignet



Ersatz: Cooper & Shuttle

- 12-Minuten Test nach Cooper (1968)
- 20m-Shuttle-Test nach Leger (1983)

Eur J Appl Physiol (1982) 49: 1–12

European Journal of
**Applied
Physiology**
and Occupational Physiology
© Springer-Verlag 1982

A Maximal Multistage 20-m Shuttle Run Test to Predict $\dot{V}O_2$ max*

Luc A. Léger¹ and J. Lambert²

¹Département d'éducation physique, Université de Montréal,
CEPSUM, C.P. 6128, Succ. "A", Montréal (Québec), Canada, H3C 3J7

²Département de Médecine sociale et préventive, Université de Montréal, Canada

Ersatz: Cooper & Shuttle

- 12-Minuten Test nach Cooper (1968)
- 20m-Shuttle-Test nach Leger (1983)
- Kann Shuttle-Test den VO_2max -Wert vorhersagen?
- Falls ja: Einfache Testmöglichkeit für breite Bevölkerung

Leger et. al., 1983:

91 Personen

Shuttle test & VO2max

