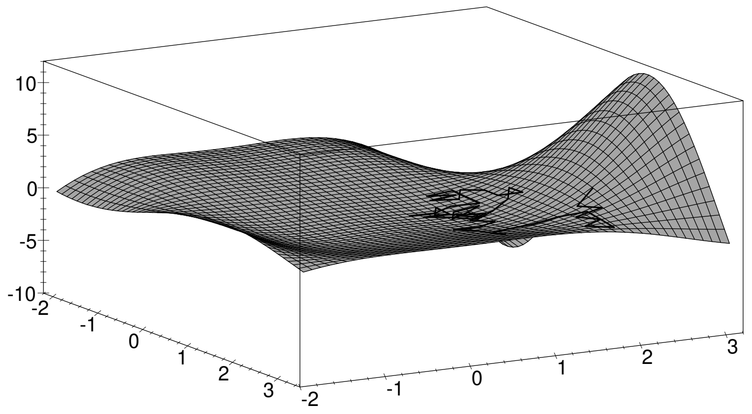


Abb. 53: Ditto mit anderen Zufallszahlen

wird es am besten sein, nach hinreichend vielen METROPOLIS-Schritten einfach ein gewöhnliches Gradientenverfahren zu starten.

Zusammenfassend läßt sich sagen, daß der METROPOLIS-Algorithmus und verwandte Verfahren (die sogenannten Monte-Carlo-Methoden) sehr nützliche Hilfsmittel zur Optimierung sind, falls man so gut wie nichts über die zu optimierende Funktion weiß. Sie funktionieren nicht nur bei kontinuierlichen Problemen, wie den hier betrachteten, sondern auch für diskrete und kombinatorische Optimierungsprobleme.

Sie haben allerdings den Nachteil, daß man nie garantieren kann, daß man ein Optimum erreichen wird, und selbst wenn man eines erreicht, kann die Methode dies nicht erkennen. (Es gibt alternative numerische Methoden, die das können.)

c) Zusammenfassung

Die nichtlineare Optimierung ist ein sehr weites Feld, von dem eine Grundvorlesung wie die *Höhere Mathematik* nur einen kleinen Ausschnitt behandeln kann. Dieser Ausschnitt besteht nicht aus den für die Praxis wichtigsten Verfahren, sondern aus denen, die sich am besten in den Stoff der Vorlesung einordnen. Sie sind zwar (in Kombination mit

dem aus der *Numerik* bekannten Simplex-Verfahren) die Grundbausteine, aus denen die meisten praktisch relevanten Verfahren zusammengesetzt sind, aber für die vielen kleinen Abwandlungen, die dazu führen, daß man ein Problem wirklich effizient lösen kann, müßte man deutlich mehr Zeit aufwenden, als hier zur Verfügung steht. Interessenten seien auf entsprechende Spezialvorlesung aus dem Bereich der Mathematik oder Operations Research verwiesen.

§4: Grundzüge der Fehler- und Ausgleichsrechnung

Physikalische Gesetze machen meist nur dann eine Aussage über ein reales System, wenn alle Umgebungsbedingungen exakt kontrolliert werden können. Das ist in der Praxis natürlich nie möglich. Insbesondere hat man bei der Anwendung physikalischer Prinzipen zur Messung von Daten keine Chance, den exakten Wert der zu messenden Größe zu bestimmen; der gemessene Wert wird immer von zahlreichen kleineren Störungen beeinflußt sein, die man bei einem gut durchgeführten Experiment für alle praktischen Zwecke als zufällig betrachten kann.

Zusätzlich kann die Messung noch durch mehr oder weniger große *systematische* Fehler verfälscht sein; diese können hervorgerufen werden durch ein falsch kalibriertes Meßgerät, Ablesen auf der falschen Skala eines Meßinstruments, durch falsche Anwendung von Meßvorschriften *usw.* Mit diesen systematischen Fehlern wollen wir uns hier nicht beschäftigen; in diesem Paragraphen soll es nur um *Zufallsfehler* gehen.

a) Das Laplacesche Fehlermodell

Der französische Mathematiker PIERRE SIMON, MARQUIS DE LAPLACE (1749–1827), dem wir in dieser Vorlesung bereits mehrfach begegnet sind, entwickelte ein extrem vereinfachtes Modell für das Zustandkommen zufälliger Meßfehler. Trotz seiner unrealistischen Annahmen ist es auch für die Praxis immer noch sehr interessant, da man inzwischen weiß, daß auch sehr viel kompliziertere realistische Fehlerquellen dasselbe Verhalten zeigen, das LAPLACE aus seinem Modell ableitete.

Die Grundannahme des LAPLACEschen Fehlermodells können wir uns so vorstellen, daß eine große Anzahl von „Dämonen“ (oder Fehlerquellen) unsere Meßergebnisse verfälschen; jeder einzelne dieser „Dämonen“ verursacht einen Fehler derselben Größe ε in positiver oder negativer Richtung, wobei die Wahrscheinlichkeit für $+\varepsilon$ *bzw.* $-\varepsilon$ für jeden der „Dämonen“ jeweils 50% sein soll und die einzelnen „Dämonen“ unabhängig voneinander handeln sollen.

Im Falle eines einzigen „Dämonen“ wäre der Fehler also mit gleicher Wahrscheinlichkeit $+\varepsilon$ oder $-\varepsilon$, bei zwei „Dämonen“ wäre er in jeweils 25% aller Fälle $+2\varepsilon$ oder -2ε, während sich in 50% der Fälle die beiden Fehler aufheben würden.

Allgemein gibt es bei n „Dämonen“ 2^n gleichwahrscheinliche Möglichkeiten für deren Verhalten; die folgende Tabelle zeigt für $n \le 5$ jeweils die Anzahl der Fälle, die zu dem in der Kopfzeile angegebenen Gesamtfehler führen:

	-5ε	-4ε	-3ε	-2ε	$-\varepsilon$	0	$+\varepsilon$	$+2\varepsilon$	$+3\varepsilon$	$+4\varepsilon$	$+5\varepsilon$
$n=0$						1					
$n=1$					1		1				
$n=2$				1		2		1			
$n=3$			1		3		3		1		
$n=4$		1		4		6		4		1	
$n=5$	1		5		10		10		5		1

Diese dreiecksförmige Anordnung von Zahlen bezeichnet man als PASCAL*sches Dreieck*. Offenbar kann man es dadurch rekursiv zeilenweise berechnen, daß man an jede Stelle die Summe der beiden links und rechts darüberstehenden Zahlen schreibt: Die n-te Störung bringt den Fehler genau dann auf $i\cdot\varepsilon$, wenn sie entweder gleich $+\varepsilon$ ist und die ersten $n-1$ Störungen einen Fehler $(i-1)\cdot\varepsilon$ produziert haben, oder aber wenn sie gleich $-\varepsilon$ ist und die ersten $n-1$ Störungen einen Fehler $(i+1)\cdot\varepsilon$ produziert haben. Entsprechend ist auch klar, daß die Summe aller Zahlen in der n-ten Zeile gleich 2^n ist, denn in der nullten Zeile haben wir Summe eins, und da die jeweils neu hinzukommende Störung genau zwei Möglichkeiten hat, verdoppelt sich die Summe von Zeile zu Zeile. Die Wahrscheinlichkeit dafür, daß sich n Störungen zu $i\cdot\epsilon$ aufsummieren, ist also gerade gleich der Zahl, die in der n-ten Spalte

unter $i\cdot\varepsilon$ steht (beziehungsweise Null, wenn dort keine Zahl steht), dividiert durch 2^n.

Bekanntlich kann man die Zahlen in diesem Dreieck auch explizit berechnen: An der n-ten Zeile stehen die $n+1$ Zahlen

$$\binom{n}{i} = \frac{n!}{i!(n-i)!} \quad \text{für} \quad i = 0,\dots,n\,.$$

Wer diese Formel nicht kennt, kann sie leicht durch vollständige Induktion beweisen: Für $n=1$ sowie allgemein für $i=0$ oder $i=n$ ist alles klar; für $n>1$ und $0<i<n$ stehen über $\binom{n}{i}$ die beiden Zahlen $\binom{n-1}{i-1}$ und $\binom{n-1}{i}$, für die in der Tat gilt

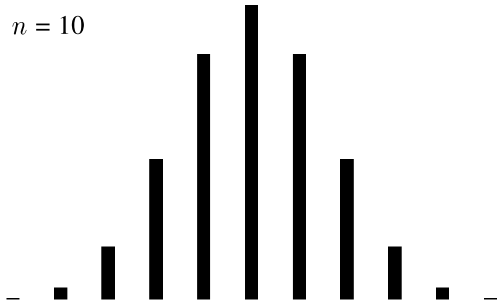
$$\begin{aligned}\binom{n-1}{i-1} + \binom{n-1}{i} &= \frac{(n-1)!}{(i-1)!(n-i)!} + \frac{(n-1)!}{i!(n-i-1)!} = \frac{(n-1)!}{i!(n-i)!}\cdot\Big(i+(n-i)\Big)\\ &= \frac{n!}{i!(n-i)!} = \binom{n}{i}\,.\end{aligned}$$

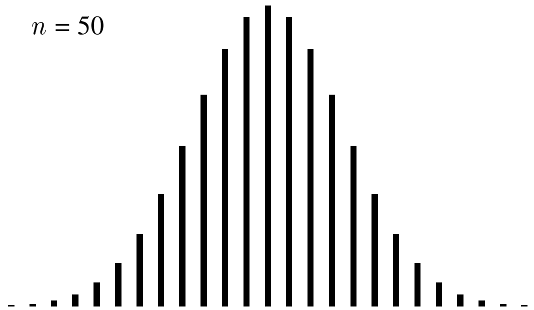
Mit dieser Formel lassen sich die Fallanzahlen für einen bestimmten Fehler im Prinzip berechnen; allerdings ist die Berechnung für große n schnell sehr aufwendig und die Binomialkoeffizienten werden auch schnell sehr groß. Um trotzdem einen Eindruck davon zu bekommen, was für größere n passiert, sind auf den beiden folgenden Seiten die Binomialkoeffizienten für $n = 5, 10, 50, 100, 500$ und 1000 graphisch dargestellt. (Die Tatsache, daß ab $n = 50$ deutlich weniger als $n+1$ Balken zu sehen sind, erklärt sich daraus, daß die restlichen Binomialkoeffizienten zu klein sind, um noch darstellbar zu sein: Die Diagramme sind so skaliert, daß der größte (mittlere) Balken jeweils eine feste Höhe hat.)

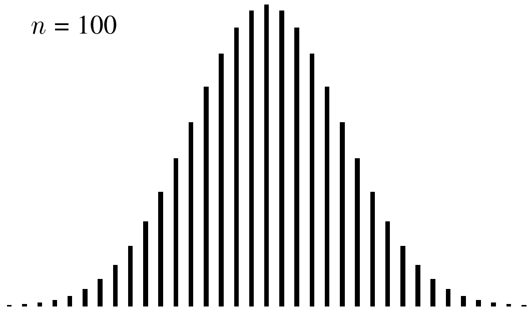
Betrachten wir als nächstes die Größe des Gesamtfehlers. Falls n gerade ist, treten nur Vielfache von 2ε auf und alle diese Vielfachen zwischen $-n\varepsilon$ und $n\varepsilon$ kommen tatsächlich vor; ensprechend sind für ungerades n nur ungeradzahlige Vielfache von ε möglich, und auch hier werden wieder alle solchen Werte zwischen $-n\varepsilon$ und $n\varepsilon$ angenommen. Wir können dies dadurch zusammenfassen, daß in beiden Fällen genau die Werte $(n-2k)\varepsilon$ mit $k = 0,\dots,n$ angenommen werden, und das PASCALsche Dreieck zeigt, daß der Fehler $(n-2k)\varepsilon$ in

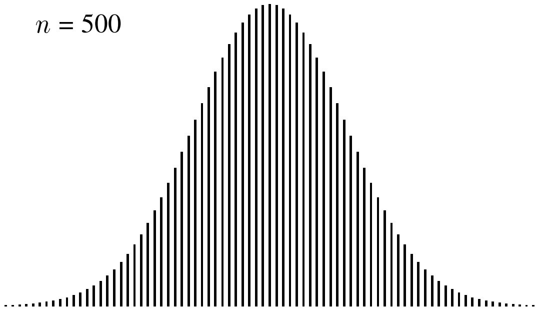
$$\binom{n}{n-k} = \binom{n}{k}$$

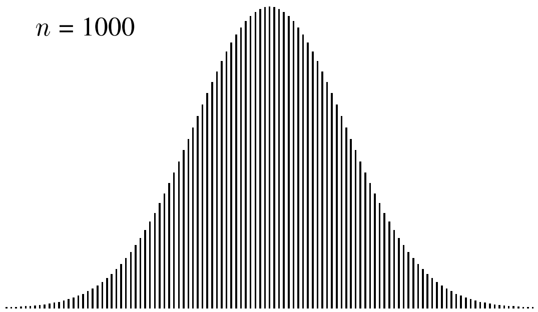

$n = 5$


$n = 10$


$n = 50$


$n = 100$


$n = 500$


$n = 1000$

Fällen auftritt. Da n „Dämonen“ insgesamt 2^n Möglichkeiten zur Fehlererzeugung haben, ist die Wahrscheinlichkeit für den Gesamtfehler $(n-2k)\varepsilon$ also

$$\binom{n}{k} \cdot 2^{-n} .$$

Diese Wahrscheinlichkeit sollte für einen festen Fehlerbetrag im wesentlichen unabhängig von n sein: Da wir nicht wirklich an Dämonen glauben, können wir deren Anzahl schließlich nicht in ein realistisches Fehlermodell einfließen lassen.

Auch die Balkendiagramme zeigen, daß sich die Verteilung der Fehlerwahrscheinlichkeiten für große n einer festen Kurve annähern sollte, der in Abbildung 54 und auf jedem Zehnmarkschein zu findenden *Glockenkurve* oder GAUSS-Kurve.

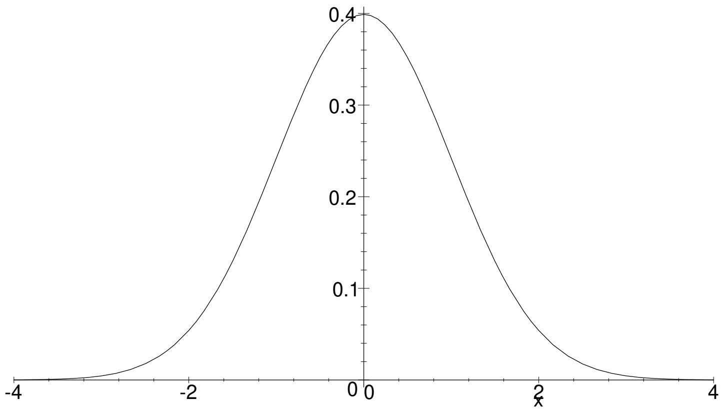


Abb. 54: Die „Glockenkurve“ $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Wenn sich Fehler oder auch beliebige Daten so veteilen, wie es dieser Kurve entspricht, redet man von *normalverteilten* Daten. Damit haben wir also zumindest graphisch gesehen, daß Meßfehler nach dem LAPLACEschen Fehlermodell normalverteilt sind. Das sagt noch nicht

unbedingt etwas über die Verteilung realer Meßfehler, da das LAPLACEsche Fehlermodell von unrealistisch einfachen Annahmen ausgeht; nach einem der fundamentale Gesetze der Statistik, dem *zentralen Grenzwertsatz,* führen aber auch realistischere Annahmen zu genau derselben Verteilung: Sind $u_1, \dots, u_n$ beliebige Quellen von Zufallsfehlern, über deren Verteilung wir (fast) nichts voraussetzen müssen, so ist ihre Summe für hinreichend großes n annähernd normalverteilt. Das eingeklammerte Wort „fast“ ist dabei für praktische Zwecke bedeutungslos, und als „groß“ kann man sich ein n ab etwa dreißig oder vierzig vorstellen.

Im nächsten Paragraphen werden wir uns überlegen, wie man zu einer Gleichung für die Glockenkurve kommt.

b) Statistische Kenngrößen

Die übliche Strategie zum Umgang mit Zufallsfehlern ist wohlbekannt: Man begnügt sich nicht mit einer einzigen Messung, sondern mißt dieselbe Größe mehrmals, so daß man eine ganze Meßreihe

$$x_1, x_2, \dots, x_N$$

erhält. Dann bildet man das *arithmetische Mittel*

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

der Meßreihe in der Hoffnung, daß sich hierbei die Fehler „ausmitteln“, so daß $\overline{x}$ dem theoretisch korrekten Wert $\hat{x}$ nahekommt.

Die Wahl des arithmetischen Mittels läßt sich auch geometrisch begründen: Eine Meßreihe $x_1, \dots, x_N$ für eine Meßgröße mit exaktem Wert $\hat{x}$ definiert einen Vektor im $\mathbb{R}^n$. Falls es keine Meßfehler gäbe, hätte dieser lauter identische Komponenten $\hat{x}$. Tatsächlich ist dies natürlich nicht der Fall; wir können aber nach einem Vektor mit identischen Komponenten suchen, der möglichst nahe am Vektor der Meßwerte liegt. Für einen Vektor, dessen sämtliche Komponenten gleich x sind, ist der EUKLIDische Abstands zum Vektor der Meßwerte gleich

$$d(x) = \sqrt{\sum_{i=1}^{N}(x - x_i)^2} = \sqrt{Nx^2 - 2x\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i^2}\,.$$

Die quadratische Funktion $d(x)^2$ hat ein eindeutig bestimmtes Minimum bei der Nullstelle ihrer Ableitung

$$2Nx - 2\sum_{i=1}^{N} x_i\,,$$

also beim arithmetischen Mittel $\overline{x}$, und dieses ist auch das einzige Minimum von $d(x)$. Wir nehmen daher das arithmetische Mittel $\overline{x}$ als besten verfügbaren Schätzwert für den unbekannten korrekten Wert $\hat{x}$.

Als Maß für die Schwankungen innerhalb der Meßreihe und damit für die Meßfehler könnte man versucht sein, den Abstand $d(\hat{x})$ zu nehmen; er hat aber den Nachteil, daß er mit steigendem N immer größer wird, d.h. die Schwankungen würden umso größer, je mehr man mißt. Das ist natürlich absurd; daher dividieren wir das Abstandsquadrat noch durch N und definieren die *mittlere quadratische Abweichung* oder *Varianz* der Meßreihe als

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{x} - x_i)^2\,.$$

Die (nichtnegative) Quadratwurzel σ hieraus heißt *Standardabweichung* der Meßreihe.

Das Ergebnis einer Messung wird meist angegeben in der Form

$$x = \overline{x} \pm \sigma\,,$$

man betrachtet also die Standardabweichung der Meßreihe als Maß für den Meßfehler. Da deren Definition allerdings vom (im allgemeinen unbekannten) korrekten Wert $\hat{x}$ abhängt, können wir sie nicht berechnen, sondern müssen im folgenden sehen, wie wir sie zumindest schätzen können.

Als einfachste Möglichkeit bietet sich an, σ^2 durch

$$\frac{1}{N}\sum_{i=1}^{N}(\overline{x} - x_i)^2$$

zu schätzen, aber das führt sicherlich zu einem zu kleinen Ergebnis: Schließlich ist $d(\overline{x})$ das eindeutig bestimmte Minimum der Abstandsfunktion d, so daß der korrekte Wert $d(\hat{x})$ für $\hat{x} \neq \overline{x}$ notwendigerweise größer sein muß.

In Abschnitt *d)* werden wir aus dem Fehlerfortpflanzungsgesetz einen besseren Schätzwert für σ herleiten.

Warum betrachten wir eigentlich quadratische Abweichungen und nicht die einfacheren linearen Abweichungen? Nun, der Mittelwert aller Abweichungen ist

$$\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i - \frac{1}{N}N\overline{x} = \overline{x} - \overline{x} = 0\,,$$

also ist dies keine geeignete Maßzahl. Möglich wäre die mittlere *betragsmäßige* Abweichung

$$\frac{1}{N}\sum_{i=1}^{N}|x_i - \overline{x}|\,,$$

allerdings wird die im allgemeinen nicht für das arithmetische Mittel $\bar{x}$ minimal, sondern, wie man sich leicht überlegen kann, für jede Zahl $\overline{x}$ mit der Eigenschaft, daß gleichviele Meßwerte größer und kleiner als $\overline{x}$ sind; einen solchen Wert $\overline{x}$ bezeichnet man als *Median* der Meßreihe. Für die Beschreibung wirtschafts- und sozialwissenschaftlicher Daten ist dieser Median meist eine aussagekräftigere Kennzahl als das arithmetische Mittel; in den Naturwissenschaften und der Technik spielt er allerdings keine große Rolle. Im nächsten Paragraphen werden wir sehen, daß auch das LAPLACEsche Fehlermodell in natürlicher Weise auf quadratische Abweichungen führt.

c) Das Fehlerfortpflanzungsgesetz

Gegeben seien zwei Größen

$$x = \hat{x} \pm \sigma_x \quad \text{und} \quad y = \hat{y} \pm \sigma_y$$

(die Verallgemeinerung auf mehr als zwei Größen erfordert, wie man sich bei der folgenden Rechnung leicht klarmacht, nur etwas mehr Schreibaufwand; sie ist nicht prinzipiell schwieriger), und eine Größe

$$w = f(x, y)\,,$$

die von diesen beiden abhängt. Um vernünftige Aussagen machen zu können, setzen wir dabei f als stetig differenzierbar voraus.

Für x seien N Meßwerte $x_1, \dots, x_N$ gegeben, und für y entsprechend M Werte $y_1, \dots, y_M$. Wenn wir echte Zufallsfehler haben, können wir davon ausgehen, daß die Fehler der x-Werte und die der y-Werte voneinander unabhängig sind, und das wollen wir im folgenden auch annehmen.

Für w haben wir dann NM Werte $w_{ij} = f(x_i, y_j)$, deren Mittelwert die beste Schätzung für den „wahren“ Wert $\widehat{w} = f(\hat{x}, \hat{y})$ ist. Dieser Mittelwert ist für komplizierte Funktionen f und/oder große Werte von n und m umständlich auszurechnen; günstiger wäre es, einfach den Mittelwert $\bar{x}$ der x_i und den Mittelwert $\bar{y}$ der y_j zu berechnen, um dann $f(\bar{x}, \bar{y})$ als Schätzung für $\widehat{w}$ zu benutzen. Zur Abschätzung des dadurch bedingten Fehler setzen wir

$$x_i = \bar{x} + h_i \quad \text{und} \quad y_j = \bar{y}_j + k_j ;$$

dann ist wegen der Differenzierbarkeit von f

$$\begin{aligned} w_{ij} &= f(x_i, y_j) = f(\bar{x} + h_i, \bar{y} + k_j) \\ &= f(\bar{x}, \bar{y}) + f_x(\bar{x}, \bar{y}) h_i + f_y(\bar{x}, \bar{y}) k_j + o\left(\sqrt{h_i^2 + k_j^2}\right) , \end{aligned}$$

wobei

$$f_x = \frac{\partial f}{\partial x} \quad \text{und} \quad f_y = \frac{\partial f}{\partial y}$$

die partiellen Ableitungen von f bezeichnen. Da die h_i und die k_j als Abweichungen vom Mittelwert die Summe null haben, ist also der Mittelwert der w_{ij} bis auf einen Fehler der Größenordnung $o\left(\sqrt{h^2 + k^2}\right)$ gleich $f(\bar{x}, \bar{y})$, wobei h, k die Batragsmaxima der h_i, k_j sind.

Als nächstes müssen wir den Fehler von $\widehat{w}$ berechnen, also den Erwartungswert der $(w_{ij} - \widehat{w})^2$. Dazu schreiben wir zunächst

$$x_i = \hat{x} + u_i \quad \text{und} \quad y_j = \hat{y}_j + v_j ,$$

betrachten also anstelle der Abweichungen vom Mittelwert die echten Meßfehler, und erhalten genau wie eben

$$w_{ij} - \widehat{w} = f(x_i, y_j) - f(\hat{x}, \hat{y}) \approx u_i \cdot f_x(\hat{x}, \hat{y}) + v_j \cdot f_y(\hat{x}, \hat{y})$$

mit Quadrat

$$\begin{aligned} (w_{ij} - \widehat{w})^2 &\approx \big(u_i \cdot f_x(\hat{x}, \hat{y}) + v_j \cdot f_y(\hat{x}, \hat{y})\big)^2 \\ &= u_i^2 \cdot f_x(\hat{x}, \hat{y})^2 + v_j^2 \cdot f_y(\hat{x}, \hat{y})^2 + 2u_i \cdot v_j \cdot f_x(\hat{x}, \hat{y}) \cdot f_y(\hat{x}, \hat{y}) . \end{aligned}$$

Hier sind die Werte u_i^2, v_j^2 und $u_i v_j$ jeweils Zufallsgrößen, über deren Werte wir nichts sagen können. Wir haben aber gewisse Erwartungen darüber, wie sie sich *im Mittel* verhalten: u_i^2 sollte, da σ_x^2 die mittlere quadratische Abweichung von $\hat{x}$ ist, im Mittel gleich σ_x^2 sein und v_j^2 entsprechend σ_y^2. Genauso sollten u_i und v_j im Mittel gleich null sein, und wenn wir annehmen, daß die Fehler u_i und v_j voneinander unabhängig sind, sollte auch ihr Produkt im Mittel verschwinden. Diese sogenannten *Erwartungswerte* sind offensichtlich die bestmöglichen Schätzwerte für die jeweiligen Größen; als beste Schätzung für σ_w^2 erhalten wir damit das GAUSS*sche Fehlerfortpflanzungsgesetz*

$$\sigma_w^2 = f_x(\hat{x}, \hat{y})^2 \cdot \sigma_x^2 + f_y(\hat{x}, \hat{y})^2 \cdot \sigma_y^2$$

oder

$$\sigma_w = \sqrt{f_x(\hat{x}, \hat{y})^2 \cdot \sigma_x^2 + f_y(\hat{x}, \hat{y})^2 \cdot \sigma_y^2} .$$

Genauso gilt dieses Gesetz auch für Funktionen von mehr als zwei Größen; für $w = f(x_1, \dots, x_n)$ ist

$$\sigma_w = \sqrt{f_{x_1}(\hat{x}_1, \dots \hat{x}_n)^2 \cdot \sigma_{x_1}^2 + \cdots + f_{x_n}(\hat{x}_1, \dots \hat{x}_n)^2 \cdot \sigma_{x_n}^2} .$$

d) Die Standardabweichung des Mittelwerts und die Schätzung der Varianz

Als einfache Anwendung des Fehlerfortpflanzungsgesetzes betrachten wir die Funktion

$$\bar{x} = f(x_1, \dots, x_N) = \frac{x_1 + \cdots + x_N}{N} ,$$

also den Mittelwert der x_i. Jede Messung x_i sei mit demselben erwarteten Fehler σ behaftet; da alle partiellen Ableitungen von f gleich $1/N$ sind, folgt für den Fehler des Mittelwerts

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} .$$

Dies bestätigt die implizit stets angewandte Regel, daß man durch mehrfaches Messen ein zuverlässigeres Ergebnis erhält; durch 25 Messungen beispielsweise läßt sich der Fehler auf ein Fünftel reduzieren, und für $N \to \infty$ geht er gegen Null *(Gesetz der großen Zahl)*.

Damit wissen wir, wie man aus den Meßwerten auf den Fehler des Mittelwerts schließen kann – sofern man die Fehler der Meßwerte kennt. Wie lassen sich diese schätzen?

Zunächst ist

$$\begin{aligned}\sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}(\hat{x}-x_i)^2 = \frac{1}{N}\sum_{i=1}^{N}\big((\hat{x}-\bar{x})+(\bar{x}-x_i)\big)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}(\hat{x}-\bar{x})^2 + \frac{1}{N}\sum_{i=1}^{N}(\bar{x}-x_i)^2 + \frac{2}{N}\sum_{i=1}^{N}(\hat{x}-\bar{x})\cdot(\bar{x}-x_i)\,.\end{aligned}$$

Die letzte dieser drei Summen ist

$$2\cdot\frac{(\hat{x}-\bar{x})}{N}\sum_{i=1}^{N}(\bar{x}-x_i) = 0\,,$$

da $\bar{x}$ der Mittelwert der x_i ist. Die zweite Summe ist der Mittelwert der $(\bar{x}-x_i)^2$, also die Varianz der Meßreihe, und von der ersten schließlich wissen wir, daß $(\hat{x}-\bar{x})^2$, das Quadrat des Fehlers des Mittelwerts, den Erwartungswert σ^2/N hat. Die gesamte erste Summe ist somit

$$\frac{1}{N}\cdot N\cdot\frac{\sigma^2}{N} = \frac{\sigma^2}{N}\,,$$

und obige Formel wird zu

$$\sigma^2 = \frac{\sigma^2}{N} + \frac{1}{N}\sum_{i=1}^{N}(\bar{x}-x_i)^2\,.$$

Bringt man hier noch den Term σ^2/N auf die linke Seite, so folgt

$$\frac{N-1}{N}\cdot\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(\bar{x}-x_i)^2$$

oder

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x}-x_i)^2\,,$$

also

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(\bar{x}-x_i)^2}{N-1}}\,.$$

Somit läßt sich auch σ aus den Meßdaten berechnen, der Meßfehler kann also ohne Kenntnis des „wahren“ Werts anhand der gemessenen Werte geschätzt werden.

e) Die Methode der kleinsten Quadrate

Oftmals ist zu gegebenen Beobachtungsdaten grundsätzlich bekannt, welcher *Art* von Gesetz sie genügen sollten, und das Problem besteht „nur“ noch darin, die in diesem Gesetz vorkommenden *Parameter* zu bestimmen. Im einfachsten Fall könnte man etwa an einen Widerstand denken, der dadurch gemessen wird, daß man verschiedene Spannungen U_i anlegt und die zugehörigen Stromstärken I_i mißt. Nach dem Ohmschen Gesetz ist dann $U_i = R\cdot I_i$, aber aufgrund der unvermeidlichen Meßfehler werden die verschiedenen Quotienten U_i/I_i natürlich nicht alle gleich sein. Die Lösung dieses Problems ist klar: Man nimmt den Mittelwert der Quotienten. Schwieriger wird es, wenn mehrere Parameter ins Spiel kommen.

Betrachten wir als Beispiel etwa die Entwicklung der Weltbevölkerung. Da (vor allem auf dem Stand von 1960) vieles für ein stärker als exponentielles Bevölkerungswachstum sprach, enwickelten damals HEINZ VON FOERSTER, PATRICIA M. MORA und LAWRENCE W. AMIOT, allesamt Elektrotechniker an der University of Illinois in Urbana, ein überexponentielles Wachstumsmodell. Die Wachstumsdifferentialgleichung

$$\dot{y}(t) = \beta y(t)^r$$

für die Weltbevölkerung $y(t)$ zum Zeitpunkt t kann durch Trennung der Veränderlichen leicht gelöst werden; die Lösung ist

$$y(t) = \left(\frac{\beta}{t_0-t}\right)^s \quad\text{mit}\quad s = \frac{1}{r-1}\,.$$

β, t_0 und r sind Parameter, die aufgrund der bekannten Zahlen zur Bevölkerungsentwicklung geschätzt werden müssen. Besonders interessant ist dabei natürlich der Parameter t_0, denn zum Zeitpunkt $t = t_0$ wird die Weltbevölkerung unendlich, was man wohl mit dem Weltuntergang gleichsetzen kann.

Für zahlreiche Zeitpunkte t_i sind die Bevölkerungszahlen y_i oder zumindest Schätzungen dafür bekannt, aber natürlich gibt es auch hier keine Parameterwerte, für die *alle* y_i gleich $f(t_i)$ sind; Ziel kann es nur sein, die Parameter so zu wählen, daß die Unterschiede zwischen y_i und $y(t_i)$ nicht zu groß werden.

FOERSTER, MORA und AMIOT taten genau das und kamen auf das Ergebnis

$$\begin{aligned} \beta &= (179 \pm 14) \cdot 10^9 \\ s &= 0{,}990 \pm 0{,}009 \\ t_0 &= 2026{,}87 \pm 5{,}5 \text{ Jahre} \\ r &= 2{,}01 \pm 0{,}009\,; \end{aligned}$$

der Weltuntergang ist also etwa 2026 fällig. Der beste Schätzwert für das Datum ist, wenn wir 0,87 Jahre in Tage, Stunden und Minuten umrechnen, Freitag, der 13. November gegen 13^{13} uhr – sofern man das bei einer Standardabweichung von fünfeinhalb Jahren überhaupt so genau wissen will. (Für Einzelheiten sei auf die Arbeit *Doomsday: Friday, 13 November, A.D. 2026*, Science **138**, November 1960, Seite 1291–1295, verwiesen.)

Da Weltuntergangsszenarien nicht zum Stoff dieser Vorlesung gehören, wollen wir auf die genaue Behandlung dieses Beispiels verzichten; wir wollen aber zumindest grundsätzlich verstehen lernen, wie man solche Parameterschätzungen durchführen kann, und wir wollen diese Techniken auch auf etwas alltäglichere Probleme anwenden.

Der allgemeine Ansatz ist folgender: Sei $u = f(a, b, \ldots; t)$ der von den Parametern $a, b, \ldots$ abhängige Zusammenhang zwischen den Größen t und u, und seien $(t_1, u_1), \ldots, (t_N, u_N)$ gegebene (Meß-)Werte, an die die Parameter angepaßt werden sollen.

Nehmen wir an, es gäbe Parameter $a, b, \ldots$, für die $u = f(a, b, \ldots; t)$ den exakten Zusammenhang zwischen den Größen u und t wiedergibt, daß aber die Werte u_i aufgrund von Beobachtungsfehlern von den $f(a, b, \ldots; t_i)$ abweichen – das ist genau die Situation, mit denen man es in den Naturwissenschaften und in der Technik meist zu tun hat.

Bezüglich der weiteren Vorgehensweise können wir wieder wie beim arithmetischen Mittel und der Standardabweichung argumentieren: Wir versuchen, den EUKLIDischen Abstand zwischen dem Vektor der tatsächlichen Beobachtungswerte und dem der vorhergesagten Werte zu minimieren, d.h. wir wählen die Parameter $a, b, \ldots$ so, daß

$$Q(a, b, \ldots) \underset{\text{def}}{=} \sum_{i=1}^{N} \bigl(u_i - f(a, b, \ldots; t_i)\bigr)^2$$

minimal wird – das ist die von CARL FRIEDRICH GAUSS eingeführte *Methode der kleinsten Quadrate.*

Obiger Ausdruck ist eine Funktion der Parameter $a, b, \ldots$; falls sie differenzierbar ist, kann sie nur dort ein Minimum annehmen, wo die partiellen Ableitungen

$$\frac{\partial Q}{\partial a}, \quad \frac{\partial Q}{\partial b}, \quad \ldots$$

allesamt verschwinden. Dies wird im allgemeinen ein nichtlineares Gleichungssystem definieren, dessen Lösung recht kompliziert sein kann; wir wollen uns hier auf den Fall beschränken, daß die Funktion f *linear* von den Parametern $a, b, \ldots$ abhängt. In diesem Fall bezeichnet man die Suche nach den optimalen Parametern als *lineare Regression;* sie hat sowohl in Naturwissenschaft und Technik als auch in den Wirtschafts- und Sozialwissenschaften zahlreiche interessante Anwendungen.

Falls alle Parameter linear in f eingehen, kommen sie in Q höchstens quadratisch vor; die partiellen Ableitungen sind also wieder linear, so daß insgesamt nur ein lineares Gleichungssystem gelöst werden muß.

Wir betrachten hier nur beispielhaft den einfachsten Fall einer Funktion, die von zwei Parametern abhängt: die lineare Funktion

$$u = f(a, b; t) = at + b\,;$$

die Rechnungen, die wir dafür im folgenden durchführen, funktionieren aber genauso auch für allgemeinere Funktionen f, solange f linear in den Parametern $a, b, \dots,$ ist; Nichtlinearität in t ist problemlos.

Gesucht sind also reelle Zahlen a und b, bei denen die Funktion

$$Q(a,b) = \sum_{i=1}^{N} (u_i - at_i - b)^2$$

minimal wird. Dort müssen dann insbesondere die beiden partiellen Ableitungen $Q_a(a,b)$ und $Q_b(a,b)$ verschwinden. Bevor wir diese ausrechnen, empfiehlt es sich, die Funktion zunächst durch Ausmultiplizieren nach Termen in a und b zu sortieren:

$$\begin{aligned} Q(a,b) = \sum_{i=1}^{N} u_i^2 + a^2 \sum_{i=1}^{N} t_i^2 + Nb^2 - 2a \sum_{i=1}^{N} u_i t_i \\ - 2b \sum_{i=1}^{N} u_i + 2ab \sum_{i=1}^{N} t_i \,. \end{aligned}$$

Damit ist

$$Q_a(a,b) = 2a \sum_{i=1}^{N} t_i^2 - 2 \sum_{i=1}^{N} u_i t_i + 2b \sum_{i=1}^{N} t_i$$

und

$$Q_b(a,b) = 2Nb - 2 \sum_{i=1}^{N} u_i + 2a \sum_{i=1}^{N} t_i \,.$$

Division durch zwei ergibt die beiden Gleichungen für das Verschwinden dieser partiellen Ableitungen:

$$a \sum_{i=1}^{N} t_i^2 + b \sum_{i=1}^{N} t_i = \sum_{i=1}^{N} t_i u_i$$

$$\text{und} \qquad a \sum_{i=1}^{N} t_i + bN = \sum_{i=1}^{N} u_i \,.$$

Um b zu eliminieren, multiplizieren wir die zweite Gleichung mit $\frac{1}{N}\sum_{i=1}^{N} t_i$

$$a \cdot \frac{1}{N} \left(\sum_{i=1}^{N} t_i \right)^2 + b \sum_{i=1}^{N} t_i = \frac{1}{N} \left(\sum_{i=1}^{N} t_i \right) \cdot \left(\sum_{i=1}^{N} u_i \right)$$

und subtrahieren das von der ersten Gleichung:

$$a \left(\sum_{i=1}^{N} t_i^2 - \frac{1}{N} \left(\sum_{i=1}^{N} t_i \right)^2 \right) = \sum_{i=1}^{N} t_i u_i - \frac{1}{N} \left(\sum_{i=1}^{N} t_i \right) \cdot \left(\sum_{i=1}^{N} u_i \right) ,$$

d.h.

$$a = \frac{\displaystyle\sum_{i=1}^{N} t_i u_i - \frac{1}{N} \left(\sum_{i=1}^{N} t_i \right) \cdot \left(\sum_{i=1}^{N} u_i \right)}{\displaystyle\sum_{i=1}^{N} t_i^2 - \frac{1}{N} \left(\sum_{i=1}^{N} t_i \right)^2} \,.$$

Entsprechend läßt sich a eliminieren, wenn wir die erste Gleichung mit $\sum t_i$ und die zweite mit $\sum t_i^2$ multiplizieren; wir erhalten

$$b = \frac{\displaystyle\left(\sum_{i=1}^{N} t_i \right) \cdot \left(\sum_{i=1}^{N} t_i u_i \right) - \left(\sum_{i=1}^{N} t_i^2 \right) \cdot \left(\sum_{i=1}^{N} u_i \right)}{\displaystyle\left(\sum_{i=1}^{N} t_i \right)^2 - N \sum_{i=1}^{N} t_i^2} \,.$$

Als Beispiel betrachten wir den Zusammenhang zwischen dem Grad der Korruption in einem Land und dessen Bruttosozialprodukt pro Einwohner. Jedes Jahr veröffentlicht die Organisation *Transparancy International* ihren *corruption perceptions index (CPI),* in dem jedem Land eine Zahl zwischen null und zehn zugeordnet wird, je nachdem, wie stark Geschäftsleute, Risikospezialisten und die Bevölkerung die Korruption im betreffenden Land einschätzen: Ein Index von zehn bedeutet, daß es praktische keine Korruption gibt, während bei null nichts läuft ohne Bimbes. Die neuesten Daten stammen von Ende 2000 und sind unter

`http://www.transparency.org/documents/cpi/index.html` zu finden. Die Zahlen werden als Mittelwerte über die letzten drei Jahren berechnet, so daß singuläre Ereignisse eines Jahres nicht zu sehr ins Gewicht fallen.

Wir vergleichen diese Zahlen mit dem Bruttosozialprodukt pro Einwohner, wie es die Weltbank für 1997 festgestellt hat. Dies sind die neuesten verfügbaren Zahlen; sie sind beispielsweise auf dem Server des Statistischen Bundesamtes unter

`http://www.statistik-bund.de/basis/d/ausl/ausllae5.htm`

zu finden. In der folgenden Tabelle sind alle Staaten aufgelistet, für die sowohl das Bruttosozialprodukt pro Einwohner als auch der CPI für 1999 oder 2000 vorliegt; das Bruttosozialprodukt pro Einwohner in US-$ ist kursiv gedruckt, der Korruptionsindex für 1999 normal und der für 2000 fett. (Fehlende Werte für 2000 bedeuten, daß es für diese Staaten nicht mehr genügend viele Quellen gibt; oft bedeutet dies, daß die Korruption dort so hoch ist, daß das Land für Geschäftsleute uninteressant wurde, so daß es kaum noch welche gibt, die man nach ihren Erfahrungen fragen kann.)

Ägypten	*1200*	3,3	**3,1**
Albanien	*760*	2,3	**–**
Argentinien	*8950*	3,0	**3.5**
Armenien	*560*	2,5	**2,5**
Aserbaidschan	*510*	1,7	**1,5**
Australien	*20650*	8,7	**8,3**
Belgien	*26730*	5,3	**6,1**
Bolivien	*970*	2,5	**2,7**
Botsuana	*3310*	6,1	**6,0**
Brasilien	*4790*	4,1	**3,9**
Chile	*4820*	6,9	**7,4**
China (ohne Hongkong & Taiwan)	*860*	3,4	**3,1**
Costa Rica	*2680*	5,1	**5,4**
Côte d'Ivoire	*710*	2,6	**2,7**
Dänemark	*34890*	10,0	**9,8**
Deutschland	*28280*	8,0	**7,6**
Ecuador	*1570*	2,4	**2,6**

El Salvador	*1810*	3,9	**4,1**
Estland	*3360*	5,7	**5,7**
Finnland	*24790*	9,8	**10,0**
Frankreich	*26300*	6,6	**6,7**
Ghana	*390*	3,3	**3,5**
Griechenland	*11640*	4,9	**4,9**
Großbritannien und Nordirland	*20870*	8,6	**8,7**
Guatemala	*1580*	3,2	**–**
Honduras	*740*	1,8	**–**
Hongkong	*25200*	7,7	**7,7**
Indien	*370*	2,9	**2,8**
Indonesien	*1110*	1,7	**1,7**
Irland	*17790*	7,7	**7,2**
Island	*26580*	9,2	**9,1**
Israel	*16180*	6,8	**6,6**
Italien	*20170*	4,7	**4,6**
Jamaika	*1550*	3,8	**–**
Japan	*38160*	6,0	**6,4**
Jordanien	*1520*	4,4	**4,6**
Kamerun	*620*	1,5	**2,0**
Kanada	*19640*	9,2	**9,2**
Kasachstan	*1350*	2,3	**3,0**
Kenia	*340*	2,0	**2,1**
Kirgisistan	*480*	2,2	**–**
Kolumbien	*2180*	2,9	**3,2**
Korea, Republik	*10550*	3,8	**–**
Kroatien	*4060*	2,7	**3,7**
Lettland	*2430*	3,4	**3,4**
Litauen	*2260*	3,8	**4,1**
Luxemburg *(BSP von 1996)*	*45360*	8,8	**8,6**
Malawi	*210*	4,1	**4,1**
Malaysia	*4530*	5,1	**4,8**
Marokko	*1260*	4,1	**4,7**
Mauritius	*3870*	4,8	**4,7**
Mazedonien	*1100*	3,3	**–**
Mexiko	*3700*	3,4	**3,3**

Mongolei	*390*	4,3	**–**
Mosambik	*140*	3,5	**2,2**
Namibia	*2110*	5,3	**5,4**
Neuseeland	*15830*	9,4	**9,4**
Nicaragua	*410*	3,1	**–**
Niederlande	*25830*	9,0	**8,9**
Nigeria	*280*	1,6	**1,2**
Norwegen	*36100*	8,9	**9,1**
Österreich	*27920*	7,6	**7,7**
Pakistan	*500*	2,2	**–**
Paraguay	*2000*	2,0	**–**
Peru	*2610*	4,5	**4,4**
Philippinen	*1200*	3,6	**2,8**
Portugal	*11010*	6,7	**6,4**
Sambia	*370*	3,5	**3,4**
Schweden	*26210*	9,4	**9,4**
Schweiz	*43060*	8,9	**8,6**
Senegal	*540*	3,4	**3,5**
Singapur	*32810*	9,1	**9,1**
Slowenien	*9840*	6,0	**5,5**
Spanien	*14490*	6,6	**7,0**
Südafrika	*3210*	5,0	**5,0**
Tansania, Vereinigte Republik	*210*	1,9	**2,5**
Thailand	*2740*	3,2	**3,2**
Tunesien	*2110*	5,0	**5,2**
Türkei	*3130*	3,6	**3,8**
Uganda	*330*	2,2	**2,3**
Uruguay	*6130*	4,4	**–**
Usbekistan	*1020*	1,8	**2,4**
Venezuela	*3480*	2,0	**2,7**
Vereinigte Staaten	*29080*	7,5	**7,8**
Vietnam	*310*	2,6	**2,5**

Der erste Augenschein zeigt, daß korruptionsärmere Länder oftmals reicher sind: Das praktisch korruptionsfreie Dänemark hat ein Bruttosozialprodukt von 34 890 \$ pro Einwohner, das deutlich korruptere

Deutschland nur 28 280 \$ und ein stark korruptes Land wie Aserbaidschan nur 510 \$. Allerdings gibt es auch Ausnahmen, denn das praktisch korruptionsfreie Finnland hat mit 24 790 \$ pro Einwohner ein deutlich geringers Bruttosozialprodukt als etwa Deutschland. Es gibt also sicherlich keinen deterministischen Zusammenhang zwischen Korruption und Wohlstand, aber es lohnt sich doch, eine Ausgleichsgerade zu berechnen. Für 1999 führen obige Formeln auf

$$\mathrm{CPI} = 3{,}177461442 + 0{,}1692187403 \cdot 10^{-3} \cdot \mathrm{BSP}\,,$$

für 2000 auf

$$\mathrm{CPI} = 3{,}369300965 + 0{,}1626838231 \cdot 10^{-3} \cdot \mathrm{BSP}\,.$$

Abbildung 55 zeigt letztere Gerade zusammen mit den Datenpunkten.

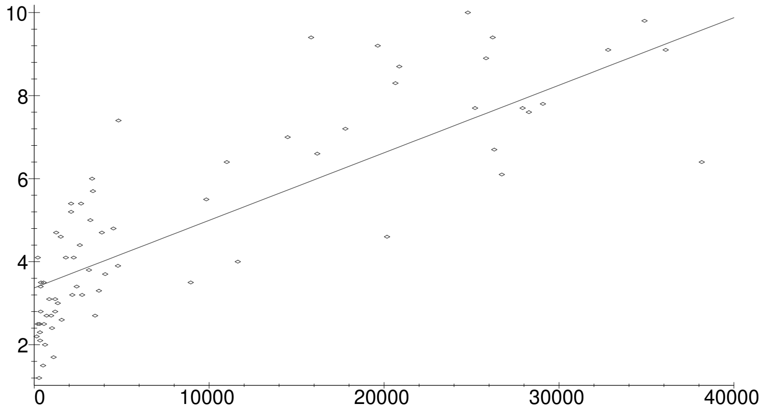


Abb. 55: Korruption 2000 über Bruttosozialprodukt 1997 für 72 Staaten

Natürlich liegen nicht alle Datenpunkte auf der Ausgleichsgeraden: Beispielsweise sind Japan, Belgien und Italien deutlich korrupter, als ihre Wirtschaftskraft vermuten läßt, wohingegen Dänemark, Finnland und Neuseeland deutlich ärmer sind, als man aufgrund der geringen Korruption erwarten sollte. Trotzdem beschreibt die Ausgleichsgerade doch einigermaßen genau einen Trend.

Für Deutschland ergibt die 1999er Formel einen erwarteten Index von 7,96, was praktisch perfekt mit 8,0 übereinstimmt; die 2000er Formel allerdings ergibt mit 7,97 zwar fast denselben Wert, aber da 1999 noch kaum jemand das Wort „Bimbes“ kannte, während es 2000 „Schwarzgeldaffäre“ zum Wort des Jahres brachte, sank der tatsächliche Index auf 7,6.

Ähnlich sieht es aus, wenn man das Klausurergebnis der letzten HM II mit den bearbeiteten Übungsaufgaben in Verbindung bringt: Hier ergibt sich die Gerade

$$\textit{Klausurpunkte} = 1{,}097325998 + 0{,}5192298020 \cdot \textit{Übungsprozente}\,,$$

die in Abbildung 56 zusammen mit den Datenpunkten eingezeichnet ist.

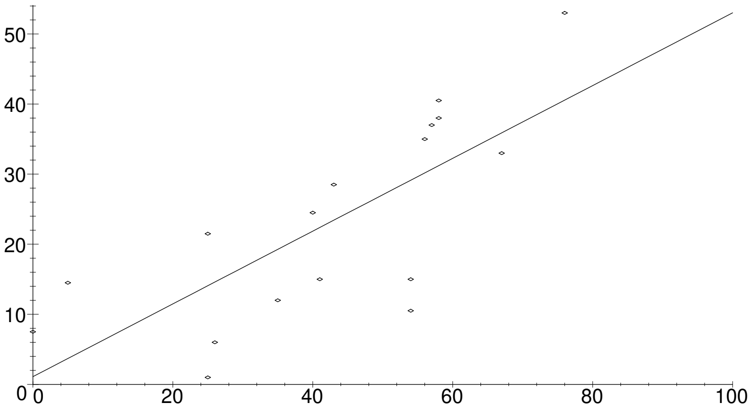


Abb. 56: Zusammenhang zwischen Klausur und Übungen im letzten Winter

Wieder sieht man, daß im großen und ganzen höhere Punktezahlen bei den Übungsaufgaben einem besseren Klausurergebnis entsprechen, aber es gibt auch recht deutliche Ausreißer, deren Interpretation dem Leser überlassen sei.

Unser nächstes Ziel ist ein Maß, mit dem wir die Qualität einer Ausgleichsgerade $u = at + b$ zu N Datenpaaren (t_i, u_i) beurteilen können.

Zu seiner Definition gehen wir ähnlich vor wie bei der Herleitung des GAUSSschen Fehlerfortpflanzungsgesetzes: $\bar t$ sei der Mittelwert der t_i und $\bar u$ der der u_i. Falls die u_i und die t_i voneinander unabhängig sind, erwarten wir, daß

$$\sum_{i=1}^{N} (t_i - \bar t)(u_i - \bar u)$$

verschwinden sollte, denn die Abweichungen $t_i - \bar t$ der t_i von ihrem Mittelwert sollten nichts mit den Abweichungen $u_i - \bar u$ der u_i von deren Mittelwert zu tun haben, und im Mittel sind beide Abweichungen null.

Tatsächlich ist obiges Argument kein Beweis, sondern die Definition des Wortes „unabhängig“: Wir bezeichnen

$$\operatorname{cov}(\boldsymbol{t}, \boldsymbol{u}) \underset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} (t_i - \bar t)(u_i - \bar u)$$

als die *Kovarianz* der Meßreihen

$$\boldsymbol{t} = (t_1, \dots, t_N) \quad \text{und} \quad \boldsymbol{u} = (u_1, \dots, u_N)$$

und bezeichnen die beiden Meßreihen als statistisch unabhängig, falls diese Kovarianz verschwindet.

Falls im anderen Extrem $u_i = at_i + b$ ist für alle i, so ist $\bar u = a\bar t + b$ und

$$\sum_{i=1}^{N} (t_i - \bar t)(u_i - \bar u) = \sum_{i=1}^{N} (t_i - \bar t)(at_i - a\bar t) = a \sum_{i=1}^{N} (t_i - \bar t)^2$$

eine (fast immer) positive Zahl, die eng mit der Standardabweichung der Verteilung der t_i zusammenhängt. Um sie weiter auszurechnen, beachten wir, daß

$$\sum_{i=1}^{N} (u_i - \bar u)^2 = a^2 \sum_{i=1}^{N} (t_i - \bar t)^2$$

ist und damit

$$\sum_{i=1}^{N}(t_i-\bar{t})^2\cdot\sum_{i=1}^{N}(u_i-\bar{u})^2=a^2\left(\sum_{i=1}^{N}(t_i-\bar{t})^2\right)^2$$

oder

$$\sqrt{\sum_{i=1}^{N}(t_i-\bar{t})^2\cdot\sum_{i=1}^{N}(u_i-\bar{u})^2}=a\sum_{i=1}^{N}(t_i-\bar{t})^2\,.$$

Damit ist also der Quotient

$$\rho=\frac{\sum_{i=1}^{N}(t_i-\bar{t})(u_i-\bar{u})}{\sqrt{\sum_{i=1}^{N}(t_i-\bar{t})^2\cdot\sum_{i=1}^{N}(u_i-\bar{u})^2}}$$

je nach Vorzeichen von a gleich ± 1, wenn alle $u_i=at_i+b$ sind, und $r=0$, wenn die u_i unabhängig von den t_i sind. Für beliebige Wertepaare (t_i,u_i) sagt uns die CAUCHY-SCHWARZsche Ungleichung aus **[HM1]**, Kap. 2, §1*a),* daß stets

$$-1\le\rho\le 1$$

ist.

Diesen Wert ρ nennen wir den *Korrelationskoeffizienten* der beiden Meßreihen $\boldsymbol{t}$ und $\boldsymbol{u}$; sein Betrag gibt an, wie gut die Beziehung zwischen den u_i und den t_i durch eine Gerade beschrieben werden kann, während sein Vorzeichen angibt, ob diese Gerade ansteigt (die Daten sind *positiv korreliert*) oder sinkt (*negative Korrelation*).

Kürzer läßt er sich auch schreiben als

$$\rho=\frac{\operatorname{cov}(\boldsymbol{t},\boldsymbol{u})}{\sigma_{\boldsymbol{t}}\sigma_{\boldsymbol{u}}}\,,$$

wobei $\sigma_{\boldsymbol{t}}$ und $\sigma_{\boldsymbol{u}}$ die Standardabweichungen der beiden Meßreihen bezeichnen.

Bei den obigen Zahlenbeispielen ergeben sich die Korrelationskoeffizienten zu

$$\rho\approx 0{,}8317249988\quad\textit{bzw.}\quad\rho\approx 0{,}8311994538$$

für den Zusammenhang zwischen Korruption 1999 *bzw.* 2000 und Bruttosozialprodukt; für die Klausur ergibt sich

$$\rho\approx 0{,}7465571442$$

In beiden Fällen ist die Korrelation also recht gut.

Zum Abschluß sei noch darauf hingewiesen, daß auch ein Korrelationskoeffizient nahe bei ± 1 nicht unbedingt bedeutet, daß ein *inhaltlicher* Zusammenhang zwischen den verglichenen Größen besteht: Die Korrelation könnte auch rein formal sein. Ein beliebtes Beispiel, zu dem leider nirgends Zahlenwerte zu finden sind, ist die gute Korrelation zwischen dem Rückgang der Storchenpopulationen und der Geburtenrückgang zur Zeit der Industrialisierung: Die meisten Bevölkerungsexperten gehen hier nicht von einem kausalen Zusammenhang aus, sondern eher davon, daß die Industrialisierung den Lebensraum der Störche zerstörte und wegen der gleichzeitigen Einführung der Rentenversicherung die Familien nicht mehr so viele Kinder zur Alterssicherung brauchten.

§5: Die Gaußsche Normalverteilung und die Maximum Likelihood Methode

a) Der Grenzfall des Laplaceschen Fehlermodells

Leider ist in dieser Vorlesung nicht genug Zeit, um auch nur annähernd die notwendigen Grundlagen für einen Beweis des zentralen Grenzwertsatzes bereitzustellen. Wir haben aber immerhin schon am Beispiel des LAPLACEschen Fehlermodell graphisch gesehen, daß für die Verteilungen mit $u_i=\pm\varepsilon$, jeweils mit Wahrscheinlichkeit $\frac{1}{2}$, die Verteilung der Summen gegen eine Normalverteilung konvergiert.

Zumindest in diesem einfachen Fall können wir dies auch rechnerisch einsehen und auf diesem Weg insbesondere auch die Gleichung der Glockenkurve herleiten. Dazu müssen wir uns zunächst überlegen, was ε tun soll, wenn n gegen unendlich geht.