Grundlagen der Mathematik II
FS 2015 – Woche 14

Marcel Dettling

Institut für Datenanalyse und Prozessdesign

Zürcher Hochschule für Angewandte Wissenschaften

marcel.dettling@zhaw.ch

http://stat.ethz.ch/~dettling

ETH Zürich, 27. Mai 2015

Grundlagen der Mathematik II
FS 2015 – Woche 14

Regression

Beispiel:

In Indien behindern basische Böden Pflanzen beim Wachstum. Es werden daher Baumarten gesucht, die eine hohe Toleranz gegen solche Umweltbedingungen haben. In einem Freilandversuch wurden auf einem Feld mit grossen lokalen Schwankungen des pH-Wert 120 Bäume einer bestimmten Art gepflanzt. Nach 3 Jahren wurde von jedem Baum die Höhe gemessen. Gleichzeitig war auch der pH-Wert des Bodens an der entsprechenden Stelle bekannt. Die Daten können in einem Scatterplot dargestellt werden.

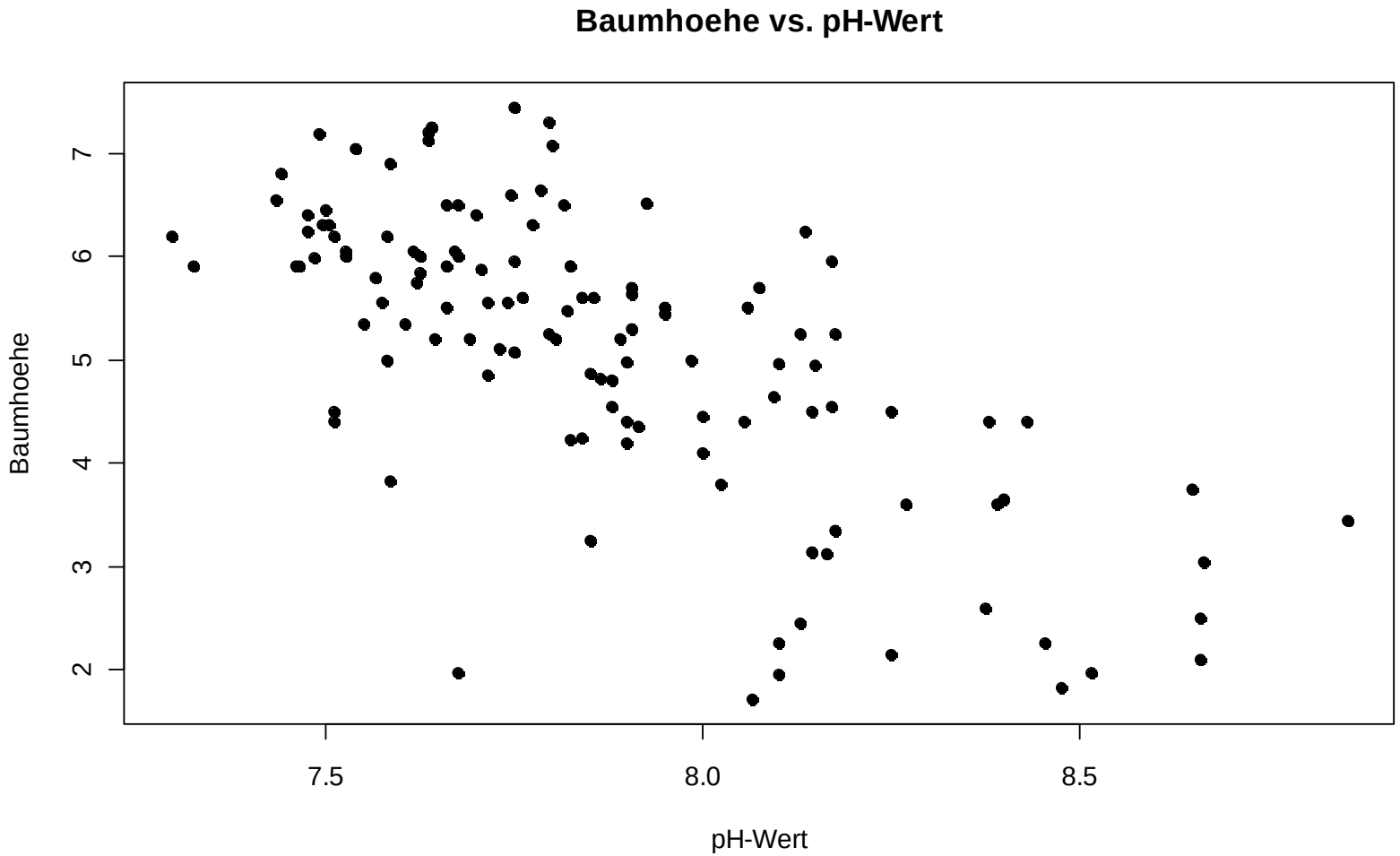
Grundlagen der Mathematik II
FS 2015 – Woche 14

Ausschnitt aus der Daten-Tabelle

Baum	Höhe	pH	SAR
1	5.91	7.325	0.0969
2	5.20	7.690	0.4393
3	4.40	7.900	1.0000
4	4.50	8.145	1.3160
5	6.05	7.615	0.0607
6	6.00	7.525	0.2041
7	…	…	…

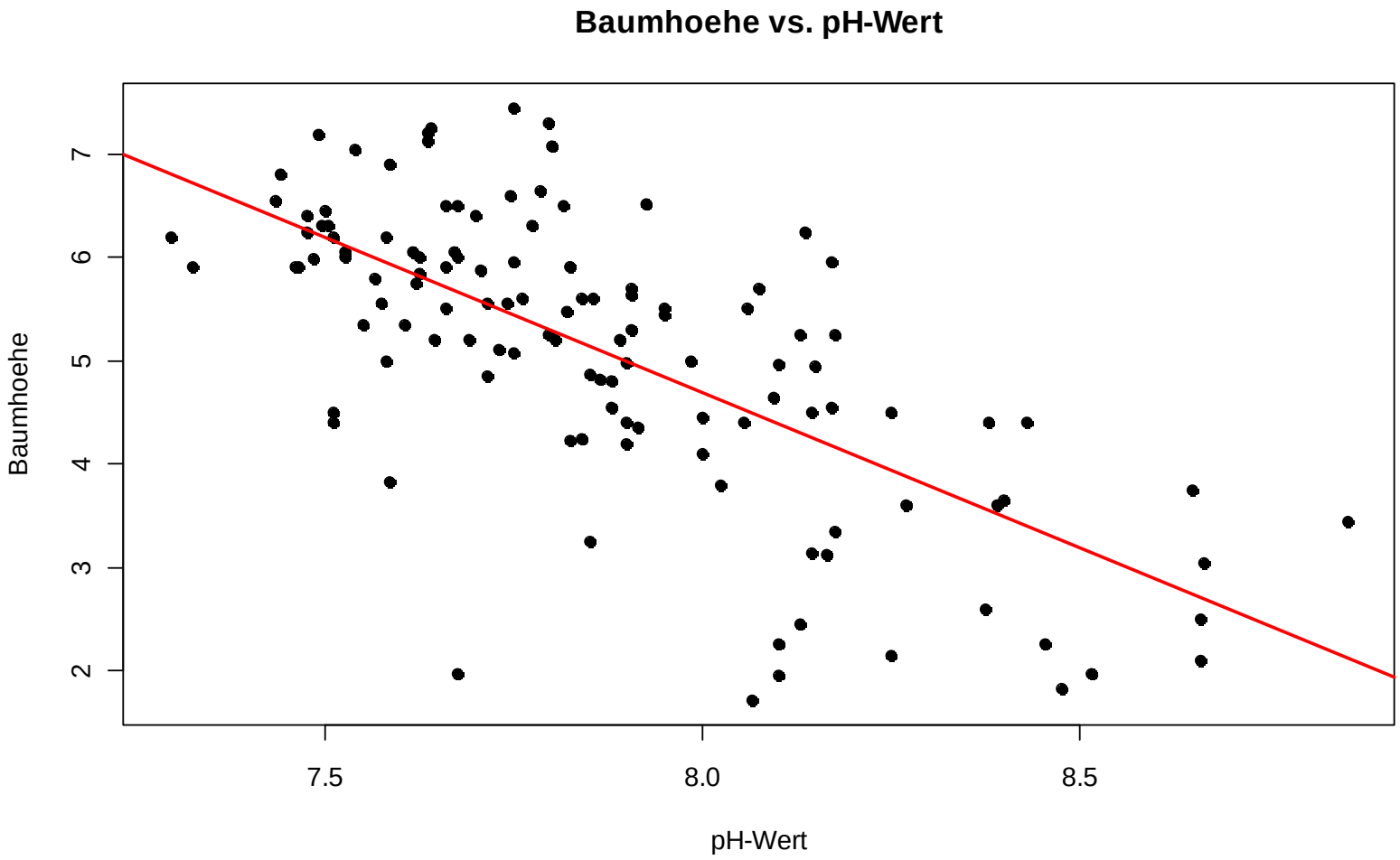
Grundlagen der Mathematik II
FS 2015 – Woche 14

Scatterplot Baumhöhe vs. pH-Wert



Grundlagen der Mathematik II
FS 2015 – Woche 14

Regressionsgerade



Einfache lineare Regression

Mit zunehmendem pH-Wert nimmt die Baumhöhe tendenziell ab. Der Zusammenhang scheint linear. Es bietet sich also an, eine Gerade zur Beschreibung zu verwenden:

$$f(x) = \beta_0 + \beta_1 x, \text{ bzw. } Höhe = \beta_0 + \beta_1 \cdot pH$$

Name/Bedeutung der Grössen in der Geradengleichung:

$\beta_0 =$ "Intercept"

$\beta_1 =$ "Slope"

Die Anpassung einer Geraden in einen 2-dimensionalen Scatterplot heisst **einfache lineare Regression,** weil:

- es nur eine erklärende Grösse gibt ("*einfach*").
- wir ein linearen Zusammenhang haben ("*linear*").

Modell & Zufallsfehler

Nun bringen wir die Daten ins Spiel. Die Gerade führt nicht durch jeden Datenpunkt, d.h. es gibt (zufällige) Abweichungen:

$$y_i = \beta_0 + \beta_1 x_i + E_i, \quad \text{für alle } i = 1, \ldots, n$$

Bedeutung der Grössen:

y_i	ist die Zielvariable (Baumhöhe) der i-ten Beobachtung.
x_i	ist die erklärende Grösse (pH) der i-ten Beobachtung.
β_0, β_1	sind die Regressionskoeffizienten, welche erst noch aus den Daten bestimmt/geschätzt werden müssen.
E_i	ist der zufällige Rest oder Fehler, d.h. die zufällige Abweichung zwischen Beobachtung und Gerade.

Grundlagen der Mathematik II
FS 2015 – Woche 14

Anpassung der Geraden

Einfache lineare Regression - was ist die Aufgabe???

Gesucht ist eine Gerade, die möglichst gut "zu den Daten passt". Wir müssen also β_0, β_1 so festlegen, dass die Abweichungen zwischen Datenpunkten und Gerade klein sind...

Intuitives Hineinlegen in den Scatterplot

→ **Welche Kriterien benützt man dabei...**

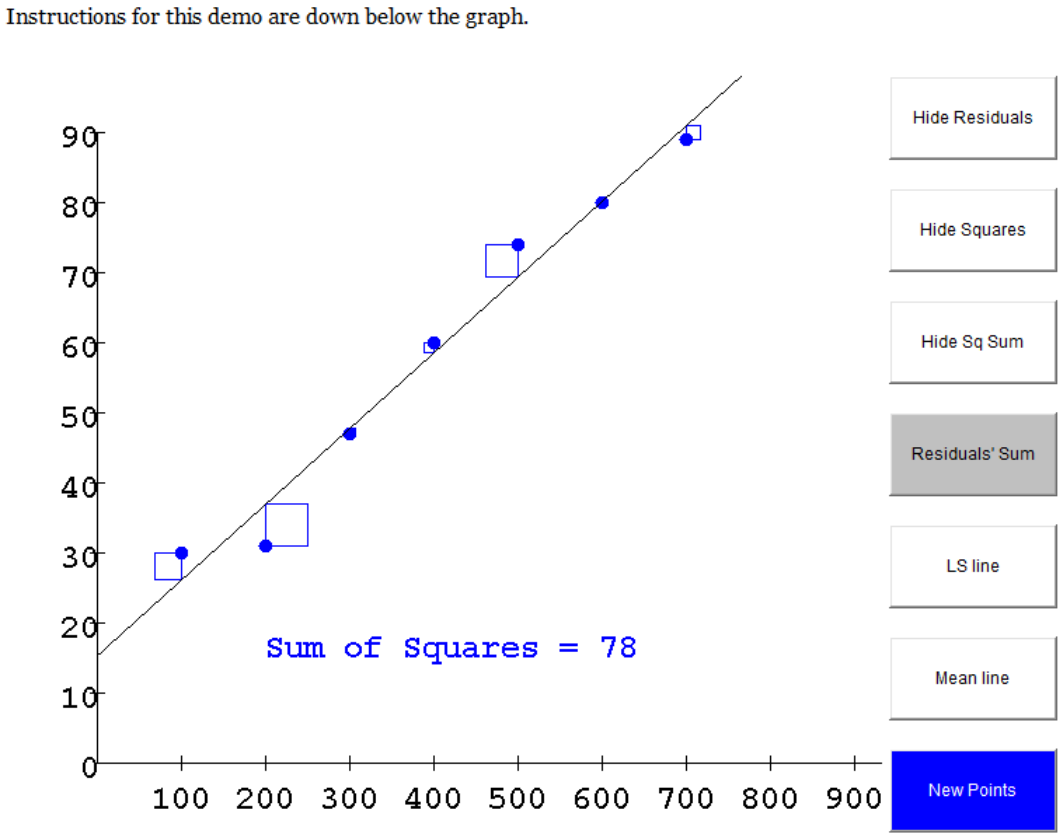
Diskussion und Erklärung

→ **Plenum / Wandtafel ...**

Grundlagen der Mathematik II
FS 2015 – Woche 14

Kleinste Quadrate: Applet

→ http://demonstrations.wolfram.com/LeastSquaresCriteriaForTheLeastSquaresRegressionLine/



Wir müssen eine Gerade durch die Punkte legen.

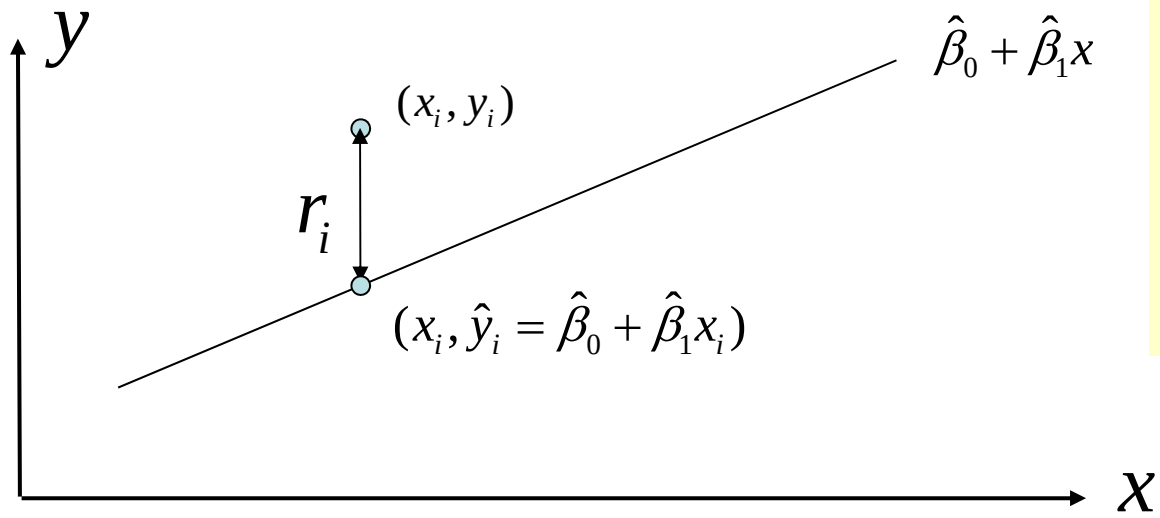
Es gibt viele Lösungen. Einige sind "gut", andere sind weniger geeignet.

Unser Paradigma: die Summe der Fehlerquadrate soll minimal sein!

Residuen vs. Fehler

Das Residuum $r_i = y_i - \hat{y}_i$ ist die Differenz zwischen dem beobachteten und dem angepassten y-Wert für den i-ten Datenpunkt. Achtung: der Fehler E_i ist ein Konzept und eine Zufallsvariable, das Residuum r_i ist ein numerischer Wert.

Illustration der Residuen



Wir bestimmen die Gerade so, dass die Quadratsumme der Residuen möglichst klein ist: $\sum_{i=1}^{n} r_i^2$

Kleinste Quadrate: Mathematisch

In Worten / Mathematisch...

Durch eine Punktwolke von Daten $(x_i, y_i)_{i=1,...,n}$ soll die Gerade so gelegt werden, dass die Summe der quadrierten Abstände r_i zwischen dem Beobachtungswert y_i und dem zugehörigen Punkt auf der Geraden $\hat{y}_i$ minimal ist. Die Funktion

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 = \min!$$

misst, wie gut die durch (β_0, β_1) definierte Gerade zu den Daten passt. Sie soll einen möglichst kleinen Wert annehmen.

Lösung: → **siehe nächste Folie...**

Lösungsidee: Partielle Ableitungen

- Wir leiten die Funktion $Q(\beta_0, \beta_1)$ partiell nach den beiden Argumenten β_0 und β_1 ab, und setzen die Ableitungen gleich null:

$$\frac{\partial Q}{\partial \beta_0} = 0 \quad \text{und} \quad \frac{\partial Q}{\partial \beta_1} = 0$$

- Es entsteht ein lineares Gleichungssystem, mit (hier) zwei Unbekannten β_0, β_1 und zwei Gleichungen. Diese Gleichungen heissen *Normalengleichungen.*

- Man kann die Lösung für β_0, β_1 *explizit* als Funktion der Datenpaare $(x_i, y_i)_{i=1,\ldots,n}$ aufschreiben, siehe nächste Folie...

Kleinste Quadrate: Lösung

Die gemäss Kleinsten Quadraten optimale Lösung für die Gerade, d.h. die aus den Daten geschätzten Regressionskoeffizienten sind:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{und} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- Für eine gegebene Punktwolke $(x_i, y_i)_{i=1,\ldots,n}$ können wir also die Gerade (mit dem TR, besser mit R) bestimmen.

- **Ergebnis für unser Beispiel "Baumhöhe":**
  $\hat{\beta}_1 = 28.7,\ \hat{\beta}_0 = -3.0$ mit Hilfe von Softwarepaket gerechnet!

→ **Probieren sie es selber aus (in den Übungen)...**

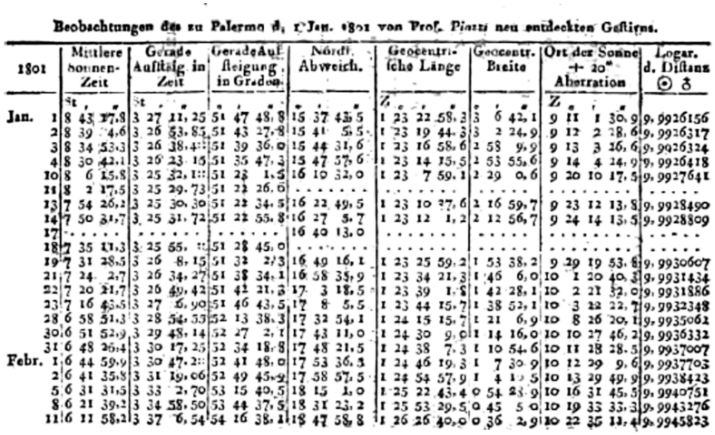
Grundlagen der Mathematik II
FS 2015 – Woche 14

Anhang: warum Kleinste Quadrate?

Historisches...

Die Methode wurde innert weniger Jahre (1801, 1805) zweimal unabhängig voneinander zum Lösen von Problemen in der Astronomie entwickelt...

Quelle: → http://de.wikipedia.org/wiki/Methode_der_kleinsten_Quadrate



Beobachtungen des zu Palermo d. 1. Jan. 1801 von Prof. Piazzi neu entdeckten Gestirns.

1801		Mittlere Sonnen-Zeit	Gerade Aufsteig. in Zeit	Gerade Aufsteigung in Graden	Nördl. Abweich.	Geocentrische Länge	Geocentr. Breite	Ort der Sonne + 20" Aberration	Logar. d. Distanz ☉ ♃
		♎ ′ ″	♎ ′ ″	° ′ ″	° ′ ″	♈ ° ′ ″	° ′ ″	♈ ° ′ ″	
Jan.	1	8 43 37,8	3 27 11, 25	51 47 48, 8	15 37 42, 5	1 23 22 58, 2	3 6 42, 1	9 11 1 30, 9	9, 9926156
	2	8 39 4,6	3 26 53, 85	51 43 27, 8	15 41 5, 5	1 23 19 44, 2	3 2 24, 9	9 12 2 28, 6	9, 9926317
	3	8 34 32,3	3 26 38, 4	51 39 36, 0	15 44 31, 6	1 23 16 58, 6	2 58 9, 9	9 13 3 26, 6	9, 9926324
	4	8 30 0,1	3 26 23, 15	51 35 47, 3	15 47 57, 6	1 23 14 15, 5	2 53 55, 6	9 14 4 24, 9	9, 9926418
	10	8 6 15,8	3 25 32, 1	51 25 1, 5	16 10 32, 0	1 23 7 59, 1	2 29 0, 6	9 20 10 17, 5	9, 9927641
	11	8 2 17,5	3 25 29, 73	51 23 26, 0	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
	13	7 54 26,2	3 25 30, 30	51 22 34, 5	16 22 49, 5	1 23 10 27, 6	2 16 59, 7	9 23 12 13, 8	9, 9928490
	14	7 50 31,7	3 25 31, 72	51 22 55, 8	16 27 5, 7	1 23 12 1, 2	2 12 56, 7	9 24 14 13, 5	9, 9928809
	17	. . . . .	. . . . .	. . . . .	16 40 13, 0	. . . . .	. . . . .	. . . . .	. . . . .
	18	7 35 13,3	3 25 55, 1	51 28 45, 0	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
	19	7 31 28,5	3 26 8, 15	51 31 2, 3	16 49 16, 1	1 23 25 59, 2	1 53 38, 2	9 29 19 53, 8	9, 9930607
	21	7 24 2,7	3 26 34, 27	51 38 34, 1	16 58 35, 9	1 23 34 21, 3	1 45 6, 0	10 1 20 40, 3	9, 9931434
	22	7 20 21,7	3 26 49, 42	51 42 21, 3	17 3 18, 5	1 23 39 1, 8	1 42 28, 1	10 2 21 32, 0	9, 9931886
	23	7 16 45,6	3 27 6, 90	51 46 43, 5	17 8 5, 5	1 23 44 15, 7	1 38 52, 1	10 3 22 22, 7	9, 9932348
	28	6 58 51,3	3 28 54, 53	52 13 38, 3	17 32 54, 1	1 24 15 15, 7	1 21 6, 9	10 8 26 20, 1	9, 9935061
	30	6 51 52,9	3 29 48, 14	52 27 2, 1	17 43 11, 0	1 24 30 9, 0	1 14 16, 0	10 10 27 36, 2	9, 9936332
	31	6 48 26,4	3 30 17, 25	52 34 18, 8	17 48 21, 5	1 24 38 7, 3	1 10 54, 6	10 11 28 28, 5	9, 9937007
Febr.	1	6 44 59,9	3 30 47, 2	52 41 48, 0	17 53 36, 2	1 24 46 19, 3	1 7 30, 9	10 12 29 9, 6	9, 9937703
	2	6 41 35,8	3 31 19, 66	52 49 45, 9	17 58 57, 5	1 24 54 57, 9	1 4 1, 5	10 13 29 49, 9	9, 9938423
	5	6 31 31,5	3 33 2, 70	53 15 40, 5	18 15 1, 0	1 25 22 43, 4	0 54 23, 9	10 16 31 45, 5	9, 9940751
	8	6 21 39,2	3 34 58, 50	53 44 37, 5	18 31 23, 2	1 25 53 29, 5	0 45 5, 0	10 19 33 33, 3	9, 9943276
	11	6 11 58,2	3 37 6, 54	54 16 38, 2	18 47 58, 8	1 26 26 40, 0	0 36 2, 9	10 22 35 12, 4	9, 9945823

Carl Friedrich Gauss

Adrien-Marie Legendre

Grundlagen der Mathematik II
FS 2015 – Woche 14

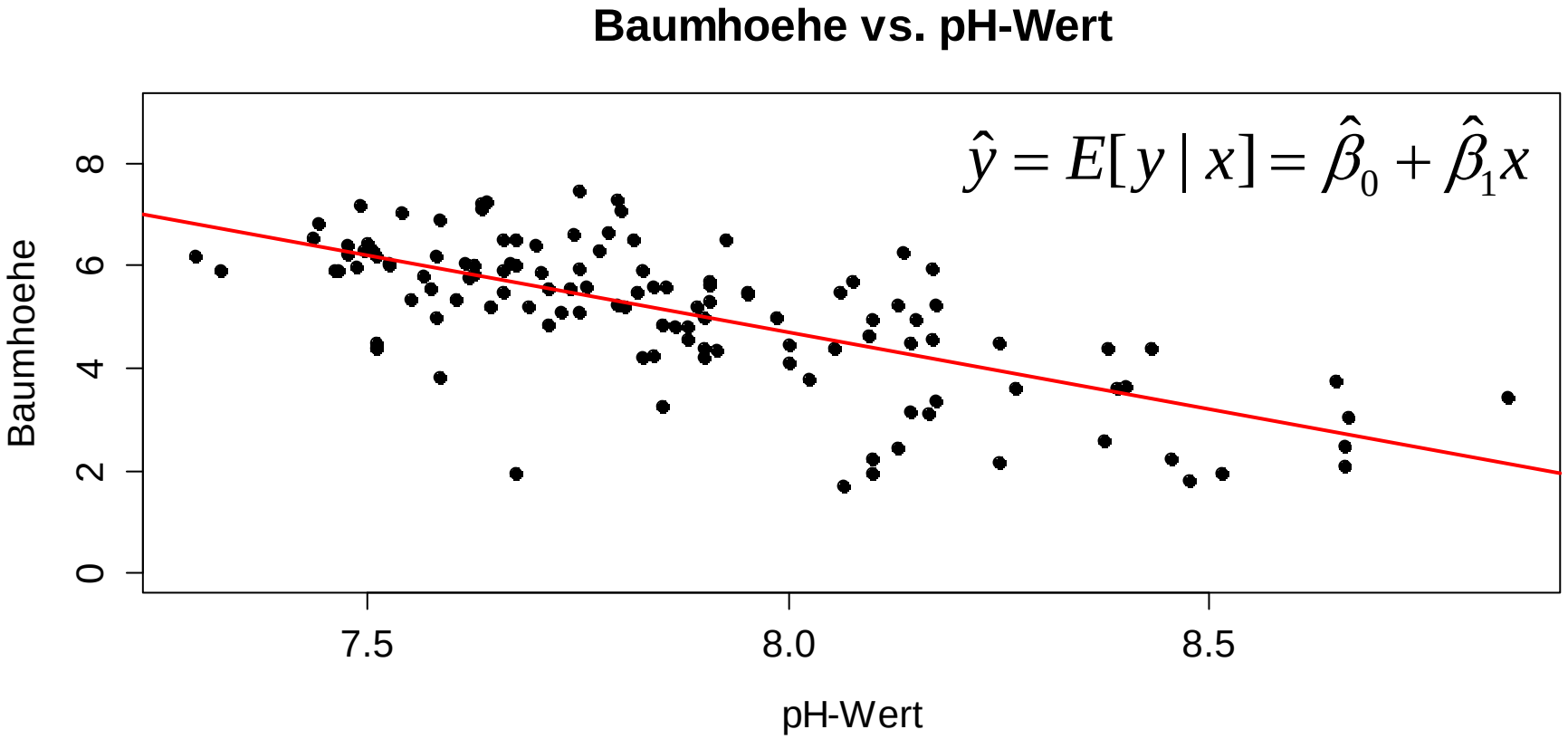
Anhang: warum Kleinste Quadrate?

Mathematisches...

- Das Verfahren ist einfach in dem Sinne, dass die Lösung explizit als Funktion von $(x_i, y_i)_{i=1,\ldots,n}$ bekannt ist.
- Die Gerade geht durch den Daten-Schwerpunkt $(\overline{x}, \overline{y})$
- Die Summe der Residuen addiert sich zu null: $\sum_{i=1}^{n} r_i = 0$
- Tiefer gehende, mathematische Optimalität lässt sich beweisen, indem man die Schätzeigenschaften von $\hat{\beta}_0, \hat{\beta}_1$ untersucht, speziell bei normalverteilten Fehlern E_i.

Fitted Values und Regressionsgerade

Die geschätzten Parameter $\hat{\beta}_0, \hat{\beta}_1$ können wir nun benützen, um die *angepassten Werte* $\hat{y}$ (engl. *Fitted Values*) anzugeben. Es handelt sich um einen bedingten Erwartungswert:



Grundlagen der Mathematik II
FS 2015 – Woche 14

Haben wir ein gutes Modell für die Baumhöhen-Vorhersage gefunden?

a) Ausserhalb der Punktwolke

Unklar, ziemlich sicher nein...

b) Innerhalb der Punktwolke?

Ja, unter den folgenden, zu prüfenden Bedingungen

- der Zusammenhang ist eine Gerade ist, d.h. $E[E_i] = 0$
- die Streuung der Fehler konstant ist, d.h. $Var(E_i) = \sigma^2$
- die Fehler unkorreliert sind (repräsentative Stichprobe!)
- die Fehler (approximativ) normalverteilt sind

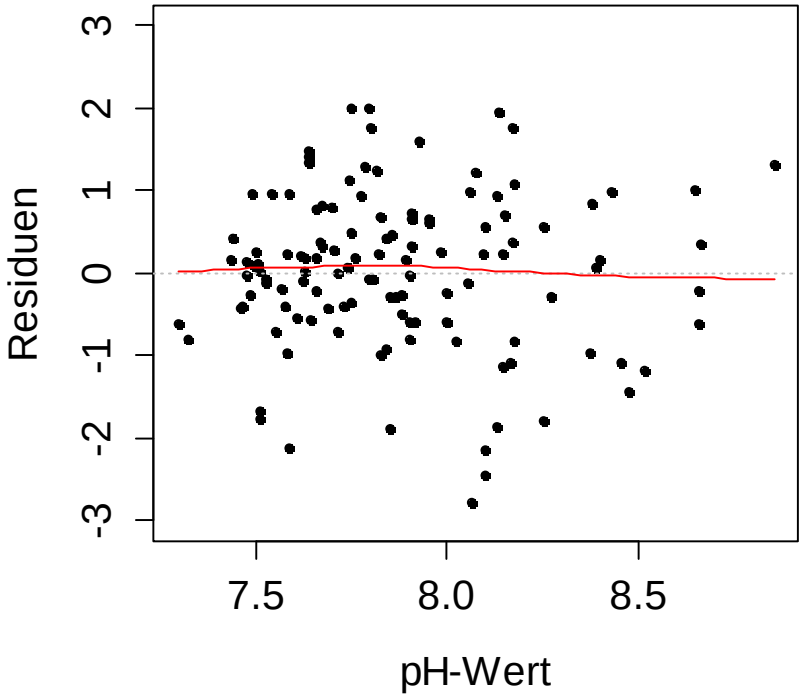
→ **Gedankenfutter: Schattige Ecke auf dem Feld?**

Grundlagen der Mathematik II
FS 2015 – Woche 14

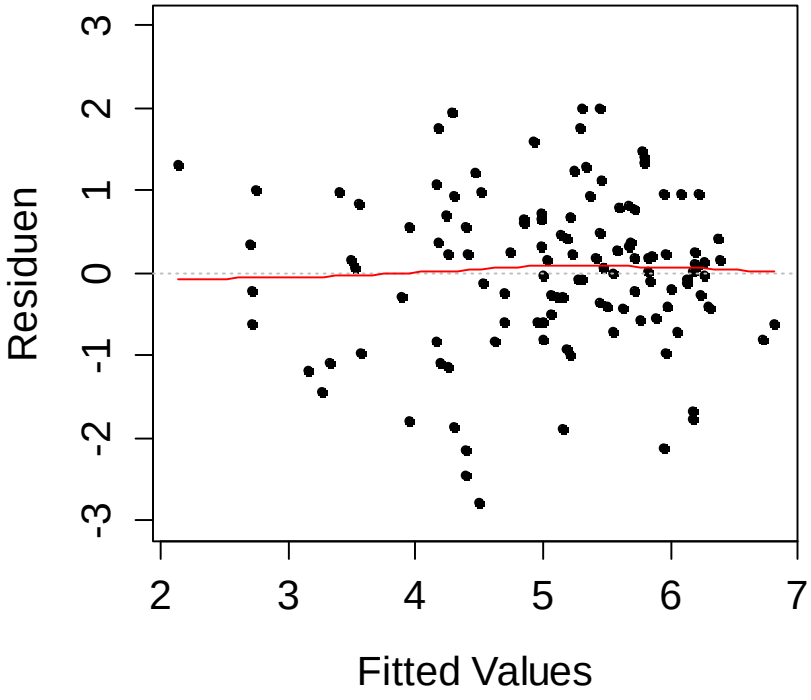
Modelldiagnostik

Um die Vertrauenswürdigkeit der Regressionsgerade zu evaluieren, müssen die getroffenen Annahmen überprüft werden. Für $E[E_i] = 0$ und $Var(E_i) = \sigma^2$ betrachten wir:

Residuen vs. Prädiktor



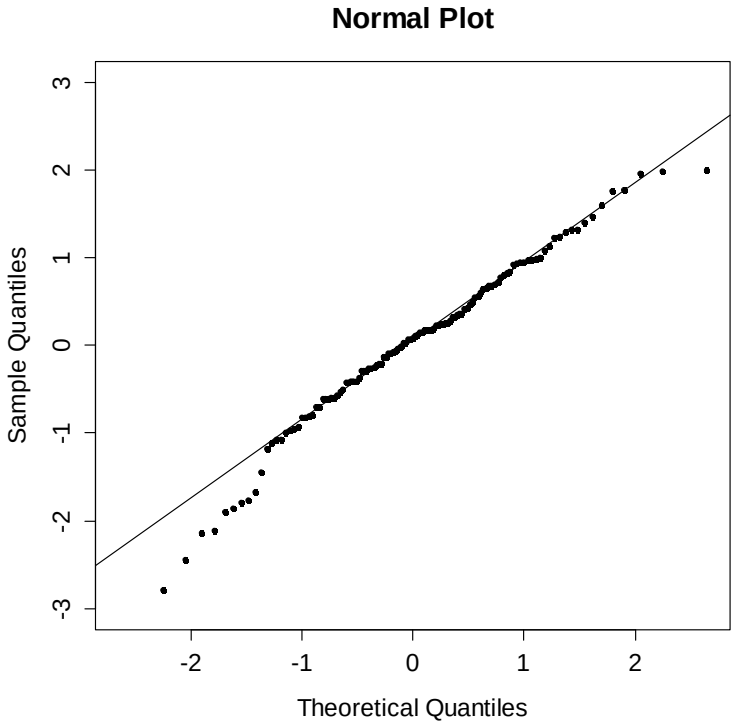
Tukey-Anscombe-Plot



Grundlagen der Mathematik II
FS 2015 – Woche 14

Modelldiagnostik

Um die Vertrauenswürdigkeit der Regressionsgerade zu evaluieren, müssen die getroffenen Annahmen überprüft werden. Für die Normalverteilung betrachten wir:



Es gibt auch noch weitere, verfeinerte Diagnoseplots.

Diese bespricht man in der Regel erst bei der multiplen linearen Regression.

Eigenschaften der Schätzer

Die KQ-Schätzer sind erwartungstreu, d.h.

$$E[\hat{\beta}_0] = \beta_0 \text{ und } E[\hat{\beta}_1] = \beta_1$$

Die Varianzen der Schätzer sind wie folgt:

$$Var(\hat{\beta}_0) = \sigma_E^2 \cdot \left(\frac{1}{n} + \frac{\bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) \text{ und } Var(\hat{\beta}_1) = \frac{\sigma_E^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Präzise Schätzungen erhält man durch:

- eine grosse Anzahl Beobachtungen n
- eine ausreichende Streuung der x_i
- einen informativen Prädiktor, so dass σ_E^2 klein ist

Schätzen der Fehlervarianz σ_E^2

Neben den Regressionskoeffizienten ist auch eine Schätzung der Fehlervarianz σ_E^2 von Interesse. Sie ist ein wichtiger Input für alle Tests und Konfidenzintervalle, die wir besprechen:

Die Schätzung basiert auf der *residual sum of squares* (RSS):

$$\hat{\sigma}_E^2 = \frac{1}{n-2} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n-2} \cdot \sum_{i=1}^{n} r_i^2$$

In unserem Beispiel ergibt sich als `Residual standard error`:

```
> summary(fit)
...
Residual standard error: 1.008 on 121 degrees of freedom
```

Nutzen von Regression

1) Untersuchen der Beziehung zwischen y und x

Die Absicht ist, genau zu verstehen, wie und wie stark die Zielvariable vom Prädiktor abhängt. Es gibt diverse Kenngrössen und statistische Tests, welche sich dieser Frage widmen.

2) Vorhersage

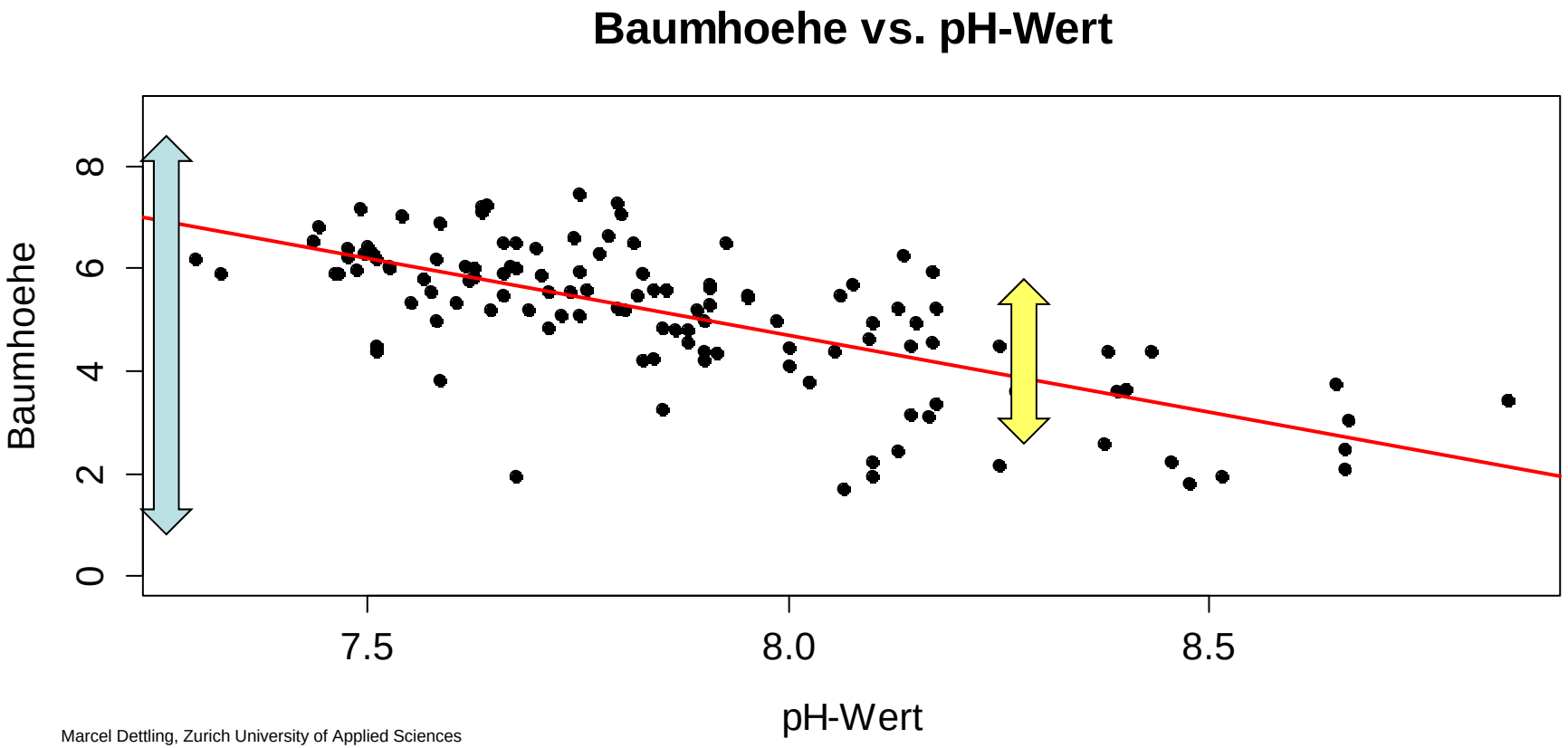
Wir können die Regressionsgleichung, bzw. –gerade benützen, um für einen beliebigen ph-Wert die Baumhöhe anzugeben.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Grundlagen der Mathematik II
FS 2015 – Woche 14

R^2: *Erklärungsgehalt der Regressionsgerade*

Intuitiv: je grösser der blaue Pfeil im Vergleich zum gelben ist, desto grösser ist der Erklärungsgehalt der Regressionsgerade



Das Bestimmtheitsmass R^2

Der Erklärungsgehalt der Regressionsgeraden wird mit R^2 gemessen. Man nennt R^2 das *Bestimmtheitsmass*, bzw. englisch *Coefficient of Determination*. Es handelt sich um das Verhältnis zwischen dem gelben und dem blauen Pfeil. Es ist der Anteil der Gesamtstreuung, welche durch die Gerade erklärt wird.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \in [0,1]$$

Je grösser R^2, desto enger streuen die Punkte um die Gerade. Es gibt aber kein formelles Kriterium, wie gross R^2 sein muss.

Vertrauensintervall für die Steigung β_1

Das 95%-VI für die Steigung β_1 ist um die Punktschätzung $\hat{\beta}_1$ zentriert und enthält alle Werte die ebenfalls plausibel sind. Die Unsicherheit ist durch die Streuung der Datenpunkte bedingt..

95%-VI für β_1 : $\hat{\beta}_1 \pm qt_{0.975;n-2} \cdot \hat{\sigma}_{\hat{\beta}_1}$, bzw.

$$\hat{\beta}_1 \pm qt_{0.975;n-2} \cdot \sqrt{\hat{\sigma}_E^2 \Big/ \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Salopp: $-3 \pm 2 \cdot 0.28 = [-3.56; -2.44]$

Exakt: $[-3.566353; -2.440355]$ aus Statistikpaket

Test für die Steigung β_1

Es gibt einen statistischen Test, mit welchem man feststellen kann, ob die Steigung der Regressionsgerade signifikant von null oder einem beliebigen anderen Wert b verschieden ist:

$$H_0 : \beta_1 = 0, \text{ bzw. } H_0 : \beta_1 = b$$

Man testet zweiseitig auf dem 95%-Niveau. Die Alternative ist:

$$H_A : \beta_1 \neq 0, \text{ bzw. } H_A : \beta_1 \neq b$$

Als Teststatistik verwenden wir:

$$T_{H_0:\beta_1=0} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}, \text{ bzw. } T_{H_0:\beta_1=b} = \frac{\hat{\beta}_1 - b}{\hat{\sigma}_{\hat{\beta}_1}}, \text{ beide haben } t_{n-2}\text{-Verteilung.}$$

Grundlagen der Mathematik II
FS 2015 – Woche 14

Lesen von Output

```
> summary(fit)

Call: lm(formula = height ~ ph, data = dat)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.7227     2.2395   12.82   <2e-16 ***
ph           -3.0034     0.2844  -10.56   <2e-16 ***
---
Residual standard error: 1.008 on 121 degrees of freedom
Multiple R-squared:  0.4797, Adjusted R-squared:  0.4754
F-statistic: 111.5 on 1 and 121 DF,   p-value: < 2.2e-16
```

→ **Man beachte die Bedeutung der Grössen!**

Test für die Steigung β_1

Praxisbeispiel:

Man nehme die Baumhöhen-Daten und überprüfe mit einem Test die Hypothese $H_0 : \beta_1 = -2$. Die Informationen von Folie 28 dürfen verwendet werden. Man beantworte auch:

a) *Formulieren sie umgangssprachlich, was sie eben getestet haben und was einem dies in der Praxis nützt.*

b) *Worin liegt der Zusammehang zwischen dem Testresultat und dem 95%-VI von Folie 25? Hätten wir das Testresultat schon vom VI vorhersehen können?*

→ **Siehe Wandtafel…**

Test für den Achsenabschnitt β_0

Für den Achsenabschnitt gibt es analoge Tests.

- Egal wie das Testresultat ausfällt, den Achsenabschnitt soll man stets im Regressionsmodell belassen!
- Der Achsenabschnitt bietet Schutz vor Nichtlinearität und Kalibrationsfehlern. Wird er weggelassen, so sind die Resultate für die Praxis meist weniger brauchbar.
- Falls einem (physikalische) Theorie diktiert, dass es keinen Achsenabschnitt geben darf, er aber trotzdem signifikant ist, dann heisst das, dass die lineare Beziehung nicht bis zum Punkt $x = 0$ extrapoliert werden darf.

Vorhersage

Mit der Regressionsgerade können wir den y-Wert an beliebiger x-Stelle vorhersagen. Wir benützen die Gleichung:

$$E[y \mid x] = \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x, \text{ a.k.a } \textbf{\textit{“fitted value”}}$$

Beispiel: Für einen pH-Wert von 8.0 erwarten wir:

$$28.7 - 3.0 \cdot 8.0 = 4.7 \text{ Meter Baumhöhe}$$

Aber Achtung:

Interpolation im Bereich der beobachteten x-Werte ist i.d.R. problemlos. Extrapolation (z.B. für pH-Werte von 1 oder 10) funktioniert in der Regel nicht und ist gefährlich!

Vertrauensintervall für $E[y \mid x]$

Wir haben gelernt, wie man den Fitted Value $\hat{\beta}_0 + \hat{\beta}_1 x$ bestimmt, d.h. die erwartete Baumhöhe für gegebenen pH-Wert. Achtung, es handelt sich um eine Schätzung mit Unsicherheit.

Ein 95%-VI für den angepassten Wert an der Stelle x ist:

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm qt_{0.975;n-2} \cdot \hat{\sigma}_E \cdot \sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}$$

Die Werte können mit Statistik-Software berechnet werden.

```
Fitted   Lower    Upper
4.695861 4.501321 4.8904
```

Prognoseintervall für y

Das 95%-VI für $E[y \mid x]$ zeigt die Variabilität des Fitted Value. Es beinhaltet jedoch nicht die Streuung der Daten um die Gerade und definiert darum nicht die Region, in der ein zukünftiger Datenpunkt zu liegen kommt. Ein 95%-Prognoseintervall an der Stelle x ist gegeben durch:

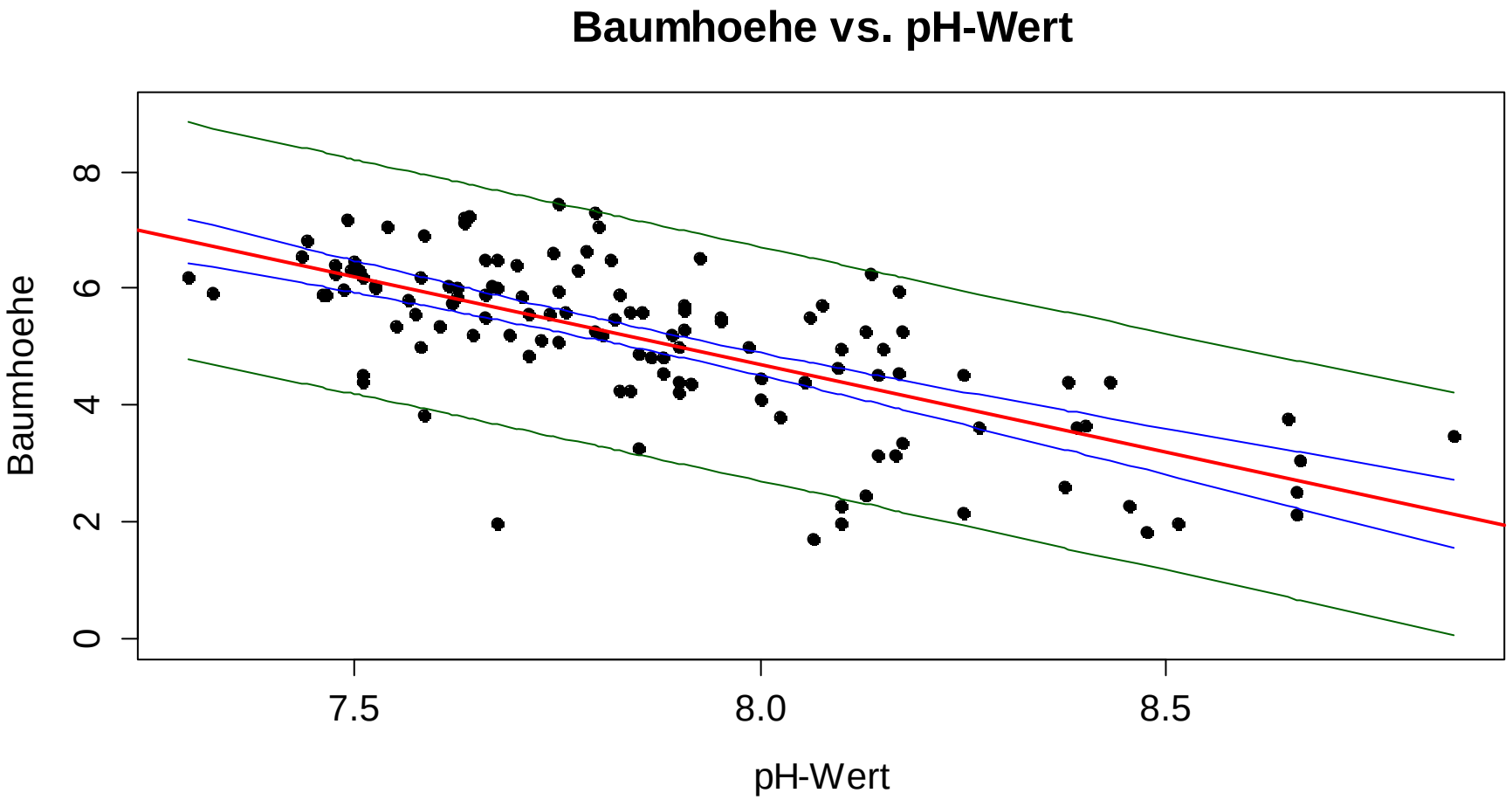
$$\hat{\beta}_0 + \hat{\beta}_1 x \pm qt_{0.975;n-2} \cdot \hat{\sigma}_E \cdot \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

Konkret:

```
Fitted    Lower     Upper
4.695861 2.690581 6.70114
```

Grundlagen der Mathematik II
FS 2015 – Woche 14

Vertrauens- und Vorhersagebereich



Ausblick: Multiple Regression

In der realen Welt wird die Zielgrösse meist von mehreren Prädiktoren gleichzeitig beeinflusst. Es lohnt sich also, den multiplen Zusammenhang zu studieren. Das Modell lautet:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + E_i$$

Aufgabe ist wiederum, die Koeffizienten $\beta_0, \beta_1, \ldots, \beta_p$ aus den zur Verfügung stehenden Daten zu schätzen. Hinweis: das Resultat fällt dabei im Allg. anders aus, wie wenn man mehrere einfache Regression von der Zielgrösse gegen jeden Prädiktor separat ausführt!

→ Zur Schätzung benützt man weiterhin die KQ-Methode.

Grundlagen der Mathematik II
FS 2015 – Woche 14

Ausblick: Multiple Regression

Output aus Statistikpaket:

```
Call: lm(height ~ ph + l.sar, data = dat)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.9466     2.7445   9.818  < 2e-16 ***
ph           -2.7558     0.3603  -7.649  5.6e-12 ***
l.sar        -0.2519     0.2255  -1.117    0.266
---
Residual standard error: 1.007 on 120 degrees of freedom
Multiple R-squared:  0.485,  Adjusted R-squared:  0.4764
F-statistic: 56.51 on 2 and 120 DF,   p-value: < 2.2e-16
```

Grundlagen der Mathematik II
FS 2015 – Woche 14

Informationen zur Prüfung

- Dauer: Neu 90min, je ~2 Aufgaben aus LinAlg und Statistik
- Stoff: alles, was in Vorlesung und Übungen vorkam
- Aufgaben: Transferleistung nötig, Verständnis wichtig
- Es sind beliebige schriftliche Hilfsmittel erlaubt
- Taschenrechner sind verboten!
- Ferienpräsenz: siehe Webpage

→ **Schöne Ferien und viel Erfolg bei der Vorbereitung!**