Ziel • Lin. Regression / Achsenskalierg. kombinieren
(reale Daten : Biophysik ; Ingenieurwissenschaften)
• Wann ist die Regr.gerade eine "gute" Näherung ?

[noch §5 lineare Regression]

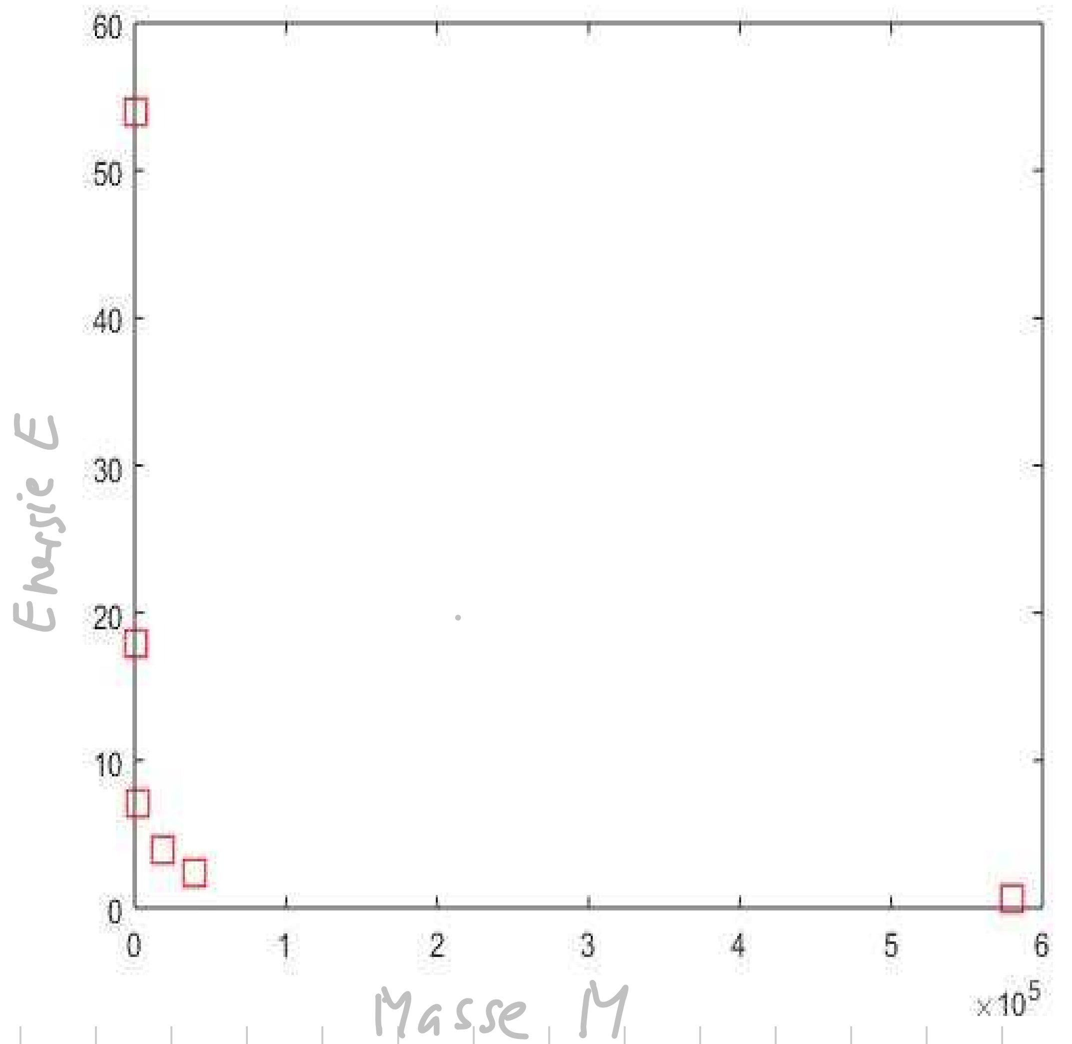
5.2. <u>Regressionskurve</u>

<u>Bsp</u>. Wieviel Energie brauchen Tiere beim Massentransport ? [1g Masse über 1 km transportieren]

	Maus	Ratte	Kl. Hund	gr. Hund	Schaf	Pferd
$M\,[g]$	21	384	2600	18 000	39 000	580 000
$E\,\left[\frac{J}{g\cdot km}\right]$	54	18	7.1	3.9	2.4	0.63

Quelle: Batschelet, Einf. in die Mathe f. Biologen, 1980

MATLAB plot der Daten (siehe VL)



Ziel: Funktion $f(M)$ sodass $E \approx f(M)$

Methode: Daten → Adimensionalisierte Daten → Reg.gerade → Rücktransformation

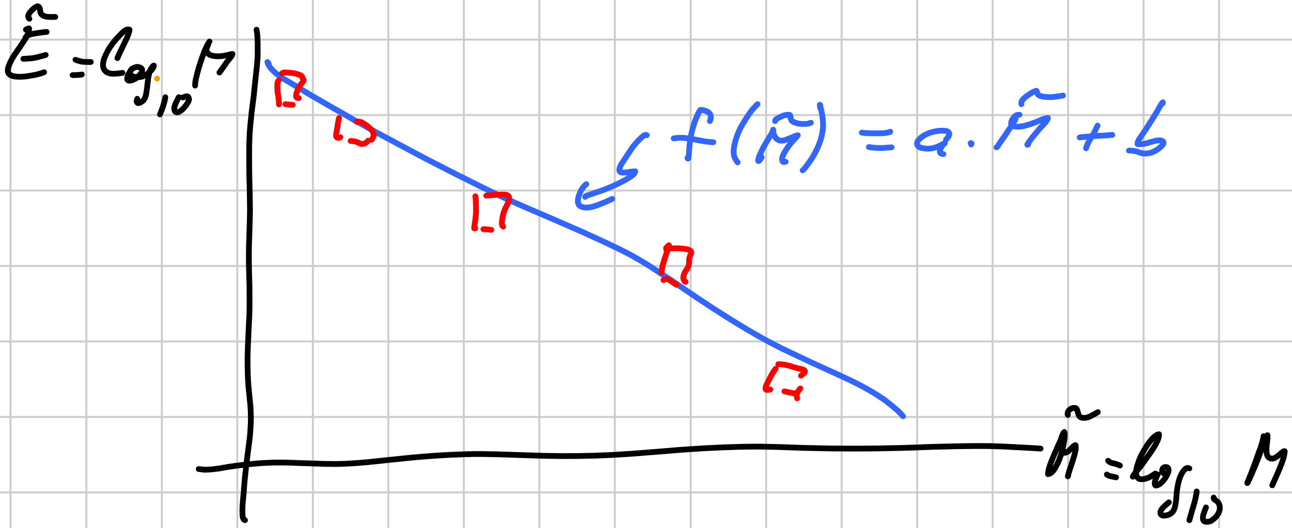
Adimensionalisierte Daten

	Maus	...
$\log_{10} M = \tilde{M}$	1.3222	...
$\log_{10} E = \tilde{E}$	1.7324	...

x, y aus §5.1

Regressionsgerade $y = a \cdot x + b$, Formeln für a & b aus VL 4

d.h. $\tilde{E} = a \cdot \tilde{M} + b$

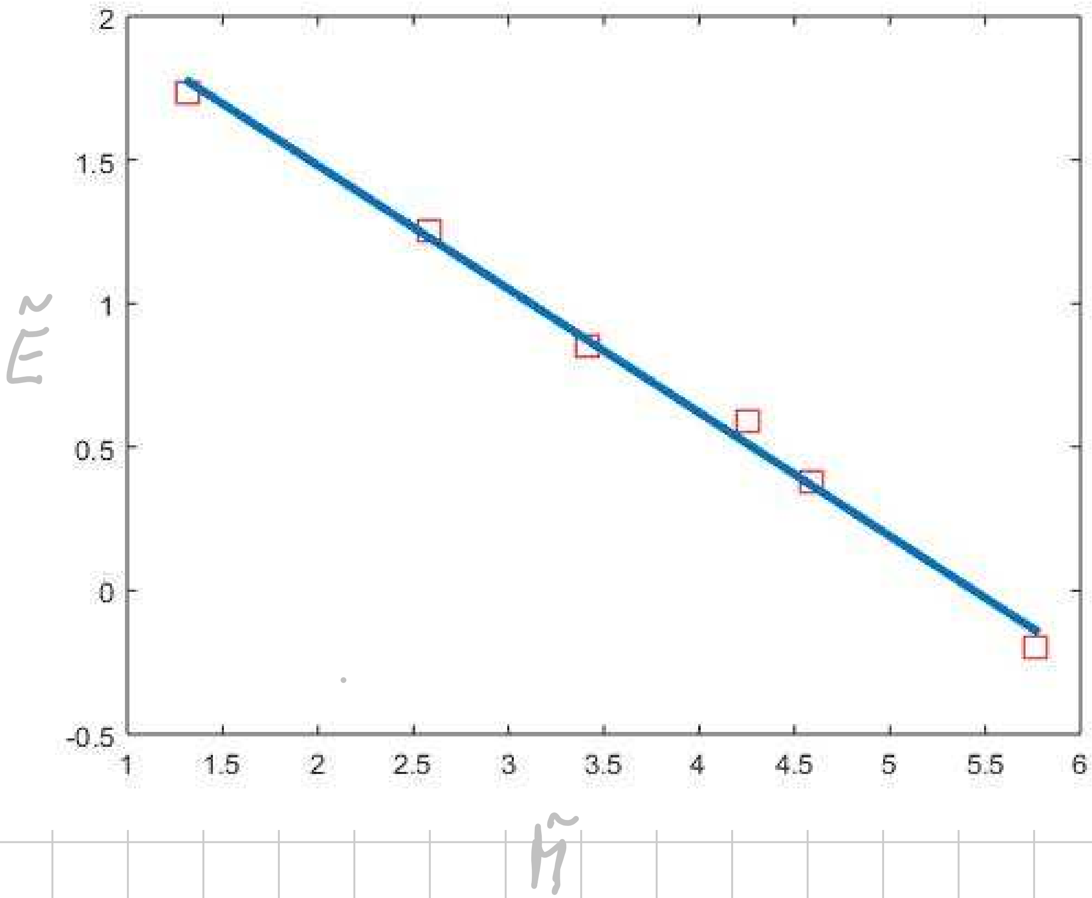


via MATLAB (siehe VL)

$a = -0.4302$

$b = 2.3406$

MATLAB Plot der logarithmierten Daten u. ihrer Regressions=
gerade (siehe VL)



Regressionskurve (= theoret. Modell für Daten)

$$\tilde{E}_{theo} = a \cdot \tilde{M} + b$$

$$\tilde{E}_{theo} = \log_{10} E_{theo}, \qquad \tilde{M} = \log_{10} M \qquad \leftarrow \text{wg. Log-Log-Plot}$$

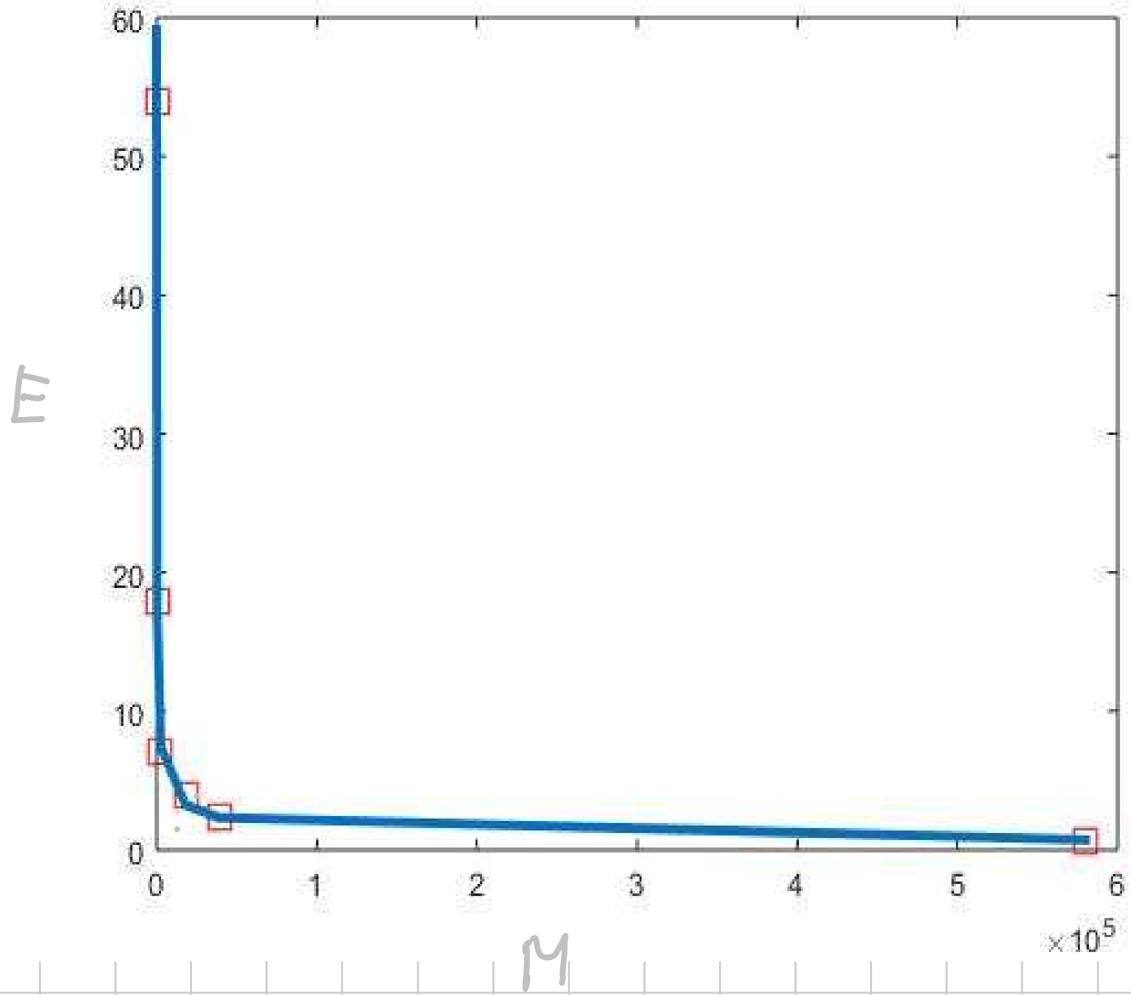
10 hoch beide Seiten $\Rightarrow$

$$
\begin{aligned}
E_{theo} &= 10^{\,a \cdot \log_{10} M + b} \\
&= 10^{\log_{10} M \cdot a} \cdot 10^{b} \\
&\overset{\text{Pot.ges.}}{=} \left(10^{\log_{10} M}\right)^{a} \cdot 10^{b} \\
&\overset{\text{Pot.ges.}}{=} M^{a} \cdot 10^{b} \\
&= A \cdot M^{\alpha} \quad \text{(Potenzfunktion)}
\end{aligned}
$$

mit $A = 10^{b}$, $\alpha = a$

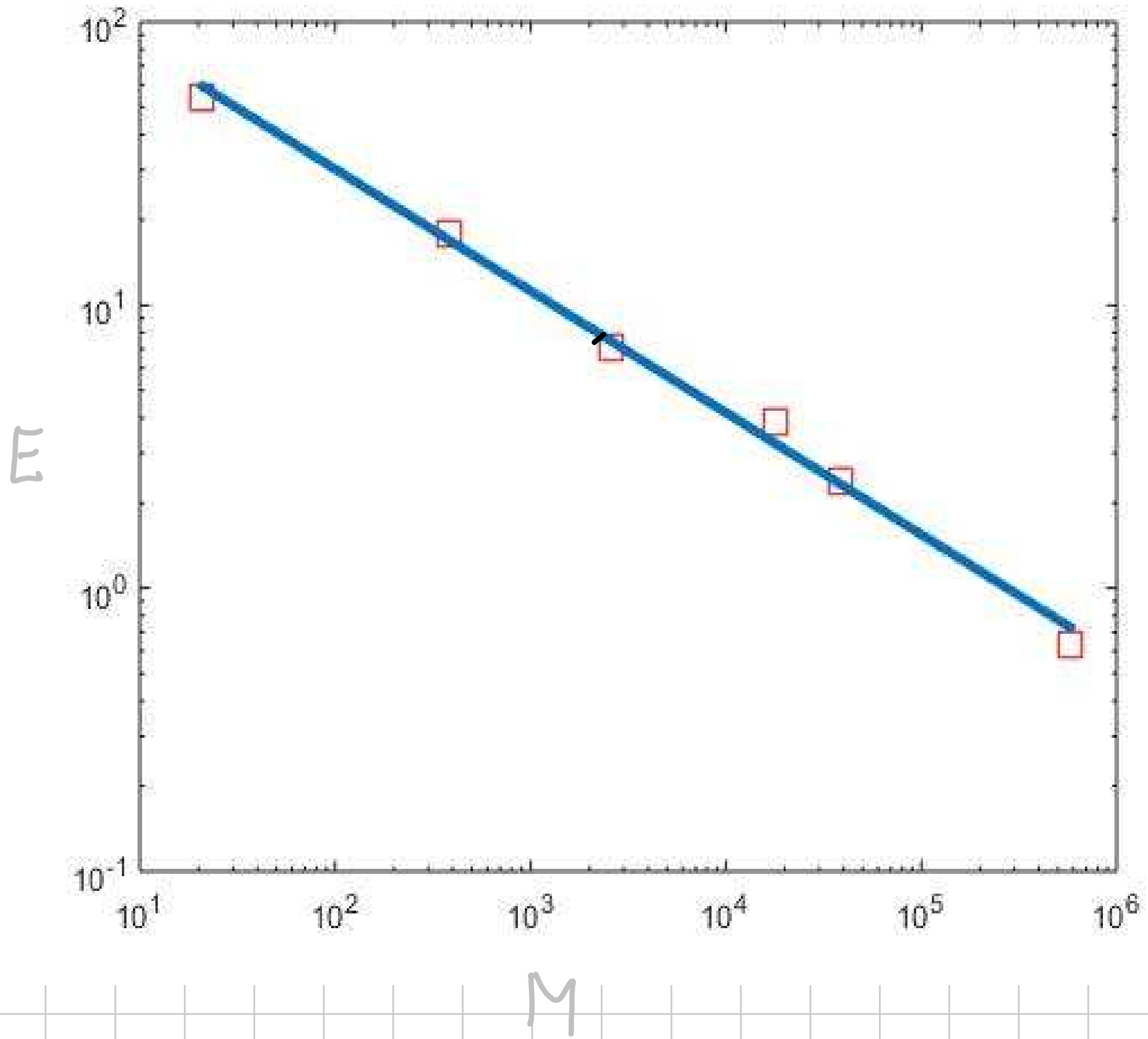
(Exponent = Steigung der Reg.gerade der logarithmierten Daten) hier: $E_{theo} = 219{,}1 \cdot M^{-0.4302}$ Regressionskurve

MATLAB Plot (siehe VL)



Daten und theoretische Kurve stimmen sehr gut überein.

MATLAB log-log-Plot von Daten u. Regressionskurve
(sieht aus wie Plot der logarithmierten Daten, aber Achsenlabels = tatsächliche Werte)



Dieser Plot ist aussagekräftiger als der vorherige, da man auch die kleinen Abweichungen erkennt (z.B. Pferd besonders effizient)

A former student of
a MINT subject
who produced the
most famous Log-Plot of all time ...

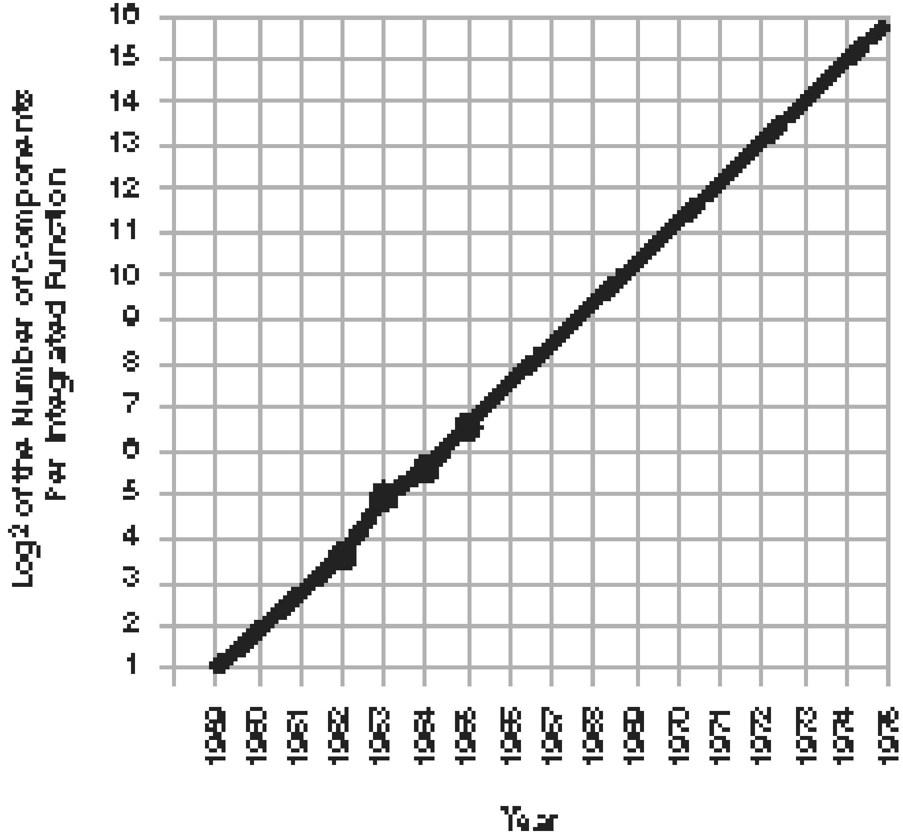
Electronics, Volume 38, Number 8, April 19, 1965

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

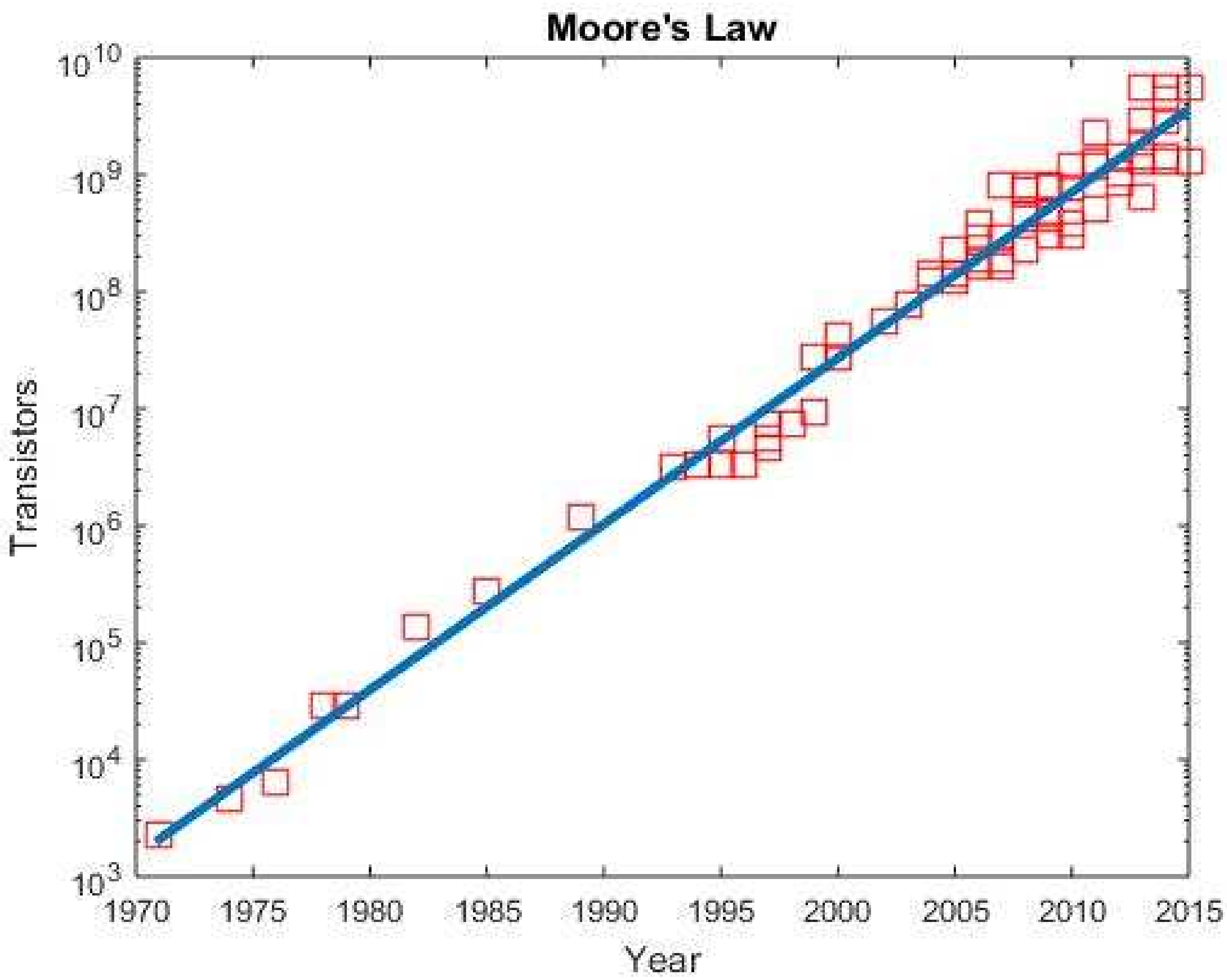
Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.


Log2 of the Number of Components
Per Integrated Function
Year

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wristwatch needs only a display to be feasible today.

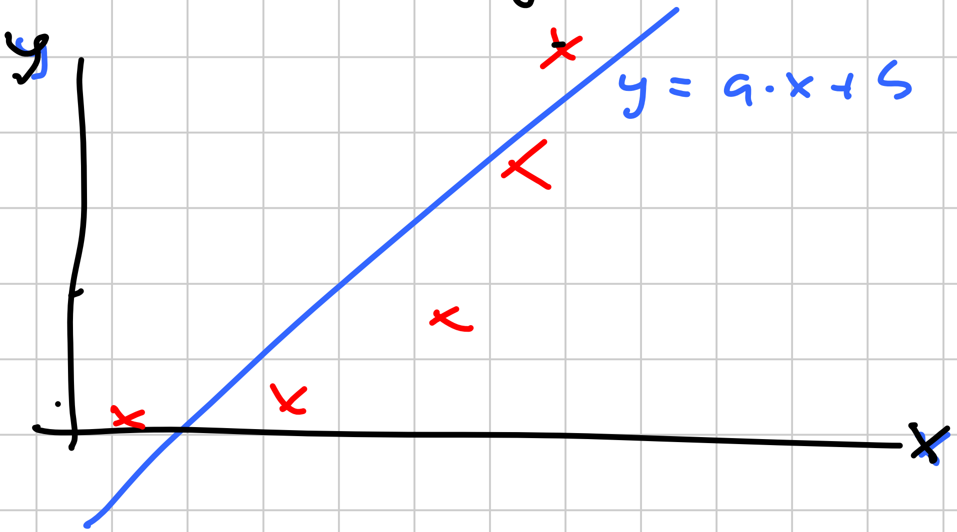
But the biggest potential lies in the production of large systems. In telephone communications, integrated circuits in digital filters will separate channels on multiplex equipment. Integrated circuits will also switch telephone circuits and perform data processing.

Datenanalyse mittels MATLAB (107 Daten=
punkte, logarithm. Darst. mit skalierter "y"-Achse, Bestim=
mung der Regressionsgeraden der skalierten Daten, siehe VL)



5.3 Relativer Standardfehler

Stärke u. gleichzeitig Schwäche der lin. Reg.:
Regr.gerade existiert für beliebige Daten, auch wenn sie nicht näherungsweise auf einer Geraden liegen.



Ziel: Fehlermaß; Toleranzgrenze

Hilfsmittel: Standardabweichung

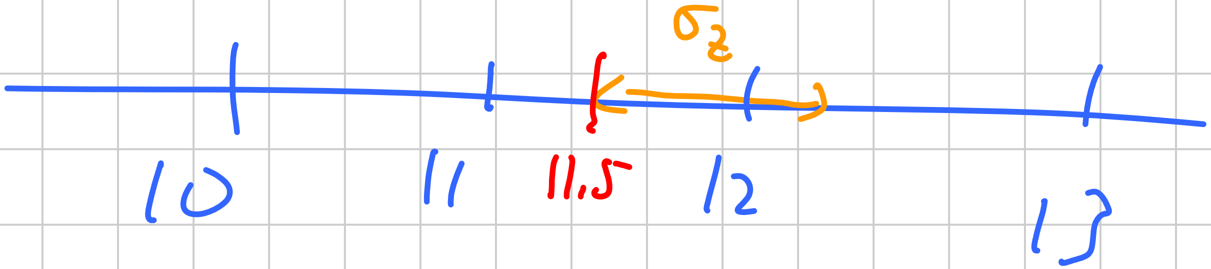
$z_1, \dots, z_n$ beliebige Zahlen

$\bar{z} = \frac{1}{n}(z_1 + \dots + z_n)$ Mittelwert

$$\sigma_z = \sqrt{\frac{1}{n}\left[(z_1-\bar{z})^2 + \dots + (z_n-\bar{z})^2\right]} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})^2}$$

Standardabweichung

Bsp $z_1, z_2, z_3, z_4 = 10, 11, 12, 13$



$$\sigma_z = \sqrt{\frac{1{,}5^2 + (0{,}5)^2 + (0{,}5)^2 + (1{,}5)^2}{4}} = \sqrt{\frac{5}{4}} \approx 1{,}1$$

- Maß f. Streuung der Daten
- genau dann 0 wenn alle Daten gleich
- Liegt zwischen kleinster u. größter Abweichung vom Mittelwert

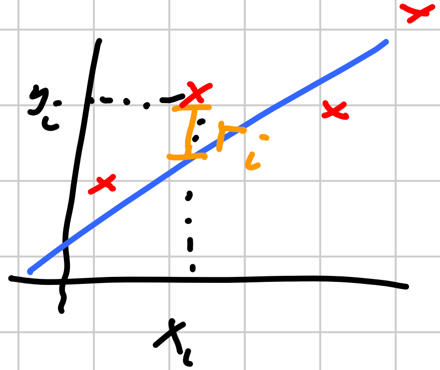
Seien

x	x_1	$\dots$	x_n
y	y_1	$\dots$	y_n

gegebene Daten,

$f(x) = ax + b$ die Regressionsgerade,

$r_i := y_i - (ax_i + b)$ Fehler im i-ten Datenpunkt (siehe VL 4)



Relativer Standardfehler

$$f = \frac{\sigma_r}{\sigma_y} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r})^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sqrt{\sum_{i=1}^{n} r_i^2}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

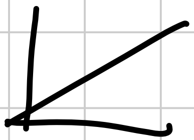
$\frac{1}{n}$ kürzen, $\bar{r}$ MW der $r_i = 0$ (Blatt 4)

Eigenschaften

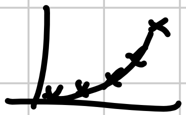
- Null genau dann wenn alle Punkte auf Geraden
- klein wenn Fehler klein im Vergl. zur Streuung der y_i
- $0 \le f \le 1$
- Unabh. v. Wahl d. Einheiten $x_i \rightsquigarrow c_1 \cdot x_i$
  $y_i \rightsquigarrow c_2 \cdot y_i$

<u>Toleranzgrenze</u>: $x_1 \dots x_n = 1 \dots n$

$y_2 \dots y_n = y(1) \dots y(n)$



$y(x) = x$ $\qquad$ $f = 0$



$y(x) = x^2$ $\qquad$ $f \approx 0.25$ wenn n groß

x^4 $\qquad$ 0.5 —"—

2^x $\qquad$ 1 —"—

Toleranzgrenze/Benchmark: $\boxed{f < 0.25}$

(sonst quadratische nicht von linearen Fkt'en unterscheidbar)

Bsp aus 5.2: $f = 0.076$ ✓