

*einbauen & alles  
echt sein*

# Stochastik für die Naturwissenschaften

Dr. C.J. Luchsinger

## 10. ANOVA/REGRESSION

### 10.1 Einfache Varianzanalyse (ANOVA (Analysis of Variance))

Einordnung der ANOVA mit Hilfe einer Kaskade von Tests, bei der man immer mehr Stichproben nimmt: 1-Stichproben-T-Test > 2-Stichproben-T-Test > ANOVA als Verallgemeinerung, sobald  $k > 2$  - wenn  $k = 2$  ist ANOVA ein quadriertes 2-Stichproben-T-Test.

Der einfachste Fall ist die einfache Varianzanalyse oder Einweg-Varianzanalyse (Single-Factor Experiment), komplizierter Zweifweg-Varianzanalysen und mehr.

Wir entwickeln die Theorie zur ANOVA anhand eines Beispiels: In einem Agro-Konzern wird das Wachstum von Pflanzen unter vier klar kontrollierten Bedingungen untersucht (GG: Mineralien von 4 Lagerstätten). Wir nennen diese vier verschiedenen Bedingungen I-IV. Diese können zum Beispiel unterschiedliche Arten von Düngemittel sein. Wir erhalten dann zum Beispiel folgende Tabelle des Wachstums der Pflanzen über einen gegebenen Zeitraum und fragen uns, ob die Mittelwerte gleich sind ( $\mathcal{H}_0$ ) oder nicht ( $\mathcal{H}_1$ ): **KLICKER**

I	II	III	IV
33.3	35.5	29.6	38.5
47.8	35.4	33.4	42.4
44.4	47.6	32.8	45.5
42.9	38.8	38.8	38.9
40.9		42.8	38.9
35.5			44.5

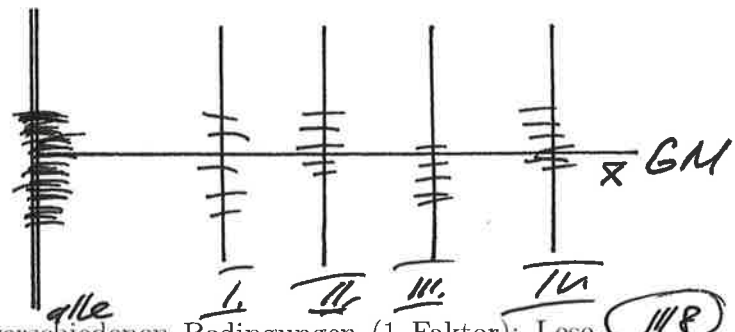


Tabelle: Wachstum von Pflanzen unter 4 verschiedenen Bedingungen (1 Faktor): Lese-  
Beispiel: die vierte Pflanze unter Bedingungen I wuchs im beobachteten Zeitraum 42.9 cm hoch.

Wir werden jetzt ein mathematisches Modell mit  $k$  Gruppen aufstellen (oben  $k = 4$ ):

*Gruppenmittelwert Grp  $j$  (unbekannt!)*

$$Y_{ij} = \mu_j + \epsilon_{ij} \quad (1 \leq j \leq k; 1 \leq i \leq n_j). \quad (10.1)$$

*Element  $i$  in Gruppe  $j$*

*Fehlerterm, Störterm, Error*

Dabei bezeichne  $n_j$  die Anzahl Messungen in Gruppe  $j$  (oben (6,4,5,6)). Die totale Anzahl Beobachtungen ist

$$n = \sum_{j=1}^k n_j.$$

*$Y_{41}(\omega) = y_{41} = 42.9$*

Die Zahl 42.9 in der Beschreibung ist dann die Realisation von  $Y_{41}$ . Wir fassen die Messgrößen also als Realisationen von Zufallsgrößen auf. Die  $\epsilon_{ij}$  sind iid  $\mathcal{N}(0, \sigma^2)$ -verteilt.

**Achtung: Gleiche Varianz in allen Gruppen!** Damit gilt

*CLT, Kapitel 7*



$$Y_{ij} \sim \mathcal{N}(\mu_j, \sigma^2)$$

und die  $n_j$  Realisationen in Gruppe  $j$  sind unabhängig voneinander. Wir interessieren uns jetzt für die Frage, ob die  $k$  Mittelwerte gleich sind oder nicht. Wir wollen in obigem Beispiel ein hohes Wachstum erzielen und den Dünger auswählen, der das höchste Wachstum hervorbringt. Die Nullhypothese lautet

$$\mu_1 = \mu_2 = \dots = \mu_k.$$

*langweilig, symmetrisch, geübt*  
( $H_0$  - Hypothese)

Sicher werden wir aus den Daten die unbekanntten Mittelwerte schätzen. Wir werden also den Mittelwert  $\mu_1$  in Gruppe 1 durch

$$\hat{\mu}_1 := \bar{Y}_{.1} := \frac{\sum_{i=1}^{n_1} Y_{i1}}{n_1}$$

$$\sim \mathcal{N}\left(\mu_1, \frac{\sigma^2}{n}\right)$$

schätzen und alle weiteren nach der allgemeinen Formel

*(Kapitel 5)*

$$\hat{\mu}_j := \bar{Y}_{.j} := \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j}$$

wo  $1 \leq j \leq k$ . Damit haben wir jetzt  $k$  geschätzte Mittelwerte - sie werden mit Wahrscheinlichkeit 1 alle verschieden sein. Sollen wir jetzt einfach den grössten nehmen und die entsprechende Gruppe zum Sieger erklären? Nein! Aus unserer bisherigen Erfahrung in Kapitel 9 wissen wir, dass auch unter der Nullhypothese immer einer am grössten sein

wird *und* dass dieser eine signifikant grösser sein muss, damit wir die Nullhypothese verwerfen. Aber was heisst signifikant? Je nachdem, wie gross  $\sigma^2$  ist, ist auch mit grösseren Abweichungen selbst bei Gültigkeit der Nullhypothese zu rechnen. Zudem kennen wir  $\sigma^2$  gar nicht. Es scheint hoffnungslos zu sein - aber: Wir können ja auch  $\sigma^2$  schätzen. Dies geschieht in 2 Schritten und wird uns (eher unerwartet) durch geschickte Umformung gleich die Lösung des Problems liefern.

1. Unter der Nullhypothese haben wir eine iid-Stichprobe und wir können alle  $n$  Datenpunkte gleichberechtigt zur Berechnung eines Grand Mean  $GM$  einsetzen: Mit  $GM := \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} Y_{ij}}{n}$  können wir in einem

2. Schritt einen Schätzer für die Varianz (genauer das  $(n-1)$ -fache davon) angeben mit:

$$\begin{aligned}
 \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - GM)^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} ((\bar{Y}_{.j} - GM) + (Y_{ij} - \bar{Y}_{.j}))^2 \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 \\
 &\quad + 2 \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)(Y_{ij} - \bar{Y}_{.j}) \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 \\
 &\quad + 2 \sum_{j=1}^k (\bar{Y}_{.j} - GM) \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j}) \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 \\
 &\quad + 2 \sum_{j=1}^k (\bar{Y}_{.j} - GM) * 0 \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2.
 \end{aligned}$$

wie in Kapitel 2:

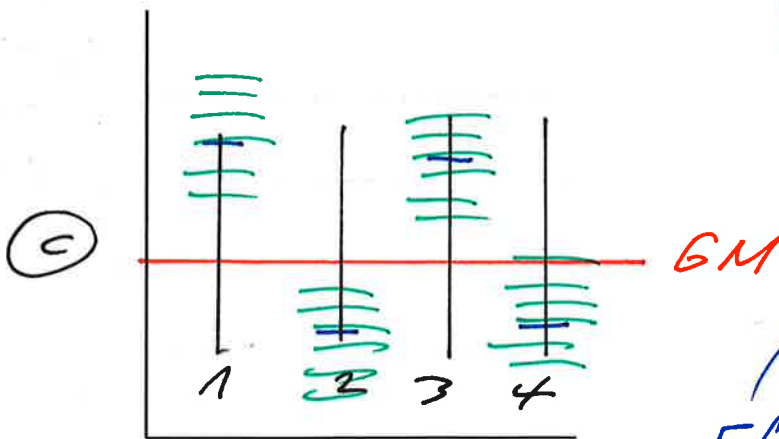
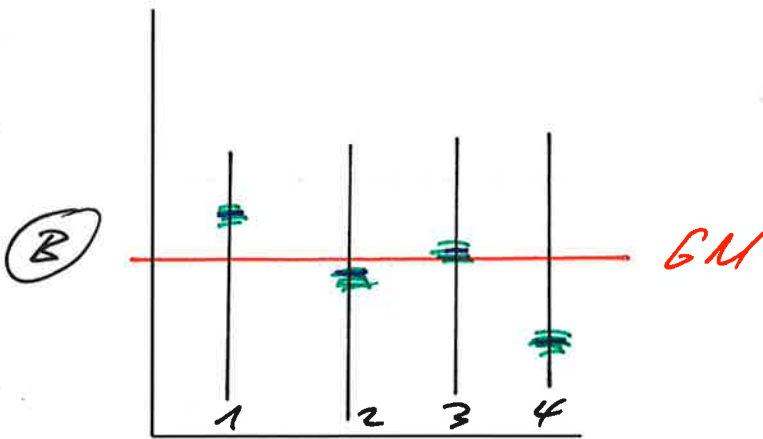
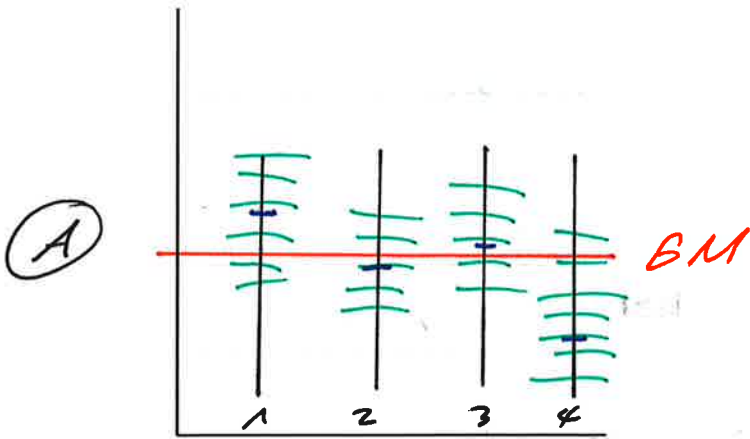
$$\sum_{i=1}^n (x_i - \bar{x})^2$$

(analog)

HA,  
Dozent  
fragen in  
Pawe  
wenn  
unklar

Resultat: Wir haben die Varianz in 2 *hochinformative* Summen aufgespalten! Deshalb ANOVA: Analysis of Variance!

Die drei wichtigsten Skizzen in ganz MAT 183:



◦ GM dreimal gleich

◦ blaue Behandlungsmittelwerte gleich in (A) & (B)

◦  $\sigma^2$  gleich in (A) & (C)

Fundi Anova:

- (A)  $11 = 5 + 6$  gleich
- (B)  $6 = 5 + 1$  gleich
- (C)  $26 = 20 + 6$

Idee: Buch aus 1. & 2. Summanden:

- (A)  $\frac{5}{6} \doteq 1$  : "H<sub>0</sub>"
- (B)  $\frac{5}{1} = 5$
- (C)  $\frac{20}{6} \doteq 3$  } "H<sub>1</sub>"

→ fehlen noch Konstanten

Effekt in's Verhältnis zur Streuung setzen!

wird *und* dass dieser eine signifikant grösser sein muss, damit wir die Nullhypothese verwerfen. Aber was heisst signifikant? Je nachdem, wie gross  $\sigma^2$  ist, ist auch mit grösseren Abweichungen selbst bei Gültigkeit der Nullhypothese zu rechnen. Zudem kennen wir  $\sigma^2$  gar nicht. Es scheint hoffnungslos zu sein - aber: Wir können ja auch  $\sigma^2$  schätzen. Dies geschieht in 2 Schritten und wird uns (eher unerwartet) durch geschickte Umformung gleich die Lösung des Problems liefern.

1. Unter der Nullhypothese haben wir eine iid-Stichprobe und wir können alle  $n$  Datenpunkte gleichberechtigt zur Berechnung eines Grand Mean  $GM$  einsetzen: Mit  $GM := \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} Y_{ij}}{n}$  können wir in einem

*Kapitel 2:  $\sum_{i=1}^n (x_i - \bar{x})^2$  entspricht (unter  $H_0$ )*

2. Schritt einen Schätzer für die Varianz (genauer das  $(n-1)$ -fache davon) angeben mit:

$$\begin{aligned}
 \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - GM)^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} ((\bar{Y}_{.j} - GM) + (Y_{ij} - \bar{Y}_{.j}))^2 \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 \\
 &\quad + 2 \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)(Y_{ij} - \bar{Y}_{.j}) \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 \\
 &\quad + 2 \sum_{j=1}^k (\bar{Y}_{.j} - GM) \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j}) \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 \\
 &\quad + 2 \sum_{j=1}^k (\bar{Y}_{.j} - GM) * 0 \\
 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{.j} - GM)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2.
 \end{aligned}$$

*HA;  
Dozent  
fragen in  
Pause wenn  
Frage*

Resultat: Wir haben die Varianz in 2 hochinformative Summen aufgespalten! Deshalb ANOVA: Analysis of Variance!

Wir haben also zusammengefasst:

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - GM)^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_j - GM)^2 + \sum_{j=1}^k \left( \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 \right). \quad (\text{Fundi ANOVA})$$

⊗  
zwischen
⊕  
innerhalb

*Gruppen- oder Behandlungsmittelwert*

Diese Gleichung wird auch als Fundamentalgleichung der Varianzanalyse bezeichnet (deshalb "Fundi"). Was sagt sie aus?

Die gesamte Summe der quadratischen Abweichungen der einzelnen Beobachtungen vom Grand Mean (linke Seite) lässt sich aufspalten ("+" ) in Summe der quadrierten Abweichungen der Behandlungsmittelwerte vom Grand Mean (zwischen den Behandlungen, engl between treatments; erster Summand) und der Summe der quadrierten Abweichungen der einzelnen Beobachtungen vom jeweiligen Behandlungsmittelwert (innerhalb der Behandlungen, engl within treatments; zweiter Summand).

Zwischenfrage ans Publikum: Angenommen, die Annahme gleicher Mittelwerte ist verletzt. Welcher der beiden Summanden auf der rechten Seite von (Fundi ANOVA) wird tendenziell grösser im Verhältnis zum anderen? Gehen Sie bei solchen Überlegungen in die Extreme! **KLICKER** *1. Summand wird grösser*

Damit bietet sich als Test-Statistik folgende Grösse an:  $(k-1) \& (n-k)$ : *Mathe*

$$V := \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_j - GM)^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 / (n-k)} = \frac{\sum_{j=1}^k n_j (\bar{Y}_j - GM)^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 / (n-k)}$$

Diese Test-Statistik hat unter  $\mathcal{H}_0$  eine  $F_{k-1, n-k}$ -Verteilung. Wir werden den Ablehnungsbereich der Nullhypothese einseitig dort wählen, wo diese Statistik gross ist. Spielt es eine Rolle, ob man zum Beispiel Gewichte in mm,  $\mu\text{m}$ , m, km eingibt? SI-Einheiten? **KLICKER**

*$\omega^2 \neq 1$  : kürzt sich heraus, also egal*

Wenn wir nur  $k = 2$  Gruppen haben, so kommen wir in eine bereits bekannte Situation. In welche und warum? *2-T-Test, quadriert!*

```
> growth<-c(33.3,47.8,44.4,42.9,40.9,...,42.4,45.5,38.9,38.9,44.5)
> dung<-rep(LETTERS[1:4],c(6,4,5,6)) # Versuche n_j, k = 4
> dung<-factor(dung)
> agro<-data.frame(dung,growth)
```

zB so eingeben wie oben, Daten spaltenweise - geht auch anders, eleganter

```
> analyse<-aov(growth ~ dung, agro) aov=Analysis of Variance
> analyse
```

Call:

```
aov(formula = growth~dung, data = agro)
```

Terms:

	dung	Residuals
Sum of Squares	113.7990	408.5105

oben: 1. und 2. Summand; unten:  $k - 1 = 4 - 1 = 3$  und  $n - k = 21 - 4 = 17$

Deg. of Freedom	3	17
-----------------	---	----

Residual standard error: 4.902043 das ist  $\hat{\sigma}$  und nicht Standard Error  $\hat{\sigma}/\sqrt{n}$  (auch in 10.2 ->)

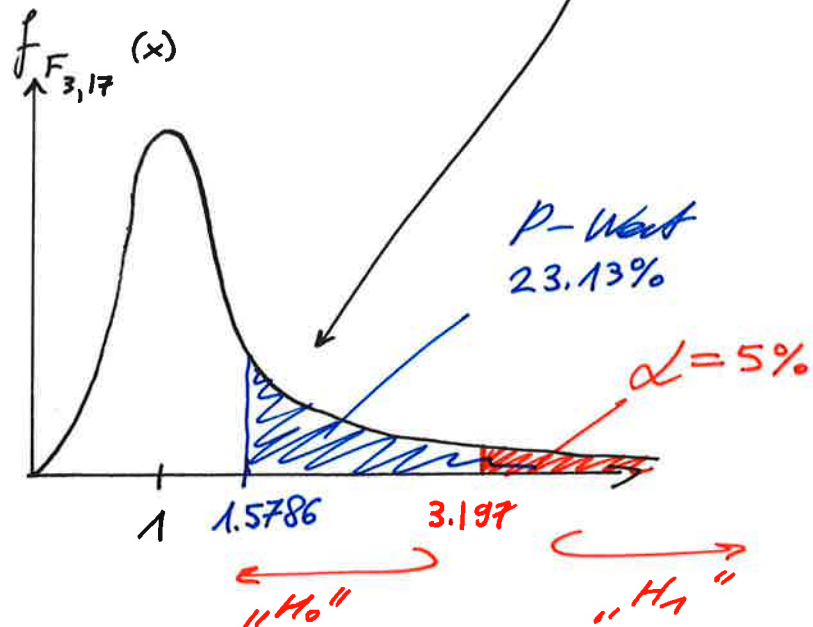
Estimated effects may be unbalanced

```
> summary(analyse) Das nennt man eine VAT: Varianz Analyse Tabelle
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dung	3	113.80	37.93	1.5786	0.2313
Residuals	17	408.51	24.03		

Mean Sq = Sum Sq / Df sind Zähler und Nenner; F value dann der Bruch

```
> summary(agro)
dung    growth
A:6    Min.  :29.60
B:4    1st Qu.:35.50
C:5    Median :38.90
D:6    Mean   :39.44
       3rd Qu.:42.90
       Max.  :47.80
```

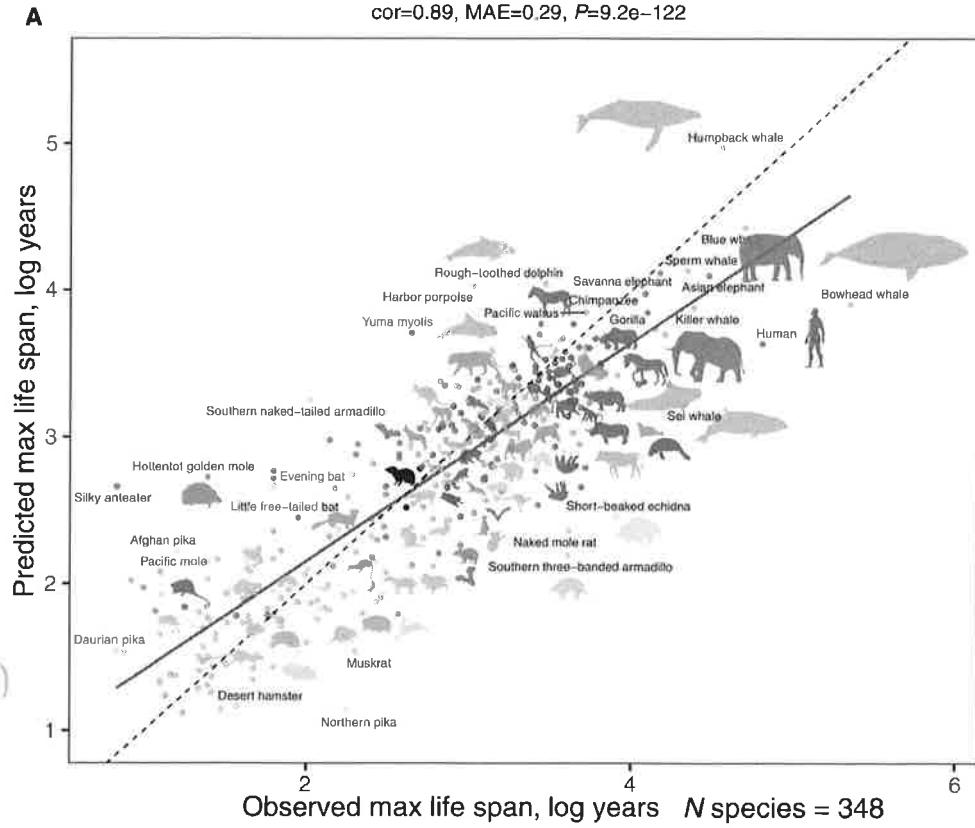


• P-Wert  $> \alpha$  }  $\Rightarrow H_0$  beibehalten  
 (23.13%  $>$  5%)

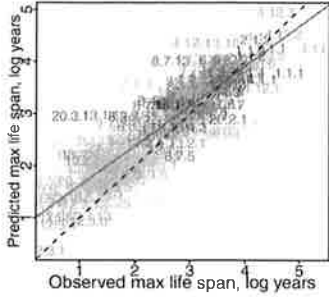
• 1.5786  $<$  3.197  $\rightarrow H_0$  beibehalten

### Epigenetic predictor of mammalian maximum life span

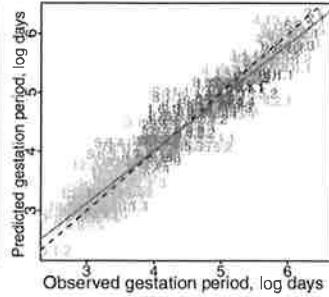
cor=0.89, MAE=0.29, P=9.2e-122



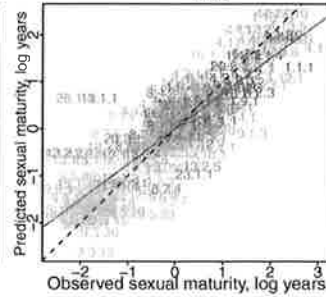
**B**  $R = 0.89$ , MAE = 0.29,  $P = 9.2e-122$



**C**  $R = 0.96$ , MAE = 0.16,  $P = 8.5e-194$



**D**  $R = 0.85$ , MAE = 0.32,  $P = 2e-99$



Taxonomic order					
● (1.) Primates	● (6.) Perissodactyla	● (11.) Diprotodontia	● (16.) Triconodontia	● (21.) Insectivora	● (26.) Paucituberculata
● (2.) Proboscidea	● (7.) Chiroptera	● (12.) Eulipotyphla	● (17.) Scandentia	● (22.) Dermoptera	
● (3.) Artiodactyla	● (8.) Rodentia	● (13.) Afrotheria	● (18.) Dasyuromorphia	● (23.) Microbiotheria	
● (4.) Carnivora	● (9.) Dasyuromorphia	● (14.) Sirenia	● (19.) Hyacoidea	● (24.) Pholidota	
		● (15.) Macroscelidea	● (20.) Pilosa		



*Kolumne: Das kleine Einmaleins*

*Ausgabe 1073 - Februar 2020*

## **Die Berechnung der Langsamkeit**

Wie der Spurt auf das Tram mit Parabeln zusammenhängt.

*von Christoph Luchsinger*

30.01.2020



**W**er morgens auf das wartende Tram spurtet, kann eine beachtliche Geschwindigkeit erreichen. Versucht man das gleiche Tempo bei einem Langstreckenlauf durchzuhalten, stösst man aber schnell an seine Grenzen. Je länger die Strecke ist, desto tiefer ist die Durchschnittsgeschwindigkeit, was unter anderem an der Ermüdung der Muskeln liegt.

Das Phänomen zeigt sich auch im Spitzensport. Der Schweizer Rekord im 100-Meter-Sprint der Männer liegt bei 10,11 Sekunden. Über 200 Meter beträgt die schnellste Zeit 20,04 Sekunden. Für 400 Meter braucht der Rekordhalter schon 44,99 Sekunden und für 800 Meter 102,6 Sekunden. Die Reihe lässt sich erweitern:

Distanz (in Metern)	Rekordzeit (in Sekunden)
1000	135,6
1609 (eine Meile)	230,4
2000	294,5
5000	787,5
10000	1664
20000	3554
25000	4735
30000	5741

&gt;&gt;

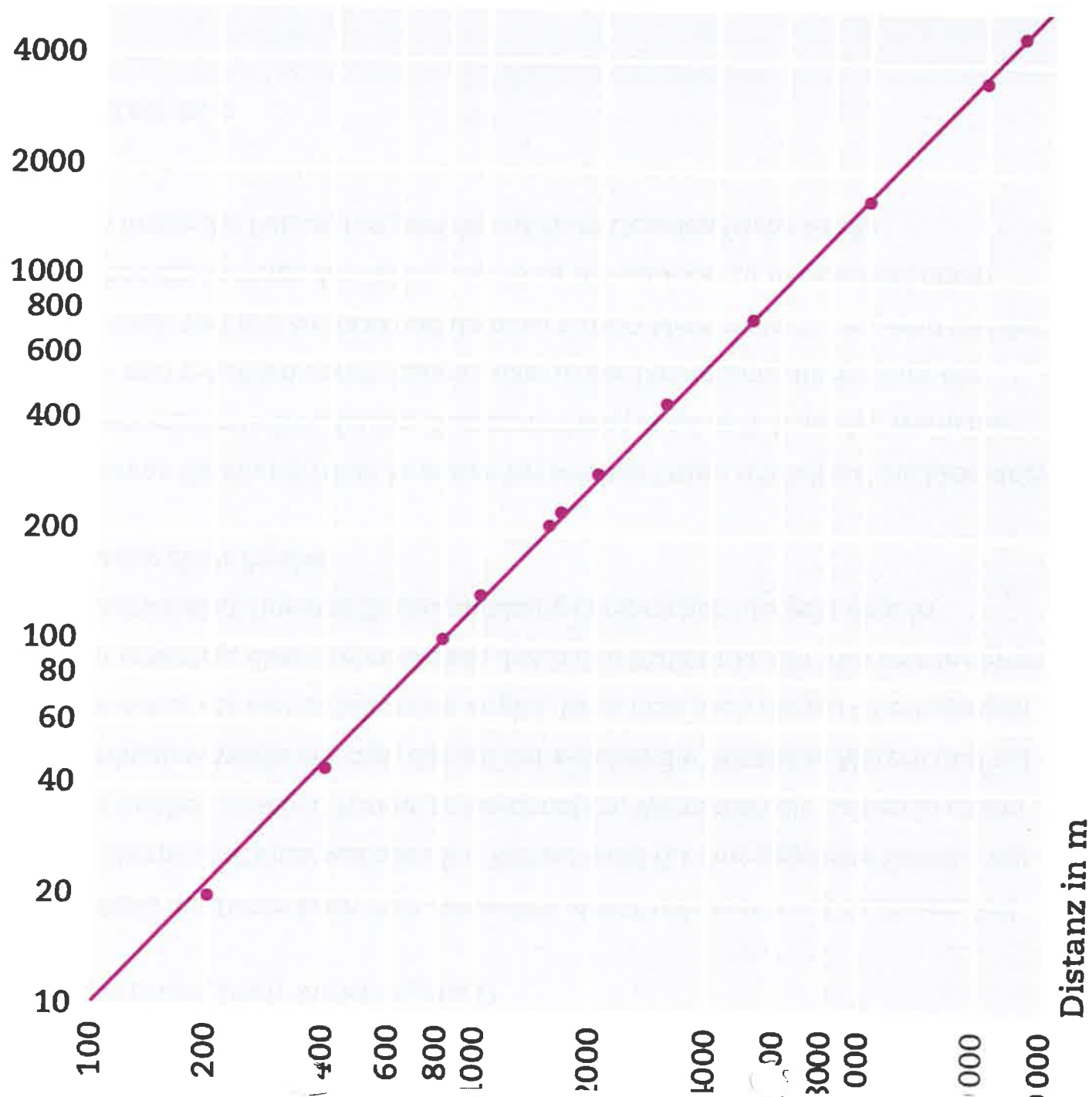
Für 800 Meter braucht man mehr als doppelt so lange wie für 400 Meter, ebenso für 2000 Meter mehr als doppelt so lange wie für 1000 Meter. Eine Ausnahme ist der Sprung von 100 auf 200 Meter, weil die zeitfressende Startsequenz auf

der kurzen Distanz stark ins Gewicht fällt. Der Zusammenhang ist also nicht proportional. Doch wie ist er genau?

Auf Basis der Daten kann man versuchen, den Zusammenhang, in diesem Fall den überproportional wachsenden Zeitaufwand für eine gegebene Strecke, mit einer mathematischen Formel zu beschreiben. Wenn man die Zahlen in einem Koordinatensystem einträgt, das auf der x-Achse die Distanz in Metern und auf der y-Achse die Zeit in Sekunden angibt, kann man nach obigen Überlegungen nicht erwarten, dass es eine Gerade durch den Nullpunkt gibt (das würde einem proportionalen, linearen Zusammenhang entsprechen). Es geht eher in Richtung einer Parabel.

Erfahrene Datenanalytiker kommen bei solchen Daten schnell auf die Idee einer Datentransformation, die den Zusammenhang anschaulich macht: Wenn man die x- und y-Achsen derart staucht, dass in der Darstellung die Sprünge der Distanzen von 100 zu 1000 und dann zu 10 000 Metern gleich lang sind (bei der y-Achse die Sprünge der Zeiten von 10 zu 100 zu 1000 zu 10 000 Sekunden), dann liegen die Punkte fast perfekt auf einer Geraden (siehe Grafik).







Wenn nun in *dieser sogenannt doppelt-logarithmischen* Darstellung eine Gerade resultiert, dann – so lässt sich mathematisch nachweisen – muss der ursprüngliche Zusammenhang ein Potenzzusammenhang sein (also etwas in der Art  $y = a \cdot x^b$ ). In der Tat gelangt man mit obigen Zahlen approximativ zu folgendem Ergebnis (probieren Sie es selber aus): Wenn  $x$  die Strecke und  $y$  die Zeit ist, dann gilt:  $y = 0,059 \cdot x^{1,12}$

Der Exponent ist 1,12; wäre er 1, so hätten wir im ursprünglichen Koordinatensystem eine Gerade, bei 2 wäre es eine Parabel. 1,12 beziffert den Leistungsabfall.



Eine kleine Warnung: Für Distanzen von 100 bis 30 000 Meter erklärt diese Formel die aktuellen Daten gut. Extrapolationen auf höhere Werte sind – nicht nur hier – mit grösster Vorsicht zu geniessen.

### Christoph Luchsinger

ist Mathematikdozent an der Universität Zürich, Gründer der Stellenbörsen

[www.math-jobs.com](http://www.math-jobs.com) und [www.acad.jobs](http://www.acad.jobs) und Redaktor beim Think Tank

[www.schatten-kabinett.ch](http://www.schatten-kabinett.ch). In seiner Kolumne kommt er alltäglichen mathematischen Geheimnissen auf die Spur.

Alles von Christoph Luchsinger lesen



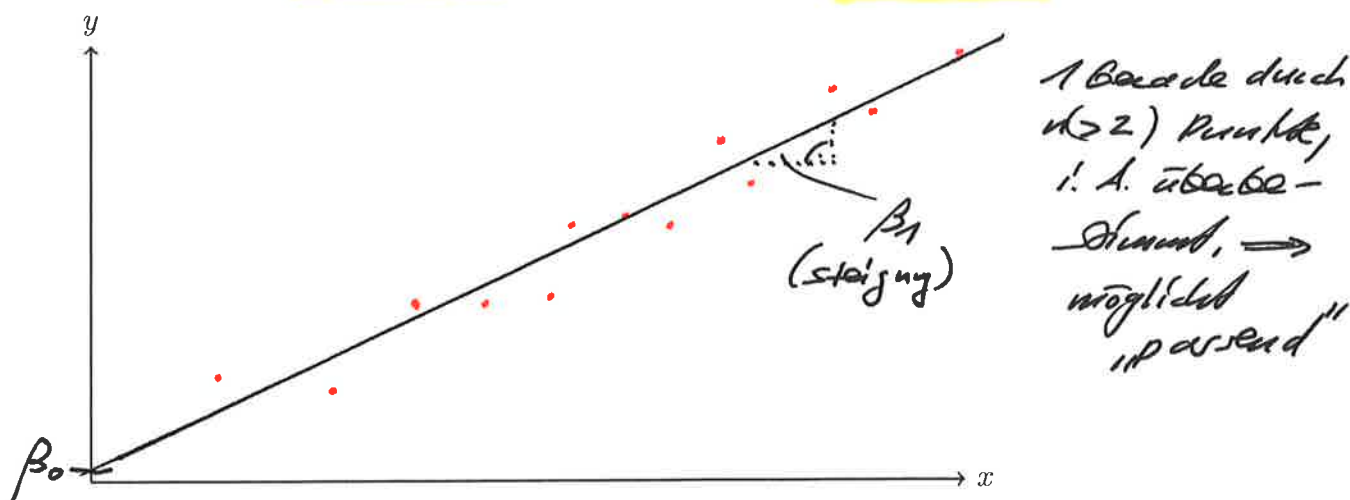
## 10.2 Regression [wieder einsteigen möglich - geht auch ohne ANOVA]

### 10.2.1 Einfache Regression (simple regression)

Die einfache Regression ist das einfachste nichttriviale Beispiel eines komplexen Modells, in dem Kapitel 8 und 9 vorkommen. Schätzen und Testen sind Hauptthema für alle komplexen Modelle.

In diesem Kapitel werden wir eine etablierte Methode kennenlernen, mit der wir eine (lineare - also "auf einer Geraden") Beziehung zwischen 2 Variablen untersuchen können (z.B. Körpergröße von Mutter ( $x$ ) und Tochter ( $y$ )). Im (theoretischen) Modell

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \leftarrow \text{Störterm} \quad ; \text{Gerade: } y = ax + b \quad (10.2)$$



werden wir die  $x$ -Variable als fest betrachten und versuchen, die  $Y$ -Variable möglichst genau durch  $x$  vorherzusagen oder durch  $x$  zu erklären (stören wird uns dabei der Störterm  $\epsilon$ ). Es wird dann mit Daten  $(x_i, y_i)$ ,  $1 \leq i \leq n$ , darum gehen,  $\beta_0$  (engl. intercept) und  $\beta_1$  (engl. slope) zu schätzen (und zu testen, ob nicht z.B.  $\beta_1 = 0$  gilt.)

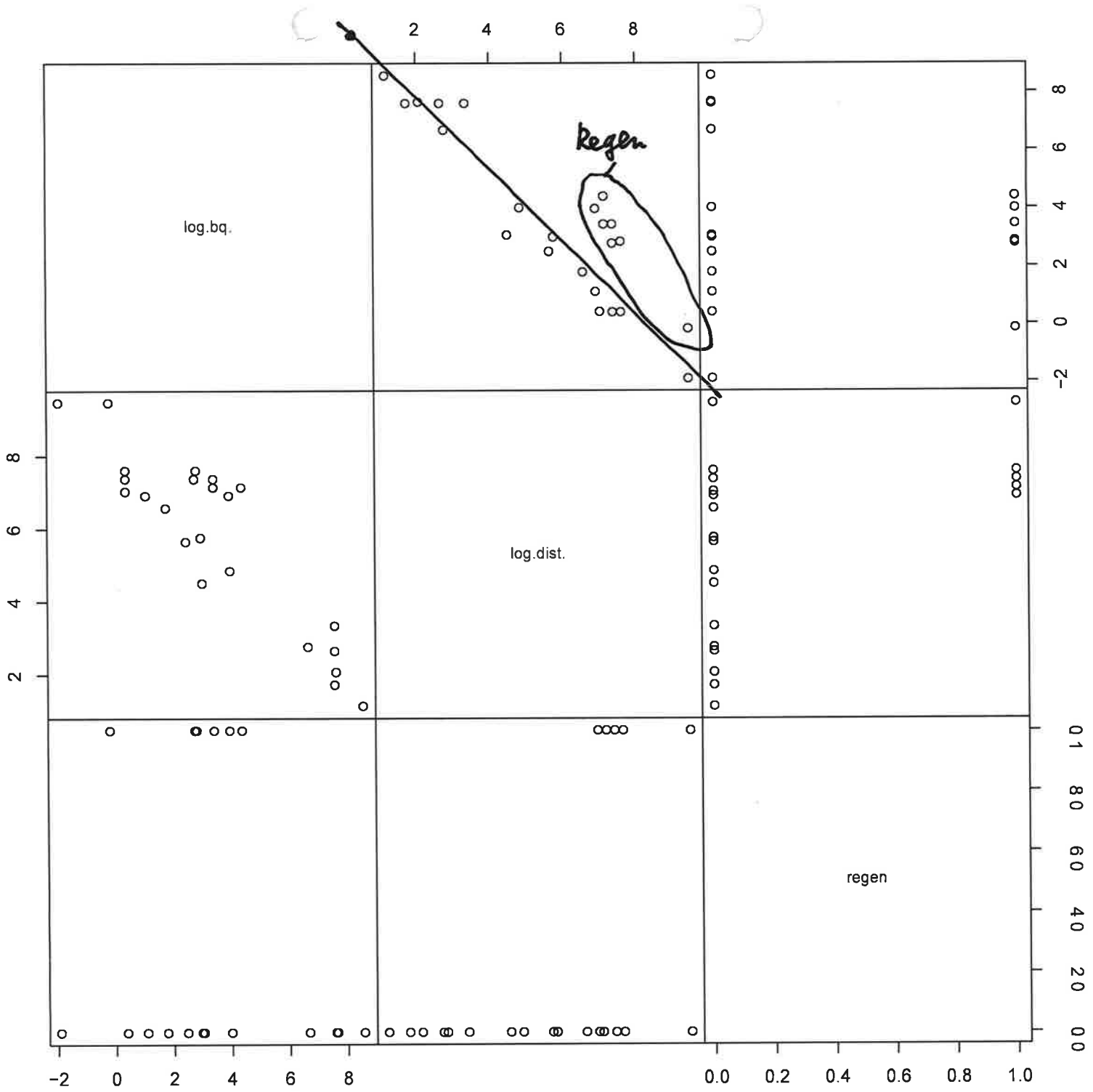
2 gute Beispiele: <https://schweizermonat.ch/die-berechnung-der-langsamkeit/> , <https://www.science.org/doi/10.1126/sciadv.adm7273>

Gründe für Regressionsanalyse sind jeweils:

\*  $Y$  vorhersagen (Interpolation (interessierendes  $x \in [x_{(1)}, x_{(n)}]$ ) und Extrapolation (interessierendes  $x \notin [x_{(1)}, x_{(n)}]$ , heikel))

\* Zusammenhang erklären anhand bisheriger Daten





### 10.2.1.1 Motivierendes Beispiel - Gefahren

Wir schauen den Tschornobyl-Datensatz (1986) mitsamt den dazugehörigen Plots an:

ort	regen	dist	kBq/m <sup>2</sup>
Prypjat	0	3.16	5300.00
Tschystohaliwka	0	5.62	2000.00
Leliw	0	7.94	2100.00
Tschornobyl	0	14.13	2000.00
Rudky	0	15.85	800.00
Aravichy	0	28.18	2000.00
Kijiv	0	89.13	21.00
Tschornykiw	0	125.89	55.00
Tscherkassy	0	281.84	12.00
Minsk	0	316.23	20.00
Donezk	0	707.95	6.00
Wien	0	1000.00	3.00
Oesterreich	1	1000.00	53.00
Stockholm	0	1122.02	1.50
Gaevle	1	1258.93	31.00
SuedBayern	1	1258.93	81.00
Konstanz	1	1584.89	31.00
Irland	1	1584.89	16.00
Stuttgart	0	1590.54	1.50
Chilton	0	1995.26	1.50
Schottland	1	1995.26	17.00
Japan	0	12589.25	0.15
Japan2	1	12589.25	0.85

*Glück, Wolke ging nach Norden*

\* **Rohdatenplot** Distanz versus Radioaktivität > nicht informativ

\* Häufig muss man in der Statistik die Daten vorgängig transformieren (zum Beispiel Kehrwert nehmen, Logarithmieren, etc; mehr dazu im **Stahel**). Welche Datentransformation könnte Abhilfe verschaffen? Wie findet man hier diese? Entweder hat man praktische Erfahrung aus der Statistik oder theoretische Kenntnisse aus dem Fachgebiet (hier Meteorologie, Ausbreitung von Gasen), oder wir machen folgende allgemeine Überlegung: global gesehen haben wir hier eine Punktquelle, deren Wirkung sich (anfänglich) radial (kugelförmig) ausbreitet. Wir erinnern an die beiden Transformationen, welche wir in MAT 182 kennengelernt haben:



Wir wenden hier also den Logarithmus an, sowohl für die  $x$ -, wie auch für die  $y$ -Achse. Schauen wir **Pairs-Plot** (R: pairs) mit den transformierten Daten an. Was wir mit einem Pairs-Plot und diversen Transformationen der Daten machen gehört zur **exploratorischen Statistik** (methodisch sehr frei). Danach macht man mit den vermuteten Zusammenhängen **konfirmatorische Statistik** (methodisch eher eng, es muss intersubjektiv Bestand haben).

Jetzt versuchen wir die Radioaktivität mit der Distanz allein zu erklären (oder vorherzusagen). Wir wollen uns vorerst auf die Schätzungen der Parameter in  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  konzentrieren. Mit einem noch zu besprechenden Verfahren (**Kleinste Quadrate** - siehe 10.2.1.3), geben wir in R ein (Befehl ist R-spezifisch)

NurDistanz <- lm(log(bq) ~ log(dist))

*linear model*

*log(bq) durch log(dist) erklären*

Dabei seien die Daten bereits als "dist", "regen" und "bq" in R eingelesen worden. "lm" steht für **Linear Model**, log(bq) nennt man die **"Response Variable"** (abhängige Variable) und log(dist) eine erklärende Variable oder einen **"Predictor"**. Wenn wir also Daten haben und in Modell (10.2) die Parameter schätzen, erhalten wir folgende Regressionsgerade:

$$\log(bq) = 9.803 - 1.091 \cdot \log(dist)$$

$$y = \beta_0 + \beta_1 \cdot x$$

Wegen des negativen Koeffizienten ( $-1.091$ ) haben wir also einen negativen (linearen) Zusammenhang: je grösser die Distanz zum Unglücksreaktor, desto kleiner die Radioaktivität. Das ist aber alles in logarithmischer Skala geschehen; wir kehren eine Seite zurück zur Besprechung der Transformation.

Wenn wir jetzt noch versuchen, die Radioaktivität mit dem Regen allein zu erklären (oder vorherzusagen), gibt es eine Überraschung; wir erhalten hier nämlich erstmals:

$$\log(bq) = 3.7784 - 0.8247 \cdot \text{regen}$$

Entgegen unseren Erwartungen haben wir auch hier einen negativen (linearen) Zusammenhang. Wieso ist das hier so (beachten Sie auch die spätere Relativierung dieses Resultats)?

- In unseren Daten nur dort Regen wo es weit weg!*
- Beobachtungsergebnis, kein Laborexperiment mit gutem Design!*

Auf dem **Computer-Ausdruck hinten** finden Sie viele Zahlen, welche wir jetzt zum ersten Mal anschauen.

*+ Korrelation hinten schauen! F5 26*

10.2.1.2 Welche Modellannahmen werden wir machen?

Wie überall in der Statistik werden wir ein intensives Wechselspiel zwischen Daten  $(x_i, y_i)$  und (vermuteten) Zufallsgrößen haben. Wenn wir uns auf der Ebene der Zufallsgrößen befinden, machen wir folgende Modellannahmen ( $1 \leq i \leq n$ ):

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (10.2)$$

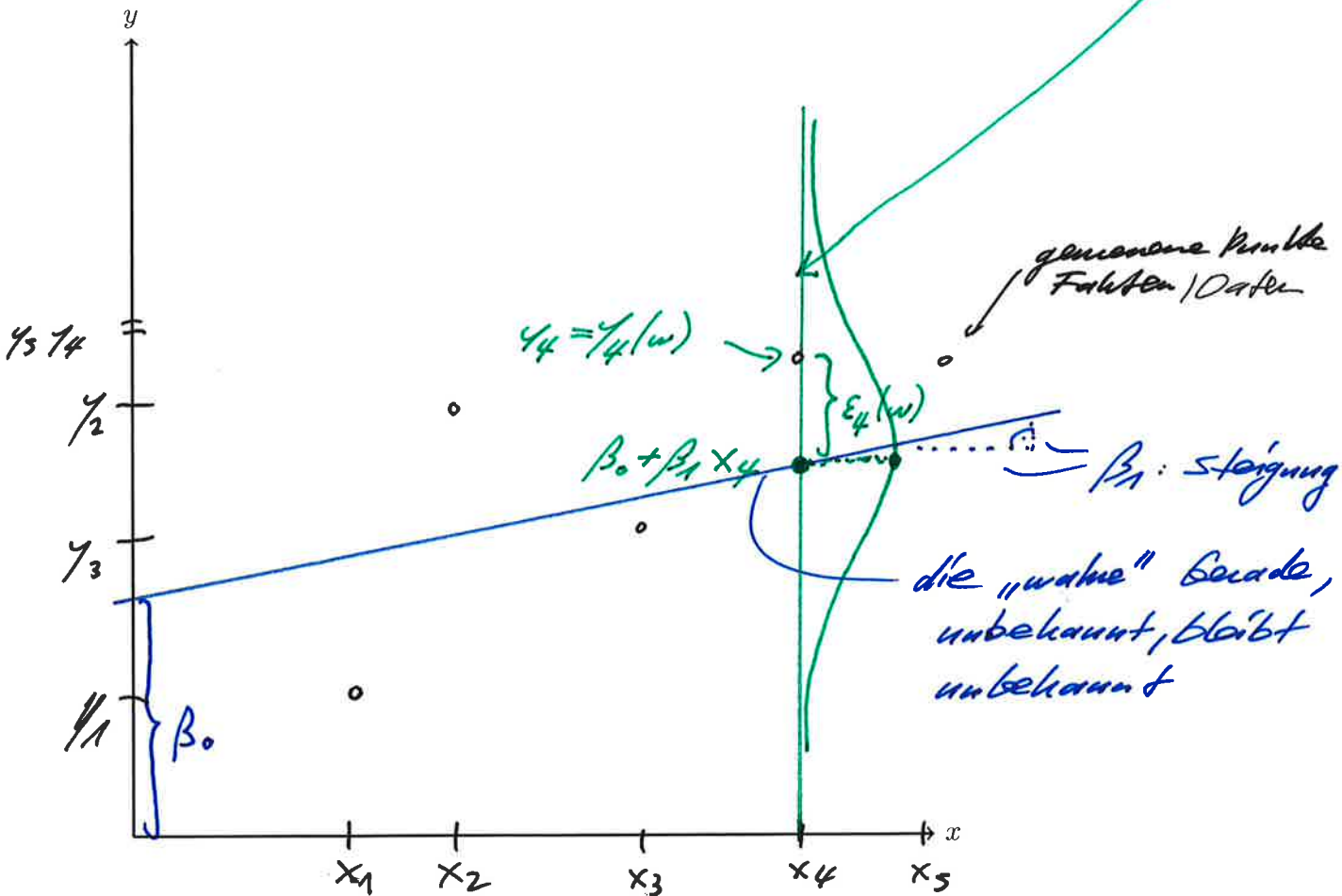
*↙ Distanz Wkt. → Realwert*

*für gegebenes  $i$  eine konstante Zahl*

CLT ↙

$$Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$

$\epsilon_1, \dots, \epsilon_n$  sind iid  $\mathcal{N}(0, \sigma^2)$ -verteilte Störterme. Auch  $\sigma^2$  ist ein zu schätzender Parameter. In vielen Anwendungen subsumiert man in der residualen Größe  $\epsilon_i$  kleine Effekte wie Messfehler, Rundungsfehler, zufällige Schwankungen, kleinste Einflüsse, welche man nicht in das Modell einbauen will, um es einfach zu halten.



*vgl. Kapitel 8*

### 10.2.1.3 Schätzen von $\beta_0$ , $\beta_1$ und $\sigma^2$ : OLS

Wir werden jetzt eine Methode kennenlernen (OLS), mit der man die unbekannt Parameter in (10.2) schätzen kann. Wir werden die Schätzungen der unbekannt Parameter mit  $\hat{\beta}_0$  und  $\hat{\beta}_1$  bezeichnen. Wir definieren damit die geschätzten Punkte

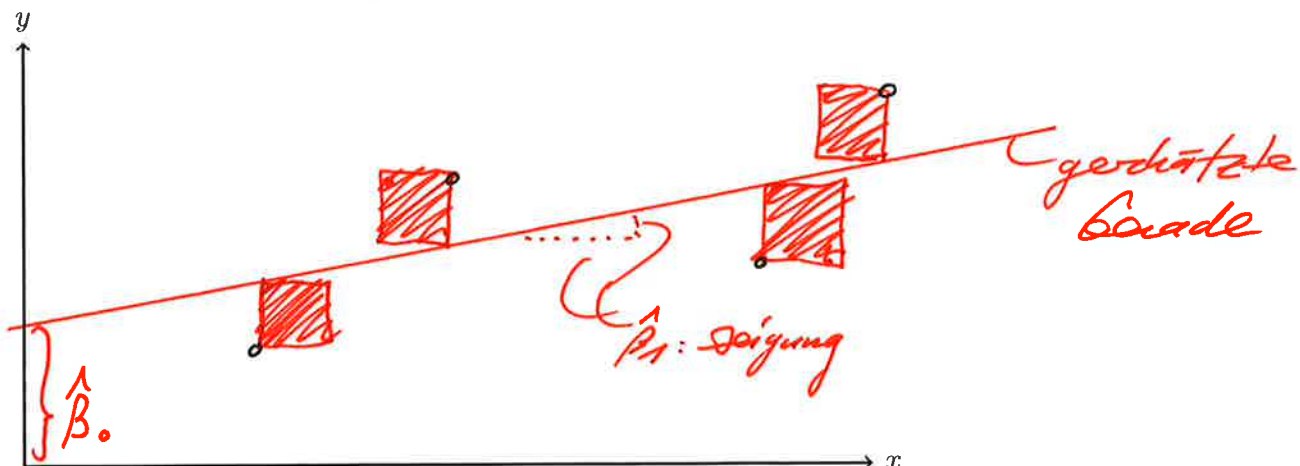
$$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, 1 \leq i \leq n,$$

die geschätzte Gerade ist dann  $y = \hat{\beta}_0 + \hat{\beta}_1 x$ . Die Aufgabe ist lediglich: wir suchen eine Gerade durch die Punktwolke  $(x_i, y_i)_{i=1}^n$  derart, dass die Summe der quadrierten Fehler (Sum of Squared Errors)

$$SSE := \sum_{i=1}^n (y_i - \hat{y}_i)^2 := \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

*hiermit spielen, um zu minimieren*

minimal ist (OLS=Ordinary Least Squares; kleinste Quadrat-Schätzung). Das ist eine sinnvolle Aufgabe, denn wir wollen den Fehler (die Abweichung der geschätzten Werte  $\hat{y}_i$  von den beobachteten Werten  $y_i$ ) klein halten.



Es gibt genau eine solche Gerade; wir werden sie gleich bestimmen.

Wir definieren wichtige Summen:

$$SSR := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2,$$

"R" steht dabei für Regression, dann noch die 3 Summen

$$SS_{xx} := \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2,$$

(entspricht etwa  $nV(X)$ )

$$SS_{yy} := \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2$$

(entspricht etwa  $nV(Y)$ )

und

$$SS_{xy} := \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}.$$

("Kreuzsumme")

Schreiten wir jetzt zur OLS-Schätzung: Wir haben von SSE die Summe

$$f(\beta_0, \beta_1) := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \quad (10.3)$$

bzgl.  $\beta_0$  und  $\beta_1$  zu minimieren (die  $x_i, y_i$  sind fest!). Von der Analysis her wissen wir, dass man bei eindimensionalen Optimierungsproblemen die erste Ableitung gleich 0 setzt (und die zweite Ableitung noch überprüft). Dies ist auch bei mehrdimensionalen Problemen so (bei der zweiten Ableitung ist es ein bisschen schwieriger - der Dozent weiss aber, dass es so gut geht).

Die Ableitung von (10.3) nach  $\beta_0$  bzw.  $\beta_1$  gleich 0 gesetzt ergibt:

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0; \quad \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0.$$

Dies ist äquivalent zu

$$\bar{y} - \beta_0 - \beta_1 \bar{x} = 0$$

und

$$SS_{xy} + n\bar{x}\bar{y} - \beta_0 n\bar{x} - \beta_1 (SS_{xx} + n\bar{x}^2) = 0.$$

Nach einfachen Umformungen erhalten wir

Kapitel 2

Neu:



HA

} 2 Gleichungen,  
2 Unbekannte

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \quad (10.4)$$

und

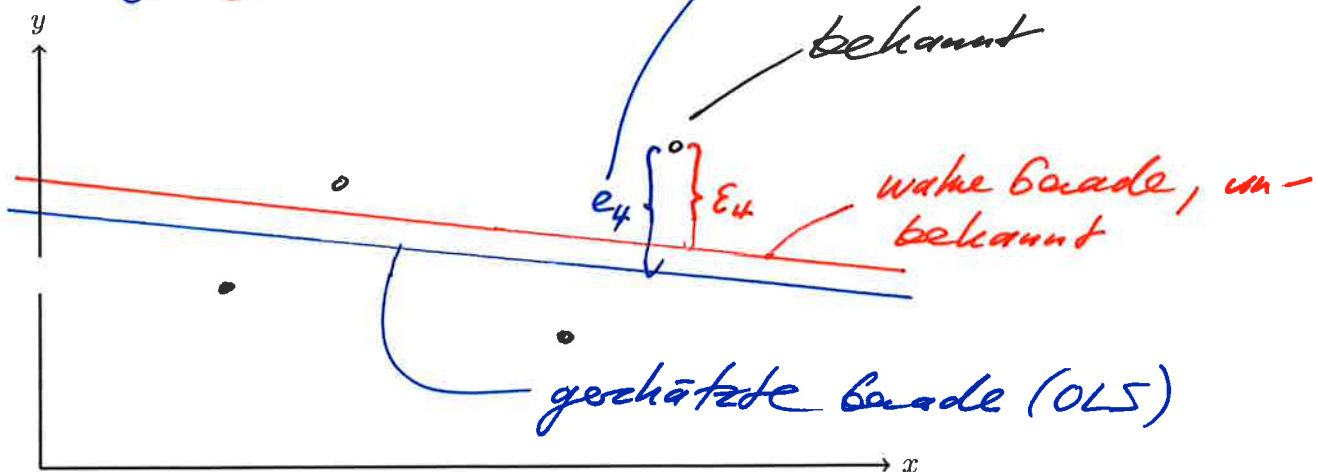
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (10.5)$$

Für die Schätzung von  $\sigma^2$  nehmen wir die *beobachteten Residuen*  $e_i := \hat{y}_i - \tilde{y}_i$ ,  $1 \leq i \leq n$ ,  
(die wahren sind ja unbeobachtet) und definieren:

$$\hat{\sigma}^2 := \frac{1}{n-2} \sum_{i=1}^n e_i^2. \quad (10.6)$$

Warum " $(n-2)$ " im Nenner? Mit dieser Wahl wird der Schätzer für  $\sigma^2$  erwartungstreu. Sie beweisen als freiwillige HA, dass auch  $\hat{\beta}_0, \hat{\beta}_1$  erwartungstreu sind. Wir besuchen nochmals unseren Computerausdruck und identifizieren dort die neu gefundenen Größen (10.4), (10.5) und (10.6).

Nachtrag zu  $e_j$  und  $\epsilon_i$ :



Eine kleine Betrachtung, welche uns für den **Modellfit mit  $R^2$  ("R-Squared")** eine leistungsfähige Kenngrösse liefern wird: Wir haben viele Summen definiert:  $SS_{xx}$ ,  $SS_{yy}$ ,  $SS_{xy}$ ; dazu SSE und SSR. Überraschenderweise gilt folgende Beziehung (den Beweis machen Sie bitte als kleine HA, wobei Sie Schätzungen (10.5) für  $\beta_0$  und (10.4) für  $\beta_1$  benutzen):

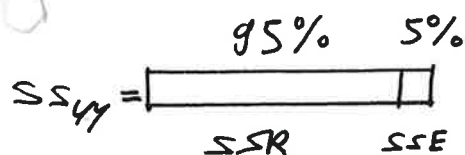
$$SS_{yy} = SSR + \underbrace{SSE}_{\substack{\text{Error} \\ \text{haben klein gemacht (OLS)}}$$

also

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (10.7)$$

*gut                      bösen*

Wie wird dies interpretiert?  $SS_{yy}$  ist ein Mass für die Variabilität in den uns interessierenden Daten  $y$  (die Radioaktivität!). Wir wollen die Radioaktivität möglichst präzise vorhersagen oder erklären. Mittel dazu sind die Anzahl (1 oder 2) und Auswahl (Distanz oder Regen) der erklärenden Variablen. Dann wollen wir die SSE, die Sum of Squared Errors, klein machen. Das haben wir mit OLS gemacht. Je nach Anzahl und Auswahl der erklärenden Variablen wird SSE mehr oder weniger klein. Wenn wir ein kleines SSE erhalten, ist es gut. Da  $SS_{yy}$  gegeben ist (es sind die Messwerte) und in der Form  **$SS_{yy} = SSR + SSE$**  aufgespalten wird, ist entsprechend ein grosses SSR gut. Statistiker sagen dann: **"Die Variation in den  $y$  ( $SS_{yy}$ ) lässt sich aufspalten in einen Anteil, der durch die Regression erklärt wird (SSR) und eine residuale Summe (SSE)".** Dies lädt ein, mit



$$R^2 := \frac{SSR}{SSR + SSE} \in [0, 1]$$

*0.95 (sehr gut)*

eine **Kenngrösse anzugeben, welche sagt, wie gut der Fit des Modells an die Daten ist.** Grosse Werte bedeuten einen guten Fit (Modell nur mit Distanz), kleine Werte einen schlechten Fit (Modell nur mit Regen). Sobald wir mehr erklärende Variablen einbauen, wird der Fit automatisch besser (siehe später und auch im vollen Modell). Deshalb hat man mit **" $R^2$ -adjusted"** noch eine weitere Kenngrösse definiert, welche viele erklärende Variablen bestraft. Auch dies suchen wir jetzt in unserem Computerausdruck. Informativ sind auch die beiden Bilder a) mit den Boxplots der Residuen und b) der Cartoon.



*vgl. Kapitel 9*

### 10.2.1.4 Testen ob $\beta_1 = b$

Wenn wir uns im Spezialfall der einfachen Regression nur für die Steigung der Geraden interessieren, gibt es eine einfache Herleitung eines statistischen Tests ob

$$H_0: \beta_1 = b$$

gilt oder nicht, wo  $b$  eine beliebige Zahl (oft 0; default in R). Wieso treten solche Fragen überhaupt auf? Dazu zwei ~~Szenarien~~  $H_0$  und  $H_1$ : *, mit  $b=0$ :*

*Szenarien*

$$(10.2) \quad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$H_0$

$$Y_i = \beta_0 + \varepsilon_i$$

$H_1$

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

*fundamentale, qualitative & nicht nur  
quantitative Unterscheid:*

*$H_1$  hat  $X$ -Struktur*

*$H_0$  nicht!*

**Kleine, notwendige Warnung:**

$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$  wird nie genau 0 (bzw  $b$ ) sein unter  $H_0$ .

Das ist genau der gleiche Fehler, wie wenn Sie bei einem t-Test naiv erwarten,  $\bar{x}$  müsse schon nahe oder gar exakt 0 (bzw  $\mu_0$ ) sein.

$\hat{\beta}_1$  reicht nicht aus zur Untersuchung, ob  $\beta_1 = 0$  oder nicht, wir brauchen auch hier die Streuung, genauer: **Effekt versus Streuung wie in Kapitel 9**. Das machen wir jetzt nachfolgend:

Eine kleine Betrachtung, welche uns für den **Modelfit mit  $R^2$  ("R-Squared")** eine leistungsfähige Kenngrösse liefern wird: Wir haben viele Summen definiert:  $SS_{xx}, SS_{yy}, SS_{xy}$ ; dazu SSE und SSR. Überraschenderweise gilt folgende Beziehung (den Beweis machen Sie bitte als kleine HA, wobei Sie Schätzungen (10.5) für  $\beta_0$  und (10.4) für  $\beta_1$  benutzen):

$$SS_{yy} = SSR + SSE,$$

*Erroz!*

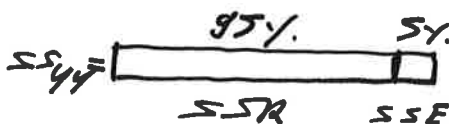
*haben möglichst klein gemacht mit OLS*

also

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (10.7)$$

*(gut) (bias)*

Wie wird dies interpretiert?  $SS_{yy}$  ist ein Mass für die Variabilität in den uns interessierenden Daten  $y$  (die Radioaktivität!). Wir wollen die Radioaktivität möglichst präzise vorhersagen oder erklären. Mittel dazu sind die Anzahl (1 oder 2) und Auswahl (Distanz und Regen) der erklärenden Variablen. Dann wollen wir die SSE, die Sum of Squared Errors, klein machen. Das haben wir mit OLS gemacht. Je nach Anzahl und Auswahl der erklärenden Variablen wird SSE mehr oder weniger klein. Wenn wir ein kleines SSE erhalten, ist es gut. Da  $SS_{yy}$  gegeben ist (es sind die Messwerte) und in der Form  **$SS_{yy} = SSR + SSE$**  aufgespalten wird, ist entsprechend ein grosses SSR gut. Statistiker sagen dann: **"Die Variation in den  $y$  ( $SS_{yy}$ ) lässt sich aufspalten in einen Anteil, der durch die Regression erklärt wird ( $SSR$ ) und eine residuale Summe ( $SSE$ )"**. Dies lädt ein, mit



$$R^2 := \frac{SSR}{SSR + SSE} \in [0, 1] \quad (0.95)$$

eine **Kenngrösse** anzugeben, welche sagt, wie gut der Fit des Modells an die Daten ist. Grosse Werte bedeuten einen guten Fit (Modell nur mit Distanz), kleine Werte einen schlechten Fit (Modell nur mit Regen). Sobald wir mehr erklärende Variablen einbauen, wird der Fit automatisch besser (siehe später und auch im vollen Modell). Deshalb hat man mit **" $R^2$ -adjusted"** noch eine weitere Kenngrösse definiert, welche **viele erklärende Variablen bestraft**. Auch dies suchen wir jetzt in unserem Computerausdruck. Informativ sind auch die beiden Bilder a) mit den Boxplots der Residuen und b) der Cartoon.

*online*

Für die jetzige Untersuchung setzen wir voraus, dass (10.2), also  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ , gilt. Dann können wir folgendermassen argumentieren:

1. Die Schätzformel für  $\beta_1$  lautet (vgl. (10.4)):

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

2. Weil  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  gilt auch  $\sum_{i=1}^n \bar{y}(x_i - \bar{x}) = 0$  und damit

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

3. Schätzer sind (vor der Realisation) Zufallsgrössen. Wir setzen für die Datenpunkte  $y$  jetzt die Zufallsgrössen  $Y$  ein: *im Bsp. Links*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x})Y_i. \quad (10.8)$$

Da  $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ ,  $1 \leq i \leq n$ , unabhängig verteilt, können wir schliessen (kleine Rechnungen für Personen, welche die Mathe lieben), dass gilt:

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right).$$

*wahre Wert*

Jetzt machen wir die Z-Transformation im Fall  $\mathcal{H}_0: \beta_1 = b$ :

*denken Sie immer an  $b=0$*

$$1. \hat{\beta}_1 - b \stackrel{\mathcal{H}_0}{\sim} \mathcal{N}\left(0, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

$$2. \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \stackrel{\mathcal{H}_0}{\sim} \mathcal{N}(0, 1)$$

Das kleine Problem ist, dass wir  $\sigma^2$  (wieder mal) nicht kennen. Aber genau wie in früheren Kapiteln können wir ja eine Schätzung von  $\sigma^2$  (hier Formel (10.6)) zu Hilfe nehmen. Unter  $\mathcal{H}_0$  haben wir  $\beta_1 = b$  und damit hat unsere Teststatistik

$$T_{n-2} := \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_1 - b}{\sqrt{\frac{1}{n-2} \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$$

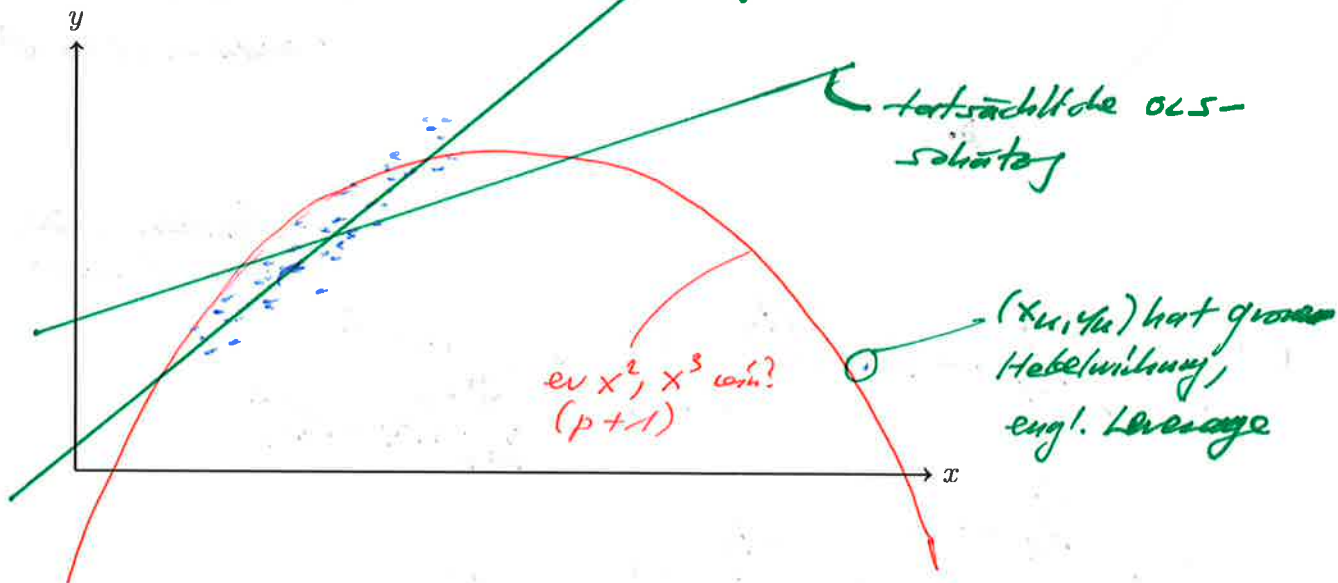
*operational*

die  $t$ -Verteilung mit  $n-2$  (Fehlerquelle!) Freiheitsgraden. Wir werden die  $\mathcal{H}_0$ -Hypothese verwerfen, wenn diese Teststatistik Werte annimmt, welche weiter als die kritischen Werte von 0 entfernt sind. Wir schauen nochmals unseren Computer-Ausdruck an (nur einfache Regression).

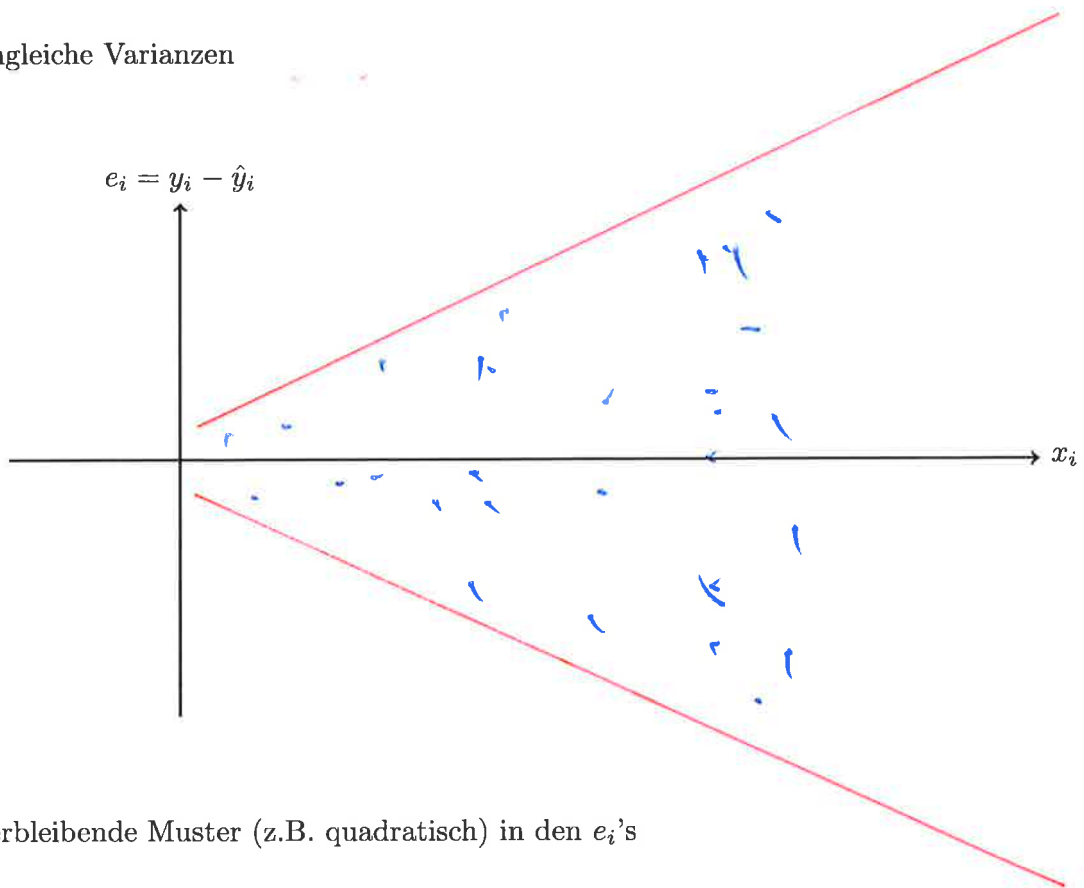
### 10.2.1.5 Probleme & Diagnostic Checking

Wir geben hier nur einen kurzen Überblick über mögliche Probleme, Gefahren und die Methoden, welche man unter "Diagnostic Checking" zusammenfasst. Zusätzliche Informationen dazu in Stachel.

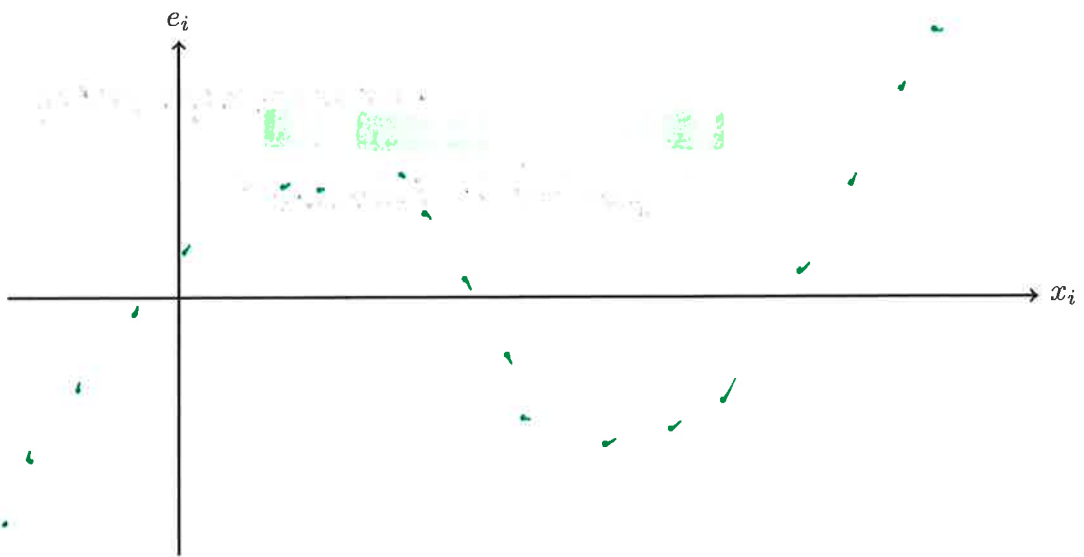
\* Ausreisser (engl. Outlier)



\* ungleiche Varianzen



\* verbleibende Muster (z.B. quadratisch) in den  $e_i$ 's



\*  $\epsilon_i$ 's nicht unabhängig

\*  $\epsilon_i$ 's nicht normalverteilt

### 10.2.1.6 Warum ist die lineare Regression mit OLS so wichtig, bekannt und erfolgreich?

- \* wird auch von Nicht-MathematikerInnen/Nicht-StatistikerInnen verstanden
- \* theoretisch einfach zu berechnen
- \* einfach auch zur multiplen Regression erweiterbar
- \* früher war in Statistik-Paketen oft nur diese Regression programmiert (heute nicht mehr als Argument relevant)

#### warum speziell linear

- \* Mensch kann nur lineare Zusammenhänge gut erfassen
- \* viele nichtlineare Abhängigkeiten können durch Transformation zu linearen Problemen gemacht werden (ist aber auch umstritten: "Man foltert die Daten bis sie gestehen"). Vor allem: viele Phänomene mit exponentiellem (oder geometrischem, falls diskret) Wachstum: Wirtschaft (gesamte Volkswirtschaft und einzelne Firmen), Pflanzen (Zellteilung), Ausbreitung von Epidemien

#### warum speziell OLS

- \* früher EDV-Probleme bei alternativen Vorschlägen (heute nicht relevant)
- \* OLS ist auch der BLUE (Best Linear Unbiased Estimator), auch BLUE im multivariaten Fall

*erwartungstreu, kein Bias*  
*„kleinste Varianz“*

### 10.2.2 Ausblick multiple Regression

Wir werden jetzt versuchen, die Radioaktivität mit Distanz *und* Regen zu erklären (oder vorherzusagen). Damit verlassen wir Modell (10.2) und gehen zu

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \quad (10.9)$$

Gierman: 182

$$z = a + bx + cy \quad (\text{Ebene im } \mathbb{R}^3)$$

Wir werden die **OLS-Schätzung** in dieser Vlsng nur für die einfache Regression anschauen.

Aber auch in der multiplen Regression kann man eine OLS-Schätzung machen; sie ist einfach komplizierter. Mit der OLS-Schätzung erhalten wir hier (vgl. R)

$$\log(\text{bq}) = 10.522 - 1.360 * \log(\text{dist}) + \underline{2.723 * \text{regen}}$$

1 ≡ Regen  
0 ≡ kein Regen

Residual standard error: 0.6166197

Dies ist *auf die Schnelle* wohl die beste Datenanalyse: der Regen hat (wie in der ersten Stunde vermutet und theoretisch erwartet) einen "Wash out" mit erhöhter Radioaktivität zur Folge.

Die geschätzte Varianz des Fehlerterms ist hier kleiner als bei der einfachen Regression, da noch ein Predictor dazugekommen ist (siehe auch bei den nachfolgenden Bemerkungen zur multiplen Regression).

Wie kann man den Koeffizienten von Regen "+2.723" verstehen?

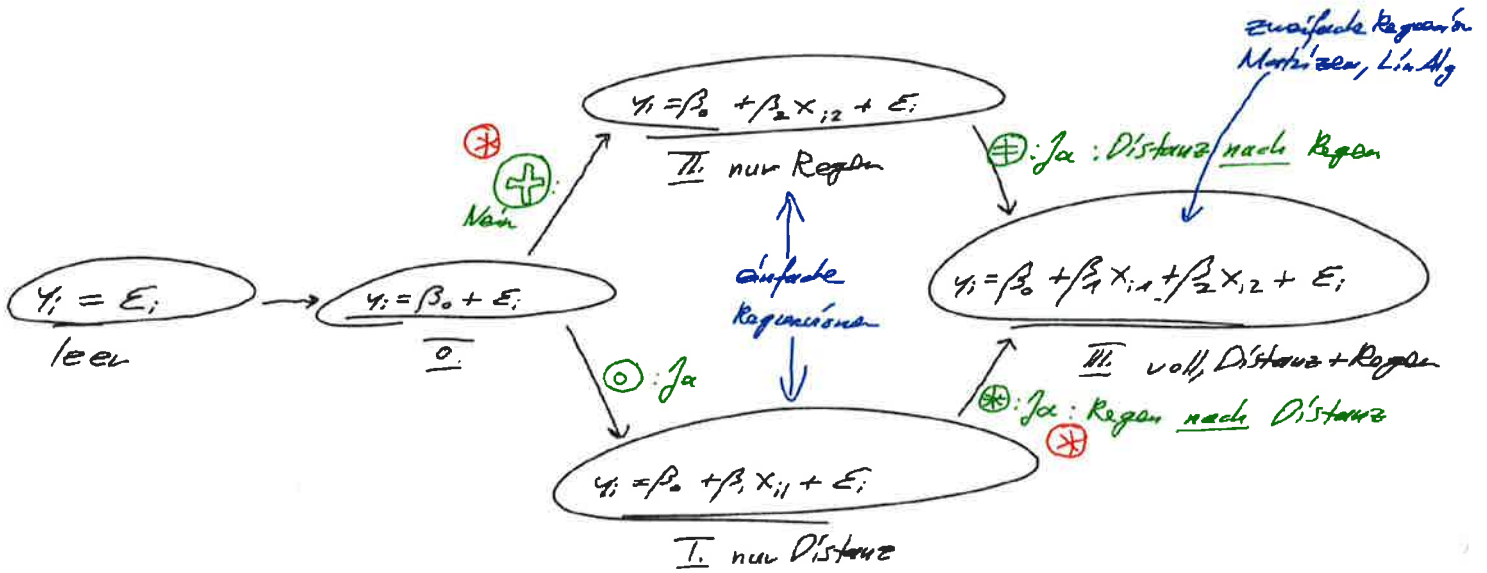
○ *undarten (ohne Logarithmus): Nicht Regen → Regen*  
*y → 10·y*  
*(neu: Abschlagsrechnung)*

(non regen)                      (regen)

$$\bullet \ln(y) \rightarrow \ln(10y) = \ln(10) + \ln(y)$$

Japan 1                      Japan 2                      ~2.3

Was wird getestet bei obiger zweifachen Regression  $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$ ?

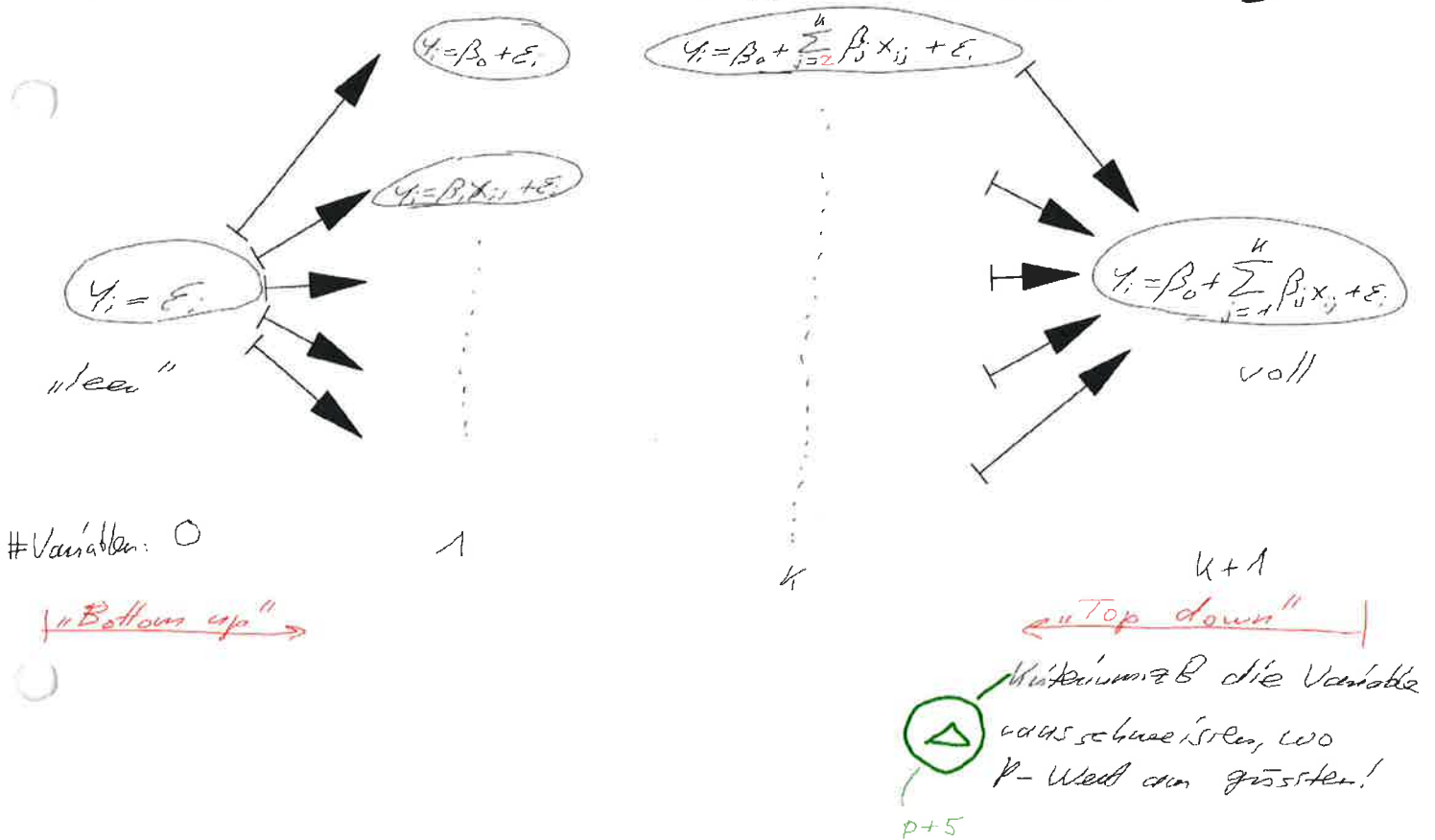


- $\otimes$ : Regen dazu oder nicht?  $\oplus$ : Nein;  $\ominus$ : Ja
- $\oplus$ : Test: Regen wichtig? P-Wert 54%  $\rightarrow$  Nein!
- $\otimes$ : Test: Regen nach Distanz? P-Wert 11%  $5.51 \cdot 10^{-8} \rightarrow$  Ja

Fazit: 1. Multiple Regression nicht so einfach  $\Rightarrow$  Beratung, statistische  
2. fähig, konkrete Analysen nachher vollziehen; vgl. R-Ausdruck III.

Wir gehen noch kurz darauf ein, wie mit noch mehr erklärenden Variablen vorgegangen wird. Dazu ein Schema zu "Top Down" und "Bottom Up":

- $(k+1)$  Variablen mit Parametern  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ ; wobei  $k=2$
- $2^{(k+1)}$  mögliche Modelle
- jede Variable kann man haben oder nicht  $\Rightarrow 2^{k+1}$



◦ Pragmatischer Grundsatz: lieber weniger Variablen im System  
„Parsimony“ (Sparsamkeit)

◦ „Alle Modelle sind falsch - ein paar sind nützlich!“

Lose Bemerkungen zur multiplen Regression:

\* wenn man in der Forschung oder Industrie Probleme untersucht, weiss man manchmal überhaupt nicht, welche Variablen (erklärenden Faktoren) relevant sind (zB Regen, Distanz als Erklärung der Radioaktivität).

1)

\* Fantasie, Kreativität, Ideen sind dann zentral wichtig; wir können dann **Hypothesen bilden (Forschungsfreiheit!)** und sollten uns selbstkritisch keine Anmassung von Wissen erlauben.

2)

\* **danach Hypothesen an Daten überprüfen (Popper: nur allfällige Falsifizierung möglich)**

\* Welche und wieviele Daten:  $n$ =Stichprobengrösse und  $k$ =Anzahl erklärende Variablen (Regen und Distanz sind  $k = 2$ , Intercept wird nicht gezählt). Ein grosses  $n$  ist erwünscht (abgesehen von Kosten/Zeitaufwand); ein grosses  $k$  ist prinzipiell auch erwünscht, aber die Analyse ist mit Fallstricken und Problemen behaftet. Ganz allgemein sollte gelten:  $n \gg k$

\* Was geschieht, wenn wir eine weitere erklärende Variable hinzufügen möchten; zum Beispiel von  $k = 1$  zu  $k = 2$ ? Anstelle von

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2, \quad (10.4)$$

welches wir bzgl.  $\beta_0$  und  $\beta_1$  minimieren tritt neu

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}))^2,$$

welches wir bzgl.  $\beta_0, \beta_1$  und  $\beta_2$  minimieren. Die <sup>minimierte</sup> neue Summe ist ganz sicher kleiner als diejenige in (10.4): wir sind in der gleich guten Situation, wenn wir  $\beta_2 = 0$  wählen, können aber neu auch noch  $\beta_2$  variieren und kommen damit ganz sicher in eine bessere Situation. Der Fit wird besser, die Residuen insgesamt kleiner. Das hat auch einen Nachteil: selbst wenn wir völlig unsinnige Grössen als erklärende Faktoren einführen, wird der Fit besser. Bei  $k = n$  haben wir im Normalfall sogar einen perfekten Fit. Um den Einbau von

# Katzen pro Haushalt

Radioaktivitaet mit Distanz allein erklaren (einfache Regression):

```
> NurDistanz<-lm(log(bq) ~ log(dist))
```

```
> aov(NurDistanz)
```

Call:

```
aov(formula = NurDistanz)
```

Terms:

log(dist) Residuals

Kommentar Luchs: SSR SSE von (10.7)

Sum of Squares 152.87486 34.39141

Deg. of Freedom 1 21

Residual standard error: 1.279721 Kommentar Luchs: das ist  $\hat{\sigma}$  mit  $\sqrt{(10.6)}$

Estimated effects may be unbalanced

```
> summary(NurDistanz)
```

Call:

```
lm(formula = log(bq) ~ log(dist))
```

Residuals:

Min 1Q Median 3Q Max
-1.8592 -1.1348 -0.1033 1.1650 2.3797

Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.8027 0.7022 13.960 4.26e-12 \*\*\*

log(dist) -1.0911 0.1129 -9.662 3.53e-09 \*\*\*

Estimate / Std. Error = t value, zweiseitiger P-Wert, Test ob 0 oder nicht

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.28 on 21 degrees of freedom

Multiple R-Squared: 0.8164, Adjusted R-squared: 0.8076

F-statistic: 93.35 on 1 and 21 DF, p-value: 3.527e-09

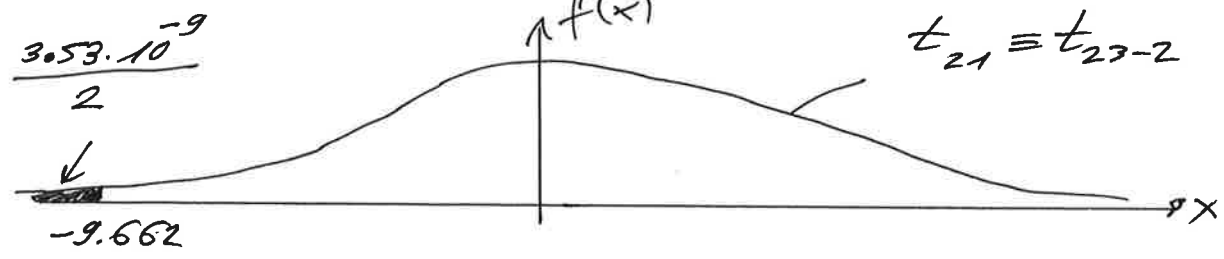
gutes, altes, einfaches Modell

(10.5)  $\hat{\beta}_0$ 
(10.4)  $\hat{\beta}_1$

$t_{n-2}$

$L_d \Rightarrow H_1$ , Distanz relevant

$(r_{xy})^2$



## // Radioaktivitaet mit Regen allein erklaren (einfache Regression):

```
> NurRegen<-lm(log(bq) ~ regen)
```

```
> aov(NurRegen)
```

```
Call:
```

```
aov(formula = NurRegen)
```

```
Terms:
```

	regen	Residuals
Sum of Squares	3.31193	183.95434

Deg. of Freedom	1	21
-----------------	---	----

```
Residual standard error: 2.959684
```

```
Estimated effects may be unbalanced
```

```
> summary(NurRegen)
```

```
Call:
```

```
lm(formula = log(bq) ~ regen)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-5.6755	-2.3332	-0.1205	2.1735	4.7970

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.7784	0.7399	5.107	4.66e-05 ***
regen	-0.8247	1.3412	-0.615	0.545 > α, nicht signifikant

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.96 on 21 degrees of freedom
```

```
Multiple R-Squared: 0.01769, Adjusted R-squared: -0.02909 (schlechtes Modell)
```

```
F-statistic: 0.3781 on 1 and 21 DF, p-value: 0.5452
```

/// Volles Modell (zweifache Regression): *(sehr gut)*

```
> FullModel<-lm(log(bq) ~ log(dist)+ regen)
```

```
> aov(FullModel)
```

Call:

```
aov(formula = FullModel)
```

Terms:

	log(dist)	regen	Residuals
Sum of Squares	152.87486	26.78701	7.60440
Deg. of Freedom	1	1	20

Residual standard error: 0.6166197

Estimated effects may be unbalanced

```
> summary(FullModel)
```

Call:

```
lm(formula = log(bq) ~ log(dist) + regen)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.37021	-0.37287	-0.05441	0.21542	1.61984

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	10.52246	0.34904	30.147	< 2e-16 ***
log(dist)	-1.36027	0.06316	-21.536	2.62e-15 ***
regen	2.72254	0.32436	8.394	5.51e-08 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6166 on 20 degrees of freedom

Multiple R-Squared: 0.9594, Adjusted R-squared: 0.9553

F-statistic: 236.3 on 2 and 20 DF, p-value: 1.219e-14

„Stolch & Kinder“

### 10.2.4 Korrelation (ein Mass für die lineare Gleichläufigkeit; kein Kausalzusammenhang)

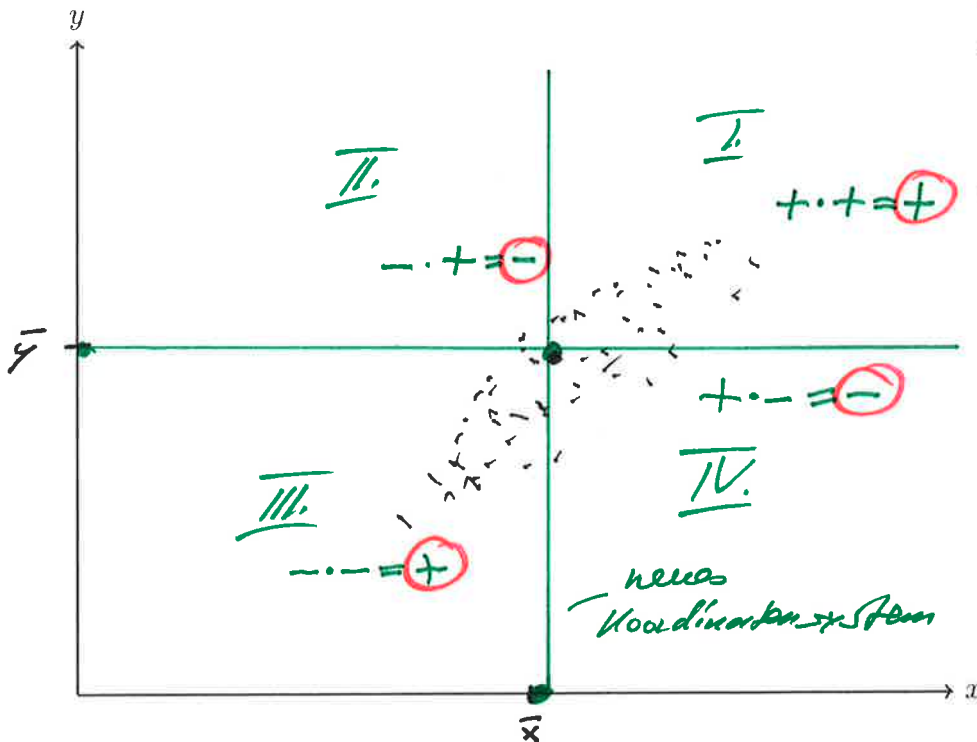
nicht: Abhängigkeit

Mit der einfachen Regression haben wir also eine Möglichkeit, den linearen Zusammenhang zwischen 2 Grössen (x, y) zu untersuchen. Dabei wird vorausgesetzt, dass das Modell

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (10.2)$$

linear, Gerade

gilt. Es gibt jedoch auch eine Masszahl, welche nicht zwingend (10.2) voraussetzt und für alle zweidimensionalen Zusammenhänge berechnet werden kann: der Korrelationskoeffizient. Wir werden in dieser Vorlesung lediglich den empirischen Korrelationskoeffizienten anschauen. Betrachten wir dazu folgende Darstellung:



3 Bsp:

- x: Futtermenge  
y: Gewicht
- x: Grösse Mutter  
y: Grösse Sohn
- x: Aktie UBS  
y: Aktie HSBC
- $x'_i := (x_i - \bar{x})$   
 $y'_i := (y_i - \bar{y})$   
(vgl. Kap 17, 182)

- Haben jetzt viele Punkte im I. & III. Quadranten & wenige im II. & IV. Quadranten.
- Deshalb bilden wir:

$$\sum_{i=1}^n x'_i \cdot y'_i = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (= \sum_{+y} \text{„Kreuzsumme“})$$

wird grosser, da viel  $\oplus$  & wenig  $\ominus$

sg. Kovarianz  
Problem: leg oder g  
⇒ 1'000 mal mehr (1'000 · 000 so oft...)

Der empirische Korrelationskoeffizient zwischen den Daten  $(x_i)_{i=1}^n$  und  $(y_i)_{i=1}^n$  ist also definiert als

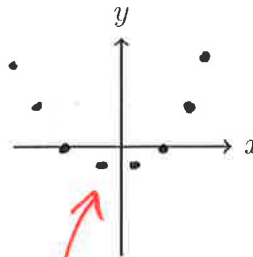
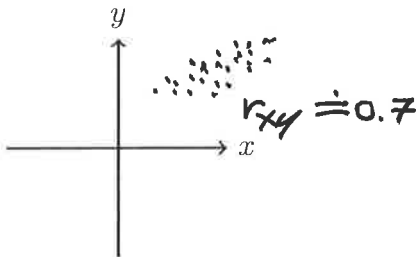
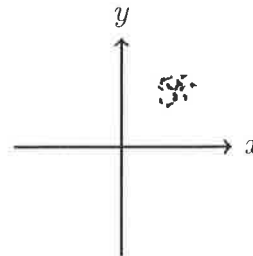
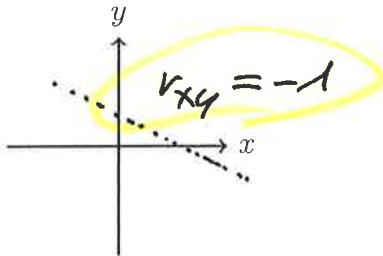
*im Zähler spielt Maszk*

*Normalisierung, Steiler  
[m cm], [km], [m]*

$$r_{xy} := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

Mathematiker/innen haben gezeigt, dass diese Masszahl zwischen  $-1$  und  $+1$  liegt.

Zudem ist  $|r_{xy}|$  genau dann 1, wenn die Punkte alle auf einer Geraden liegen. Ist  $r_{xy} = -1$ , so hat diese Gerade negative Steigung - ist  $r_{xy} = 1$ , so hat diese Gerade positive Steigung. Wir machen dazu jetzt noch ein paar Bilder, um ein bisschen ein Gefühl für diese Masszahl zu erhalten: **KLICKER**



*Abhängig, aber nicht linear!*

**Wichtig:**

1. Lesen Sie jetzt im Buch Kapitel 10.
2. Gehen Sie in die Übungsstunde. Drucken Sie das Übungsblatt dazu *vorher* aus, lesen Sie *vorher* die Aufgaben durch und machen sich erste Gedanken dazu (zum Beispiel, wie man sie lösen könnte).
3. Dann lösen Sie das Übungsblatt: zuerst immer selber probieren, falls nicht geht: Tipp von Mitstudi benutzen, falls immer noch nicht geht: Lösung von Mitstudi anschauen, 1 Stunde warten, versuchen, aus dem Kopf heraus wieder zu lösen, falls immer noch nicht geht: Lösung von Mitstudi abschreiben (und verstehen - also sollte man insbesondere keine Fehler abschreiben!).
4. Lösen Sie die entsprechenden Prüfungsaufgaben im Archiv.

**Lessons Learnt:**

ANOVA-Teststatistik

$$V := \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_j - GM)^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 / (n-k)} = \frac{\sum_{j=1}^k n_j (\bar{Y}_j - GM)^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 / (n-k)}.$$

Unter  $\mathcal{H}_0$   $F_{k-1, n-k}$ -verteilt. Ablehnungsbereich einseitig wo gross. Wenn wir nur  $k = 2$  Gruppen haben, so ist das ein quadrierter 2-Stichproben-T-Test.

Modell einfache Regression:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (10.2)$$

Geschätzte Punkte

$$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, 1 \leq i \leq n,$$

geschätzte Gerade  $y = \hat{\beta}_0 + \hat{\beta}_1 x$ 

Summe der quadrierten Fehler (Sum of Squared Errors)

$$SSE := \sum_{i=1}^n (y_i - \hat{y}_i)^2 := \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

$$SSR := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2,$$

"R" steht dabei für Regression, dann noch die 3 Summen

$$SS_{xx} := \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2, \quad (\text{entspricht etwa } nV(X))$$

$$SS_{yy} := \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 \quad (\text{entspricht etwa } nV(Y))$$

und

$$SS_{xy} := \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}. \quad (\text{"Kreuzsumme"})$$

Schätzung der Parameter:

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \quad (10.4)$$

und

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (10.5)$$

Mit *beobachteten Residuen*  $e_i := y_i - \hat{y}_i, 1 \leq i \leq n$  definieren wir:

$$\hat{\sigma}^2 := \frac{1}{n-2} \sum_{i=1}^n e_i^2. \quad (10.6)$$

Güte der Regression:

$$SS_{yy} = SSR + SSE,$$

also

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (10.7)$$

$$R^2 := \frac{SSR}{SSR + SSE} \in [0, 1]$$

Teststatistik für die Steigung:

$$T_{n-2} := \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$$

$t_{n-2}$ -Verteilung.  $\mathcal{H}_0$  verwerfen, wenn Teststatistik Werte annimmt, welche weiter als die kritischen Werte von 0 entfernt sind.

Der empirische Korrelationskoeffizient zwischen Daten  $(x_i)_{i=1}^n$  und  $(y_i)_{i=1}^n$

$$r_{xy} := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}.$$

### Prüfungstipps MAT 183 (nicht einklagbar wenn missverstanden)

- FS26
- \* Sigma geschätzt > t-Verteilung (KI, Tests, ~~Regression~~)
  - \* F, Chiquadrat und t: df abziehen, ausser expl Frage zB nach  $t_n$
  - \* Vorsicht: Delta alter Storrer und R:  $\mathcal{N}(\mu, \sigma)$  vs Luchsinger  $\mathcal{N}(\mu, \sigma^2)$
  - \* Mehr als 2 Behandlungen in dieser Vlsg > immer ANOVA
  - \* alpha nicht 5 % > keine ANOVA in dieser Vlsg; ausser P-Wert gegeben
  - \* Gleich viele  $x_i$  wie  $y_i$ : paired t-test od Regression, Korrelation
  - \* Teststatistik ANOVA mit  $k=2$  ist 2-Stichproben-T-Test quadriert
  - \* A2:  $X^2$  und  $|X|$ : negative Beiträge nicht vergessen
  - \* A3: diskret (isolierte Punkte) od stetig (durchgezogen)
  - \* alpha/2 oder nicht - ein/zweiseitig? KI bei alpha/2
  - \*  $\chi^2$ -Test Unabhängigkeit oder 2-Stichproben-Proportionentest? wird angegeben
  - \* ANOVA und Chiquadrattests sind in MAT 183 immer einseitig;  $\mathcal{H}_1$  oben

### Klickerfragen zum Aufwärmen:

Frage 1: Ein Mediziner untersucht die Wirkung von drei verschiedenen Schmerzmitteln auf die Schmerzreduktion bei Patienten. Er misst den Schmerzlevel der Patienten nach der Einnahme in drei Gruppen. Welche Methode sollte er verwenden, um zu prüfen, ob sich die Schmerzmittel signifikant in ihrer Wirkung unterscheiden? (a) ANOVA, (b) t-Test, (c) Chi-Quadrat-Test

Frage 2: Was ist das Hauptziel der ANOVA (Analysis of Variance)? (a) Den Unterschied zwischen zwei Mittelwerten zu testen. (b) Zu überprüfen, ob die Mittelwerte mehrerer Gruppen gleich sind. (c) Die beste Gerade zur Beschreibung eines Zusammenhangs zwischen zwei Variablen zu finden.

Frage 3: Wann ist es sinnvoll, eine ANOVA anstelle eines t-Tests zu verwenden? (a) Wenn es genau zwei Gruppen gibt. (b) Wenn es mehr als zwei Gruppen gibt. (c) Wenn die Daten diskret sind.

Frage 4: Was ist die Nullhypothese ( $\mathcal{H}_0$ ) bei einer ANOVA? (a) Alle Gruppen haben den gleichen Mittelwert. (b) Die Gruppen haben unterschiedliche Mittelwerte. (c) Die Gruppen haben identische Standardabweichungen.

Frage 5: In einer ANOVA vergleicht man die zwischen den Gruppen liegende Varianz mit der innerhalb der Gruppen liegenden Varianz. Was bedeutet ein grosser F-Wert? (a) Es gibt vermutlich einen signifikanten Unterschied zwischen den Gruppen. (b) Alle Gruppen haben die gleiche Mittelwertdifferenz. (c) Die Gruppen unterscheiden sich nicht signifikant.

Frage 6: Welche Verteilung folgt der ANOVA-Teststatistik unter der Nullhypothese? (a) Normalverteilung, (b) t-Verteilung, (c) F-Verteilung

Frage 7: Was ist das Hauptziel der linearen Regression? (a) Die Beziehung zwischen einer unabhängigen und einer abhängigen Variable zu modellieren. (b) Die Standardabweichung einer Stichprobe zu berechnen. (c) Mehrere Gruppen auf Gleichheit der Mittelwerte zu testen.

Frage 8: Welche allgemeine Gleichung beschreibt eine einfache lineare Regression? (a)  $y = ax + b$ , (b)  $y = \beta_0 + \beta_1x + \epsilon$ , (c)  $y = x^2 + bx + c$ .

Frage 9: In der Regressionsgleichung  $y = \beta_0 + \beta_1x + \epsilon$  beschreibt der Parameter  $\beta_1$ : (a) Die Steigung der wahren Regressionsgeraden. (b) Den Wert von  $y$ , wenn  $x = 0$  ist. (c) Die Summe aller Fehlerterme.

Frage 10: Ein Wissenschaftler möchte untersuchen, ob ein Zusammenhang zwischen der Körpergrösse der Eltern und der Körpergrösse ihrer Kinder besteht. Welche Methode sollte er verwenden? (a) ANOVA; (b) Lineare Regression; (c) Chi-Quadrat-Test.

Frage 11: Was ist das Ziel der Ordinary Least Squares Approximation (OLS) in der linearen Regression? (a) Die Gerade zu finden, die durch alle Datenpunkte exakt verläuft. (b) Die Parameter einer linearen Funktion so zu bestimmen, dass die Summe der quadrierten Abweichungen der vorhergesagten Werte von den beobachteten Werten minimiert wird. (c) Die Parameter einer linearen Funktion so zu bestimmen, dass die Summe der absoluten Abweichungen der vorhergesagten Werte von den beobachteten Werten minimiert wird.

Frage 12: Ein Professor untersucht den Zusammenhang zwischen der Anzahl Punkten in den Übungsblättern ( $x$ ) und der Punktzahl in einer Prüfung ( $y$ ). Er erhält die Regressionsgleichung:  $y = 50 + 0.5x$ . Was bedeutet die Steigung (0.5) in diesem Zusammenhang? (a) Die maximale erreichbare Punktzahl in der Prüfung beträgt 0.5 Punkte. (b) Die Punktzahl in der Prüfung verdoppelt sich mit jedem zusätzlichen Punkt in den Übungsblättern. (c)

Mit jedem zusätzlichen Punkt in den bungsbilättern steigt die erwartete Punktzahl um 0.5 Punkte.

Frage 13: Eine Firma untersucht den Zusammenhang zwischen der Anzahl an Werbeanzeigen ( $x$ ) und dem monatlichen Umsatz ( $y$ ) in tausend CHF. Sie erhält die Regressionsgleichung:  $y = 20 + 3x$ . Was bedeutet die Steigung (3) in diesem Zusammenhang? (a) Der Umsatz steigt insgesamt um 3'000 CHF, unabhängig von der Anzahl der Werbeanzeigen. (b) Jede zusätzliche Werbeanzeige verdreifacht den monatlichen Umsatz. (c) Mit jeder zusätzlichen Werbeanzeige steigt der erwartete Umsatz um 3'000 CHF.

Frage 14: Ein Sportwissenschaftler untersucht die Wirkung von drei verschiedenen Trainingsprogrammen auf die Sprintgeschwindigkeit von Athleten. Jede Gruppe trainiert mit einem anderen Programm, und am Ende der Studie wird die Sprintzeit der Athleten gemessen. Welche Methode sollte er verwenden, um zu prüfen, ob sich die Programme signifikant in ihrer Wirkung unterscheiden? (a) t-Test, (b) ANOVA, (c) Lineare Regression.

#### **Lösungen zu Klickerfragen:**

Frage 1: (a) ANOVA; Frage 2: (b) Zu überprüfen, ob die Mittelwerte mehrerer Gruppen gleich sind; Frage 3: (b) Wenn es mehr als zwei Gruppen gibt; Frage 4: (a) Alle Gruppen haben den gleichen Mittelwert; Frage 5: (a) Es gibt vermutlich einen signifikanten Unterschied zwischen den Gruppen; Frage 6: (c) F-Verteilung; Frage 7: (a) Die Beziehung zwischen einer unabhängigen und einer abhängigen Variable zu modellieren; Frage 8: (b)  $y = \beta_0 + \beta_1x + \epsilon$ ; Frage 9: (a) Die Steigung der wahren Regressionsgeraden; Frage 10: (b) Lineare Regression; Frage 11: (b) Die Parameter einer linearen Funktion so zu bestimmen, dass die Summe der quadrierten Abweichungen der vorhergesagten Werte von den beobachteten Werten minimiert wird; Frage 12: (c) Mit jedem zusätzlichen Punkt in den bungsbilättern steigt die erwartete Punktzahl um 0.5 Punkte; Frage 13: (c) Mit jeder zusätzlichen Werbeanzeige steigt der erwartete Umsatz um 3'000 CHF; Frage 14: (b) ANOVA.