

Herbert Paukert
Schulmathematik in 8 Bänden
Version 6.0, 2016

MATHE 6

Rechnen mit Matrizen
Statistische Berechnungen
Wahrscheinlichkeitsrechnung

MATHE, Band 1

Arithmetik - Unterstufe

MATHE, Band 2

Geometrie - Unterstufe

MATHE, Band 3

**Logik
Zahlenmengen
Algebra**

MATHE, Band 4

Differenzialrechnung

MATHE, Band 5

Integralrechnung

MATHE, Band 6

**Matrizenrechnung
Statistik
Wahrscheinlichkeit**

MATHE, Band 7

Trigonometrie

MATHE, Band 8

**Analytische Geometrie
Kegelschnittslinien
Geometrische Abbildungen**

Inhaltsverzeichnis

| | |
|--|-----------------|
| (1) Rechnen mit Matrizen | Seite 05 |
| (2) Statistische Berechnungen | Seite 19 |
| (3) Wahrscheinlichkeitsrechnung | Seite 71 |

Hinweis: In Dezimalzahlen wird anstelle eines Kommas ein Dezimalpunkt geschrieben.

Hinweis: Auf seiner Homepage www.paukert.at stellt der Autor viele weitere Lernhilfen aus unterschiedlichen Fachgebieten zur Verfügung.

MATRIZEN

In dieser Einführung werden die wichtigsten theoretischen Konzepte des Rechnens mit Matrizen kurz dargestellt. Dieser Text stammt aus dem Rechenprogramm *MATRIX.EXE* des Autors, das von www.paukert.at heruntergeladen werden kann.

----- (1) MATRIZEN -----

Eine Matrix A ist eine Anordnung von Elementen $a(i,j)$ in n Zeilen und m Spalten. Dabei ist i der Zeilenindex ($1 \leq i \leq n$) und j ist der Spaltenindex ($1 \leq j \leq m$).

$$\begin{array}{cccc} a(1,1) & a(1,2) & \dots & a(1,m) \\ a(2,1) & a(2,2) & \dots & a(2,m) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a(n,1) & a(n,2) & \dots & a(n,m) \end{array}$$

Matrizen mit n Zeilen und n Spalten heißen quadratisch. Matrizen mit n Zeilen und 1 Spalte, oder 1 Zeile und n Spalten heißen Vektoren. Sie werden klein geschrieben.

Beispiel eine Vektors $\vec{x} = [a(1) \ a(2) \ \dots \ a(n)]$

i-ter Zeilenvektor der Matrix: $[a(i,1) \ a(i,2) \ \dots \]$
j-ter Spaltenvektor der Matrix: $[a(1,j) \ a(2,j) \ \dots \]$

Hauptdiagonale: $a(1,1) \ a(2,2) \ \dots \ a(n,n)$
Nebendiagonale: $a(1,m) \ a(2,m-1) \ \dots \ a(n,1)$

Einheitsmatrix E: $a(i,j) = 1$ für alle $i = j$
 $a(i,j) = 0$ für alle $i \neq j$
Nullmatrix O: $a(i,j) = 0$ für alle i und alle j

Eine Dreiecksmatrix ist eine quadratische Matrix, deren Elemente unter- oder oberhalb der Hauptdiagonale alle Null sind. Eine Diagonalmatrix ist eine Matrix, bei der alle Elemente außerhalb der Hauptdiagonale Null sind.

Matrizen-Addition $C = A + B$: $c(i,j) = a(i,j) + b(i,j)$.
Multiplikation mit Zahl β , $C = \beta \cdot A$: $c(i,j) = \beta \cdot a(i,j)$.

----- (2) VEKTOREN -----

Gegeben sind zwei Vektoren x und y.

$$\vec{x} = [x(1) \ x(2) \ \dots \ x(n)]$$

$$\vec{y} = [y(1) \ y(2) \ \dots \ y(n)]$$

Für das Skalarprodukt zweier Vektoren $s = \vec{x} \cdot \vec{y}$ gilt:
 $s = x(1) \cdot y(1) + x(2) \cdot y(2) + \dots + x(n) \cdot y(n)$

Zwei Vektoren heißen orthogonal, wenn ihr Skalarprodukt Null ergibt.

Für den Betrag (Länge) $|\vec{x}|$ eines Vektors \vec{x} gilt:

$$|\vec{x}| = \sqrt{x(1)^2 + x(2)^2 + \dots + x(n)^2}$$

Ein Vektor \vec{x} heißt normiert, wenn $|\vec{x}| = 1$.

Gegeben ist eine Menge von Vektoren: $\vec{w}, \vec{x}, \vec{y}, \vec{z}$.
Vektor \vec{w} ist eine Linearkombination (Vielfachensumme) der Vektoren $\vec{x}, \vec{y}, \vec{z}$, wenn es reelle Zahlen a, b, c gibt, sodass gilt: $\vec{w} = a\vec{x} + b\vec{y} + c\vec{z}$.

Vektoren heißen voneinander linear abhängig (l.a.), wenn mindestens einer von ihnen eine Linearkombination der restlichen anderen Vektoren ist.

Vektoren heißen voneinander linear unabhängig (l.u.a.), wenn keiner von ihnen eine Linearkombination der restlichen anderen Vektoren ist.

----- (3) DAS PRODUKT VON MATRIZEN $C = A * B$ -----

A ist eine (n,m)-Matrix und B ist eine (m,k)-Matrix:

```
a(1,1) a(1,2) . . . a(1,m)
a(2,1) a(2,2) . . . a(2,m)
. . . . .
. . . . .
a(i,1) a(i,2) . . . a(i,m)
. . . . .
. . . . .
a(n,1) a(n,2) . . . a(n,m)
```

```
b(1,1) b(1,2) . . b(1,j) . . b(1,k)
b(2,1) b(2,2) . . b(2,j) . . b(2,k)
. . . . .
. . . . .
b(m,1) b(m,2) . . b(m,j) . . b(m,k)
```

Das Matrizenprodukt $C = A * B$ ist eine (n,k)-Matrix in der jedes Element $c(i,j)$ das Skalarprodukt des i-ten Zeilenvektors von A mit dem j-ten Spaltenvektor von B ist:

$$c(i,j) = a(i,1)*b(1,j)+a(i,2)*b(2,j)+\dots+a(i,m)*b(m,j)$$

Ein wichtiger Spezialfall ist die Multiplikation einer quadratischen Matrix A mit einem Vektor \vec{x} . Das ergibt dann einen neuen Vektor $\vec{y} = A * \vec{x}$.

```
a(1,1) a(1,2) . . . a(1,n)
a(2,1) a(2,2) . . . a(2,n)
. . . . .
. . . . .
a(i,1) a(i,2) . . . a(i,n)
. . . . .
. . . . .
a(n,1) a(n,2) . . . a(n,n)
```

Vektor $\vec{x} = [x(1) \ x(2) \ . \ . \ . \ x(n)]$

$\vec{y} = A * \vec{x} = [y(1) \ y(2) \ . \ . \ y(n)]$ mit folgenden Werten:
 $y(i) = a(i,1)*x(1) + a(i,2)*x(2) + . \ . \ . + a(i,n)*x(n)$

Für die Matrizen-Multiplikation gelten folgende Regeln:

$A * E = E * A = A$ (mit Einheitsmatrix E)
 $A * B \neq B * A$ (antikommunikativ)
 $(A*B)*C = A*(B*C)$ (assoziativ)
 $A*(B+C) = A*B + A*C$ (distributiv)
 $(B+C)*A = B*A + C*A$ (distributiv)

Vertauscht man in einer Matrix A ihre Zeilen mit ihren Spalten, dann erhält man die transponierte Matrix A^t .
 Die Transponierte eines Produktes: $(A * B)^t = B^t * A^t$.
 Eine Matrix heißt symmetrisch, wenn $A = A^t$.

(4) DER RANG EINER MATRIX, rank(A)

Unter dem Rang einer Matrix "rank(A)" versteht man die maximale Anzahl von linear unabhängigen Zeilenvektoren der Matrix A. Zeilenrang und Spaltenrang stimmen immer überein. Also ist $\text{rank}(A) = \text{rank}(A^t)$.

Rang einer quadratischen (n,n)-Matrix kann maximal n sein. Eine (n,n)-Matrix mit Rang n heißt regulär. Andernfalls heißt sie singular, wenn $\text{rank}(A) < n$.

(5) DIE DETERMINANTE EINER MATRIX, det(A)

Wenn man in einer Matrix A die i-te Zeile und auch die j-te Spalte streicht, dann erhält man die so genannte Streichungsmatrix $A|ij$.

Für die Determinante $\det(A)$ einer (n,n)-Matrix gilt:

$\det(A) = \text{Summe} ((-1)^{(i+j)} * a(i,j) * \det(A|ij))$
 für laufenden Zeilenindex i von 1 bis n,
 für einen beliebigen festen Spaltenindex j.

Beispielsweise mit Spalte $j = 1$ gilt dann: $\det(A) = a(1,1)*\det(A|11) - a(2,1)*\det(A|21) + a(3,1)*\det(A|31) \dots$

Anstatt die Determinante nach den Zeilen zu entwickeln, kann sie auch nach den Spalten entwickelt werden. Das Ergebnis ist immer gleich (Entwicklungssatz von Laplace).

Beispiel einer (2,2)-Determinante:

$a_{11} \ a_{12}$
 $a_{21} \ a_{22}$

$\det(A) = a_{11}*a_{22} - a_{21}*a_{12}$

Geometrisch ist $\det(A)$ die Fläche des Parallelogramms in der Ebene, das von den zwei Zeilenvektoren $\vec{x} = [a_{11} \ a_{12}]$ und $\vec{y} = [a_{21} \ a_{22}]$ aufgespannt wird.

Beispiel einer (3,3)-Determinante:

```
a11 a12 a13
a21 a22 a23
a31 a32 a33
```

$$\det(A) = a_{11} \cdot \det(A|11) - a_{21} \cdot \det(A|21) + a_{31} \cdot \det(A|31)$$

$$\begin{aligned} \det(A) &= a_{11} \cdot (a_{22} \cdot a_{33} - a_{32} \cdot a_{23}) - \\ &\quad a_{21} \cdot (a_{12} \cdot a_{33} - a_{32} \cdot a_{13}) + \\ &\quad a_{31} \cdot (a_{12} \cdot a_{23} - a_{22} \cdot a_{13}) \\ &= a_{11} \cdot a_{22} \cdot a_{33} + a_{21} \cdot a_{32} \cdot a_{13} + a_{31} \cdot a_{12} \cdot a_{23} \\ &\quad - (a_{11} \cdot a_{32} \cdot a_{23} + a_{21} \cdot a_{12} \cdot a_{33} + a_{31} \cdot a_{22} \cdot a_{13}) \end{aligned}$$

Geometrisch ist $\det(A)$ das Volumen des schiefen Quaders im Raum (Parallelepipeds), der von den 3 Zeilenvektoren aufgespannt wird. Dieses Volumen heißt auch Spatprodukt.

Es gelten folgende vier wichtige Determinantenregeln, welche sich direkt aus der Definition ergeben:

- [D1] Vertauscht man zwei Zeilen (Spalten) der Matrix, dann ändert sich nur das Vorzeichen, aber nicht der Betrag der Determinante.
- [D2] Addiert man zu einer Zeile (Spalte) der Matrix ein Vielfaches einer anderen Zeile (Spalte), dann ändert sich die Determinante nicht.
- [D3] Die Determinante einer Dreiecksmatrix oder einer Diagonalmatrix ist das Produkt der Elemente der Hauptdiagonale.
- [D4] Sind die Zeilen (Spalten) einer (n,n) -Matrix A linear abhängig ($\text{rank}(A) < n$, A singulär), dann ist die Determinante gleich Null.

(6) DIE INVERSE EINER MATRIX (A_i)

Die inverse Matrix A_i zu einer (n,n) -Matrix A ist eine Matrix mit folgender Eigenschaft:

$$A \cdot A_i = A_i \cdot A = E \quad (E = \text{Einheitsmatrix})$$

Nur zu quadratischen, nicht singulären Matrizen gibt es eindeutig bestimmte inverse Matrizen.

Für die inverse Matrix eines Produktes gilt:

$$\begin{aligned} (A \cdot B)_i &= (A \cdot B) \cdot A_i = E \\ (A \cdot B)_i \cdot A &= E \cdot B_i \\ (A \cdot B)_i &= E \cdot B_i \cdot A_i \\ (A \cdot B)_i &= B_i \cdot A_i \end{aligned}$$

Der Lösungsvektor x' wird mit dem Eliminationsverfahren von Gauß ermittelt. Dabei werden entsprechende Vielfache von geeigneten Zeilenvektoren $Z(i)$ der Matrix A_e von den restlichen Zeilenvektoren so lange subtrahiert bis die Systemmatrix A eine Dreiecksform hat. Weil es sich um Äquivalenz-Umformungen handelt, wird die Lösung dabei nicht verändert. Beginnend mit der letzten Zeile können dann die einzelnen Koeffizienten des Lösungsvektors x schrittweise durch Rückwärtseinsetzen berechnet werden.

Beispiel mit allen Eliminationsschritten.

Zuerst mit dem Gauß-Verfahren bis zur Dreiecksmatrix und dann weiter mit dem Gauß-Jordan-Verfahren bis zur Diagonalmatrix.

Alle Zeilen-Hinweise rechts beziehen sich auf die jeweils vorangehende Matrix.

$$\begin{aligned} 3x_1 + 5x_2 + 1x_3 &= 14 \\ 1x_1 + 2x_2 + 2x_3 &= 10 \\ 2x_1 + 4x_2 + 5x_3 &= 23 \end{aligned}$$

$$\begin{aligned} 3x_1 + 5x_2 + 1x_3 &= 14, Z(1) \text{ bleibt unverändert} \\ 0 + 1/3x_2 + 5/3x_3 &= 16/3, Z(2) = Z(2) - (1/3)Z(1) \\ 0 + 2/3x_2 + 13/3x_3 &= 41/3, Z(3) = Z(2) - (2/3)Z(1) \end{aligned}$$

$$\begin{aligned} 3x_1 + 5x_2 + 1x_3 &= 14, Z(1) \text{ bleibt unverändert} \\ 0 + 1x_2 + 5x_3 &= 16, Z(2) = Z(2) * 3 \\ 0 + 2x_2 + 13x_3 &= 41, Z(3) = Z(3) * 3 \end{aligned}$$

$$\begin{aligned} 3x_1 + 5x_2 + 1x_3 &= 14, Z(1) \text{ bleibt unverändert} \\ 0 + 1x_2 + 5x_3 &= 16, Z(2) \text{ bleibt unverändert} \\ 0 + 0 + 3x_3 &= 9, Z(3) = Z(3) - (2/1)Z(2) \end{aligned}$$

Durch diese Umformungen ist eine Systemmatrix A in Dreiecksform entstanden. Aus der letzten Gleichung folgt $x_3 = 3$. Durch Einsetzen in die zweite Gleichung erhält man $x_2 = 1$. Einsetzen in die erste Gleichung ergibt dann $x_1 = 2$. Damit ist das Eliminationsverfahren nach Gauß beendet.

Beim Verfahren von Gauß-Jordan wird der Eliminationsprozess nicht nur auf alle, der Eliminationszeile $Z(i)$ NACHFOLGENDEN Zeilen angewendet, sondern auf ALLE Gleichungszeilen der Matrix, ausgenommen der Zeile $Z(i)$.

Nach dem Erreichen der Dreiecksform werden die Umformungen so lange fortgesetzt bis schließlich die Systemmatrix eine Diagonalform aufweist, mit lauter 1 in der Hauptdiagonale (Einheitsmatrix).

$$\begin{aligned} 3x_1 + 0 - 24x_3 &= -66, Z(1) = Z(1) - (5/1)Z(2) \\ 0 + 1x_2 + 5x_3 &= 16, Z(2) \text{ bleibt unverändert} \\ 0 + 0 + 1x_3 &= 3, Z(3) = Z(3)/3 \end{aligned}$$

$$\begin{aligned} 1x_1 + 0 - 8x_3 &= -22, Z(1) = Z(1)/3 \\ 0 + 1x_2 + 5x_3 &= 16, Z(2) \text{ bleibt unverändert} \\ 0 + 0 + 1x_3 &= 3, Z(3) \text{ bleibt unverändert} \end{aligned}$$

$$\begin{array}{rcl}
 1 \cdot x_1 + 0 - 0 = 2, & Z(1) = Z(1) - (-8/1) \cdot Z(3) \\
 0 + 1 \cdot x_2 + 0 = 1, & Z(2) = Z(2) - (5/1) \cdot Z(3) \\
 0 + 0 + 1 \cdot x_3 = 3, & Z(3) \text{ bleibt unverändert}
 \end{array}$$

Nach den letzten Äquivalenz-Umformungen sind die Lösungen aus der Diagonalform der Matrix direkt ablesbar. Damit ist das Eliminationsverfahren von Gauß-Jordan beendet.

(8) BERECHNUNG DER INVERSEN MATRIX

A_i ist die inverse Matrix zur quadratischen Matrix A .

Es gilt $A \cdot A_i = E$.

Wir verwandeln die Matrix A mit dem Eliminationsverfahren von Gauß-Jordan in die Einheitsmatrix E . Alle dabei ausgeführten Umformungen wenden wir nun auch auf E an. Dadurch entsteht eine neue Matrix B . Diese Matrix B ist die gesuchte inverse Matrix A_i , weil $A \cdot B = E$ ist, und daher $B = A_i$ sein muss.

Zur Demonstration verwenden wir die Systemmatrix A des ersten Beispiels und schreiben daneben auch die entsprechend umgeformte Einheitsmatrix E .

Alle Zeilen-Hinweise rechts beziehen sich auf die jeweils vorangehende Matrix.

$$\begin{array}{rcl}
 \begin{array}{ccc|ccc}
 3 & 5 & 1 & 1 & 0 & 0 \\
 1 & 2 & 2 & 0 & 1 & 0 \\
 2 & 4 & 5 & 0 & 0 & 1
 \end{array} \\
 \\
 \begin{array}{rcl}
 \begin{array}{ccc|ccc}
 3 & 5 & 1 & 1 & 0 & 0 \\
 0 & 1/3 & 5/3 & -1/3 & 1 & 0 \\
 0 & 2/3 & 13/3 & -2/3 & 0 & 1
 \end{array} & , & \begin{array}{l}
 Z(1) \text{ bleibt unverändert} \\
 Z(2) = Z(2) - (1/3) \cdot Z(1) \\
 Z(3) = Z(3) - (2/3) \cdot Z(1)
 \end{array}
 \end{array} \\
 \\
 \begin{array}{rcl}
 \begin{array}{ccc|ccc}
 3 & 5 & 1 & 1 & 0 & 0 \\
 0 & 1 & 5 & -1 & 3 & 0 \\
 0 & 2 & 13 & -2 & 0 & 3
 \end{array} & , & \begin{array}{l}
 Z(1) \text{ bleibt unverändert} \\
 Z(2) = Z(2) \cdot 3 \\
 Z(3) = Z(3) \cdot 3
 \end{array}
 \end{array} \\
 \\
 \begin{array}{rcl}
 \begin{array}{ccc|ccc}
 3 & 5 & 1 & 1 & 0 & 0 \\
 0 & 1 & 5 & -1 & 3 & 0 \\
 0 & 0 & 3 & 0 & -6 & 3
 \end{array} & , & \begin{array}{l}
 Z(1) \text{ bleibt unverändert} \\
 Z(2) \text{ bleibt unverändert} \\
 Z(3) = Z(3) - (2/1) \cdot Z(2)
 \end{array}
 \end{array} \\
 \\
 \begin{array}{rcl}
 \begin{array}{ccc|ccc}
 3 & 0 & -24 & 6 & -15 & 0 \\
 0 & 1 & 5 & -1 & 3 & 0 \\
 0 & 0 & 1 & 0 & -2 & 1
 \end{array} & , & \begin{array}{l}
 Z(1) = Z(1) - (5/1) \cdot Z(2) \\
 Z(2) \text{ bleibt unverändert} \\
 Z(3) = Z(3)/3
 \end{array}
 \end{array} \\
 \\
 \begin{array}{rcl}
 \begin{array}{ccc|ccc}
 1 & 0 & -8 & 2 & -5 & 0 \\
 0 & 1 & 5 & -1 & 3 & 0 \\
 0 & 0 & 1 & 0 & -2 & 1
 \end{array} & , & \begin{array}{l}
 Z(1) = Z(1)/3 \\
 Z(2) \text{ bleibt unverändert} \\
 Z(3) \text{ bleibt unverändert}
 \end{array}
 \end{array} \\
 \\
 \begin{array}{rcl}
 \begin{array}{ccc|ccc}
 1 & 0 & 0 & 2 & -21 & 8 \\
 0 & 1 & 0 & -1 & 13 & -5 \\
 0 & 0 & 1 & 0 & -2 & 1
 \end{array} & , & \begin{array}{l}
 Z(1) = Z(1) - (-8/1) \cdot Z(3) \\
 Z(2) = Z(2) - (5/1) \cdot Z(3) \\
 Z(3) \text{ bleibt unverändert}
 \end{array}
 \end{array}
 \end{array}$$

Damit gilt für die inverse Matrix A_i von Matrix A :

$$\begin{array}{ccc} 2 & -21 & 8 \\ -1 & 13 & -5 \\ 0 & -2 & 1 \end{array}$$

Zur Überprüfung kann noch $A * A_i = E$ gerechnet werden.

Gleichungen mit einer unbekanntem Matrix heißt Matrizen-Gleichungen: $A * X = B$. Wenn die Matrizen quadratisch und invertierbar sind, dann gilt $X = A_i * B$, wobei A_i die inverse Matrix zu A ist, die von links multipliziert wird.
Beispiel: $A + B * X = C * X$, Lösung: $X = (C - B)_i * A$

(9) WEITERE BEISPIELE ZUM ÜBEN

Zwischen den Elementen der Matrizen wird im Folgenden immer der Strichpunkt als Separator-Zeichen geschrieben.

BSP 1: $A * x' = b'$, $x' = A_i * b'$

$$\begin{array}{cccc} 3; & 2; & 6; & 1 \\ 1; & 1; & 3; & 0 \\ -3; & -2; & -5; & -2 \end{array}$$

Lösungen: $x_1 = 1$, $x_2 = 2$, $x_3 = -1$
Inverse Systemmatrix A_i :

$$\begin{array}{ccc} 1; & -2; & 0 \\ -4; & 3; & -3 \\ 1; & 0; & 1 \end{array}$$

BSP 2: $A * x' = b'$, $x' = A_i * b'$

$$\begin{array}{cccc} 1; & 1; & 1; & 0 \\ 4; & 2; & 1; & 1 \\ 9; & 3; & 1; & 3 \end{array}$$

Lösungen: $x_1 = 0.5$, $x_2 = -0.5$, $x_3 = 0$
Inverse Systemmatrix A_i :

$$\begin{array}{ccc} 0.50; & -1.00; & 0.50 \\ -2.50; & 4.00; & -1.50 \\ 3.00; & -3.00; & 1.00 \end{array}$$

BSP 3: $A * x' = b'$, $x' = A_i * b'$

$$\begin{array}{cccc} 2; & -2; & -1; & 2 \\ 1; & 3; & -2; & 4 \\ 1; & -3; & 1; & 2 \end{array}$$

Lösungen: $x_1 = 4.67$, $x_2 = 2.00$, $x_3 = 3.33$
 Inverse Systemmatrix A_i :

-0.50; 0.83; 1.17
 -0.50; 0.50; 0.50
 -1.00; 0.67; 1.33

 BSP 4: $A * \vec{x} = \vec{b}$, $\vec{x} = A_i * \vec{b}$

3; 2; 1; 7
 0; 1; 2; 8
 3; 1; 1; 3

Lösungen: $x_1 = -1$, $x_2 = 4$, $x_3 = 2$
 Inverse Systemmatrix A_i :

-0.17; -0.17; 0.50
 1.00; 0.00; -1.00
 -0.50; 0.50; 0.50

BEISPIEL einer QUADRATISCHEN REGRESSION: 5 Messpaare $(x;y)$ sind gegeben: $(x_1;y_1)$, $(x_2;y_2)$, $(x_3;y_3)$, $(x_4;y_4)$, $(x_5;y_5)$. Gesucht ist eine Polynomfunktion $f(x) = a_0 + a_1*x + a_2*x^2$, welche sich optimal an die Messpunkte anpasst. Zur Lösung wird die Methode der kleinsten Fehlerquadrate angewendet.

Diese Aufgabe wird mit Hilfe der Matrizenrechnung gelöst. Im Menü <Bearbeiten> sind zur Ermittlung der inversen Matrix acht Nachkommastellen zu wählen. Es gilt folgende Theorie:

1; x_1 ; x_1^2 Die links stehende Systemmatrix sei M.
 1; x_2 ; x_2^2 $(y_1; y_2; y_3; y_4; y_5)$ sei der Vektor \vec{y} .
 1; x_3 ; x_3^2 $(a_0; a_1; a_2)$ sei der Vektor \vec{a} . Es gilt
 1; x_4 ; x_4^2 dann: $M * \vec{a} = \vec{y}$. Das ist ein lineares
 1; x_5 ; x_5^2 Gleichungssystem mit 5 Gleichungen und
 den 3 Unbekannten a_0, a_1, a_2 . Das System
 ist daher überbestimmt.

Weiters gelten folgende Umformungen: $M_t * M * \vec{a} = M_t * \vec{y}$
 und $\vec{a} = (M_t * M)^{-1} * M_t * \vec{y}$. Dabei ist M_t die Transponierte
 von M, und $(M_t * M)^{-1}$ ist die Inverse Matrix von $(M_t * M)$.
 Die Umformungen sind sinnvoll, weil $(M_t * M)$ eine quadratische
 Matrix ist. Die Umformungen liefern das sog. Normalsystem NS.

Beispiel: 5 Messpaare: $(-1;14)$, $(0;5)$, $(1;4)$, $(2;-1)$, $(8;26)$

System-Matrix: (5,3)-Matrix M

1; -1; 1
 1; 0; 0
 1; 1; 1
 1; 2; 4
 1; 8; 64

Vektor $\vec{y} = (14; 5; 4; -1; 26)$

Transponierte Matrix: (3,5)-Matrix Mt

1; 1; 1; 1; 1
 -1; 0; 1; 2; 8
 1; 0; 1; 4; 64

Matrizenprodukt: (3,3)-Matrix Mt * M

5; 10; 70
 10; 70; 520
 70; 520; 4114

Inverse Matrix (Mt * M)⁻¹:

0.28585366; -0.07707317; 0.00487805
 -0.07707317; 0.25479675; -0.03089431
 0.00487805; -0.03089431; 0.00406504

Matrizenprodukt (Mt * M)⁻¹ * Mt:

0.36780488; 0.28585366; 0.21365854; 0.15121952; -0.01853650
 -0.36276423; -0.07707317; 0.14682927; 0.30894309; -0.01593501
 0.03983740; 0.00487805; -0.02195122; -0.04065041; 0.01788613

Lösungsvektor $\vec{a} = (a_0; a_1; a_2) = (Mt * M)^{-1} * Mt * \vec{y}$
 6.80000000; -5.60000000; 1.00000000

Regressionsfunktion $f(x): f(x) = 6.8 - 5.6 * x + x^2$

(10) QR-Zerlegung einer Matrix

Für jede Matrix A gibt es eine so genannte QR-Zerlegung

$$A = Q * R$$

Dabei ist Q eine orthogonale Matrix und R eine obere Dreiecksmatrix. In einer orthogonalen Matrix Q sind alle Spaltenvektoren paarweise orthogonal, und zusätzlich sind sie noch normiert. Daher gilt: $Q * Q^t = E$ (Einheitsmatrix), d.h. die transponierte Matrix ist auch die inverse Matrix. In der oberen Dreiecksmatrix R sind alle Matrixelemente unterhalb der Hauptdiagonale gleich Null. Die QR-Zerlegung kann mit Hilfe verschiedener Verfahren durchgeführt werden, beispielsweise mit dem GIVENS-Verfahren.

Mit Hilfe der QR-Zerlegung können lineare Gleichungssysteme folgendermaßen gelöst werden.

Beispiel:

$$\begin{aligned} 3*x_1 + 5*x_2 + 1*x_3 &= 14 \\ 1*x_1 + 2*x_2 + 2*x_3 &= 10 \\ 2*x_1 + 4*x_2 + 5*x_3 &= 23 \end{aligned}$$

Systemmatrix A:

3; 5; 1
 1; 2; 2
 2; 4; 5

Vektor $\vec{x} = [x_1; x_2; x_3]$
 Vektor $\vec{b} = [14; 10; 23]$

Das lineare Gleichungssystem $A * \vec{x} = \vec{b}$ wird zuerst zum Normalsystem NS umgeformt, weil dadurch die inverse Matrix von einer quadratischen Matrix ($A^t * A$) gebildet werden kann. Das gilt auch, wenn die Matrix A selbst keine quadratische Matrix ist, wie beispielsweise bei den Regressionsaufgaben.

$A * \vec{x} = \vec{b}$
 $A^t * A * \vec{x} = A^t * \vec{b}$
 $\vec{x} = (A^t * A)^{-1} * A^t * \vec{b}$ (Normalsystem NS)

$A = Q * R$ (QR-Zerlegung von A)
 $A^t = R^t * Q^t$
 $\vec{x} = (A^t * A)^{-1} * A^t * \vec{y}$
 $\vec{x} = (R^t * Q^t * Q * R)^{-1} * R^t * Q^t * \vec{y}$
 $\vec{x} = (R^t * E * R)^{-1} * R^t * Q^t * \vec{y}$
 $\vec{x} = (R^t * R)^{-1} * R^t * Q^t * \vec{y}$
 $\vec{x} = R^{-1} * (R^t)^{-1} * R^t * Q^t * \vec{y}$
 $\vec{x} = R^{-1} * E * Q^t * \vec{y}$
 $\vec{x} = R^{-1} * Q^t * \vec{y}$

Die Inverse R^{-1} zur oberen Dreiecksmatrix R wird mit dem Gauß-Jordan-Verfahren ohne einen Zeilentausch berechnet, weil alle Elemente in der Diagonale von R nicht Null sind.

Das Beispiel kann als Übung mit dem Gauß-Verfahren, mit dem Normalsystem NS und mit der QR-Zerlegung durchgerechnet werden. Als Lösungsvektor \vec{x} erhält man in allen Fällen $\vec{x} = [2; 1; 3]$. Mit Hilfe der QR-Zerlegung der Matrix A kann auch die Inverse A^{-1} berechnet werden: $A^{-1} = R^{-1} * Q^t$, weil $Q^{-1} = Q^t$.

----- (11) Eigenwerte und Eigenvektoren -----

Es sei A eine quadratische Matrix. Die Zahl k und ein dazugehöriger Vektor \vec{x} heißen Eigenwert und Eigenvektor der Matrix, wenn folgende Gleichung gilt:

$$A * \vec{x} = k * \vec{x}$$

Die Multiplikation des Eigenvektors \vec{x} mit der Matrix A ergibt das k-Fache des Eigenvektors. Falls die Matrix symmetrisch ist, sind alle Eigenwerte k reelle Zahlen.

$$A * \vec{x} = k * \vec{x}$$

$$A * \vec{x} - k * \vec{x} = \vec{0} \quad (\text{Nullvektor } \vec{0})$$

$$(A - k * E) * \vec{x} = \vec{0} \quad (\text{Einheitsmatrix } E)$$

$$\begin{array}{cccccccc} a(1,1)-k & a(1,2) & a(1,3) & \dots & \dots & \dots & \dots & a(1,n) \\ a(2,1) & a(2,2)-k & a(2,3) & \dots & \dots & \dots & \dots & a(2,n) \\ a(3,1) & a(3,2) & a(3,3)-k & \dots & \dots & \dots & \dots & a(3,n) \\ \dots & \dots \\ \dots & \dots \\ a(n,1) & a(n,2) & a(n,3) & \dots & \dots & \dots & \dots & a(n,n)-k \end{array}$$

Falls es zu dieser Matrix $(A - k \cdot E)$ eine inverse Matrix gibt, dann gilt:

$$(A - k \cdot E) \cdot \vec{x} = \vec{0}$$

$$\vec{x} = (A - k \cdot E)^{-1} \cdot \vec{0} = \vec{0}$$

Der Eigenvektor \vec{x} ist hier der Nullvektor. Das ist aber sinnlos. Daher darf $(A - k \cdot E)$ nicht invertierbar sein. Dann ist diese Matrix nicht regulär, d.h. ihre Zeilenvektoren sind linear abhängig und ihre Determinante ist gleich Null, $\det(A - k \cdot E) = 0$.

Rechnet man die Determinante der Matrix aus, so erhält man ein charakteristisches Polynom n-ten Grades in den Eigenwerten k , welches den Wert Null annimmt. $P_n(k) = 0$. Somit sind die Eigenwerte die Nullstellen des charakteristischen Polynoms.

Die zu den ermittelten Eigenwerten k dazugehörigen Eigenvektoren \vec{x} erhält man, wenn man das lineare Gleichungssystem $(A - k \cdot E) \cdot \vec{x} = \vec{0}$ auflöst. Das ergibt dann den Lösungsvektor $\vec{x} = [x(1); x(2); x(3); \dots; x(n)]$.

Beispiel: (2×2) -Matrix A

$$\begin{matrix} 1; & -2 \\ 1; & 4 \end{matrix}$$

$$\begin{matrix} 1-k; & -2 \\ 1; & 4-k \end{matrix}$$

$$\det(A - k \cdot E) = P_2(k) = (1-k) \cdot (4-k) - 1 \cdot (-2) = 0$$

Das liefert die quadratische Gleichung: $k^2 - 5k + 6 = 0$.
Die zwei Lösungen davon sind: $k_1 = 2$ und $k_2 = 3$.

Den ersten Eigenvektor $\vec{x} = [x(1), x(2)]$ zum Eigenwert $k = 2$ erhält man durch Einsetzen in $(A - 2 \cdot E) \cdot \vec{x} = \vec{0}$. Das ergibt ein lineares Gleichungssystem mit linear abhängigen Zeilen:

$$\begin{aligned} -1 \cdot x(1) - 2 \cdot x(2) &= 0 \\ 1 \cdot x(1) + 2 \cdot x(2) &= 0 \end{aligned}$$

$x(1) = -2 \cdot x(2)$. Wir wählen für $x(2) = 1$. Dann ist $x(1) = -2$.
Zum Eigenwert 2 existiert ein Eigenvektor $\vec{x} = [-2; 1]$.

In analoger Weise findet man den zweiten Eigenvektor \vec{y} zum Eigenwert $k = 3$: $(A - 3 \cdot E) \cdot \vec{y} = \vec{0}$.

$$\begin{aligned} -2 \cdot y(1) - 2 \cdot y(2) &= 0 \\ 1 \cdot y(1) + 1 \cdot y(2) &= 0 \end{aligned}$$

$y(1) = -1 \cdot y(2)$. Wir wählen für $y(2) = 1$. Dann ist $y(1) = -1$.
Zum Eigenwert 3 existiert ein Eigenvektor $\vec{y} = [-1; 1]$.

Mittels Menüeintrag <Eigenwerte (2,3)> können die Eigenwerte und Eigenvektoren von beliebigen quadratischen Matrizen, deren Zeilenanzahl 2 oder 3 ist, ermittelt werden. Dabei erfolgen die Berechnungen durch Auflösung des charakteristischen Polynoms.

Mittels Menüeintrag <Eigenwerte (n)> werden die Eigenwerte und Eigenvektoren von symmetrischen (n,n)-Matrizen ermittelt. Die Berechnungen erfolgen mit Hilfe des so genannten JACOBI-Näherungsverfahrens. Es werden hier nur symmetrische Matrizen behandelt, weil diese Matrizen in der Praxis wichtig sind.

Beim JACOBI-Verfahren wird die symmetrische Matrix A in eine Diagonalmatrix Ad umgeformt, in deren Diagonale nur mehr die Eigenwerte von A stehen.

Die diagonalisierte Matrix Ad erhält man durch $Ad = U^t * A * U$. Dabei ist die Matrix U eine orthogonale Matrix, die selbst ein Produkt von orthogonalen Matrizen Uk (Jacobi-Rotationen) ist, welche die Eigenwerte von A nicht ändern: $U = U_1 * U_2 * \dots * U_k$.

Die Rotationen werden nun so lange ausgeführt bis man eine gute Näherung für die Diagonalmatrix erhält. In der Diagonale dieser Matrix Ad stehen dann die Eigenwerte von A. Das Verfahren liefert auch noch die zugehörigen Eigenvektoren, die dann in den Spalten der Matrix U stehen (bzw. in den Zeilen von Ut).

Beispiel:

Symmetrische (4*4)-Matrix A:

```
4; 1; -2; 2
1; 2; 0; 1
-2; 0; 3; -2
2; 1; -2; -1
```

Diagonal-Matrix Ad:

```
6.8446; 0.0000; 0.0000; -0.0000
0.0000; 1.0844; -0.0000; 0.0000
0.0000; 0.0000; 2.2685; 0.0000
0.0000; 0.0000; 0.0000; -2.1975
```

Eigenwerte:

```
6.8446; 1.0844; 2.2685; -2.1975
```

Eigenvektoren (Zeilen):

```
0.7180; 0.2212; -0.5574; 0.3534
-0.6423; 0.5442; -0.5202; 0.1440
0.2017; 0.7894; 0.5796; 0.0103
-0.1767; -0.1781; 0.2877; 0.9243
```

Hinweis:

Ausführliche Herleitungen des Gauß-Jordan-Verfahrens, der QR-Zerlegung, des Givens- und des Jacobi-Verfahrens findet man in den einschlägigen Fachbüchern über numerische Mathematik und lineare Algebra.

STATISTIK

| | |
|-----------------------------|--------|
| Beschreibende Statistik I | [20] |
| Beschreibende Statistik II | [23] |
| Beschreibende Statistik III | [27] |
| Regressionsanalyse | [33] |
| Faktorenanalyse | [52] |
| Häufigkeitsverteilungen | [56] |
| Normalverteilung | [58] |
| Fünf Rechenbeispiele | [61] |
| Beurteilende Statistik | [65] |
| Varianzanalyse | [68] |

Beschreibende Statistik I

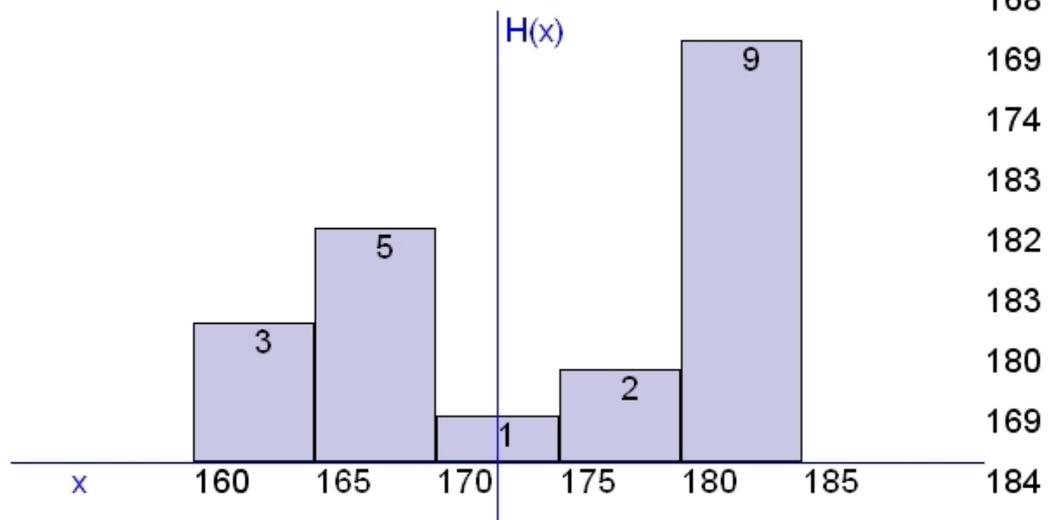
Bei der statistischen Datenauswertung wird zuerst das Datenmerkmal x (Körpergröße, Schulnote, usw.) bestimmt. Dann wird aus der Population der Merkmalsträger eine repräsentative Stichprobe (Anzahl n) herausgegriffen.

Im nächsten Schritt wird die Häufigkeit $H(x)$ der Merkmalswerte x gezählt. Zur besseren Übersicht können die Werte x auch zu wenigen Klassen zusammengefasst werden. Diese Klassen werden in der Zeichenebene auf der x -Achse aufgetragen. Ihre Häufigkeiten werden parallel zur y -Achse aufgetragen. Das ergibt die sogenannte Häufigkeitsverteilung.

Zur Beschreibung der Häufigkeitsverteilung eines Merkmals werden verschiedene Kenngrößen berechnet. Die wichtigsten sind der Mittelwert m und die Varianz v . Der Mittelwert m ist der durchschnittliche Merkmalswert und wird berechnet nach der Formel $m = \text{Datensumme} / \text{Datenanzahl} = \text{sum}(x) / n$.

Die Varianz v gibt an, wie stark die Werte x vom Mittelwert m abweichen. $v = \text{sum}((x-m)^2) / n$. Diese Formel kann umgeformt werden: $v = \text{sum}(x^2)/n - m^2$. Oft wird zur Beschreibung der Streuung die Standardabweichung $s = \text{sqrt}(v)$ verwendet.

Stichprobe von 20 Zufallszahlen zwischen 160 und 185
 Statistik mit einer Häufigkeitsverteilung mit 5 Klassen



n = Anzahl aller Datenwerte x

$H(x)$ = Häufigkeit eines Wertes x in der Stichprobe

$h(x) = H(x) / n$ = Relative Häufigkeit

$p(x) = 100 * h(x)$ = Prozentuelle Häufigkeit

$KH(a)$ = Kumulative Häufigkeit eines Wertes a ,

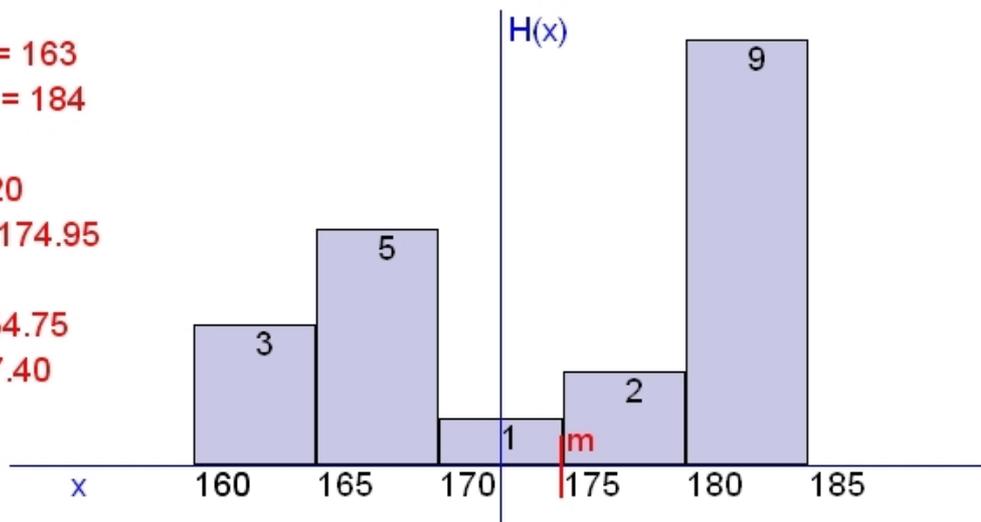
d.h. Anzahl aller Werte x mit $x \leq a$.

Stichprobe von 20 Zufallszahlen zwischen 160 und 185
 Statistik mit einer Häufigkeitsverteilung mit 5 Klassen

min = 163
 max = 184

n = 20
 m = 174.95

v = 54.75
 s = 7.40



n = Anzahl aller Datenwerte x
 sum(x) = Summe aller Werte x
 m = Mittelwert = sum(x) / n

sum(x²) = Summe aller Wertequadrate x²
 v = Varianz = Mittelwert der Abweichungsquadrate (x-m)²
 Dafür gilt: v = sum((x - m)²) / n = sum(x²) / n - m²
 s = Streuung = Wurzel aus der Varianz = sqrt(v)

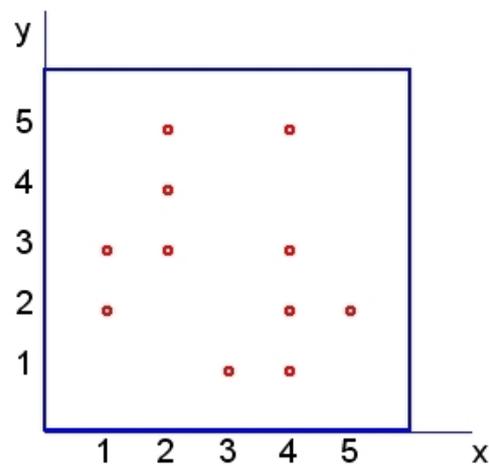
Beschreibende Statistik II

Bei der statistischen Datenauswertung können auch zwei Merkmale $(x; y)$ von n Merkmalsträgern erfasst und als Punkte eingezeichnet werden (Streuungsdiagramm). Dann berechnet man Mittelwert, Varianz und Streuung der beiden Merkmalsverteilungen (m_x, v_x, s_x) und (m_y, v_y, s_y) .

Zur Beschreibung des statistischen Zusammenhangs der beiden Merkmale werden weitere Kenngrößen benötigt. Zuerst bildet man die Abweichungsprodukte $(x - m_x) \cdot (y - m_y)$, summiert diese und dividiert durch die Anzahl n . Das liefert die sogenannte Kovarianz $v_{xy} = \text{sum}((x - m_x) \cdot (y - m_y)) / n$. Sie wird dann umgeformt zu $v_{xy} = \text{sum}(x \cdot y) / n - m_x \cdot m_y$. Diese Kovarianz beschreibt den Zusammenhang der Merkmale. Sie wird noch normiert zum Korrelationskoeffizienten r , welcher dann zwischen -1 und $+1$ liegt: $r = v_{xy} / (s_x \cdot s_y)$. r gibt an, wie gut die Merkmalspunkte $(x; y)$ im Streuungsdiagramm einem geraden (linearen) Verlauf angepasst sind.

Zuletzt wird eine Gerade $y_1 = k \cdot x + d$ ermittelt, die optimal an die Punkte $(x; y)$ angepasst ist. Für die Regressionsgerade gilt: $k = v_{xy} / v_x$ (Regressionskoeffizient) und $d = m_y - k \cdot m_x$. Mit y_1 wird zu einem bekannten x -Wert der y -Wert geschätzt.

Stichprobe von 20 Zufallspaaren von Schulnoten (x;y)
 Statistik mit einem 2-dimensionalen Streuungsdiagramm



$n =$ Anzahl aller Wertepaare (x;y)

$\text{sum}(x) =$ Summe aller Werte x

$\text{sum}(x^2) =$ Summe aller Wertequadrate x^2

$\text{sum}(y) =$ Summe aller Werte y

$\text{sum}(y^2) =$ Summe aller Wertequadrate y^2

$\text{sum}(x \cdot y) =$ Summe aller Werteprodukte $x \cdot y$

4;2

5;2

4;1

4;2

2;5

4;3

1;2

1;3

2;5

1;2

2;4

2;5

5;2

3;1

4;1

1;3

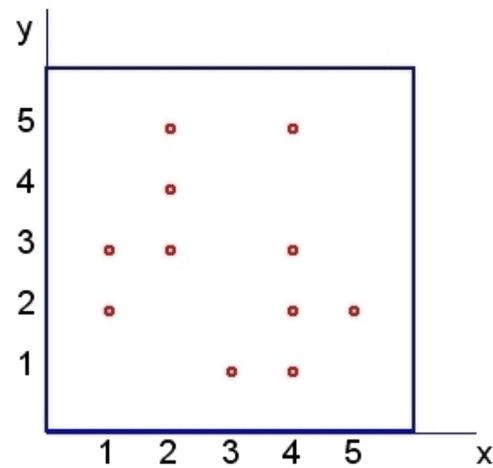
2;3

4;5

4;1

4;1

Stichprobe von 20 Zufallspaaren von Schulnoten (x;y)
 Statistik mit einem 2-dimensionalen Streuungsdiagramm



n = Anzahl der Wertepaare (x;y)

$m_x = \text{sum}(x)/n$ = Mittelwert von x

$v_x = \text{sum}(x^2)/n - m_x^2$ = Varianz von x

$s_x = \text{sqrt}(v_x)$ = Streuung von x

$m_y = \text{sum}(y)/n$ = Mittelwert von y

$v_y = \text{sum}(y^2)/n - m_y^2$ = Varianz von y

$s_y = \text{sqrt}(v_y)$ = Streuung von y

$v_{xy} = \text{Mittelwert der Abweichungsprodukte } (x-m_x)(y-m_y) =$

$= \text{Kovarianz} = \text{sum}((x-m_x)(y-m_y))/n = \text{sum}(x*y)/n - m_x*m_y$

$r = \text{Korrelationskoeffizient} = v_{xy} / (s_x*s_y)$

$k = \text{Regressionskoeffizient} = v_{xy} / v_x$

$d = \text{Regressionsabschnitt} = m_y - k*m_x$

$y_1 = k*x + d$, Regressionsgerade

4;2

5;2

4;1

4;2

2;5

4;3

1;2

1;3

2;5

1;2

2;4

2;5

5;2

3;1

4;1

1;3

2;3

4;5

4;1

4;1

Stichprobe von 20 Zufallspaaren von Schulnoten (x;y)
 Statistik mit einem 2-dimensionalen Streudiagramm

n = 20

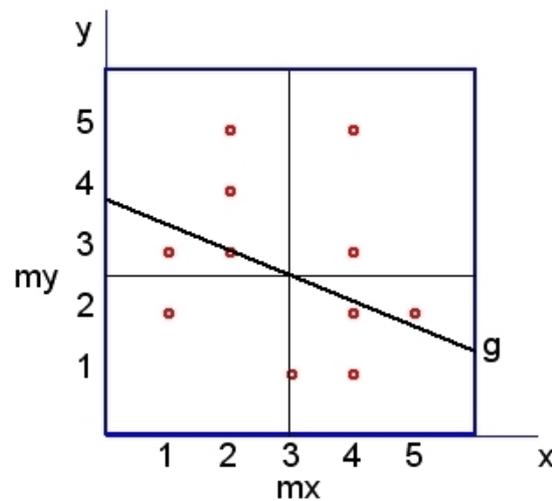
sum(x) = 59
 sum(x²) = 211
 sum(y) = 53
 sum(y²) = 181
 sum(x*y) = 141

mx = 2.95
 vx = 1.85
 sx = 1.36

my = 2.65
 vy = 2.03
 sy = 1.42

Korrelation r = -0.40
 Regression k = -0.42
 y-Abschnitt d = 3.88

Regressionsgerade g: $y_1 = -0.42 * x + 3.88$



4;2

5;2

4;1

4;2

2;5

4;3

1;2

1;3

2;5

1;2

2;4

2;5

5;2

3;1

4;1

1;3

2;3

4;5

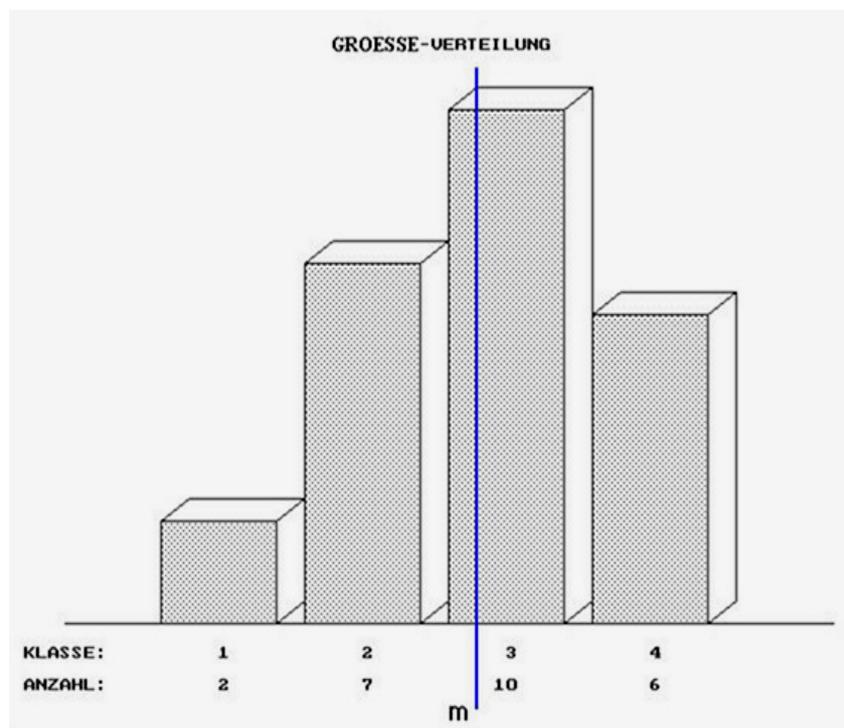
4;1

4;1

Beschreibende Statistik III

Bei statistischen Datenauswertungen wird zuerst ein Datenmerkmal x (Größe, Gewicht, ...) bestimmt. Dann wird aus der Population der Merkmalsträger eine Stichprobe (Anzahl N) herausgegriffen. Im nächsten Schritt werden die **Häufigkeiten** $H(x)$ der Merkmalswerte x gezählt. Zur besseren Übersicht können die Werte x auch zu wenigen Klassen zusammengefasst werden. Diese Klassen werden in der Zeichenebene auf der x -Achse aufgetragen. Ihre Häufigkeiten werden dann parallel zur y -Achse aufgetragen. Dadurch entsteht eine so genannte **Häufigkeitsverteilung**.

Die **kumulative Häufigkeit** $H[x \leq x_0]$ gibt an, wie viele Werte x kleiner/gleich dem Wert x_0 sind. Statt der absoluten Häufigkeit $H(x)$ kann die **relative Häufigkeit** $h(x) = H(x) / N$ betrachtet werden. Für sie gilt $0 \leq h(x) \leq 1$, und $100 * h(x)$ ist die prozentuelle Häufigkeit.



Stichprobe aus Körpergrößen von 25 Schülern

187,167,160,171,167,
182,178,178,185,156,
160,170,178,178,185,
169,167,157,165,173,
171,184,170,178,182.

Vier Merkmalsklassen
(K1, K2, K3 und K4)

$150 \leq x < 160$: 2 (8%)
 $160 \leq x < 170$: 7 (28%)
 $170 \leq x < 180$: 10 (40%)
 $180 \leq x < 190$: 6 (24%)

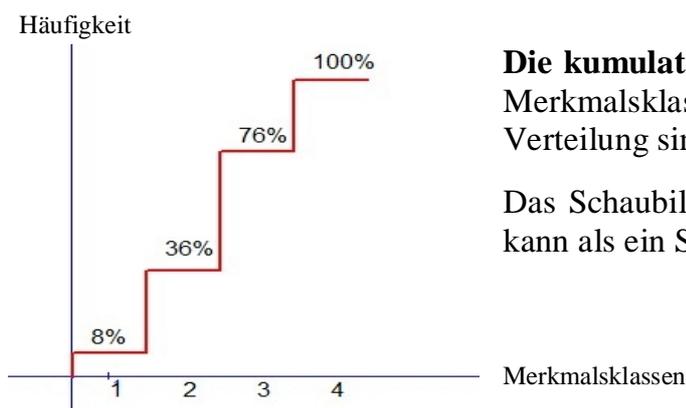
Mittelwert $m = 172.72$ cm
 Streuung $s = 8.92$ cm

Zur Beschreibung der Häufigkeitsverteilung werden bestimmte Kennwerte berechnet. die wichtigsten sind der **Mittelwert** m und die **Varianz** v .

Der **Mittelwert** m ist der durchschnittliche Wert und wird berechnet nach der Formel:
 $m = \text{Datensumme} / \text{Datenanzahl} = \text{sum}(x) / N$.

Das Symbol $\text{sum}(x)$ ist eine Abkürzung für Summe $\sum(x_i) = x_1 + x_2 + \dots + x_N$.

Die **Varianz** v gibt an, wie stark die Werte x vom Mittelwert m abweichen. Sie wird als mittleres Abweichungsquadrat vom Mittelwert m definiert: $v = \text{sum}((x - m)^2) / N$. Oft wird auch zur Beschreibung der Streuung die Standardabweichung $s = \text{sqrt}(v)$ verwendet. (sqrt = square root = Quadratwurzel).



Die kumulativen Häufigkeiten $h[x \leq x_0]$ für die vier Merkmalsklassen K_1 , K_2 , K_3 und K_4 der Größenverteilung sind: 8%, 36%, 76% und 100%.

Das Schaubild der kumulativen Häufigkeitsverteilung kann als ein Stufendiagramm dargestellt werden.

Neben dem Mittelwert m wird oft der Modus d und die drei Quartile q_1 , q_2 und q_3 zur Beschreibung einer Verteilung herangezogen. Der Modus ist der am häufigsten vorkommende Wert $d = 178$ mit $H(178) = 5$. Unterhalb des unteren Quartils q_1 liegen 25%, d.h. $h(x \leq q_1) = 1/4$. Unterhalb des mittleren Quartils q_2 (auch **Median** genannt) liegen 50%, d.h. $h(x \leq q_2) = 1/2$. Unterhalb des oberen Quartils q_3 liegen 75%, d.h. $h(x \leq q_3) = 3/4$.

Zur Ermittlung der Quartile müssen die 25 Daten der Stichprobe zuerst sortiert werden: **156, 157, 160 (2), 165, 167 (3), 169, 170 (2), 171 (2), 173, 178 (5), 182 (2), 184, 185 (2), 187.**

Ein Viertel von 25 ist 6.25. Durch einfaches Abzählen und entsprechende Interpolation erhält man dann folgende Quartile:

$$q_1 = 166.00 \text{ mit } H(x \leq 166.00) = 6.25, \text{ weil } H(x \leq 165) = 5 \text{ und } H(x \leq 167) = 8$$

$$q_2 = 170.50 \text{ mit } H(x \leq 170.50) = 12.50, \text{ weil } H(x \leq 170) = 12 \text{ und } H(x \leq 171) = 14$$

$$q_3 = 175.50 \text{ mit } H(x \leq 175.50) = 18.75, \text{ weil } H(x \leq 173) = 15 \text{ und } H(x \leq 178) = 20$$

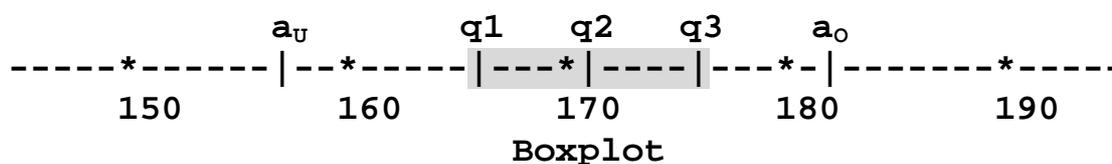
Der Boxplot

Zur Beschreibung einer Verteilung wird oft auch der so genannte Boxplot verwendet. Das ist ein Rechteck, dem 50% der Verteilung entsprechen, d.h. es reicht vom unteren Quartil q_1 bis zum oberen Quartil q_3 . Deren Abstand ($q_3 - q_1$) wird als Interquartilsabstand IQR bezeichnet. $IQR = 175.50 - 166.00 = 9.5$. Zusätzlich werden noch eine „untere Antenne“ a_U und eine „obere Antenne“ a_O bestimmt. Das sind jene zwei Datenwerte, die gerade noch innerhalb des Bereichs $[q_2 - 1.5 \cdot IQR, q_2 + 1.5 \cdot IQR]$ liegen.

$$a_U \geq q_2 - 1.5 \cdot IQR = 170.50 - 1.5 \cdot 9.5 = 170.50 - 14.25 = 156.25, \quad a_U = 157$$

$$a_O \leq q_2 + 1.5 \cdot IQR = 170.50 + 1.5 \cdot 9.5 = 170.50 + 14.25 = 184.75, \quad a_O = 182$$

Daten innerhalb der Box (50%) werden als normalwertig bezeichnet, Daten innerhalb der beiden Antennen als grenzwertig und Daten außerhalb der beiden Antennen als Ausreißer.



Herleitung der wichtigsten Formeln

Der Rechenoperator $\text{sum}(x) = \sum(x_i)$ für i von 1 bis N

Der Operator symbolisiert die Summierung von N Werten x : $\text{sum}(x) = x_1 + x_2 + \dots + x_N$.
Dafür gelten alle Regeln der Addition (assoziativ, distributiv, kommutativ), beispielsweise

$\text{sum}(k \cdot x) = k \cdot \text{sum}(x)$ mit k als konstanten Faktor oder auch
 $\text{sum}(x+y) = \text{sum}(x) + \text{sum}(y)$ mit x und y als variable Werte.

Beweis: $\text{sum}(x+y) = (x_1+y_1) + (x_2+y_2) + (x_3+y_3) + \dots + (x_N+y_N)$.
 $\text{sum}(x+y) = (x_1+x_2+\dots+x_N) + (y_1+y_2+\dots+y_N) = \text{sum}(x) + \text{sum}(y)$.

Der Mittelwert $m = \text{sum}(x) / N$

Mittelwert $m = \text{Datensumme} / \text{Datenanzahl} = \text{sum}(x) / N$. Man kann zeigen, dass die Summe der Abweichungen vom Mittelwert $\text{sum}(x - m)$ Null ist.

Beweis: $\text{sum}(x - m) = \text{sum}(x) - \text{sum}(m) = N \cdot m - N \cdot m = 0$.

Neben dem Mittelwert ist der Median ein anderes Zentralmaß. Unterhalb und oberhalb des Median liegen 50% aller Werte x .

Die Varianz $v = \text{sum}((x - m)^2) / N$

Die Varianz $v = \text{sum}((x - m)^2) / N$ ist der Mittelwert aller Abweichungsquadrate $(x - m)^2$ vom Mittelwert m . Sie ist umso größer, je stärker die Werte x vom Mittelwert m abweichen. Um sie in obiger Form zu berechnen, muss der Mittelwert m bekannt sein - was aber bei der Datenerfassung der Werte x nicht der Fall ist. Durch Umformung kann die Varianz v in die äquivalente Formel $v = \text{sum}(x^2) / N - m^2$ umgewandelt werden. In der Praxis werden bei der Datenerfassung $\text{sum}(x)$ und $\text{sum}(x^2)$ schrittweise gebildet und zum Schluss die Varianz berechnet. Als Streuungsmaß gilt auch die Standardabweichung $s = \text{sqrt}(v)$.

Beweis der Äquivalenz der Formeln:

$\text{sum}((x - m)^2) = \text{sum}(x^2 - 2 \cdot m \cdot x + m^2) = \text{sum}(x^2) - 2 \cdot m \cdot \text{sum}(x) + N \cdot m^2$.
 $\text{sum}((x - m)^2) = \text{sum}(x^2) - 2 \cdot m \cdot N \cdot m + N \cdot m^2 = \text{sum}(x^2) - N \cdot m^2$.

Die Kovarianz $v_{xy} = \text{sum}((x - m_x) \cdot (y - m_y)) / N$

Wenn von den Merkmalsträgern in einer Stichprobe 2 Merkmale erfasst werden (x und y), dann erhält man reelle Messpaare (x/y) , die als Punkte in der Zeichenebene aufgetragen werden. Die Kovarianz ein Maß für den Zusammenhang der beiden Merkmale. Sie ist der Mittelwert der Abweichungsprodukte $(x - m_x) \cdot (y - m_y)$.

Die Kovarianz ist: $v_{xy} = \text{sum}((x - m_x) \cdot (y - m_y)) / N$.

Ist der Betrag von v_{xy} groß und das Vorzeichen positiv, ändern sich die Merkmale stark in der gleichen Richtung.

Ist der Betrag von v_{xy} groß und das Vorzeichen negativ, ändern sich die Merkmale stark in der gegensätzlichen Richtung.

Ist der Betrag von v_{xy} klein, dann liegen die Punkte (x/y) regellos ohne Zusammenhang in der Koordinatenebene.

So wie die Varianz wird die Kovarianz umgeformt in die **äquivalente Formel**:

$$v_{xy} = \text{sum}((x - m_x) \cdot (y - m_y)) / N = \text{sum}(x \cdot y) / N - m_x \cdot m_y.$$

Beweis:

$$\begin{aligned} \text{sum}((x - m_x) \cdot (y - m_y)) &= \text{sum}(x \cdot y - m_x \cdot y - m_y \cdot x + m_x \cdot m_y) = \\ &= \text{sum}(x \cdot y) - m_x \cdot \text{sum}(y) - m_y \cdot \text{sum}(x) + N \cdot m_x \cdot m_y = \\ &= \text{sum}(x \cdot y) - N \cdot m_x \cdot m_y - N \cdot m_x \cdot m_y + N \cdot m_x \cdot m_y. \text{ Also gilt} \\ v_{xy} &= \text{sum}((x - m_x) \cdot (y - m_y)) / N = \text{sum}(x \cdot y) / N - m_x \cdot m_y \end{aligned}$$

Die lineare Regression $k = v_{xy} / v_x$ von y auf x

Die Regressionsgerade $g: y_1 = k \cdot x + d$ ist eine Gerade, die so verläuft, dass sie optimal an die Merkmalspunkte (x/y) angepasst ist, und y_1 ein optimaler Schätzwert für y ist.

Man quadriert alle Schätzfehler $(y_1 - y)$ und summiert sie. $F = \text{sum}((y_1 - y)^2) =$ Summe der Fehlerquadrate. Diese soll möglichst klein (ein Minimum) sein.

Summe aller Fehlerquadrate = $F = \text{sum}((y_1 - y)^2) = \text{Minimum}$

$F = \text{sum}((k \cdot x + d - y)^2)$. Die Gerade $y_1 = k \cdot x + d$ soll durch den Punkt $P(m_x/m_y)$ gehen, so dass sie möglichst zentral in der Menge der Merkmalspunkte (x,y) verankert ist. Daraus folgt die Nebenbedingung $m_y = k \cdot m_x + d$ und $d = m_y - k \cdot m_x$.

$$\begin{aligned} F &= \text{sum}((k \cdot x + m_y - k \cdot m_x - y)^2) = \text{sum}((k \cdot (x - m_x) - (y - m_y))^2). \\ F &= \text{sum}(k^2(x - m_x)^2 - 2 \cdot k \cdot (x - m_x) \cdot (y - m_y) + (y - m_y)^2). \\ F &= k^2 \cdot \text{sum}((x - m_x)^2) - 2 \cdot k \cdot \text{sum}((x - m_x) \cdot (y - m_y)) + \text{sum}((y - m_y)^2). \\ F &= k^2 \cdot N \cdot v_x - 2 \cdot k \cdot N \cdot v_{xy} + N \cdot v_y. \text{ Das ist eine Funktion in } k. \end{aligned}$$

$$F(k) = k^2 \cdot N \cdot v_x - 2 \cdot k \cdot N \cdot v_{xy} + N \cdot v_y.$$

Diese Funktion wird nun nach der Variablen k differenziert.

$F'(k) = 2 \cdot N \cdot v_x \cdot k - 2 \cdot N \cdot v_{xy}$ und $F'(k) = 0$ liefert dann mit $k = v_{xy} / v_x$ den so genannten Regressionskoeffizient k , der den Anstieg der Regressionsgeraden $y_1 = k \cdot x + d$ bestimmt.

Die Regressionsgerade $y_1 = k \cdot x + d$ ist vollständig bestimmt durch $k = v_{xy} / v_x$ und $d = m_y - k \cdot m_x$.

Auf diese Weise wurde eine Gerade g in der Menge der Merkmalspunkte (x/y) ermittelt, die optimal angepasst ist.

Der Korrelationskoeffizient $r = v_{xy} / (s_x * s_y)$

Setzt man den Koeffizienten $k = v_{xy} / v_x$ in die Summe der Fehlerquadrate

$F(k) = k^2 * N * v_x - 2 * k * N * v_{xy} + N * v_y$ ein, dann erhält man:

$$F(k) = N * v_{xy}^2 / v_x - 2 * N * v_{xy}^2 / v_x + N * v_y.$$

$$F = N / v_x * (v_x * v_y - v_{xy}^2).$$

Als Summe von Quadraten ist $F \geq 0$. Also muss $(v_x * v_y - v_{xy}^2) \geq 0$, d.h. $v_x * v_y \geq v_{xy}^2$.

Dann gilt $v_{xy}^2 / (v_x * v_y) \leq 1$. Nun auf beiden Seiten die Wurzel ziehen. $|v_{xy} / (s_x * s_y)| \leq 1$.

Das ist der Korrelationskoeffizient r , wo $s_x = \sqrt{v_x}$ und $s_y = \sqrt{v_y}$ die Streuungen sind.

Diese Normierung der Kovarianz v_{xy} zum Korrelationskoeffizient r , erzeugt ein Zusammenhangsmaß, für das gilt $|r| \leq 1$, d.h. $-1 \leq r \leq +1$.

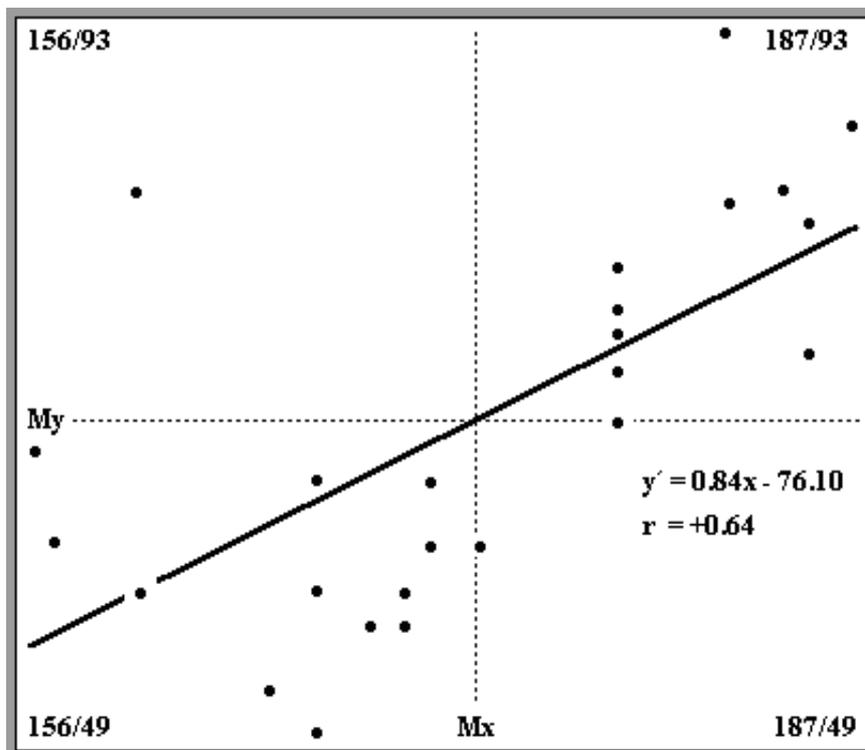
Für den Korrelationskoeffizient $r = v_{xy} / (s_x * s_y)$ gilt $|r| \leq 1$.

Wenn alle Merkmalspunkte (x,y) auf der Regressionsgeraden liegen, dann muss die Summe der Fehlerquadrate F gleich 0 sein. Also $v_x * v_y - v_{xy}^2 = 0$, d.h. $v_{xy}^2 / (v_x * v_y) = 1$ und nach dem Wurzelziehen gilt für den Korrelationskoeffizient $|r| = 1$.

Es liegt ein **idealer linearer** Zusammenhang (Korrelation) vor.

Wenn die Merkmalspunkte (x,y) völlig regellos in der Ebene liegen, dann ist die Kovarianz $v_{xy} = 0$ und dann gilt $|r| = 0$.

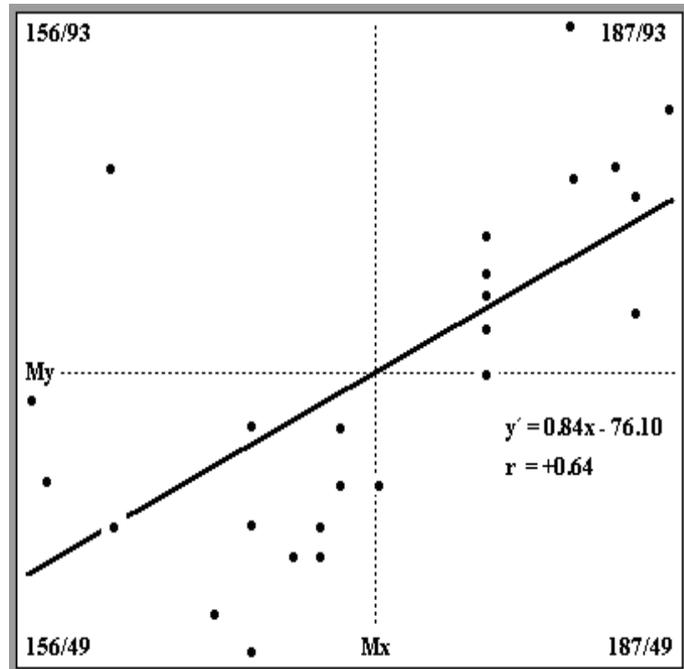
Es liegt **kein linearer** Zusammenhang (Korrelation) vor.



Beispiel eines Streuungsdiagramms mit Korrelation und Regressionsgeraden.

Von $N = 25$ Schülern einer Klasse werden die Körpergröße x (cm) und das Gewicht y (kg) erfasst. Zuerst werden die Verteilungsparameter berechnet und dann eine Regressionsanalyse durchgeführt. Zum Schluss werden in einem zweidimensionalen Streuungsdiagramm die 25 Datenpaare eingetragen und die Regressionsgerade gezeichnet.

| X | Y | X ² | Y ² | XY |
|-----|----|----------------|----------------|-------|
| 187 | 87 | 34969 | 7569 | 16269 |
| 167 | 65 | 27889 | 4225 | 10855 |
| 160 | 83 | 25600 | 6889 | 13280 |
| 171 | 61 | 29241 | 3721 | 10431 |
| 167 | 49 | 27889 | 2401 | 8183 |
| 182 | 82 | 33124 | 6724 | 14924 |
| 178 | 74 | 31684 | 5476 | 13172 |
| 178 | 76 | 31684 | 5776 | 13528 |
| 185 | 81 | 34225 | 6561 | 14985 |
| 156 | 67 | 24336 | 4489 | 10452 |
| 160 | 58 | 25600 | 3364 | 9280 |
| 170 | 58 | 28900 | 3364 | 9860 |
| 178 | 78 | 31684 | 6084 | 13884 |
| 178 | 72 | 31684 | 5184 | 12816 |
| 185 | 73 | 34225 | 5329 | 13505 |
| 169 | 56 | 28561 | 3136 | 9464 |
| 167 | 58 | 27889 | 3364 | 9686 |
| 157 | 61 | 24649 | 3721 | 9577 |
| 165 | 52 | 27225 | 2704 | 8580 |
| 173 | 61 | 29929 | 3721 | 10553 |
| 171 | 65 | 29241 | 4225 | 11115 |
| 184 | 83 | 33856 | 6889 | 15272 |
| 170 | 56 | 28900 | 3136 | 9520 |
| 178 | 69 | 31684 | 4761 | 12282 |
| 182 | 93 | 33124 | 8649 | 16926 |



Anzahl: $N = 25$

Summen: $\sum x = 4318$, $\sum y = 1718$

Summen: $\sum x^2 = 747792$, $\sum y^2 = 121462$, $\sum xy = 298399$

Mittelwerte: $m_x = 172.72$, $m_y = 68.72$

Streuungen: $s_x = 8.92$, $s_y = 11.66$

Kovarianz: $v_{xy} = 66.56$

Ergebnis:

Korrelation: $r = v_{xy} / (s_x \cdot s_y) = +0.64$

Regression: $k = v_{xy} / (s_x)^2 = +0.84$

Gerade: $y_1 = 0.84 \cdot x - 76.10$

Beispiel einer Regressionsschätzung:

Wirkliches Wertepaar: $(x/y) = (157,61)$. d.h. $y = 61$

Geschätzter Y-Wert: $y_1 = 0.84 \cdot 157 - 76.10 = 58$

REGRESSIONSANALYSE

- (I) Methode der kleinsten Fehlerquadrate
- (II) Polynomiale Regression mit Matrizen
- (III) Lineare Regression
- (IV) Exponentielle Regression

(I) Methode der kleinsten Fehlerquadrate

Gegeben sind n Wertepaare $(X_i; Y_i)$ mit $i = 1, \dots, n$.
 Gesucht ist eine Funktion $f(X) = f(X, a_0, a_1, \dots, a_m)$, welche von $m+1$ Parametern abhängt, beispielsweise die Polynomfunktion $f(X) = a_0 + a_1 \cdot X + a_2 \cdot X^2 + \dots + a_m \cdot X^m$. Diese Funktion soll optimal an die n Punkte $(X_i; Y_i)$ angepasst werden.

Dazu bildet man die Summe S aller Abweichungsquadrate $S = \sum (f(X_i) - Y_i)^2$. Dabei ist $(f(X_i) - Y_i)$ die i -te Abweichung (Fehler). Die Summe S soll nun möglichst klein sein. Das ist der Fall, wenn ihre $m+1$ partiellen Ableitungen $\delta S / \delta a_k$ nach den unbekanntem Parametern a_0, a_1, \dots, a_m gleich Null sind.

Es muss daher gelten: $\delta S / \delta a_k = 0$, für $k = 0, 1, \dots, m$.
 Bei der partiellen Ableitungen $\delta S / \delta a_k$ ist nur a_k die Variable.

$$S = \sum (f(X_i) - Y_i)^2 = \sum (a_0 + a_1 \cdot X_i + \dots + a_m \cdot X_i^m - Y_i)^2$$

$\delta S / \delta a_k = \sum (2 \cdot (f(X_i) - Y_i) \cdot \delta f(X_i) / \delta a_k)$, wegen Kettenregel.
 Für die partielle Ableitung von S nach a_k gilt: $\delta f(X_i) / \delta a_k = X_i^k$

$$\delta S / \delta a_k = 2 \cdot \sum (a_0 + a_1 \cdot X_i + \dots + a_m \cdot X_i^m - Y_i) \cdot X_i^k = 0$$

Das liefert $m+1$ lineare Gleichungen:

$$\delta S / \delta a_0 = 2 \cdot \sum (a_0 + a_1 \cdot X_i + \dots + a_m \cdot X_i^m - Y_i) \cdot X_i^0 = 0$$

$$\delta S / \delta a_1 = 2 \cdot \sum (a_0 + a_1 \cdot X_i + \dots + a_m \cdot X_i^m - Y_i) \cdot X_i^1 = 0$$

$$\dots$$

$$\dots$$

$$\dots$$

$$\delta S / \delta a_m = 2 \cdot \sum (a_0 + a_1 \cdot X_i + \dots + a_m \cdot X_i^m - Y_i) \cdot X_i^m = 0$$

Durch Umformungen erhält man das lineare Gleichungssystem:

$$\begin{aligned} a_0 \cdot \Sigma(X_i^0) + a_1 \cdot \Sigma(X_i^1) + a_2 \cdot \Sigma(X_i^2) + a_3 \cdot \Sigma(X_i^3) + \dots + a_m \cdot \Sigma(X_i^m) &= \Sigma(X_i^0 Y_i) \\ a_0 \cdot \Sigma(X_i^1) + a_1 \cdot \Sigma(X_i^2) + a_2 \cdot \Sigma(X_i^3) + a_3 \cdot \Sigma(X_i^4) + \dots + a_m \cdot \Sigma(X_i^{m+1}) &= \Sigma(X_i^1 Y_i) \\ \dots & \\ \dots & \\ a_0 \cdot \Sigma(X_i^m) + a_1 \cdot \Sigma(X_i^{m+1}) + a_2 \cdot \Sigma(X_i^{m+2}) + \dots + a_m \cdot \Sigma(X_i^{2m}) &= \Sigma(X_i^m Y_i) \end{aligned}$$

Man führt nun folgende zwei neue Bezeichnungen s_k und t_k ein:

$$(I) \quad s_k = \Sigma(X_i^k) \quad \text{mit } i = 1, \dots, n \quad \text{und } k = 0, \dots, 2m$$

$$(II) \quad t_k = \Sigma(X_i^k \cdot Y_i) \quad \text{mit } i = 1, \dots, n \quad \text{und } k = 0, \dots, m$$

Damit ergibt sich nachfolgendes lineares Gleichungssystem für die $m+1$ unbekanntenen Polynomkoeffizienten a_k ($k = 0, \dots, m$).

Diese Gleichungen werden auch **Normalgleichungen NGL** genannt.

$$\begin{aligned} s_0 \cdot a_0 + s_1 \cdot a_1 + \dots + s_m \cdot a_m &= t_0 \\ s_1 \cdot a_0 + s_2 \cdot a_1 + \dots + s_{(m+1)} \cdot a_m &= t_1 \\ s_2 \cdot a_0 + s_3 \cdot a_1 + \dots + s_{(m+2)} \cdot a_m &= t_2 \\ \dots & \\ \dots & \\ s_m \cdot a_0 + s_{(m+1)} \cdot a_1 + \dots + s_{2m} \cdot a_m &= t_m \end{aligned}$$

Damit ist das Näherungspolynom $f(x)$ vom m -ten Grad, das optimal durch die n Wertepaare $(X_i; Y_i)$ gelegt werden kann, vollständig bestimmt. Die Formeln (I) und (II) werden zur schrittweisen Berechnung der polynomialen Regression verwendet.

(II) Polynomiale Regression mit Matrizen

Gegeben sind n Wertepaare $(X_i; Y_i)$ mit $i = 1, \dots, n$.
 Gesucht ist ein Polynom m -ten Grades $f(X)$ mit $m < n$,
 $f(X) = a_0 + a_1 \cdot X + a_2 \cdot X^2 + a_3 \cdot X^3 + \dots + a_m \cdot X^m$.

Dabei soll die Summe der Fehlerquadrate $\Sigma((f(X_i) - Y_i)^2)$
 ein Minimum werden.

Einsetzen der n Wertepaare in ein lineares Gleichungssystem
 mit den $m+1$ unbekanntem Koeffizienten $a_0, a_1, a_2, \dots, a_m$.

$$\begin{aligned} Y_1 &= a_0 + a_1 \cdot X_1 + a_2 \cdot X_1^2 + a_3 \cdot X_1^3 + \dots + a_m \cdot X_1^m \\ Y_2 &= a_0 + a_1 \cdot X_2 + a_2 \cdot X_2^2 + a_3 \cdot X_2^3 + \dots + a_m \cdot X_2^m \\ &\dots \\ &\dots \\ Y_n &= a_0 + a_1 \cdot X_n + a_2 \cdot X_n^2 + a_3 \cdot X_n^3 + \dots + a_m \cdot X_n^m \end{aligned}$$

Dieses Gleichungssystem ist überbestimmt, weil $m < n$ ist.

x' ist der Vektor $[X_1, X_2, \dots, X_n]$

y' ist der Vektor $[Y_1, Y_2, \dots, Y_n]$

f' ist der Vektor der Näherungswerte $[f(X_1), f(X_2), \dots, f(X_n)]$,
 die man dann erhält, wenn man die gegebenen X -Werte in das
 Polynom $f(X)$ einsetzt.

a' ist der Vektor der Polynomkoeffizienten $[a_0, a_1, a_2, \dots, a_m]$

M ist folgende $n \cdot (m+1)$ - Matrix:

$$\begin{aligned} &1 \quad X_1 \quad X_1^2 \quad X_1^3 \quad \dots \quad X_1^m \\ &1 \quad X_2 \quad X_2^2 \quad X_2^3 \quad \dots \quad X_2^m \\ &\dots \\ &\dots \\ &1 \quad X_n \quad X_n^2 \quad X_n^3 \quad \dots \quad X_n^m \end{aligned}$$

Das lineare Gleichungssystem lautet dann: $M \cdot a' = y'$

Die Lösungen des Gleichungssystems $M \cdot a' = y'$ sind Lösungen des so genannten Normalsystems $M^T \cdot M \cdot a' = M^T \cdot y'$, wobei die Matrix M^T die transponierte Matrix von M ist. Die transponierte Matrix erhält man, wenn man in einer Matrix die Zeilen mit den Spalten vertauscht. Das Matrizen-Produkt $M^T \cdot M$ ist eine quadratische, symmetrische Matrix. Das System $M^T \cdot M \cdot a' = M^T \cdot y'$ hat dann eine eindeutige Lösung a' , wenn nicht alle Wertepaare $(X_i; Y_i)$ auf einer zur X-Achse senkrechten Geraden liegen. Dann gilt die Lösungsformel: $a' = (M^T \cdot M)^{-1} \cdot M^T \cdot y'$. Dabei ist $(M^T \cdot M)^{-1}$ die inverse Matrix zu $(M^T \cdot M)$. Die inverse Matrix einer quadratischen, regulären Matrix wird am einfachsten mit dem Gauß-Jordan-Algorithmus berechnet.

Eine Aufspaltung der Lösungsgleichung $a' = (M^T \cdot M)^{-1} \cdot M^T \cdot y'$ in die $m+1$ Gleichungszeilen, liefert nach Ausrechnung dann die bereits oben hergeleiteten Normalgleichungen NGL.

Die höhere Mathematik geht nun noch einen Schritt weiter. Für jede Matrix M gibt es eine so genannte **QR-Zerlegung** $M = Q \cdot R$. Dabei ist Q eine orthogonale Matrix und R eine obere Dreiecksmatrix. In einer orthogonalen Matrix Q sind alle Spaltenvektoren paarweise orthogonal und können zusätzlich auch normiert sein. Es gilt $Q \cdot Q^T = E$, wobei E die Einheitsmatrix ist. In der oberen Dreiecksmatrix R sind alle Elemente unterhalb der Hauptdiagonale gleich Null. Die QR-Zerlegung wird mit Hilfe verschiedenartiger Verfahren durchgeführt, beispielsweise mit der Givens-Rotation. Man kann die QR-Zerlegung bei Regressionsaufgaben (bzw. bei linearen Gleichungssystemen) anwenden:

$$\begin{aligned} M &= Q \cdot R \\ M^T &= R^T \cdot Q^T \end{aligned}$$

$$\begin{aligned} a' &= (M^T \cdot M)^{-1} \cdot M^T \cdot y' \\ a' &= (R^T \cdot Q^T \cdot Q \cdot R)^{-1} \cdot R^T \cdot Q^T \cdot y' \\ a' &= (R^T \cdot E \cdot R)^{-1} \cdot R^T \cdot Q^T \cdot y' \\ a' &= (R^T \cdot R)^{-1} \cdot R^T \cdot Q^T \cdot y' \\ a' &= R^{-1} \cdot (R^T)^{-1} \cdot R^T \cdot Q^T \cdot y' \\ a' &= R^{-1} \cdot E \cdot Q^T \cdot y' \end{aligned}$$

$$a' = R^{-1} \cdot Q^T \cdot y'$$

Die Anwendung der QR-Zerlegung anstelle der Normalgleichungen NGL führt bei Rechnungen mit dem Computer häufig zu genaueren Resultaten.

(III) Lineare Regression

Gegeben sind n Wertepaare $(X_i; Y_i)$ mit $i = 1, \dots, n$.
Gesucht ist eine lineare Funktion $f(X) = a_0 + a_1 \cdot X$, welche sich optimal an die n Punkte $(X_i; Y_i)$ anpasst.

Die zwei unbekannt Parameter a_0 und a_1 werden mit Hilfe der Methode der kleinsten Fehlerquadrate ermittelt. Die entsprechenden zwei Normalgleichungen lauten dann:

$$(G1) \quad a_0 \cdot \sum(X_i^0) + a_1 \cdot \sum(X_i^1) = \sum(X_i^0 Y_i)$$

$$(G2) \quad a_0 \cdot \sum(X_i^1) + a_1 \cdot \sum(X_i^2) = \sum(X_i^1 Y_i)$$

$$(G1) \quad a_0 \cdot n + a_1 \cdot \sum(X_i) = \sum Y_i$$

$$(G2) \quad a_0 \cdot \sum(X_i) + a_1 \cdot \sum(X_i^2) = \sum(X_i \cdot Y_i)$$

In der beschreibenden Statistik gelten folgende Kenngrößen:

$$\text{Mittelwerte } m_X = \sum(X_i) / n, \quad m_Y = \sum(Y_i) / n$$

$$\text{Varianzen } v_X = \sum(X_i^2) / n - m_X^2, \quad v_Y = \sum(Y_i^2) / n - m_Y^2$$

$$\text{Kovarianz } v_{XY} = \sum(X_i \cdot Y_i) / n - m_X \cdot m_Y$$

Setzt man die statistischen Kenngrößen in die zwei Normalgleichungen ein und ermittelt die Lösungen, dann erhält man:

$$(G1) \quad a_0 \cdot n + a_1 \cdot n \cdot m_X = n \cdot m_Y$$

$$a_0 + a_1 \cdot m_X = m_Y$$

$$a_0 = m_Y - a_1 \cdot m_X$$

$$(G2) \quad a_0 \cdot n \cdot m_X + a_1 \cdot n \cdot (v_X + m_X^2) = n \cdot (v_{XY} + m_X \cdot m_Y)$$

$$a_0 \cdot m_X + a_1 \cdot (v_X + m_X^2) = v_{XY} + m_X \cdot m_Y$$

$$(m_Y - a_1 \cdot m_X) \cdot m_X + a_1 \cdot (v_X + m_X^2) = v_{XY} + m_X \cdot m_Y$$

$$m_X \cdot m_Y - a_1 \cdot m_X^2 + a_1 \cdot v_X + a_1 \cdot m_X^2 = v_{XY} + m_X \cdot m_Y$$

$$a_1 \cdot v_X = v_{XY}$$

$$a_1 = v_{XY} / v_X$$

Für die gesuchte lineare Funktion $f(X) = a_0 + a_1 \cdot X$ (Regressionsgerade) gilt dann: $a_1 = v_{XY} / v_X$ (Regressionskoeffizient) und $a_0 = m_Y - a_1 \cdot m_X$. Diese Gerade passt sich optimal an die gegebenen n Punkte $(X_i; Y_i)$ an.

Der Korrelationskoeffizient $r_{XY} = v_{XY} / \sqrt{v_X \cdot v_Y}$ ist ein Maß für die Stärke des linearen Zusammenhanges der beiden Messgrößen X_i und Y_i . Für den Korrelationskoeffizienten gilt: $-1 \leq r_{XY} \leq +1$.

(IV) Exponentielle Regression

Gegeben sind n Wertepaare $(X_i; Y_i)$ mit $i = 1, \dots, n$.
Gesucht ist eine Exponentialfunktion $f(X) = a \cdot \exp(b \cdot X)$,
welche sich optimal an die n Punkte $(X_i; Y_i)$ anpasst.

Unter der Voraussetzung, dass alle $Y_i > 0$ sind, wird durch eine
logarithmische Transformation eine lineare Funktion $g(x)$ erzeugt:

$$\begin{aligned}f(X) &= a \cdot \exp(b \cdot X) \\g(X) &= \ln(f(X)) = \ln(a) + b \cdot X \\g(X) &= a_0 + a_1 \cdot X \\ \text{Es gilt: } a_0 &= \ln(a) \text{ und } a_1 = b\end{aligned}$$

Zuerst müssen alle Y_i durch $\ln(Y_i)$ ersetzt werden. Dann wird die
lineare Regressionsfunktion $g(X) = a_0 + a_1 \cdot X$ mit der Methode der
kleinsten Fehlerquadrate ermittelt. Dann erfolgt eine Rücktrans-
formation durch $a = \exp(a_0)$ und $b = a_1$. Zuletzt wird damit die
exponentielle Regressionsfunktion $f(x) = a \cdot \exp(b \cdot X)$ gebildet.

Zusätzlich kann bei allen Regressionen auch der mittlere absolute
Fehler $e = 1/n \cdot \sum |f(X_i) - Y_i|$ berechnet werden. Dieser ist ein Maß
für die Güte der Anpassung von $f(x)$ an die Punkte $(X_i; Y_i)$.

Hinweis: Das Programm **STATIS.EXE** ermöglicht die Ermittlung
von beliebigen Regressionsfunktionen zu eingegebenen Zahlen-
paaren von Messwerten. In dem Programm **MATRIX.EXE** können
alle grundlegenden Operationen mit Matrizen einfach und bequem
durchgeführt werden.

Beispiel: Lineare Regression

Zu n Zahlenpaaren $(x ; y)$
wird eine optimal angepasste
Polynomfunktion 1. Grades
 $f(x) = k \cdot x + d$ ermittelt.
(Regressionsgerade)

Ein Beispiel mit 10 Zahlenpaaren: $(5;10)$, $(10;25)$, $(20;30)$, $(40;28)$, $(55;20)$,
 $(60;20)$, $(70;30)$, $(80;40)$, $(85;50)$, $(95;80)$.
Ausrechnung mit dem Programm „**STATIS.EXE**“.

Zweidimensionale Statistik von n Zahlenpaaren

Anzahl $n = 10$

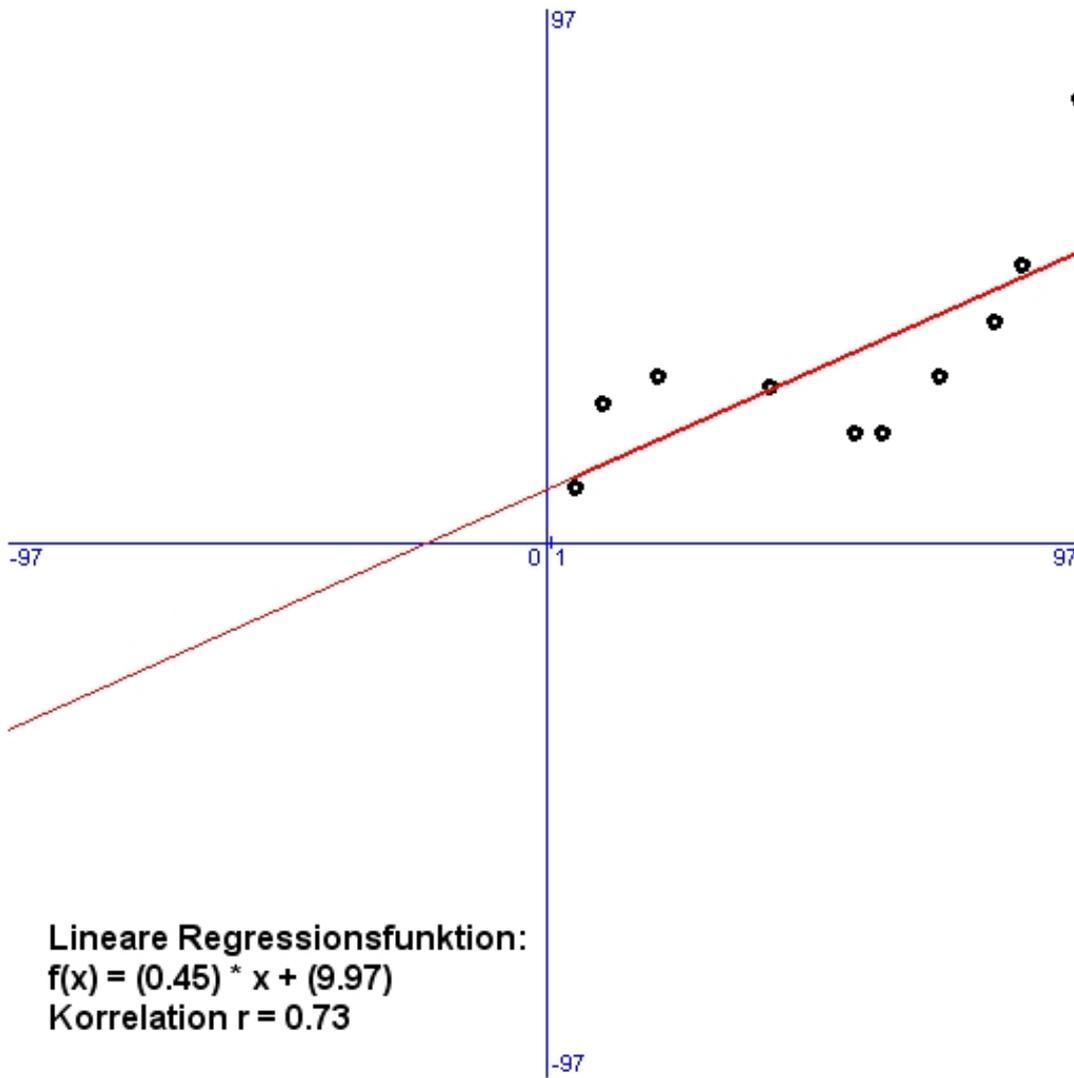
$\text{sum}(x) = 520$
 $\text{sum}(x^2) = 36300$
 $\text{sum}(y) = 333$
 $\text{sum}(y^2) = 14609$
 $\text{sum}(x \cdot y) = 21470$

Mittelwert $M_x = 52$
Varianz $V_x = 926$
Streuung $S_x = 30.43$

Mittelwert $M_y = 33.30$
Varianz $V_y = 352.01$
Streuung $S_y = 18.76$

Kovarianz $V_{xy} = 415.40$

Korrelation $r = 0.73$
Regression $k = 0.45$
 y -Abschnitt $d = 9.97$
Regressionsgerade $g: f(x) = (0.45) \cdot x + (9.97)$



Beispiel: Quadratische Regression

Zu n Zahlenpaaren $(x ; y)$
wird eine optimal angepasste
Polynomfunktion 2. Grades
 $f(x) = a \cdot x^2 + b \cdot x + c$ ermittelt.

Ein Beispiel mit 10 Zahlenpaaren: $(5;10)$, $(10;25)$, $(20;30)$, $(40;28)$, $(55;20)$,
 $(60;20)$, $(70;30)$, $(80;40)$, $(85;50)$, $(95;80)$.
Ausrechnung mit dem Programm „**STATIS.EXE**“.

Quadratische Regression von 10 Zahlenpaaren Lösungsmethode mit 3 Normalgleichungen

$$\begin{aligned}10 \cdot x + 520 \cdot y + 36300 \cdot z &= 333 \\520 \cdot x + 36300 \cdot y + 2782000 \cdot z &= 21470 \\36300 \cdot x + 2782000 \cdot y + 223462500 \cdot z &= 1678300\end{aligned}$$

Lösungen:

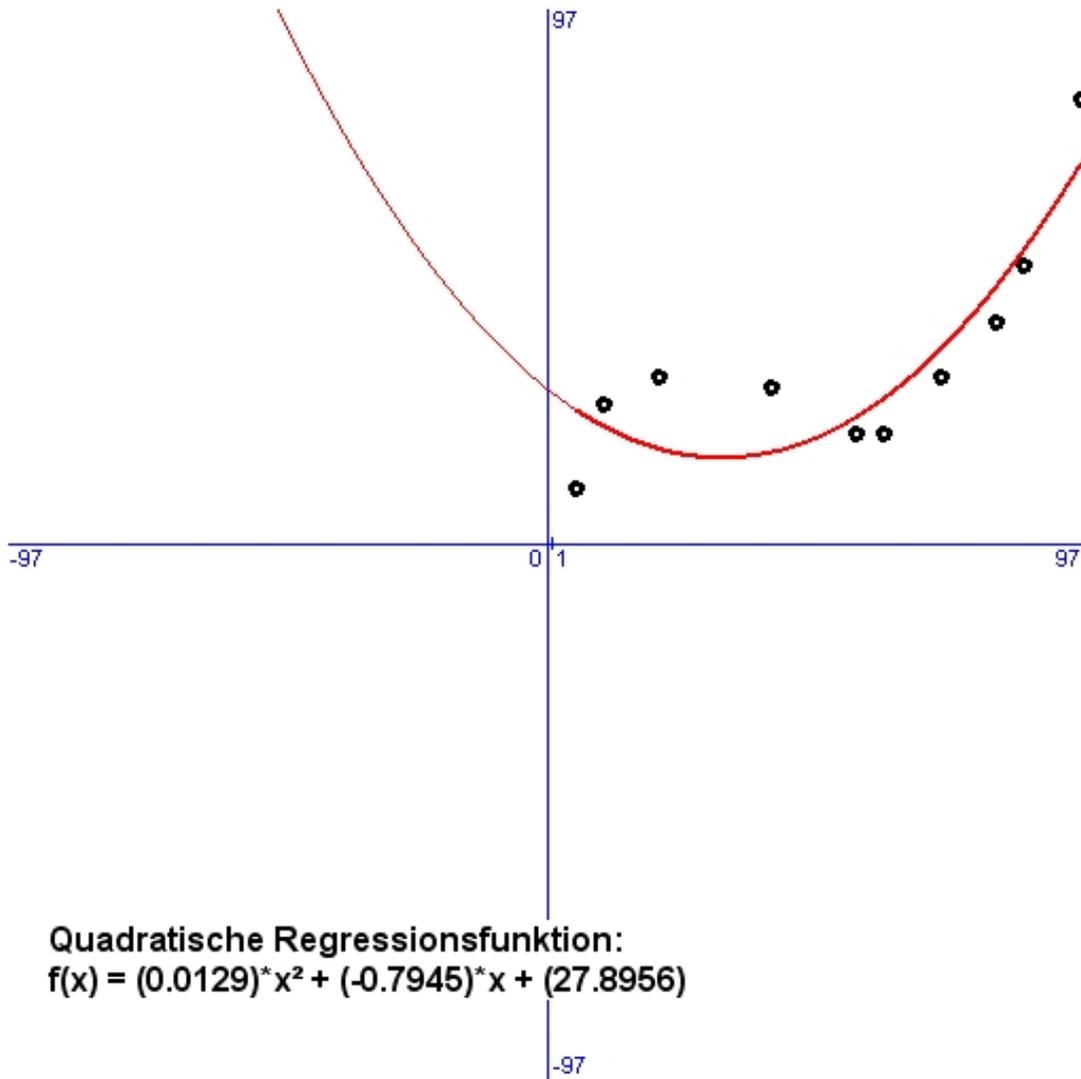
$$\begin{aligned}x &= 27.8956 \\y &= -0.7945 \\z &= 0.0129\end{aligned}$$

Probe:

$$\begin{aligned}333.0001 &= 333 \\21470.0060 &= 21470 \\1678300.4854 &= 1678300\end{aligned}$$

Quadratische Regressionsfunktion:

$$f(x) = (0.0129) \cdot x^2 + (-0.7945) \cdot x + (27.8956)$$



Ausrechnung des Beispiels mit dem Programm "MATRIX.EXE".
Zur Auflösung der Normalgleichungen wird der Gauß-Jordan-Algorithmus verwendet.

Gegeben sind 10 Paare von Messwerten (x;y):

(5;10),(10;25),(20;30),(40;28),(55;20),
(60;20),(70;30),(80;40),(85;50),(95;80)

Gesucht ist eine quadratische Polynomfunktion

$$f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2,$$

die sich optimal an die Messpunkte anpasst.

Vektor $\hat{x} = (5; 10; 20; 40; 55; 60; 70; 80; 85; 95)$

Vektor $\hat{y} = (10; 25; 30; 28; 20; 20; 30; 40; 50; 80)$

Vektor $\hat{a} = (a_0; a_1; a_2)$

Systemmatrix M, (10*3-Matrix):

1; 5; 25
1; 10; 100
1; 20; 400
1; 40; 1600
1; 55; 3025
1; 60; 3600
1; 70; 4900
1; 80; 6400
1; 85; 7225
1; 95; 9025

Transponierte Matrix M^T , (3*10-Matrix):

```
1; 1; 1; 1; 1; 1; 1; 1; 1; 1
5; 10; 20; 40; 55; 60; 70; 80; 85; 95
25; 100; 400; 1600; 3025; 3600; 4900; 6400; 7225; 9025
```

Matrizenprodukt $M^T * M$, (3*3-Matrix):

```
10; 520; 36300
520; 36300; 2782000
36300; 2782000; 223462500
```

Inverse Matrix $(M^T * M)^{-1}$, (3*3-Matrix):

(Die Berechnung erfolgt auf 12 Dezimalstellen gerundet)

```
0.757503238826; -0.030966335541; 0.000262464520
-0.030966335541; 0.001866327298; -0.000018204596
0.000262464520; -0.000018204596; 0.000000188478
```

Matrizenprodukt $(M^T * M)^{-1} * M^T$, (3*10-Matrix):

0.609233174121; 0.474086335416; 0.243162336006; -0.061206950814;
-0.151690042929; -0.155604621634; -0.124064101044; -0.040030676454;
0.021670874841; 0.184443655431

-0.022089813951; -0.014123522161; -0.000921627981; 0.014559402779;
0.016612762949; 0.015476756739; 0.010474054919; 0.001830433899;
-0.003856721311; -0.017961721131

0.000176153490; 0.000099266360; -0.000026236200; -0.000164154520;
-0.000168642310; -0.000151290440; -0.000088315000; 0.000012356040;
0.000076827410; 0.000234041850

Lösungsvektor $\hat{a} = (a_0; a_1; a_2) = (M^T * M)^{-1} * M^T * \hat{y}$:

27.895558379788; -0.794516113893; 0.012870636440

Regressionsfunktion $f(x)$:

$$f(x) = 27.8956 - 0.7945 * x + 0.0129 * x^2$$

Beispiel: Kubische Regression

Zu n Zahlenpaaren $(x ; y)$
wird eine optimal angepasste
Polynomfunktion 3. Grades
 $f(x) = a \cdot x^3 + b \cdot x^2 + c \cdot x + d$
ermittelt.

Ein Beispiel mit 10 Zahlenpaaren: (5;10), (10;25), (20;30), (40;28), (55;20),
(60;20), (70;30), (80,40), (85;50), (95;80).

Ausrechnung mit dem Programm „**STATIS.EXE**“.

Kubische Regression von 10 Zahlenpaaren Lösungsmethode mit 4 Normalgleichungen

$$10 \cdot w + 520 \cdot x + 36300 \cdot y + 2782000 \cdot z = 333$$

$$520 \cdot w + 36300 \cdot x + 2782000 \cdot y + 223462500 \cdot z = 21470$$

$$36300 \cdot w + 2782000 \cdot x + 223462500 \cdot y + 18518950000 \cdot z = 1678300$$

$$2782000 \cdot w + 223462500 \cdot x + 18518950000 \cdot y + 1570532062500 \cdot z = 139772000$$

Lösungen:

$$w = 2.409291$$

$$x = 2.501610$$

$$y = -0.067911$$

$$z = 0.000530$$

Probe:

$$333.007465 = 333$$

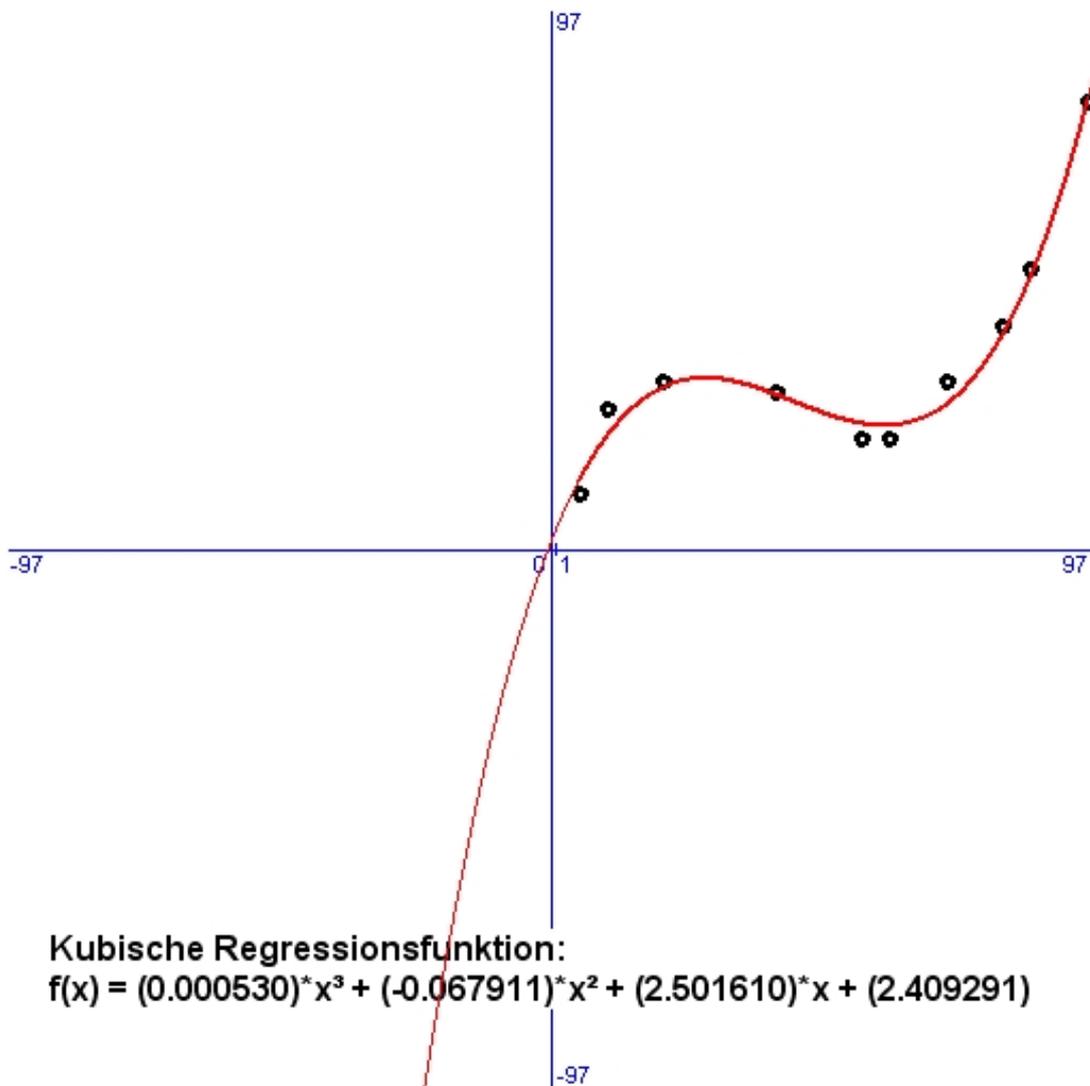
$$21470.599261 = 21470$$

$$1678349.645214 = 1678300$$

$$139776209.200785 = 139772000$$

Kubische Regressionsfunktion:

$$f(x) = (0.000530) \cdot x^3 + (-0.067911) \cdot x^2 + (2.501610) \cdot x + (2.409291)$$



Ausrechnung des Beispiels mit dem Programm "MATRIX.EXE".
Zur Auflösung der Normalgleichungen wird der Gauß-Jordan-Algorithmus verwendet.

Gegeben sind 10 Paare von Messwerten (x;y):

(5;10),(10;25),(20;30),(40;28),(55;20),
(60;20),(70;30),(80;40),(85;50),(95;80).

Gesucht ist eine kubische Polynomfunktion

$$f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3,$$

die sich optimal an die Messpunkte anpasst.

Vektor $\vec{x} = (5; 10; 20; 40; 55; 60; 70; 80; 85; 95)$

Vektor $\vec{y} = (10; 25; 30; 28; 20; 20; 30; 40; 50; 80)$

Vektor $\vec{a} = (a_0; a_1; a_2; a_3)$

Systemmatrix M, (10*4-Matrix):

1; 5; 25; 125
1; 10; 100; 1000
1; 20; 400; 8000
1; 40; 1600; 64000
1; 55; 3025; 166375
1; 60; 3600; 216000
1; 70; 4900; 343000
1; 80; 6400; 512000
1; 85; 7225; 614125
1; 95; 9025; 857375

Transponierte Matrix M^T , (4*10-Matrix):

```

1;    1;    1;    1;    1;    1;    1;    1;    1;    1
5;   10;   20;   40;   55;   60;   70;   80;   85;   95
25;  100;  400; 1600; 3025; 3600; 4900; 6400; 7225; 9025
125; 1000; 8000; 64000; 166375; 216000; 343000; 512000; 614125; 857375

```

Matrizenprodukt $M^T * M$, (4*4-Matrix):

```

10;      520;      36300;      2782000
520;     36300;     2782000;     223462500
36300;   2782000;   223462500;   18518950000
2782000; 223462500; 18518950000; 1570532062500

```

Inverse Matrix $(M^T * M)^{-1}$, (4*4-Matrix):

(Die Berechnung erfolgt auf 12 Dezimalstellen gerundet)

```

1.655296453121; -0.147077480541;  0.003108128337; -0.000018654829
-0.147077480541;  0.016882923301; -0.000386232867;  0.000002412620
 0.003108128337; -0.000386232867;  0.000009208152; -0.000000059129
-0.000018654829;  0.000002412620; -0.000000059129;  0.000000000388

```

Matrizenprodukt $(M^T * M)^{-1} * M^T$, (4*10-Matrix)

0.995280405216; 0.476679652411; -0.192240454899; -0.448706485319;
 -0.135573932084; -0.009533430139; 0.191095319551; 0.229846918641;
 0.153540982336; -0.260389970724

-0.072017108211; -0.014458914231; 0.055388798679; 0.064674544299;
 0.014528530839; -0.003414483681; -0.030285237771; -0.033072525261;
 -0.020911206531; 0.039568680879

0.001399776677; 0.000107485867; -0.001406300203; -0.001392399143;
 -0.000117606923; 0.000311639517; 0.000910525447; 0.000867623777;
 0.000494635717; -0.001176148603

-0.000008021454; -0.000000053529; 0.000009049971; 0.000008075571;
 -0.000000272454; -0.000002954029; -0.000006419529; -0.000005414829;
 -0.000002508654; 0.000009566346

Lösungsvektor $\hat{a} = (a_0; a_1; a_2) = (M^T * M)^{-1} * M^T * \hat{y}$:

2.409240673123; 2.501664206217; -0.067950004669; 0.000532228643

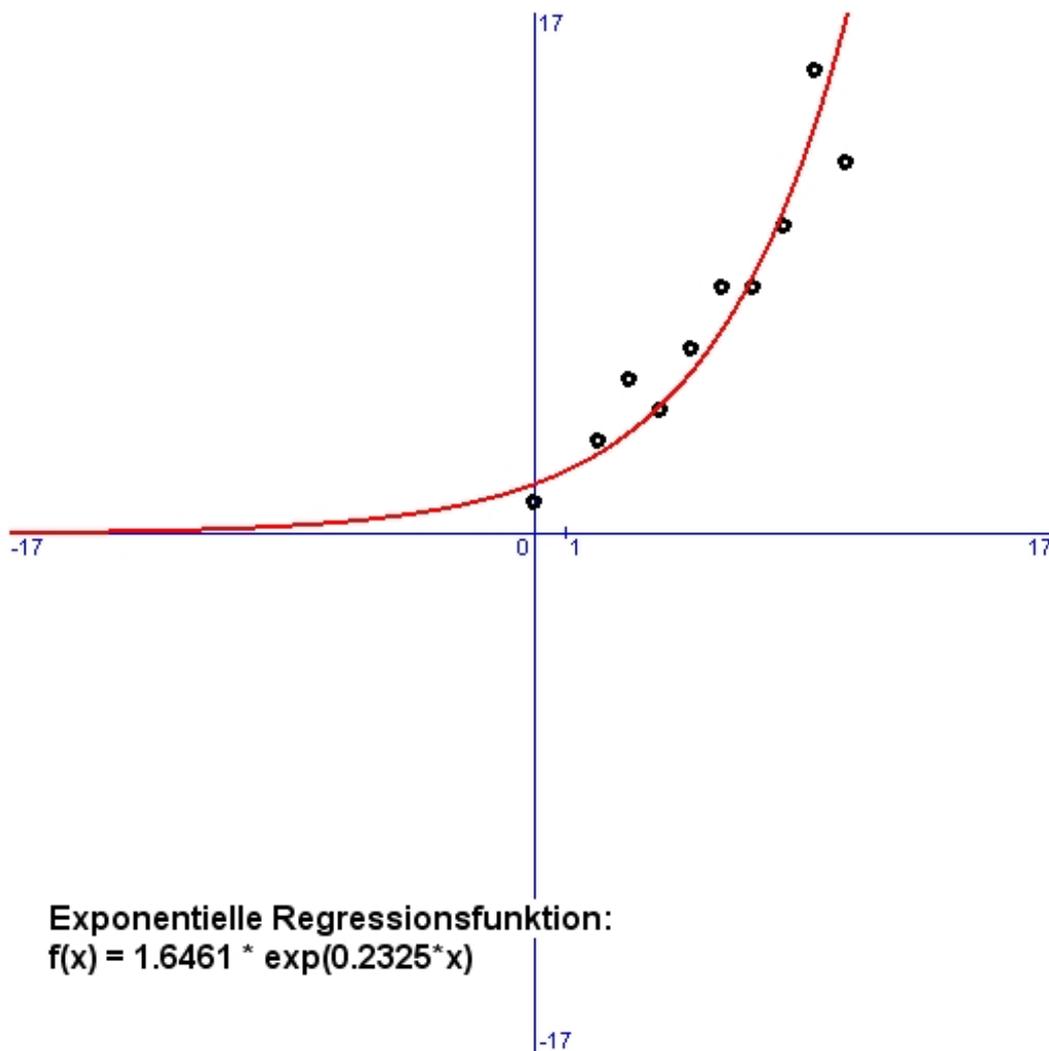
Regressionsfunktion $f(x)$:

$f(x) = 2.409241 + 2.501664 * x - 0.067950 * x^2 + 0.000532 * x^3$

Beispiel: Exponentielle Regression

Zu n Zahlenpaaren $(x ; y)$
wird eine optimal angepasste
natürliche Exponentialfunktion
 $f(x) = a \cdot \exp(b \cdot x)$ ermittelt.

Ein Beispiel mit 10 Zahlenpaaren: $(0;1)$, $(2;3)$, $(3;5)$, $(4;4)$, $(5;6)$,
 $(6;8)$, $(7;8)$, $(8;10)$, $(9;15)$, $(10;12)$.
Ausrechnung mit dem Programm „**STATIS.EXE**“.



Grundbegriffe der Faktorenanalyse

In einer Stichprobe mit N Versuchspersonen werden M Merkmale gemessen. Die Messergebnisse können in einem zweidimensionalen Schema angeordnet werden (Datenmatrix). X_{ij} ist dabei der Messwert der j -ten Versuchsperson (Spalte) im i -ten Merkmal (Zeile).

Datenmatrix ((X_{ij})):

| | | | | | | |
|----------|----------|-------|----------|-------|----------|--------------------|
| X_{11} | X_{12} | | | | X_{1N} | |
| X_{21} | X_{22} | | | | X_{2N} | |
| | | | | | | |
| | | | X_{ij} | | | → i -tes Merkmal |
| | | | | | | |
| X_{M1} | X_{M2} | | | | X_{MN} | |

↑
j-te Versuchsperson

Nun werden zwischen sämtlichen Merkmalen paarweise die Korrelationen r_{ik} berechnet. Das ergibt eine **Matrix von Interkorrelationen ((r_{ik}))**:

| | | | | | |
|----------|----------|-------|----------|-------|----------|
| r_{11} | r_{12} | | | | r_{1M} |
| r_{21} | r_{22} | | | | r_{2M} |
| | | | | | |
| | | | r_{kk} | | |
| | | | | | |
| r_{M1} | r_{M2} | | | | r_{MM} |

Diese Matrix der M Interkorrelationen ist quadratisch und symmetrisch zur Hauptdiagonale, weil ja $r_{ik} = r_{ki}$ ist.

Alle Elemente r_{kk} in der Hauptdiagonale sind genau 1, weil die Korrelation eines Merkmals X_k mit sich selbst 1 sein muss.

Mit der so genannten **Faktorenanalyse** werden mehrere Merkmalsvariable gemäß ihren korrelativen Beziehungen in wenige, voneinander unabhängige Gruppen eingeteilt. Merkmale in derselben Gruppe korrelieren hoch miteinander, Merkmale aus verschiedenen Gruppen korrelieren niedrig. Die Faktorenanalyse dient somit einer datenreduzierenden Klassifizierung. Ihr Ergebnis sind wechselseitig unabhängige Faktoren, welche das System der Merkmalsvariablen möglichst erschöpfend erklären. Dadurch wird eine Hypothese über Dimensionalität und Struktur komplexer Merkmals-Systeme erzeugt.

In der Faktorenanalyse geht man davon aus, dass jedes quantitative Merkmal X als eine Funktion von wenigen, voneinander unabhängigen Faktoren f_k aufzufassen ist. Dabei wird angenommen, dass sich X als eine gewichtete Summe dieser Faktoren darstellen lässt, d.h. als lineare Funktion:

$$X = a_1f_1 + a_2f_2 + a_3f_3 + \dots + a_Rf_R + q_X$$

(R = Anzahl der verschiedenen Faktoren)

Die Koeffizienten a_k (Gewichtszahlen) werden auch als Faktorenladungen des Merkmals X bezeichnet. Je größer a_k ist, umso stärker ist X im betreffenden Faktor f_k aufgeladen. Ein Faktor ist offenbar ohne Einfluss auf X , wenn seine Ladungszahl Null ist. Die additive Konstante q_X ist ein merkmalspezifischer Restwert (in ihm ist auch ein möglicher spezifischer Fehler enthalten). Die Faktorenwerte f_k charakterisieren eine bestimmte Versuchsperson. Sie sind jene Werte, welche die Versuchsperson in den einzelnen Faktoren des Merkmals aufweist (Personenparameter). Hingegen charakterisieren die Ladungszahlen a_k ein bestimmtes Merkmal (Merkmalsparameter). Sie geben an, wie gewichtig ein Faktor auf ein Merkmal Einfluss nimmt. Die Merkmalsausprägung X einer Person wird bestimmt durch die gewichtete Summe der Faktorenwerte der Person.

Obige Gleichung wird als **Grundgleichung** der reduzierten Faktorenlösung bezeichnet, weil die Anzahl R der Faktoren kleiner als die Anzahl M der Merkmale ist. Das ist durchaus sinnvoll, denn es soll ja eine größere Menge von verschiedenen Merkmalen auf einige, wenige Grundfaktoren zurückgeführt werden. Aus dieser Tatsache, dass $R < M$ ist, erklärt sich auch der merkmalspezifische Restwert q_X . Dieser enthält jenen Anteil vom Merkmalswert X , welcher nicht durch die Faktoren erklärt werden kann. Weil es M Merkmale gibt, wird daher ein Gleichungssystem von insgesamt M solcher Grundgleichungen aufgestellt.

Sind X , Y zwei Merkmale aus dem vorliegenden System von M Merkmalen, dann gelten folgende Grundgleichungen:

$$X = a_1f_1 + a_2f_2 + a_3f_3 + \dots + a_Rf_R + q_X$$

$$Y = b_1f_1 + b_2f_2 + b_3f_3 + \dots + b_Rf_R + q_Y$$

Die Faktorenanalyse standardisiert zunächst alle Merkmalsvariable $Z=(X-M_X)/S_X$ und setzt voraus, dass alle Faktoren voneinander paarweise unabhängig sind, d.h. ihre Korrelationen jeweils Null ergeben, also $r(f_k, f_i) = 0$. Unter diesen beiden Voraussetzungen lässt sich mithilfe der höheren Mathematik folgende wichtige Gleichung herleiten (**Garnett'sche Gleichung**):

$$r(X,Y) = a_1b_1 + a_2b_2 + a_3b_3 + \dots + a_Rb_R + e_{XY}$$

Dabei sind a_k und b_k die k-ten Faktorenladungen der beiden Merkmale X und Y. Der Wert e_{XY} ist ein Restglied, das sich aus den Restwerten q_X und q_Y ergibt.

Diese Garnett'sche Gleichung besagt, dass die Interkorrelationen der Merkmale im Wesentlichen als die skalaren Produkte ihrer Faktorenladungen aufgefasst werden können.

Das System der Garnett'schen Gleichungen ist nun der Ausgangspunkt für die eigentliche Faktorenextraktion. Dabei sind die Interkorrelationen auf der linken Gleichungsseite bekannt, nicht aber die Faktorenladungen auf der rechten Seite. Unter **Faktorenextraktion** versteht man ein mathematisches Verfahren, mit dessen Hilfe auf Grund der Garnett'schen Gleichungen aus den gegebenen Interkorrelationen der Merkmale die unbekanntes Faktorenladungen berechnet werden. Die zwei wichtigsten Extraktionsverfahren sind die Hauptkomponentenmethode und die Zentroidmethode. Letztere wurde von dem amerikanischen Psychologen **Louis Thurstone** (1887-1957) entwickelt. Bei jeder Faktorenextraktion werden die Faktoren so bestimmt, dass die von ihnen nicht erklärte Restvarianz der Merkmale minimal ist.

Grundsätzlich kann man sich die R voneinander unabhängigen Faktoren f_k als Achsen eines rechtwinkligen Koordinatensystems vorstellen, welches einen R-dimensionalen Raum aufspannt. Jeder der M Merkmalsvariablen X entspricht ein Pfeil (Vektor), der im Koordinatenursprung beginnt und dessen Endpunkt als Koordinaten die entsprechenden Ladungszahlen in den einzelnen Faktoren hat (a_1, a_2, \dots, a_R).

Bei der Faktorenextraktion kommen die M Merkmalsvektoren nicht regellos verteilt im R-dimensionalen Raum zu liegen, sondern in Bündeln. Je höher zwei Merkmale korrelieren, umso gleichgerichteter sind ihre Vektoren. Um eine möglichst einfache Darstellung der Merkmale durch die Faktoren zu erreichen („simple structure“ - Prinzip von Thurstone), wird mithilfe der so genannten **Faktorenrotation** das Koordinatensystem so gedreht, dass möglichst viele Faktorenachsen jeweils möglichst zentral durch diese Bündeln von Merkmalsvektoren verlaufen. Dabei wird sehr häufig die Varimax-Methode (Kaiser, 1958) verwendet. Nach den entsprechenden Rotationen stehen jedenfalls die endgültigen Faktorenladungen (a_1, a_2, \dots, a_R) der Merkmale X fest.

Auf die mathematischen Formeln der analytischen Geometrie, insbesondere auf die Theorie der Eigenvektoren und Eigenwerte, sowie auf das Modell schiefwinkliger Faktorenachsen kann hier nicht näher eingegangen werden. Der interessierte Leser sei auf die einschlägige statistische Fachliteratur verwiesen.

Als Richtlinien des „simple structure“-Kriteriums für die Faktorengewinnung gibt Louis Thurstone folgende durchaus plausible Regeln an:

- (1) Jedes Merkmal soll wenigstens in einem Faktor eine Null-Ladung besitzen.
- (2) Jeder Faktor soll wenigstens in so vielen Merkmalen fehlen, als Faktoren vorhanden sind.
- (3) Je zwei Faktoren sollen so festgelegt werden, dass möglichst viele Merkmale mit dem einen Faktor hoch und mit dem anderen Faktor hingegen möglichst niedrig geladen sind.
- (4) Einige Merkmale sollen in zumindest zwei Faktoren zugleich Null-Ladungen aufweisen.
- (5) Einige Merkmale sollen in zumindest zwei Faktoren zugleich sehr hohe Ladungen aufweisen.

Diese Richtlinien stellen weitgehend sicher, dass eine sinnvolle Reduktion der Gesamtheit der Merkmale auf eine Menge von wenigen, unabhängigen Faktoren erreicht wird.

Im Zeitalter der elektronischen Datenverarbeitung fällt ein großer Rechenaufwand nicht mehr so ins Gewicht wie früher. Aus diesem Grund wird anstelle der einfacheren Zentroid-Methode heute häufiger die so genannte Hauptkomponenten-Methode (PCA, principle components analysis) angewendet, welche ursprünglich auf Hotelling (1933) zurückgeht, aber in der Zwischenzeit oftmals modifiziert und verbessert wurde. Die Bezeichnung „Faktorenanalyse“ ist so zum Sammelbegriff für viele, zum Teil unterschiedliche Techniken der multivariaten Datenanalyse geworden. Sie finden heute in den unterschiedlichsten Bereichen ihre Anwendung, beispielsweise in der Persönlichkeitsforschung und in der Intelligenzforschung.

Faktorenanalyse und Persönlichkeitsforschung

Nach Kurt Pawlik sind zumindest sechs gesicherte, voneinander unabhängige Hauptfaktoren der Persönlichkeit herausgefunden worden.

- F1: Extraversion (versus Introversion)
- F2: Emotionale Stabilität (versus Labilität/Angst)
- F3: Gefühlsbetontheit (versus kühle Ausgeglichenheit)
- F4: Unabhängigkeit (versus Abhängigkeit der Meinungsbildung)
- F5: Kooperationsbereitschaft (versus Kooperationsmangel)
- F6: Durchsetzungsvermögen (versus Willensschwäche)

Häufigkeitsverteilungen

Häufigkeitsverteilungen und Wahrscheinlichkeiten

Es gibt verschiedene Formen von Häufigkeitsverteilungen, durch welche bestimmte Zufallsereignisse dargestellt werden können. In der Natur begegnet man oft der so genannten Normalverteilung. Diese wird weiter unten ausführlich beschrieben. Bei einer nicht zu geringen Anzahl N von Würfeln mit einem Spielwürfel sind die Häufigkeiten der sechs Würfelaugen gleich groß, $H(x) = N / 6$. In diesem Fall spricht man von einer Gleichverteilung. Diese ist die Grundlage für die meisten Spiele.

Die relative Häufigkeit $h(x)$ eines Ereignisses x ist definiert durch den Quotienten von absoluter Häufigkeit $H(x)$ durch die Gesamtanzahl N : $h(x) = H(x) / N$. Den Grenzwert der relativen Häufigkeit für den Fall, dass N unendlich groß wird, bezeichnet man als Wahrscheinlichkeit (p , probability) des Ereignisses.

$p(x) = \lim h(x) = \lim H(x) / N$ für N gegen Unendlich.

Das wird auch "Gesetz der großen Zahlen" genannt.

Offensichtlich liegt $p(x)$ immer zwischen 0 und 1. Bei $p = 0$ ist das Ereignis x unmöglich, bei $p(x) = 1$ ist das Ereignis sicher.

Mit einfachen und zusammengesetzten Ereignissen und deren Wahrscheinlichkeiten beschäftigt sich die Wahrscheinlichkeitstheorie. Die Kombinatorik ist eine spezielle Theorie, bei welcher die Gleichverteilung der Grundereignisse vorausgesetzt wird.

Zwei Zufallsexperimente

Zufallsexperiment "Werfen einer Münze"

Ereignisraum: $R = \{ \text{Zahl Z, Kopf K} \}$

Wahrscheinlichkeiten der Ereignisse e:

$$p(Z) = p(K) = 1/2 = 0.50$$

Anzahl n von 100: 100

Absolute Häufigkeiten $H(e)$

$$H(Z) = 52$$

$$H(W) = 48$$

Relative Häufigkeiten $h(e) = H(e) / n$

$$h(Z) = 0.52$$

$$h(W) = 0.48$$

Zufallsexperiment "Würfeln"

Ereignisraum: $R = \{ \text{Augen: 1, 2, 3, 4, 5, 6} \}$

Wahrscheinlichkeiten der Ereignisse e:

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6 = 0.17$$

Anzahl n von 100: 100

Absolute Häufigkeiten $H(e)$

$$H(1) = 20$$

$$H(2) = 15$$

$$H(3) = 13$$

$$H(4) = 11$$

$$H(5) = 19$$

$$H(6) = 22$$

Relative Häufigkeiten $h(e) = H(e) / n$

$$h(1) = 0.20$$

$$h(2) = 0.15$$

$$h(3) = 0.13$$

$$h(4) = 0.11$$

$$h(5) = 0.19$$

$$h(6) = 0.22$$

Normalverteilung

Viele quantitativ messbare Merkmale x (Körpergröße, Gewicht, usw.), die von mehreren verschiedenen Faktoren abhängen und die in nicht zu kleinen Stichproben erfasst werden, haben eine Häufigkeitsverteilung mit glockenförmigem Aussehen.

Diese Verteilungskurve verläuft symmetrisch zu einem Gipfel (Mittelwert) und nähert sich auf beiden Seiten davon asymptotisch der x -Achse. Je extremer ein Merkmalswert ist, umso seltener kommt er vor. Die relativen Häufigkeiten $h(x)$ können durch eine Formel beschrieben werden, die vom Mittelwert m und der Streuung s abhängt:

$$h(x) = 1/\sqrt{2\pi s^2} * \exp(-(x-m)^2/(2s^2)).$$

Die Formel wird viel einfacher, wenn man folgende Ersetzung vornimmt: $z = (x-m) / s$. Diese Standardisierung bedeutet eine Verschiebung des Mittelwertes in den Nullpunkt und Stauchung der verschobenen x -Werte durch Faktor s . Dann erhält man die so genannte *standardisierte Normalverteilung* mit Mittelwert 0 und Streuung 1.

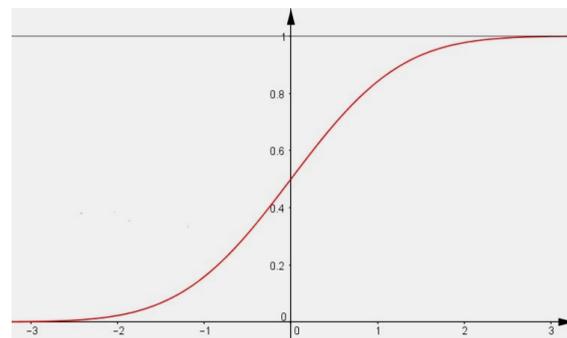
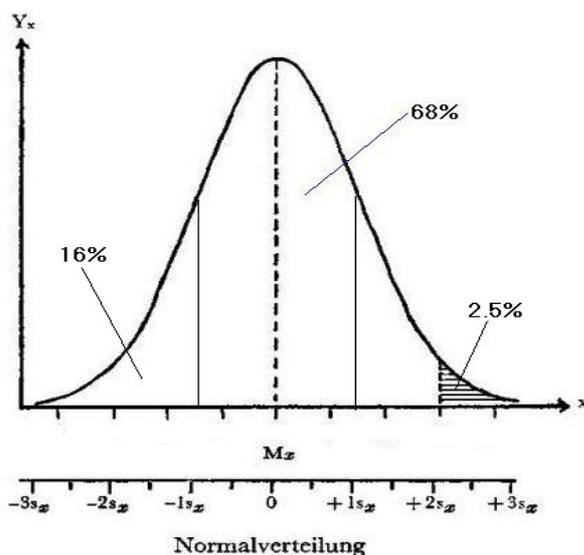
$$h(z) = 1/\sqrt{2\pi} * \exp(-z^2/2).$$

Die *kumulative Häufigkeit* $h[z \leq z_0]$ bzw. $100 * h[z \leq z_0]$ der standardisierten Normalverteilung gibt an, wie viele Prozente kleiner oder gleich einem bestimmten Wert z_0 sind.

$$h[z \leq z_0] = 1/\sqrt{2\pi} * \int_{-\infty}^{z_0} \exp(-z^2/2) dz = \Phi(z_0)$$

(Die Summierung bzw. Integration erfolgt dabei von $-\infty$ bis z_0 .)

Für diese Prozentrangwerte ($\Phi(z_0)$, Verteilungsfunktion "phi") gibt es fertige Tabellen. Die Verteilungsfunktion für die Normalverteilung ist unten rechts abgebildet.



Für jede Normalverteilung gilt:

Im Bereich $[m - 1s ; m + 1s]$ liegen 68.00 % aller z -Werte (Normalbereich).

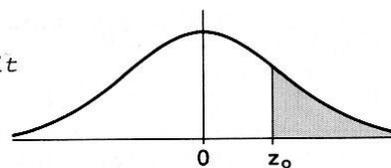
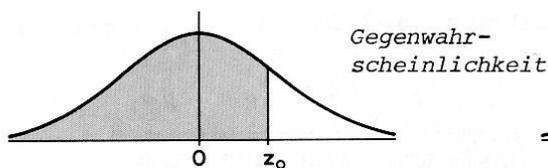
Im Bereich $[m - 2s ; m + 2s]$ liegen 95.00 % aller z -Werte.

Im Bereich $[m - 3s ; m + 3s]$ liegen 99.75 % aller z -Werte.

In der theoretischen Statistik wird die relative Häufigkeit $h(x)$ auch als die Wahrscheinlichkeit $p(x)$ bezeichnet ("probability" des Merkmalswertes x). Die Wahrscheinlichkeit $p(x)$ ist der Grenzwert der relativen Häufigkeit $h(x)$ für den theoretischen Fall, dass die Stichprobe unendlich groß wird. Dem entsprechend schreibt man dann für die kumulative Häufigkeit $h[x \leq x_0]$ auch $p[x \leq x_0]$. Handelt es sich dabei um die standardisierten Werte $z = (x - m) / s$ dann schreibt man $p[z \leq z_0]$. Dieser Wahrscheinlichkeit entspricht eine bestimmte Fläche unter der standardisierten Normalverteilungskurve, und sie wird auch Verteilungsfunktion "phi" $\Phi(z_0)$ genannt. Dafür gibt es bereits fertige Tabellen zum Nachschlagen (siehe nächste Seite). In den folgenden Grafiken ist zu beachten, dass nur für den Fall A die Werte direkt aus der Tabelle entnommen werden können.

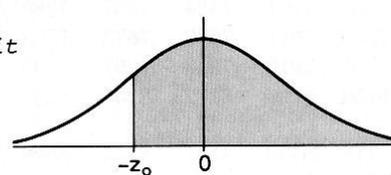
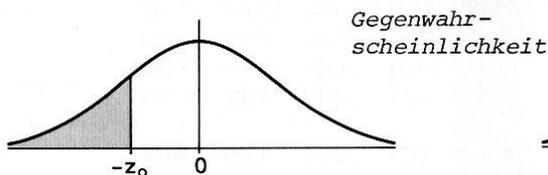
$$\frac{z_0 \geq 0}{\text{A: } P(z \leq z_0) = \Phi(z_0)}$$

$$\text{B: } P(z > z_0) = 1 - \Phi(z_0)$$



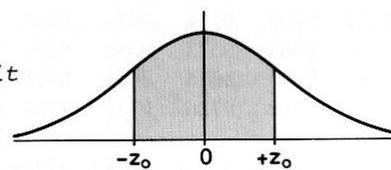
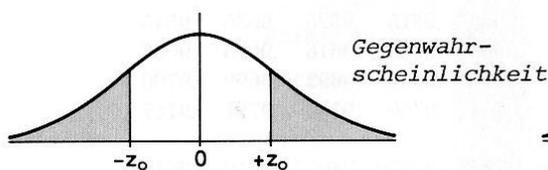
$$\text{C: } P(z \leq -z_0) = \Phi(-z_0) = 1 - \Phi(z_0)$$

$$\text{D: } P(z > -z_0) = \Phi(z_0)$$



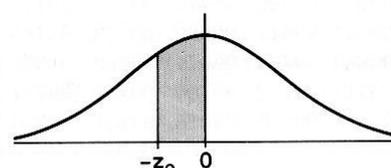
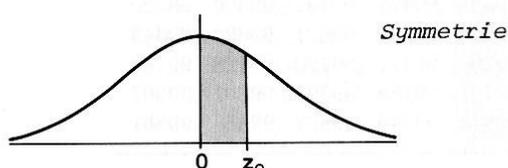
$$\text{E: } P(|z| > z_0) = 2(1 - \Phi(z_0))$$

$$\text{F: } P(-z_0 < z < z_0) = 2\Phi(z_0) - 1$$



$$\text{G: } P(0 < z < z_0) = \Phi(z_0) - 0,5$$

$$\text{H: } P(-z_0 < z < 0) = \Phi(z_0) - 0,5$$



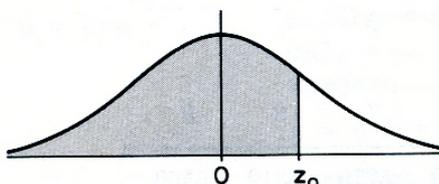
Die nachfolgende Tabelle der Verteilungsfunktion liefert die Wahrscheinlichkeiten $p[z \leq z_0]$ für die standardisierte Normalverteilung:

| z_0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------|---------|--------|---------|--------|---------|--------|---------|--------|--------|--------|
| 0,0 | 0,5000 | ,5040 | ,5080 | ,5120 | ,5160 | ,5199 | ,5239 | ,5279 | 5319 | ,5359 |
| 0,1 | 0,5398 | ,5438 | ,5478 | ,5517 | ,5557 | ,5596 | ,5636 | ,5675 | ,5714 | ,5753 |
| 0,2 | 0,5793 | ,5832 | ,5871 | ,5910 | ,5948 | ,5987 | ,6026 | ,6064 | ,6103 | ,6141 |
| 0,3 | 0,6179 | ,6217 | ,6255 | ,6293 | ,6331 | ,6368 | ,6404 | ,6443 | ,6480 | ,6517 |
| 0,4 | 0,6554 | ,6591 | ,6628 | ,6664 | ,6700 | ,6736 | ,6772 | ,6808 | ,6844 | ,6879 |
| 0,5 | 0,6915 | ,6950 | ,6985 | ,7019 | ,7054 | ,7088 | ,7123 | ,7157 | ,7190 | ,7224 |
| 0,6 | 0,7257 | ,7291 | ,7324 | ,7357 | ,7389 | ,7422 | ,7454 | ,7486 | ,7517 | ,7549 |
| 0,7 | 0,7580 | ,7611 | ,7642 | ,7673 | ,7703 | ,7734 | ,7764 | ,7794 | ,7823 | ,7852 |
| 0,8 | 0,7881 | ,7910 | ,7939 | ,7967 | ,7995 | ,8023 | ,8051 | ,8078 | ,8106 | ,8133 |
| 0,9 | 0,8159 | ,8186 | ,8212 | ,8238 | ,8264 | ,8289 | ,8315 | ,8340 | ,8365 | ,8389 |
| 1,0 | 0,8413 | ,8438 | ,8461 | ,8485 | ,8508 | ,8531 | ,8554 | ,8577 | ,8599 | ,8621 |
| 1,1 | 0,8643 | ,8665 | ,8686 | ,8708 | ,8729 | ,8749 | ,8770 | ,8790 | ,8810 | ,8830 |
| 1,2 | 0,8849 | ,8869 | ,8888 | ,8907 | ,8925 | ,8944 | ,8962 | ,8980 | ,8997 | ,9015 |
| 1,3 | 0,9032 | ,9049 | ,9066 | ,9082 | ,9099 | ,9115 | ,9131 | ,9147 | ,9162 | ,9177 |
| 1,4 | 0,9192 | ,9207 | ,9222 | ,9236 | ,9251 | ,9265 | ,9279 | ,9292 | ,9306 | ,9319 |
| 1,5 | 0,9332 | ,9345 | ,9357 | ,9370 | ,9382 | ,9394 | ,9406 | ,9418 | ,9429 | ,9441 |
| 1,6 | 0,9452 | ,9463 | ,9474 | ,9484 | ,9495 | ,9505 | ,9515 | ,9525 | ,9535 | ,9545 |
| 1,7 | 0,9554 | ,9564 | ,9573 | ,9582 | ,9591 | ,9599 | ,9608 | ,9616 | ,9625 | ,9633 |
| 1,8 | 0,9641 | ,9649 | ,9656 | ,9664 | ,9671 | ,9678 | ,9686 | ,9693 | ,9699 | ,9706 |
| 1,9 | 0,9713 | ,9719 | ,9726 | ,9732 | ,9738 | ,9744 | ,9750 | ,9756 | ,9761 | ,9767 |
| 2,0 | 0,97725 | ,97778 | ,97831 | ,97882 | ,97932 | ,97982 | ,98030 | ,98077 | ,98124 | ,98169 |
| 2,1 | 0,98214 | ,98257 | ,98300 | ,98341 | ,98382 | ,98422 | ,98461 | ,98500 | ,98537 | ,98574 |
| 2,2 | 0,98610 | ,98645 | ,98679 | ,98713 | ,98745 | ,98778 | ,98809 | ,98840 | ,98870 | ,98899 |
| 2,3 | 0,98928 | ,98956 | ,98983 | ,99010 | ,99036 | ,99061 | ,99086 | ,99111 | ,99134 | ,99158 |
| 2,4 | 0,99180 | ,99202 | ,99224 | ,99245 | ,99266 | ,99286 | ,99305 | ,99324 | ,99343 | ,99361 |
| 2,5 | 0,99379 | ,99396 | ,99413 | ,99430 | ,99446 | ,99461 | ,99477 | ,99492 | ,99506 | ,99520 |
| 2,6 | 0,99534 | ,99547 | ,99560 | ,99573 | ,99585 | ,99598 | ,99609 | ,99621 | ,99632 | ,99643 |
| 2,7 | 0,99653 | ,99664 | ,99674 | ,99683 | ,99693 | ,99702 | ,99711 | ,99720 | ,99728 | ,99736 |
| 2,8 | 0,99744 | ,99752 | ,99760 | ,99767 | ,99774 | ,99781 | ,99788 | ,99795 | ,99801 | ,99807 |
| 2,9 | 0,99813 | ,99819 | ,99825 | ,99831 | ,99836 | ,99841 | ,99846 | ,99851 | ,99856 | ,99861 |
| 3,0 | 0,99865 | 3,1 | 0,99903 | 3,2 | 0,99931 | 3,3 | 0,99952 | | | |

Die Werte in der Tabelle entsprechen der unten abgebildeten Fläche:

$$z_0 \geq 0$$

$$P(z \leq z_0) = \Phi(z_0)$$



Fünf Rechenbeispiele

1. Beispiel (Eindimensionale Statistik)

Als erstes Beispiel soll aus einer Population aller 16-jährigen eine Stichprobe von $N = 25$ Personen zufällig ausgewählt werden. Die Messung der Körpergrößen X liefert folgende Werte in cm: 187, 167, 160, 171, 167, 182, 178, 178, 185, 156, 160, 170, 178, 178, 185, 169, 167, 157, 165, 173, 171, 184, 170, 178, 182.

Es wird gezählt wie häufig jede Maßzahl X unter den N Messungen vorkommt. Das ergibt die absoluten Häufigkeiten $H(X)$ der Merkmalsausprägungen X . Wenn viele Werte X nur eine geringe Frequenz $H(X)$ aufweisen, ist eine Zusammenfassung aufeinander folgender Maßzahlen zu Klassen (Intervallen) zweckmäßig, deren Intervallmitten dann als neue Maßzahlen dienen. Die Häufigkeit H eines derartigen Intervalls gibt nun an, wie oft in der Stichprobe ein Messergebnis X angetroffen wird, welches in das betreffende Intervall fällt ($L \leq X < R$). Dabei sind L und R linker und rechter Intervallrand. Die 25 Messwerte werden in 4 gleich breite Klassen eingeteilt und die entsprechenden Häufigkeiten ausgezählt. Dividiert man die absoluten Häufigkeiten $H(X)$ durch den Stichprobenumfang N , erhält man die relativen Häufigkeiten $h(X) = H(X) / N$. Die Multiplikation mit 100 liefert die prozentuellen Häufigkeiten, $p(X) = 100 * H(X) / N$.

Anzahl der Häufigkeitsklassen (2...10) : 4.00
 Unterste Datengrenze von allen Klassen : 150.00
 Oberste Datengrenze von allen Klassen : 190.00
 Konstante Breite der Häufigkeitsklassen : 10.00

1. Klasse: $150 \leq X < 160 \rightarrow 8.00\%$ (2)
2. Klasse: $160 \leq X < 170 \rightarrow 28.00\%$ (7)
3. Klasse: $170 \leq X < 180 \rightarrow 40.00\%$ (10)
4. Klasse: $180 \leq X < 190 \rightarrow 24.00\%$ (6)

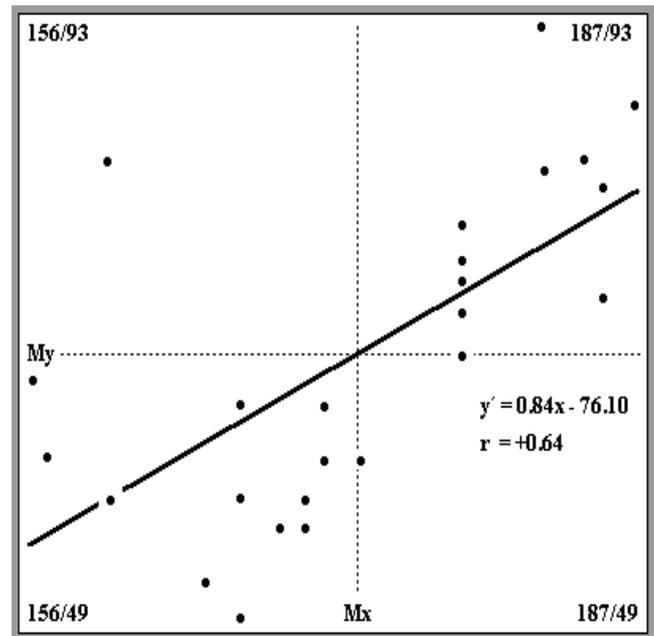
Mittelwert $m_x = \sum(X) / N = 4381 / 25 = 172.72$
 Varianz $v_x = \sum(X^2) / N - m^2 = 747792 / 25 - (172.72)^2 = 79.48$
 Streuung $s_x = \text{sqrt}(v_x) = \text{sqrt}(79.48) = 8.92$

Neben dem Mittelwert ist der so genannte *Median* ein wichtiges Zentralmaß. Dabei liegen unterhalb und oberhalb des Medians 50% der gegebenen Daten X . Um ihn zu bestimmen, müssen die Daten zuerst der Größe nach geordnet werden. Dann wird ganz einfach bis zur Hälfte der Datenanzahl gezählt und der erreichte Wert ist der Median. Im obigen Beispiel liegt der Median bei 171 cm.

2. Beispiel (Zweidimensionale Statistik)

Von $N = 25$ Schülern einer Klasse werden die Körpergröße X (cm) und das Gewicht Y (kg) erfasst. Zuerst werden eine Verteilungsanalyse und dann eine Regressionsanalyse durchgeführt. Zum Abschluss werden in einem zweidimensionalen Diagramm alle 25 Datenpaare eingetragen und die Regressionsgerade gezeichnet.

| X | Y | X^2 | Y^2 | XY |
|-----|----|-------|-------|-------|
| 187 | 87 | 34969 | 7569 | 16269 |
| 167 | 65 | 27889 | 4225 | 10855 |
| 160 | 83 | 25600 | 6889 | 13280 |
| 171 | 61 | 29241 | 3721 | 10431 |
| 167 | 49 | 27889 | 2401 | 8183 |
| 182 | 82 | 33124 | 6724 | 14924 |
| 178 | 74 | 31684 | 5476 | 13172 |
| 178 | 76 | 31684 | 5776 | 13528 |
| 185 | 81 | 34225 | 6561 | 14985 |
| 156 | 67 | 24336 | 4489 | 10452 |
| 160 | 58 | 25600 | 3364 | 9280 |
| 170 | 58 | 28900 | 3364 | 9860 |
| 178 | 78 | 31684 | 6084 | 13884 |
| 178 | 72 | 31684 | 5184 | 12816 |
| 185 | 73 | 34225 | 5329 | 13505 |
| 169 | 56 | 28561 | 3136 | 9464 |
| 167 | 58 | 27889 | 3364 | 9686 |
| 157 | 61 | 24649 | 3721 | 9577 |
| 165 | 52 | 27225 | 2704 | 8580 |
| 173 | 61 | 29929 | 3721 | 10553 |
| 171 | 65 | 29241 | 4225 | 11115 |
| 184 | 83 | 33856 | 6889 | 15272 |
| 170 | 56 | 28900 | 3136 | 9520 |
| 178 | 69 | 31684 | 4761 | 12282 |
| 182 | 93 | 33124 | 8649 | 16926 |



| | |
|-------------|---|
| Anzahl | $N = 25,$ |
| Summen | $\sum X = 4318, \quad \sum Y = 1718,$ |
| Summen | $\sum X^2 = 747792, \quad \sum Y^2 = 121462, \quad \sum XY = 298399,$ |
| Mittelwerte | $m_X = 172.72, \quad m_Y = 68.72,$ |
| Streuungen | $s_X = 8.92, \quad s_Y = 11.66,$ |
| Kovarianz | $s_{XY} = 66.56,$ |

Ergebnis:

| | |
|-------------|--|
| Korrelation | $r = S_{XY} / (S_X \cdot S_Y) = +0.64$ |
| Regression | $k = S_{XY} / (S_X)^2 = +0.84$ |
| Gerade | $Y_1 = 0.84 \cdot X - 76.10$ |

Beispiel einer Regressionsschätzung: Wirkliches Wertepaar $(X, Y) = (157, 61)$.
Geschätzter Y-Wert: $Y_1 = 0.84 \cdot 157 - 76.10 = 58$ kg.

3. Beispiel (Rangreihen-Korrelation)

Zwölf verschiedene Gemälde werden von zwei Kritikern (X,Y) durch Rangreihen bewertet, indem jedem Gemälde ein Rangplatz zwischen 1 und 12 zugeordnet wird. Wie hoch ist die Übereinstimmung der beiden Rangreihen ?

Liegen in einer Stichprobe nicht quantitativ gemessene Daten vor, sondern nur geschätzte Rangpositionen, so kann die statistische Übereinstimmung zweier solcher Rangreihen (X,Y) berechnet werden. Die Formel für diese Rangreihen-Korrelation r_R wird aus dem Korrelationskoeffizienten hergeleitet: $r_R = 1 - 6\sum(D_i^2) / (N(N^2-1))$. Dabei ist N die Anzahl der Objekte und D_i der Unterschied zwischen dem Rangplatz X_i und dem Rangplatz Y_i des i-ten Objektes in den beiden Rangreihen ($D_i = X_i - Y_i$). Auch für diese Korrelation gilt: $-1 \leq r_R \leq +1$.

| Gemälde | Kritiker X | Kritiker Y | D | D_i^2 |
|---------|------------|------------|----|---------|
| 1 | 8 | 6 | 2 | 4 |
| 2 | 7 | 9 | -2 | 4 |
| 3 | 3 | 1 | 2 | 4 |
| 4 | 11 | 12 | -1 | 1 |
| 5 | 4 | 5 | -1 | 1 |
| 6 | 1 | 4 | -3 | 9 |
| 7 | 5 | 8 | -3 | 9 |
| 8 | 6 | 3 | 3 | 9 |
| 9 | 10 | 11 | -1 | 1 |
| 10 | 2 | 2 | 0 | 0 |
| 11 | 12 | 10 | 2 | 4 |
| 12 | 9 | 7 | 2 | 4 |

Anzahl N = 12
Summe $\sum(D_i^2) = 50$

Ergebnis:

$$r_R = 1 - (6 \cdot 50) / (12 \cdot (12^2 - 1))$$

$$= +0.83$$

d.h. diese Korrelation ist relativ hoch!

4. Beispiel (Normalverteilung)

Im nachfolgenden Beispiel werden in einer Stichprobe von Versuchspersonen die physischen Leistungswerte X an einem Kraftmesser (Dynamometer) ermittelt. Dabei ergibt sich eine Normalverteilung mit Mittelwert $m = 8.0$ kg und Streuung $s = 1.5$ kg.

Für die kumulative Häufigkeit $\Phi(Z)$ der Standard-Werte $Z = (X - m) / s$ gibt es Tabellen, welche angeben, wie viele Prozent aller Merkmalswerte Z kleiner oder gleich einem bestimmten Wert Z_1 sind. Mithilfe einer solchen Tabelle und der Standardisierungsformel können zwei wichtige Fragestellungen der angewandten Statistik beantwortet werden. Als Vorlage soll das obige Dynamobeispiel dienen.

Frage 1: *Wie viel Prozent aller Fälle erreichen eine bessere Leistung als 11 kg ?*

Durch Einsetzen in die Standardisierungsformel $Z = (11 - 8) / 1.5 = 2$ erhält man für den X-Wert 11 den standardisierten Z-Wert 2. Durch Nachschauen in der Tabelle erhält man die kumulative Häufigkeit $\Phi(2) = 0.9773$. Die fehlende Differenz auf 1 (bzw. 100%) ergibt dann den gesuchten Prozentsatz von 2.27%, der eine größere Leistung als 11 kg am Dynamometer erbringt.

Frage 2: *Welche Merkmalswerte X begrenzen die Hälfte aller Fälle derart, dass je 25% oberhalb und unterhalb des Mittelwertes liegen ?*

Aus der Tabelle der kumulativen Häufigkeiten liest man jenen Z_1 -Wert ab, unter dem 75% aller Fälle liegen. Das ergibt $Z_1 = 0.68$ mit $\Phi(Z_1) = 0.75$. Wegen der Symmetrie der Normalverteilung liegen zwischen -0.68 und $+0.68$ genau 50% von allen Fällen, weil ja $50 = (100 - 2 \cdot 25)$. Durch Einsetzen in die Standardisierungsformel von $-0.68 = (X_1 - 8) / 1.5$ und $+0.68 = (X_2 - 8) / 1.5$ erhält man dann die gesuchten Grenzen $X_1 = -0.68 \cdot 1.5 + 8 = 6.98$ und $X_2 = +0.68 \cdot 1.5 + 8 = 9.02$. Also liegen zwischen den Grenzen von 6.98 kg und 9.02 kg genau 50% aller Merkmalswerte.

5. Beispiel (Normalverteilung)

Glühlampen werden maschinell erzeugt. Die Brenndauer der Glühlampen ist normal verteilt. Repräsentative Stichproben ergeben dabei einen Mittelwert von $m = 1200$ Stunden und eine Streuung von $s = 200$ Stunden.

Frage 1: *Wie groß ist die Wahrscheinlichkeit, dass die Brenndauer einer Glühlampe zwischen 1300 und 1400 Stunden liegt ?*

Eine Standardisierung $Z = (X - m) / s$ der normalverteilten Zufallsvariable X liefert folgende Werte: $Z(1300) = 100 / 200 = 0.5$ und $Z(1400) = 200 / 200 = 1.0$. Aus der Tabelle erhält man dann die kumulative Häufigkeiten $\Phi(1) = 0.8413$ und $\Phi(0.5) = 0.6915$. Die gesuchte Wahrscheinlichkeit $h[0.5 < Z \leq 1.0]$ erhält man mittels $\Phi(1.0) - \Phi(0.5) = 0.15$. Die Wahrscheinlichkeit, dass die Brenndauer einer Glühlampe zwischen 1300 und 1400 Stunden liegt, beträgt 15%.

Frage 2: *Welche Toleranzgrenzen muss man setzen, wenn nur 5% der Produktion als Ausschuss deklariert werden ?*

Außerhalb des Toleranzintervalls $[-g ; +g]$ liegen 5%. $\Phi(g) = 1 - 0.025 = 0.975$. Der Wert g wird der Tabelle der kumulativen Häufigkeiten entnommen: $g = 1.96$. Für X erhält man $X = s \cdot g + m = 392 + m = 1592$. Daraus ergibt sich dann der gesuchte Toleranzbereich in Stunden: $[m - 392 ; m + 392] = [808 ; 1592]$.

Beurteilende Statistik

Die *beschreibende Statistik* beschäftigt sich mit der Verteilung von Häufigkeiten in Stichproben, die aus Populationen gewonnen werden. Diese Häufigkeitsverteilungen werden mit bestimmten Kennzahlen (Parametern) beschrieben: Mittelwert, Varianz, Korrelation, usw.

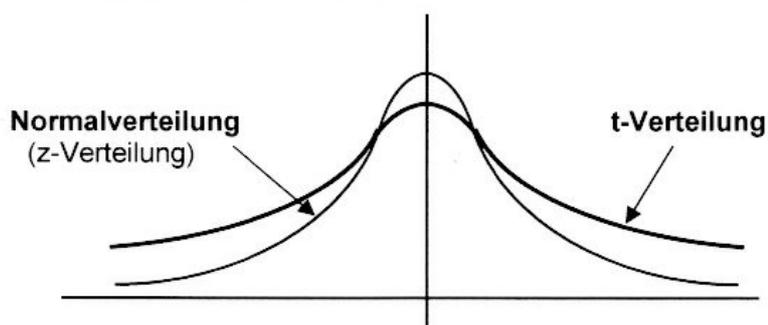
Die *beurteilende Statistik* beschäftigt sich mit der Frage, wie sich die Häufigkeitsverteilung eines Merkmals ändert, wenn man von Stichproben auf die zugrunde liegende Population übergeht. Dabei werden Schätzverfahren mit speziellen Testverteilungen angewendet, die zum Teil mathematisch kompliziert sind.

Typische Fragestellungen der beurteilenden, schlussfolgernden Statistik sind:

- (a) *Schätzen der Populationsparameter* mit Hilfe von Intervallen, in denen sie sich mit einer bestimmten Wahrscheinlichkeit befinden (Konfidenzbereich).
- (b) *Prüfen von Hypothesen* über den Unterschied von Mittelwerten oder über das Verhältnis von Varianzen, usw. (Signifikanztest).

Ist eine Hypothese auf dem **5%-Fehlerniveau signifikant**, dann bedeutet das anschaulich Folgendes: Wenn man aus einer Population 100 zufällige Stichproben herausgreift, dann wird die Hypothese in 95% aller Fälle zutreffen.

In vielen Fragestellungen wird untersucht, ob der Unterschied der Mittelwerte von zwei unabhängigen Stichproben signifikant ist, d.h. ob er mit einer hohen Wahrscheinlichkeit auch beim Übergang auf die zwei zugrunde liegenden Populationen erhalten bleibt. Aus der theoretischen Statistik weiß man, dass die Mittelwertsunterschiede in normalverteilten Populationen mit unbekanntem Streuungen einer Testverteilung vom t-Typ gehorchen. Die t-Verteilung ist so wie die Normalverteilung symmetrisch, jedoch etwas flacher.



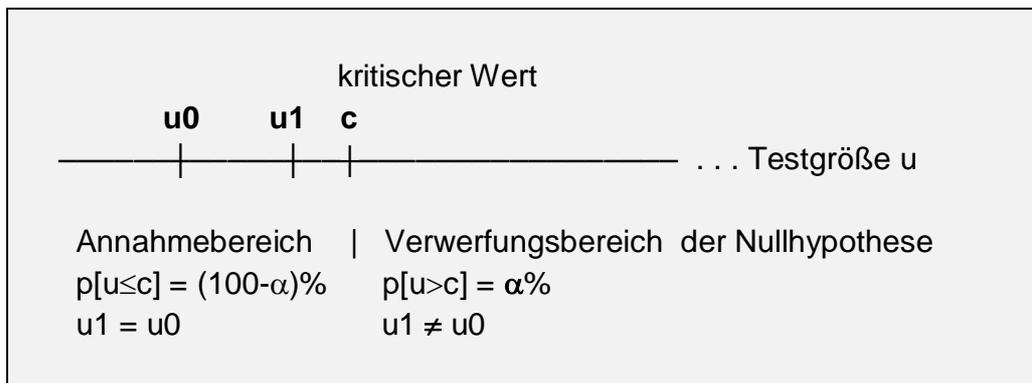
Zur Lösung von diesen Grundaufgaben ist die Kenntnis der Häufigkeitsverteilung jener statistischen Kennzahlen erforderlich, auf die sich eine Hypothese bezieht. Entnimmt man als Denkmodell einer Population eine große Anzahl unabhängiger, gleich großer, zufälliger Stichproben und berechnet für jede die gewünschte statistische Kennzahl (z.B. den Mittelwert m), so werden sich diese Stichprobenwerte um den entsprechenden Populationsparameter (also μ) verteilen. Die Häufigkeitsverteilung der Stichprobenwerte oder bestimmter mathematischer Testgrößen u , welche die Stichprobenwerte enthalten, heißt **Testverteilung**. Ist deren Funktionsverlauf im Voraus bekannt, dann kann man (ähnlich wie für normal verteilte Maßzahlen) aus Tabellen für die kumulative Häufigkeit jene Wahrscheinlichkeit ermitteln, mit der die Stichprobenkennzahl vom Populationsparameter um eine bestimmte Intervallbreite abweicht. In der angewandten Statistik sind neben der **Normalverteilung** die ebenfalls symmetrische, aber flachere **t-Verteilung**, die asymmetrische χ^2 -Verteilung (CHI-Quadrat) und die davon abgeleitete **F-Verteilung** die wichtigsten Testverteilungen. So sind z.B. die Stichproben-Mittelwerte einer normalverteilten Zufallsvariablen selbst wieder normalverteilt. So gehorchen z.B. die Summen der Abweichungsquadrate vom Stichproben-Mittelwert einer normalverteilten Zufallsvariablen der χ^2 -Verteilung. Schließlich kann gezeigt werden, dass die Verhältnisse der Varianzen normalverteilter Merkmale einer F-Verteilung entsprechen.

Für die kumulativen Häufigkeiten (Verteilungsfunktionen) von diesen Testverteilungen gibt es detaillierte Tabellen, wo abgelesen werden kann, mit welcher Wahrscheinlichkeit p der entsprechende Testwert u größer als ein vorgegebener kritischer Wert c ist. Dadurch kann man bei der Schätzung von Populationsparametern durch Stichprobenwerte ein Intervall angeben $[c_1, c_2]$, in dem der Parameter sich mit einer bestimmten Wahrscheinlichkeit p befindet. Diese so bestimmte Wahrscheinlichkeit wird auch Sicherheit der Schätzung oder Konfidenz genannt. Das ermittelte Intervall heißt **Konfidenzintervall**.

Bei **statistischen Hypothesen** geht es um Aussagen über die Verteilung von bestimmten **Prüfgrößen** u (z.B. Mittelwerts-Unterschieden). Unter der Nullhypothese H_0 versteht man die Annahme, dass in der Grundgesamtheit der Prüfgrößenwert u_1 gleich einem Sollwert u_0 ist, d.h. etwaige Abweichungen von diesem Sollwert in erhobenen Stichproben sind zufällig und verschwinden beim Übergang in die Grundgesamtheit, d.h. ($u_1 = u_0$). Als Alternativhypothese H_1 wird das Gegenteil dieser Annahme bezeichnet, nämlich ($u_1 \neq u_0$). Beispielsweise ermittelt man zum Testen der einseitigen Alternativhypothese ($u_1 > u_0$) aus der Testverteilung der Prüfgröße u jenen kritischen Wert c , sodass die Wahrscheinlichkeit p für $u > c$ gleich einem bestimmten Betrag α ist. Man schreibt diese Bedingung folgendermaßen an: $p[u > c] = \alpha$. Dabei heißt α die Irrtumswahrscheinlichkeit (Fehlerniveau).

Zumeist wird α mit 5%, manchmal sogar mit 1% vorgegeben. Die Differenz $(100 - \alpha)$ heißt **Signifikanz** (statistische Sicherheit).

Hat sich nun herausgestellt, dass ein vorliegender Prüfwert u_1 tatsächlich größer als der kritische Wert c ist, dann wird die Nullhypothese H_0 mit einer Irrtumswahrscheinlichkeit von α (z.B. 5%) verworfen und die Alternativhypothese H_1 mit der Signifikanz von $100 - \alpha$ (z.B. 95%) angenommen. Andernfalls, wenn $u_1 \leq c$ ist, dann muss die Nullhypothese H_0 beibehalten werden.



u_0 = Sollwert, u_1 = Prüfwert, Nullhypothese: $u_1 = u_0$.

So kann beispielsweise in einem Versuch über die Wirksamkeit von einer leistungssteigernden Droge ein Leistungsanstieg nach der Einnahme des Präparates ein Anzeichen für eine tatsächliche chemische Wirkung sein (Alternativhypothese H_1) oder ein zufallsbedingtes Ergebnis der Fehlerstreuung der Messwerte (Nullhypothese H_0). Zur Prüfung wird man zwei parallelisierte Stichproben heranziehen, d.h. die Leistungsmittelwerte sind in den Gruppen nicht unterschiedlich.

Diese Parallelisierung kann durch geeignete Vortests sichergestellt werden. Die Versuchsgruppe erbringt die eigentliche Testleistung unter Drogeneinfluss, die Kontrollgruppe hingegen ohne Drogeneinwirkung. Das ergibt die Leistungsmittelwerte m_2 und m_1 . Als statistische Prüfgröße u wird der Mittelwertsunterschied ($m_2 - m_1$) genommen. Die Frage ist nun, ob diese Differenz auch erhalten bleibt (signifikant ist), wenn von den Stichproben zu den Populationen übergegangen wird. Nullhypothese H_0 lautet $\mu_2 = \mu_1$, Alternativhypothese hingegen lautet $\mu_2 > \mu_1$. Dabei sind μ_1 , μ_2 die Mittelwerte in den zwei zugehörigen Populationen. Aus der theoretischen Statistik weiß man nun, dass die Mittelwertsunterschiede in normalverteilten Populationen mit unbekanntem Streuungen einer Testverteilung vom t -Typ gehorchen.

Aus der Tabelle der t -Verteilung ermittelt man den kritischen Wert c derart, dass die Wahrscheinlichkeit für $t > c$ gleich 5% ist. Liegt nun die vorliegende Prüfgröße t_1 (das ist der bestehende Mittelwertsunterschied $m_2 - m_1$) oberhalb des kritischen Wertes ($t_1 > c$), dann muss die Nullhypothese ($\mu_2 = \mu_1$) verworfen werden.

In diesem Fall hat sich der Mittelwertsunterschied als signifikant herausgestellt, die Droge zeigt tatsächlich eine leistungssteigernde Wirkung. Im anderen Fall muss die Nullhypothese beibehalten und von einer leistungssteigernden Drogenwirkung abgesehen werden.

Das obige Beispiel zeigt den prinzipiellen Entscheidungsweg bei der Prüfung von statistischen Hypothesen. Dabei können grundsätzlich zwei Arten von Fehlern auftreten. Ein so genannter Fehler erster Art wird begangen, wenn die Hypothese H_0 zu Unrecht verworfen wird (ein echter Drogeneffekt wird gefolgert, obwohl die Leistungsdifferenz nur zufallsbedingt ist). Ein so genannter Fehler zweiter Art liegt dann vor, wenn die Nullhypothese zu Unrecht beibehalten wird (die Leistungsdifferenz wird als eine bloß zufällige angesehen, obwohl sie jedoch eine echte Drogenauswirkung ist).

Neben den Mittelwertsunterschieden werden hauptsächlich Streuungsverhältnisse und Korrelationen statistisch auf ihre Signifikanz getestet. Die hier kurz dargestellten Methoden der statistischen Überprüfung von Hypothesen sind ein unentbehrliches Werkzeug in sämtlichen empirischen Wissenschaften.

Varianzanalyse

Die Varianzanalyse ist ein statistisches Verfahren zur Überprüfung der Verschiedenheit von mehr als nur zwei Mittelwerten. Ein Merkmal X wird von einer Einflussgröße A beeinflusst. Die Einflussgröße A kommt dabei in R verschiedenen Ausprägungsgraden (A_1, A_2, \dots, A_R) vor. Zu jeder dieser Intensitätsstufen der Einflussgröße A gibt es eine Klasse von N_i Individuen, sodass die gesamte Stichprobe in R unterschiedliche Klassen mit jeweils N_1, N_2, \dots, N_R Individuen eingeteilt wird. Die entsprechenden Werte des Merkmals X sollen in den einzelnen Klassen normalverteilt sein und alle die gleiche Varianz σ^2 besitzen. Die Nullhypothese H_0 behauptet nun, dass alle R Klassenmittelwerte $\mu_1, \mu_2, \dots, \mu_R$ der entsprechenden Populationen gleich groß sind, d.h. die verschiedenen Ausprägungsgrade der Größe A haben keinen Einfluss auf das untersuchte Merkmal X . Die Alternativhypothese H_1 behauptet hingegen das Gegenteil, nämlich dass die Größe A eine echte Veränderung des Merkmals X bewirkt.

| | | | | |
|-----------|-----------|----------|--|-----------|
| A_1 | A_2 | | | A_R |
| X_{11} | X_{21} | | | X_{R1} |
| X_{12} | X_{22} | | | X_{R2} |
| | | | | |
| | | X_{ij} | | |
| | | | | |
| X_{1N1} | X_{2N2} | | | X_{RNR} |

Datentabelle für eine einfache Varianzanalyse.

X_{ij} = Merkmal der j -ten Person in der i -ten Einflussklasse A_i .

Zur Überprüfung der Nullhypothese zerlegt man die "**totale**" Gesamtvarianz (QT, Summe der Abweichungsquadrate aller Stichprobenwerte X_{ij} vom Gesamtmittelwert m , $QT = \sum_i \sum_j (X_{ij} - m)^2$) in eine Varianz "**zwischen**" (QZ, Summe der Abweichungsquadrate der Klassenmittelwerte m_i vom Gesamtmittelwert m) und in eine Varianz "**innerhalb**" (QI, Summe der Abweichungsquadrate der Stichprobenwerte X_{ij} von ihren Klassenmittelwerten m_i der einzelnen Klassen). Ohne große Schwierigkeiten kann hergeleitet werden, dass $QT = QZ + QI$ gilt. Die Varianz „zwischen“ ist ein Maß für eine durch die Einflussgröße A bedingte Variabilität, die Varianz „innerhalb“ ist ein Maß für die zufallsbedingte Variabilität. Diese beiden Varianzanteile werden nun auf einen signifikanten Unterschied hin miteinander verglichen (Signifikanzniveau α %).

Aus der theoretischen Statistik weiß man, dass das Verhältnis von Varianzen normalverteilter Zufallsvariablen $V = QZ / QI$ als statistische Prüfgröße der F-Verteilung gehorcht. Aus der Tabelle der kumulativen F-Verteilung entnimmt man den kritischen Wert c , sodass die Wahrscheinlichkeit für $V > c$ dem gewünschten Signifikanzniveau α entspricht.

Wenn nun der tatsächlich berechnete Wert $V_1 \leq c$ ist, dann wird die Nullhypothese ($\mu_1 = \mu_2 = \dots = \mu_R$) beibehalten. Ist hingegen $V_1 > c$, so wird die Nullhypothese verworfen, d.h. es gibt echte Mittelwertsunterschiede zwischen den einzelnen Klassen und die Einflussgröße A zeigt eine signifikante Auswirkung auf das Merkmal X.

Beispielsweise soll in einem Experiment der Effekt von verschiedenen Massenmedien (Zeitung, Radio, Fernsehen) auf die Einstellungsänderung von Personen zu einem aktuellen Thema untersucht werden. Dazu erhalten drei unabhängige Stichproben inhaltlich dieselbe Information: die erste Stichprobe in Form eines Zeitungsartikels (Z), die zweite in Form einer Radiosendung (R) und die dritte in Form einer Fernsehsendung (F). Den Versuchspersonen wird ein Einstellungsfragebogen vorgelegt, einmal vor der Informationsübermittlung (Ergebnis X_1) und einmal nach der Informationsübermittlung (Ergebnis X_2). Daraus lässt sich eine mögliche Einstellungsänderung quantitativ bestimmen ($X = X_1 - X_2$) und für jede der drei Stichproben die mittlere Einstellungsänderung (m) berechnen. Die Frage ist nun, ob sich diese drei Mittelwerte (m_Z, m_R, m_F) signifikant voneinander unterscheiden. Die Nullhypothese behauptet, dass die Mittelwerte in den zugehörigen Populationen aller Zeitungsleser, Radiohörer und Fernseher gleich sind.

Es besteht die nahe liegende Alternativhypothese, dass der Effekt des Fernsehens wesentlich größer ist als jener von Radio bzw. Zeitung, weil das Fernsehen sowohl optische als auch akustische Wahrnehmungen ermöglicht (Multimedia). Die **Varianzanalyse** ermöglicht es nun, die Signifikanz solcher Hypothesen zu überprüfen.

Neben dieser einfachen Varianzanalyse gibt es auch multiple Varianzanalysen zur Feststellung der Einwirkung von mehreren Einflussgrößen A, B, C, auf ein Merkmal X. Als abschließendes Beispiel sei die Fragestellung einer doppelten Varianzanalyse beschrieben.

Ein Merkmal X wird von zwei Einflussgrößen A, B beeinflusst. Die zwei Einflussgrößen treten in mehreren Stärkegraden auf, z.B. Einflussgröße A in drei $\{A_1, A_2, A_3\}$ und Einflussgröße B in zwei $\{B_1, B_2\}$. Von der gesamten Stichprobe N sollen auf die einzelnen Kombinationen der beiden Einflussgrößen {sechs Zellen: $A_1B_1, A_1B_2, A_2B_1, A_2B_2, A_3B_1, A_3B_2$ } gleich viele Individuen R entfallen, d.h. $N = 6 * R$.

| | A ₁ | A ₂ | A ₃ |
|----------------|---|---|---|
| B ₁ | X ₁₁₁ X _{11R} | X ₁₂₁ X _{12R} | X ₁₃₁ X _{13R} |
| B ₂ | X ₂₁₁ X _{21R} | X ₂₂₁ X _{22R} | X ₂₃₁ X _{23R} |

Datentabelle für eine doppelte Varianzanalyse.

X_{kij} = Merkmal der j-ten Person unter Einfluss von A_i und B_k.

Es wird vorausgesetzt, dass die beiden Einflussgrößen A, B voneinander unabhängig sind, und dass alle Variablen in den sechs Zellen normalverteilt sind und auch die gleiche Varianz σ^2 besitzen. Folgende Alternativhypothesen können nun formuliert und statistisch überprüft werden:

- H1: Die einzelnen Zellenmittelwerte unterscheiden sich signifikant.
- H2: Die Klassenmittelwerte von der Größe A unterscheiden sich signifikant.
- H3: Die Klassenmittelwerte von der Größe B unterscheiden sich signifikant.
- H4: Es besteht eine signifikante Wechselwirkung zwischen den Größen A und B auf das Merkmal X, d.h. die Größe A zeigt bei verschiedenen Ausprägungsraden der Größe B unterschiedliche Einwirkungen auf X.

Sowie bei der einfachen Varianzanalyse wird auch hier die Gesamtvarianz in die verschiedenen Varianzanteile aufgespaltet und deren Unterschiede auf statistische Signifikanz hin überprüft.

Ausführliche Darstellungen von Varianzanalysen findet man in den einschlägigen Lehrbüchern der Statistik.

WAHRSCHEINLICHKEITSRECHNUNG

| | |
|--------------------------------|--------|
| Kombinatorik | [72] |
| Binomialverteilung | [77] |
| Normalverteilung | [80] |
| Zwei Zufallsexperimente | [83] |
| Wahrscheinlichkeitstheorie I | [85] |
| Wahrscheinlichkeitstheorie II | [88] |
| Wahrscheinlichkeitstheorie III | [93] |
| 23 Übungsaufgaben | [97] |

Kombinatorik

Permutationen

Bildet man aus n wohl unterscheidbaren Elementen alle möglichen Anordnungen, so dass jedes Element in jeder Anordnung genau einmal vorkommt, dann nennt man eine solche Anordnung eine Permutation.

Beispiel: $G = \{a, b, c\}$. Durch zyklisches Vertauschen der drei Elemente erhält man folgende 6 Anordnungen:

$(a,b,c), (a,c,b), (b,a,c), (b,c,a), (c,a,b), (c,b,a)$.

Wie viele solche Permutationen von n Elementen gibt es?

Die Anzahl der Permutation von $(n-1)$ Elementen sei $P(n-1)$. Eine solche Permutation sei

_ a_1 _ a_2 _ a_3 _ . . . _ $a_{(n-1)}$ _

Auf jede Leerstelle _ in dieser Anordnung kann das n -te Element $a(n)$ gestellt werden. Es sind dann $n * P(n-1)$ Anordnungen möglich. Also gilt $P(n) = n * P(n-1)$. Beginnt man diese Formel mit $n = 1$ und setzt sie schrittweise fort, dann gilt:

$P(n) = 1 * 2 * 3 * \dots * (n-1) * n$

Man schreibt dafür $n!$ und sagt dazu "n Faktorielle".

Für die Anzahl der Permutationen von n Elementen $P(n)$ gilt:

$P(n) = n! = 1 * 2 * 3 * \dots * (n-1) * n$

Gegeben ist eine Menge mit 5 Elementen $G = \{a, b, c, d, e\}$.

Es gibt damit $5! = 1*2*3*4*5 = 120$ verschiedene Permutationen.

Sind nun beispielsweise von 5 Elementen $\{a, b, c, d, e\}$ genau drei Elemente gleich, z.B. $a = b = c$, so unterscheiden sich die $3! = 1*2*3 = 6$ Permutationen, welche mit diesen drei Elementen gebildet werden können, nicht voneinander. Die Gesamtanzahl der Permutationen beträgt jetzt nur den sechsten Teil der früheren.

Wenn wir mit $P(n|k)$ die Anzahl der Permutationen von genau n Elementen bezeichnen, wo sich k Elemente wiederholen, dann sind das Permutationen MIT Wiederholung: **$P(n|k) = n! / k!$**

Variationen

Greift man aus einer Menge von n verschiedenen Elementen genau k Elemente heraus und berücksichtigt ihre Reihenfolge, so erhält man eine Variation der n Elemente zur Klasse k . Die Anzahl der Variationen ist $V(n,k)$.

$G = \{a, b, c, d\}$. Es gibt hier 12 Variationen von 4 Elementen zur Klasse 2: $ab, ba, ac, ca, ad, da, bc, cb, bd, db, cd, dc$.

Wie viele Variationen von n Elementen zur Klasse k gibt es?

$G = \{a_1, a_2, \dots, a_n\}$ mit n wohl unterschiedenen Elementen. Variationen zur Klasse 1: a_1, a_2, \dots, a_n . Es gilt $V(n,1) = n$.

Variationen zur Klasse 2 werden gebildet, indem man zu jeder Variation der Klasse 1 die restlichen $(n-1)$ Elemente dazustellen: $a_1a_2, a_1a_3, \dots, a_1a_n; a_2a_1, a_2a_3, \dots, a_2a_n; a_3a_1, a_3a_2, \dots$. Das ergibt dann $V(n,2) = V(n,1) \cdot (n-1) = n \cdot (n-1)$.

Variationen zur Klasse 3 werden gebildet, indem man zu jeder Variation der Klasse 2 die restlichen $(n-2)$ Elemente dazustellen: $a_1a_2a_3, a_1a_3a_2, \dots, a_1a_n a_{n-1}; a_2a_1a_3, a_2a_3a_1, \dots, \dots$. Das ergibt dann $V(n,3) = V(n,2) \cdot (n-2) = n \cdot (n-1) \cdot (n-2)$.

Variationen zur Klasse k werden gebildet, indem man zu jeder Variation der Klasse $k-1$ die restlichen $(n-k+1)$ Elemente stellt. Das ergibt dann: $V(n,k) = V(n,k-1) \cdot (n-k+1) =$

Für Variationen gilt: $V(n,k) = n \cdot (n-1) \cdot \dots \cdot (n-k+1)$

Erlaubt man, dass in jeder Variationen von n Elementen zur Klasse k jedes Element sich auch wiederholen darf, dann sind das Variationen mit Wiederholung $V(n,k|k) = n \cdot n \cdot \dots \cdot n = n^k$. z.B. $G = \{a, b, c\}$: $V(3,2|2) = 3^2 = 9$ mit folgenden Variationen: $aa, ab, ac, ba, bb, bc, ca, cb, cc$.

Für Variationen mit Wiederholung gilt: $V(n,k|k) = n^k$

Kombinationen

Wieviele Möglichkeiten gibt, es aus einer Menge von n verschiedenen Elementen genau k Elemente herauszugreifen, wobei es auf die Reihenfolge der Elemente nicht ankommt ?

Kombinationen sind also Variationen von n Elementen zur Klasse k mit nur einer Reihenfolge der k Elemente einer Klasse. Ihre Anzahl bezeichnet man mit $K(n,k)$ und sagt " n über k ". Verwendet wird häufig folgendes Symbol:

$$\binom{N}{K}$$

Die Anzahl solcher Kombinationen von n Elementen zur Klasse k erhält man, indem man die Anzahl der Variationen $V(n,k)$ durch die Anzahl der Permutationen $P(k)$ dividiert. Daraus folgt: $K(n,k) = V(n,k)/P(k) = (n \cdot (n-1) \cdot \dots \cdot (n-k+1)) / (1 \cdot 2 \cdot \dots \cdot k)$. Erweitert man mit $(n-k)! = (n-k) \cdot (n-k-1) \cdot \dots \cdot 2 \cdot 1$, dann gilt: $K(n,k) = (n \cdot (n-1) \cdot \dots \cdot (n-k+1)) / (1 \cdot 2 \cdot \dots \cdot k) = n! / (k! \cdot (n-k)!)$.

Beispiel: $G = \{a, b, c, d, e\}$. Gesucht sind alle Kombinationen der 5 Elemente zur Klasse 3. Es ist $K(5,3) = 5 \cdot 4 \cdot 3 / 1 \cdot 2 \cdot 3 = 10$. Die Kombinationen sind: ab, ac, ad, ae, bc, bd, be, cd, ce, de.

Kombinationen ohne Wiederholung: $K(n,k) = n! / (k! \cdot (n-k)!)$

Aus der Bauweise dieser Formel folgen die Eigenschaften:

$$K(n,k) = K(n,n-k).$$

$$K(n,n) = K(n,0) = 1.$$

$$K(n,k) + K(n,k+1) = K(n+1,k+1). \text{ Lehrsatz [KA].}$$

Beweis des Lehrsatzes [KA] von der Kombinationenaddition:

$$K(n,k) = n \cdot (n-1) \cdot \dots \cdot (n-k+1) / 1 \cdot 2 \cdot \dots \cdot k.$$

$$K(n,k+1) = n \cdot (n-1) \cdot \dots \cdot (n-k+1) \cdot (n-k) / 1 \cdot 2 \cdot \dots \cdot k \cdot (k+1).$$

$K(n,k) + K(n,k+1)$ auf gleichen Nenner bringen

$$K(n,k) + K(n,k+1) = (n+1) \cdot n \cdot (n-1) \cdot \dots \cdot (n-k+1) / 1 \cdot 2 \cdot \dots \cdot (k+1).$$

$$K(n,k) + K(n,k+1) = K(n+1,k+1).$$

Kombinationen mit Wiederholung

Erlaubt man, dass in jeder Kombination von n Elementen zur Klasse k jedes Element innerhalb der Klasse sich wiederholen darf, dann wirkt das so, als hätte man noch $(k-1)$ Elemente hinzugefügt: z.B. $G = \{a, b, c\}$ mit $K(3,2) = 3$. Aus ab, ac, bc wird bei Wiederholung ab, ac, "aa", bc, "bb", "cc". Das ergibt $6 = 4 \cdot 3 / 1 \cdot 2 = K(4,2)$. Die Kombinationen mit Wiederholung sind mit $K(n,k|k)$ bezeichnet und es gilt: $K(n,k|k) = K(n+k-1,k)$.

Der binomische Lehrsatz

Ein Binom $(a+b)$ soll fortgesetzt potenziert werden:

$$(a+b)^0 = 1$$

$$(a+b)^1 = 1 \cdot a + 1 \cdot b$$

$$(a+b)^2 = 1 \cdot a^2 + 2 \cdot a \cdot b + 1 \cdot b^2$$

$$(a+b)^3 = 1 \cdot a^3 + 3 \cdot a^2 \cdot b + 3 \cdot a \cdot b^2 + 1 \cdot b^3$$

Wir sehen am Beispiel von $(a+b)^3$, dass die Koeffizienten 1, 3, 3, 1 gleich den Kombinationen "3 über k " sind:

$$K(3,0) = 1, K(3,1) = 3, K(3,2) = 3 \text{ und } K(3,3) = 1.$$

Allgemein gilt folgender "binomische Lehrsatz":

Das Binom $(a+b)$ wird zur n -ten Potenz erhoben: $(a+b)^n$.

Durch schrittweises Ausmultiplizieren erhält man als Ergebnis

$$K(n,0) \cdot a^n +$$

$$K(n,1) \cdot a^{(n-1)} \cdot b +$$

..... +

$$K(n,n-1) \cdot a \cdot b^{(n-1)} +$$

$$K(n,n) \cdot b^n$$

Die Koeffizienten $K(n,k)$ sind die Kombinationen "n über k".
Der Beweis des Satzes erfolgt durch vollständige Induktion.

Der Satz hat viele Anwendungen. So ist die Summe von allen
Kombinationen $\sum(K(n,k)) = 2^n$ für k von 0 bis n. Das folgt
direkt aus dem binomischen Lehrsatz, weil $2^n = (1+1)^n$.

Die Binomialkoeffizienten $K(n,k)$ für $n = 8$ bis 1.

| | | | | | | | |
|----|----|----|----|---|---|---|---|
| 1 | | | | | | | |
| | 1 | | | | | | |
| 8 | | 1 | | | | | |
| | 7 | | 1 | | | | |
| 28 | | 6 | | 1 | | | |
| | 21 | | 5 | | 1 | | |
| 56 | | 15 | | 4 | | 1 | |
| | 35 | | 10 | | 3 | | 1 |
| 70 | | 20 | | 6 | | 2 | |
| | 35 | | 10 | | 3 | | 1 |
| 56 | | 15 | | 4 | | 1 | |
| | 21 | | 5 | | 1 | | |
| 28 | | 6 | | 1 | | | |
| | 7 | | 1 | | | | |
| 8 | | 1 | | | | | |
| | 1 | | | | | | |
| 1 | | | | | | | |

Permutationen, Variationen und Kombinationen

Anzahl $n = 5$

Klasse $k = 2$

Permutationen ohne Wiederholung $P(n) = 120$

$$P(n) = n!$$

Permutationen mit Wiederholung $P(n|k) = 60$

$$P(n|k) = n! / k!$$

Variationen ohne Wiederholung $V(n,k) = 20$

$$V(n,k) = n! / (n-k)!$$

Variationen mit Wiederholung $V(n,k|k) = 25$

$$V(n,k|k) = n^k$$

Kombinationen ohne Wiederholung $K(n,k) = 10$

$$K(n,k) = n! / (k!(n-k)!)$$

Kombinationen mit Wiederholung $K(n,k|k) = 15$

$$K(n,k|k) = K(n+k-1, k)$$

Binomialverteilung

Gegeben ist ein Ereignisraum $G = \{e, \neg e\}$ mit nur zwei Elementarereignissen e und $\neg e$.

e hat die Grundwahrscheinlichkeit $p = p(e)$ und sein Gegenereignis $\neg e$ hat die Wahrscheinlichkeit $q = p(\neg e) = 1 - p$.

Ein Zufallsexperiment mit diesem Ereignisraum wird n Mal in unabhängiger Weise wiederholt. A sei dann jenes Ereignis, dass e bei diesen n Wiederholungen genau k Mal auftritt. Gesucht ist $p(A)$, also die Wahrscheinlichkeit von A .

Wir bilden eine Zufallsvariable X , welche den Elementarereignissen die Zahlenwerte 1 und 0 zuordnet. Der Ereigniswahrscheinlichkeit $p(A)$ entspricht dann $p(X = k)$. In einer Menge von n Elementen lassen sich $K(n, k)$ Kombinationen der Klasse k bilden. Wegen ihrer gegenseitigen Unabhängigkeit werden die Wahrscheinlichkeiten multipliziert, sodass gilt $p(X = k) = K(n, k) * p^k * q^{(n-k)}$, weil beim Ereignis A das Elementarereignis e genau k Mal vorkommt, sein Gegenereignis aber genau $(n - k)$ Mal. Die Verteilung von $p(X = k)$ heißt "Binomialverteilung" mit $p(X = k)$ als "Verteilungsdichte".

Die Wahrscheinlichkeit, dass Ereignis e mindestens a Mal und höchstens b Mal auftritt, ist $p(a \leq k \leq b)$. Wegen der Unvereinbarkeit der einzelnen Ereignisse gilt der Additionssatz, so dass $p(a \leq k \leq b) = \sum(K(n, k) * p^k * q^{(n-k)})$ für k von a bis b .

Weiters gilt $p(0 \leq k \leq n) = \sum(K(n, k) * p^k * q^{(n-k)})$ für k von 0 bis n . Wegen dem Binomialsatz ist das $(p + q)^n$.
 $p(0 \leq k \leq n) = (p + 1 - p)^n = 1^n = 1$.

Beispiel 1: Wie groß ist die Wahrscheinlichkeit, dass das Elementarereignis e unter den n Versuchen mindestens einmal vorkommt?

$$p(1 \leq k \leq n) = p(0 \leq k \leq n) - p(k = 0) = 1 - q^n.$$

Beispiel 2: Bei Kreuzungsversuchen von Lebewesen sind 5% erfolgreich. Wie viele Versuche muss man durchführen, damit die Wahrscheinlichkeit w , dass ein Versuch gelingt, größer als 99% ist?

$e =$ "Versuch gelingt" und $\neg e =$ "Versuch gelingt nicht".
 $p(e) = 0.05$ und $p(\neg e) = 0.95$. Für die Wahrscheinlichkeit w , dass mindestens ein Versuch gelingt, gilt: $w = 1 - q^n > 0.99$.
 $q^n < 0.01$, $0.95^n < 0.01$, $n * \lg(0.95) < \lg(0.01)$. Daraus folgt $n > \lg(0.01) / \lg(0.95)$, d.h. $n > 89$.

Satz: Eine Binomialverteilung mit n Versuchen und den Grundwahrscheinlichkeiten p und q hat den Mittelwert $m = n * p$.

Beweis: Es sei $g(t) = (q + p*t)^n$ eine Hilfsfunktion. Dann gilt $g(t) = (q + p*t)^n = \sum(K(n,k)*q^{(n-k)}*p^k*t^k)$ für k von 0 bis n . $g(t) = \sum(f(k)*t^k)$ mit $f(k)$ als Wahrscheinlichkeitsdichte der Binomialverteilung. Für die erste Ableitung gilt dann $g'(t) = n*p*(q+p*t)^{(n-1)} = \sum(f(k)*k*t^{(k-1)})$. Mit $t = 1$ folgt $g'(1) = n*p*(q+p)^{(n-1)} = n*p*1 = \sum(k*f(k)) = m$, d.h. $m = n*p$.

Satz: Eine Binomialverteilung mit n Versuchen und den Grundwahrscheinlichkeiten p und q hat die Varianz $v = n * p * q$.

Beweis: Es sei $g(t) = (q + p*t)^n$ eine Hilfsfunktion. Dann gilt $g(t) = (q + p*t)^n = \sum(K(n,k)*q^{(n-k)}*p^k*t^k)$ für k von 0 bis n . $g(t) = \sum(f(k)*t^k)$ mit $f(k)$ als Wahrscheinlichkeitsdichte der Binomialverteilung. Für die zweite Ableitung gilt nun $g''(t) = n*(n-1)*p^2*(q + p*t)^{(n-2)} = \sum(f(k)*k*(k-1)*t^{(k-1)})$. Mit $(t=1)$ und $(q+p)=1$ folgt:

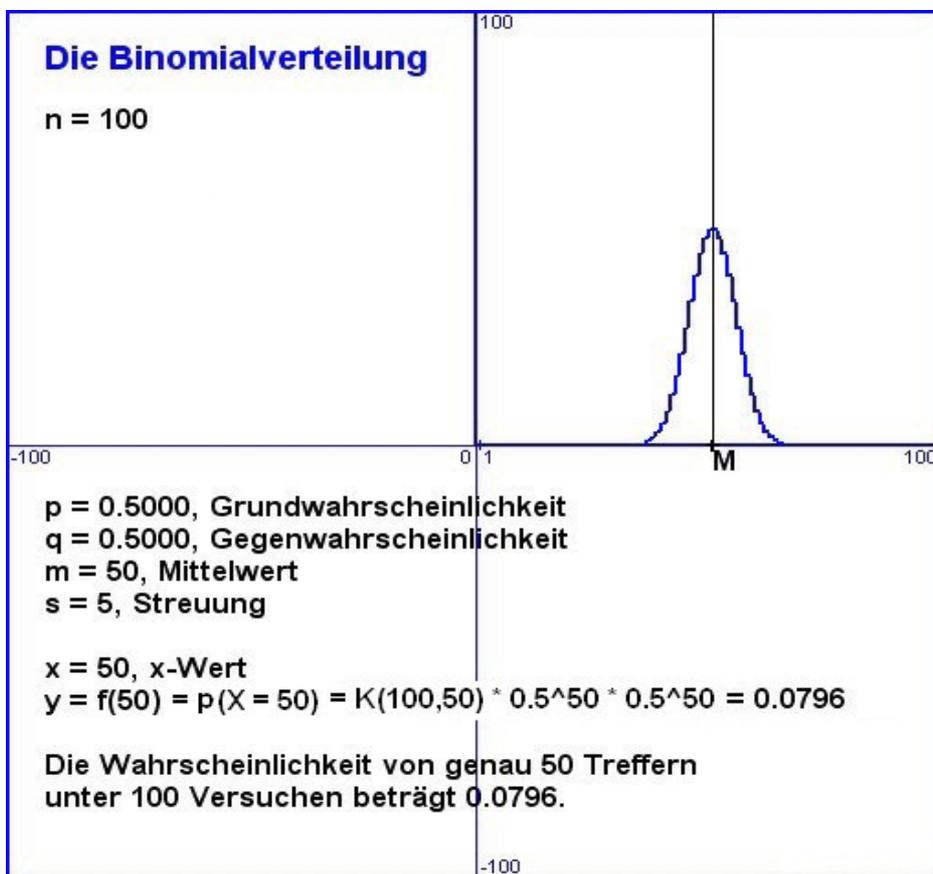
$$g''(1) = n*(n-1)*p^2 = \sum(f(k)*(k^2-k)).$$

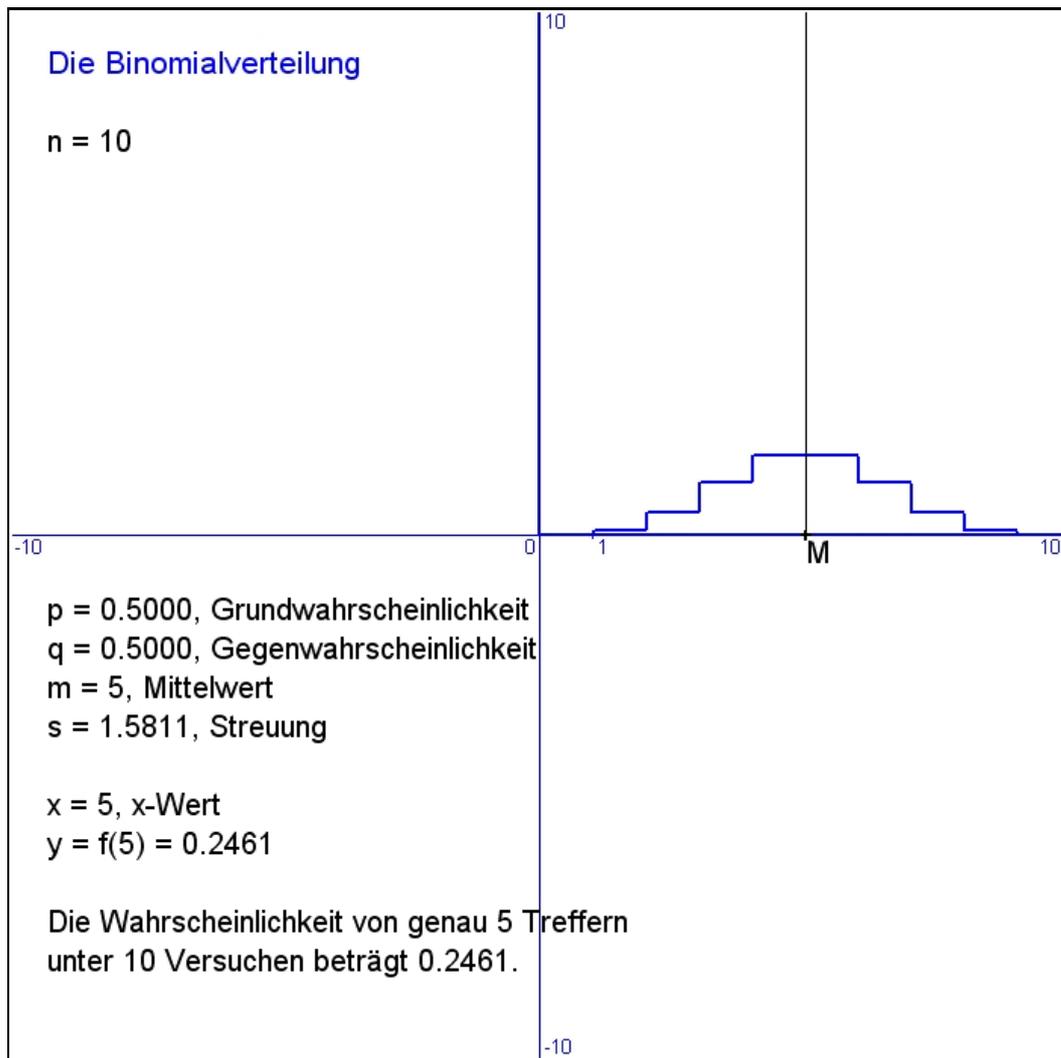
$$g''(1) = (n*p)^2 - n*p^2 = \sum(k^2*f(k)) - \sum(k*f(k)).$$

Aus der Statistik wissen wir, dass $v = \sum(k^2*f(k)) - m^2$.

$$(n*p)^2 - n*p^2 = v + m^2 - m, \text{ d.h. } m^2 - m * p = v + m^2 - m.$$

Daraus folgt $v = m - m * p = m * (1 - p)$. Also ist $v = n*p*q$.





Falls die **Laplace-Bedingung** $v = n * p * (1 - p) > 9$ für eine **Binomialverteilung** erfüllt ist, dann kann die Binomialverteilung gut angenähert werden durch eine **Normalverteilung** mit Mittelwert $m = n * p$ und Varianz $v = n * p * (1 - p)$. Das ist für Aufgaben mit einer großen Anzahl n rechentechnisch sehr praktisch.

Beispiel: Ein fairer Würfel wird 400 Mal aufgeworfen und die Häufigkeit der Zahl 5 wird dabei gezählt. Gesucht ist jener Bereich $[X_1; X_2]$ um den Mittelwert, in welchem die Anzahl der gewürfelten Fünfer mit 90%-iger Wahrscheinlichkeit liegt.

Lösung: Binomialverteilung, $n = 400$, $p = 1/6$, $q = (1 - p) = 5/6$, $m = 400 * p = 66.7$, $v = 400 * p * q = 55.5$, $s = \text{sqrt}(v) = 7.5$. Die Laplace-Bedingung $v = 55.5 > 9$ ist erfüllt. Daher kann die Binomialverteilung durch eine Normalverteilung $NV(66.7; 7.5)$ gut angenähert werden. Aus der Tabelle der Verteilungsfunktion der normierten $NV(0; 1)$ $Z = (X - m) / s$ folgt: $p(Z \leq -1.64) = 95\%$, d.h. $P(Z \leq -1.64 \text{ und } Z \geq +1.64) = 10\%$, d.h. $p(-1.64 < Z < +1.64) = 90\%$, d.h. $X_1 = m - 1.64 * s = 54.5$ und $X_2 = m + 1.64 * s = 78.9$.
Ergebnis: Mit 90%-iger Wahrscheinlichkeit liegt die Anzahl der gewürfelten Fünfer zwischen 54 und 79.

Normalverteilung

Viele quantitativ messbare Merkmale x (Körpergröße, Gewicht, usw.), die von mehreren verschiedenen Faktoren abhängen und die in nicht zu kleinen Stichproben erfasst werden, haben eine Häufigkeitsverteilung mit glockenförmigem Aussehen.

Diese Verteilungskurve verläuft symmetrisch zu einem Gipfel (Mittelwert) und nähert sich auf beiden Seiten davon asymptotisch der x -Achse. Je extremer ein Merkmalswert ist, umso seltener kommt er vor. Die relativen Häufigkeiten $h(x)$ können durch eine Formel beschrieben werden, die vom Mittelwert m und der Streuung s abhängt:

$$h(x) = 1/\sqrt{2\pi s^2} * \exp(-(x-m)^2/(2s^2)).$$

Die Formel wird viel einfacher, wenn man folgende Ersetzung vornimmt: $z = (x-m) / s$. Diese Standardisierung bedeutet eine Verschiebung des Mittelwertes in den Nullpunkt und Stauchung der verschobenen x -Werte durch Faktor s . Dann erhält man die so genannte *standardisierte Normalverteilung* mit Mittelwert 0 und Streuung 1.

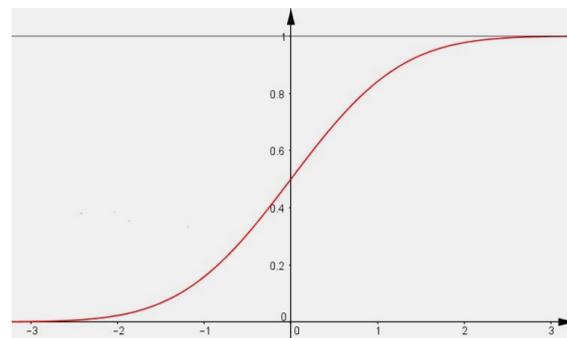
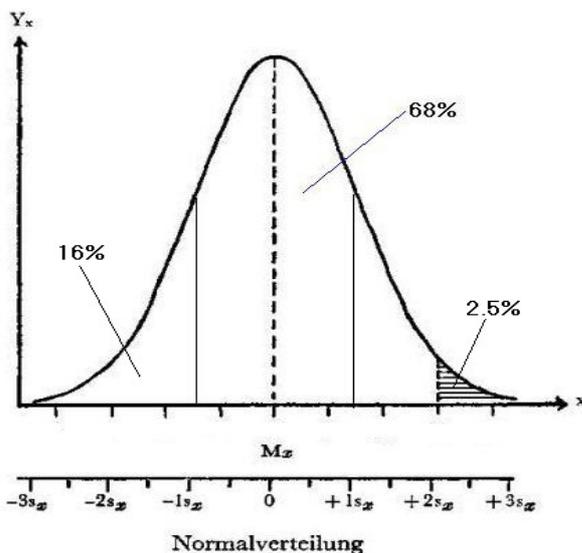
$$h(z) = 1/\sqrt{2\pi} * \exp(-z^2/2).$$

Die *kumulative Häufigkeit* $h[z \leq z_0]$ bzw. $100 * h[z \leq z_0]$ der standardisierten Normalverteilung gibt an, wie viele Prozente kleiner oder gleich einem bestimmten Wert z_0 sind.

$$h[z \leq z_0] = 1/\sqrt{2\pi} * \int_0^{z_0} \exp(-z^2/2) = \Phi(z_0)$$

(Die Summierung bzw. Integration erfolgt dabei von 0 bis z_0 .)

Für diese Prozentrangwerte ($\Phi(z_0)$, Verteilungsfunktion "phi") gibt es fertige Tabellen. Die Verteilungsfunktion für die Normalverteilung ist unten rechts abgebildet.



Für jede Normalverteilung gilt:

Im Bereich $[m - 1s ; m + 1s]$ liegen 68.00 % aller z -Werte (Normalbereich).

Im Bereich $[m - 2s ; m + 2s]$ liegen 95.00 % aller z -Werte.

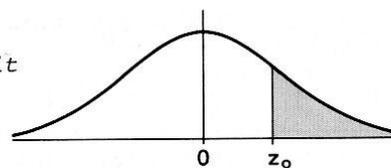
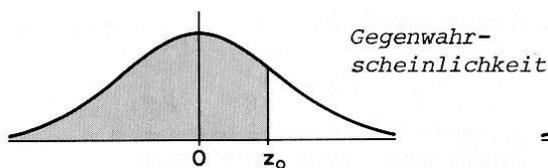
Im Bereich $[m - 3s ; m + 3s]$ liegen 99.75 % aller z -Werte.

In der theoretischen Statistik wird die relative Häufigkeit $h(x)$ auch als die Wahrscheinlichkeit $p(x)$ bezeichnet ("probability" des Merkmalswertes x). Die Wahrscheinlichkeit $p(x)$ ist der Grenzwert der relativen Häufigkeit $h(x)$ für den theoretischen Fall, dass die Stichprobe unendlich groß wird. Dem entsprechend schreibt man dann für die kumulative Häufigkeit $h[x \leq x_0]$ auch $p[x \leq x_0]$. Handelt es sich dabei um die standardisierten Werte $z = (x - m) / s$ dann schreibt man $p[z \leq z_0]$. Dieser Wahrscheinlichkeit entspricht eine bestimmte Fläche unter der standardisierten Normalverteilungskurve, und sie wird auch Verteilungsfunktion "phi" $\Phi(z_0)$ genannt. Dafür gibt es bereits fertige Tabellen zum Nachschlagen (siehe nächste Seite). In den folgenden Grafiken ist zu beachten, dass nur für den Fall A die Werte direkt aus der Tabelle entnommen werden können.

$$\frac{z_0 \geq 0}{\text{A: } P(z \leq z_0) = \Phi(z_0)}$$

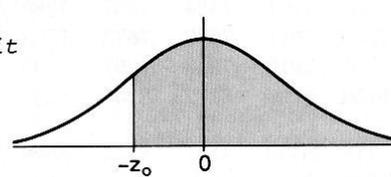
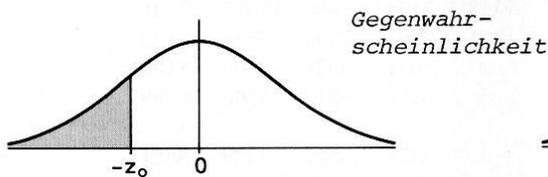
$$\text{A: } P(z \leq z_0) = \Phi(z_0)$$

$$\text{B: } P(z > z_0) = 1 - \Phi(z_0)$$



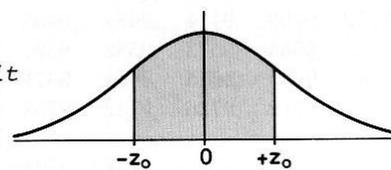
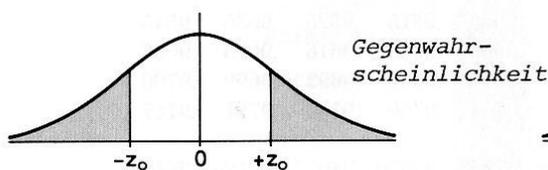
$$\text{C: } P(z \leq -z_0) = \Phi(-z_0) = 1 - \Phi(z_0)$$

$$\text{D: } P(z > -z_0) = \Phi(z_0)$$



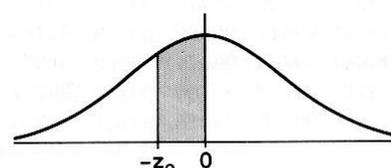
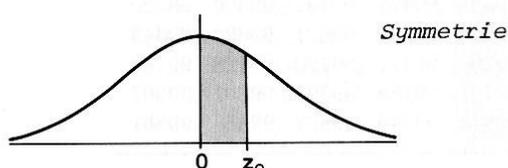
$$\text{E: } P(|z| > z_0) = 2(1 - \Phi(z_0))$$

$$\text{F: } P(-z_0 < z < z_0) = 2\Phi(z_0) - 1$$



$$\text{G: } P(0 < z < z_0) = \Phi(z_0) - 0,5$$

$$\text{H: } P(-z_0 < z < 0) = \Phi(z_0) - 0,5$$



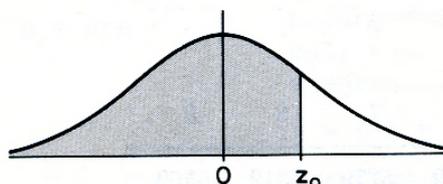
Die nachfolgende Tabelle der Verteilungsfunktion liefert die Wahrscheinlichkeiten $P[z \leq z_0]$ für die standardisierte Normalverteilung:

| z_0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------|---------|--------|---------|--------|---------|--------|---------|--------|--------|--------|
| 0,0 | 0,5000 | ,5040 | ,5080 | ,5120 | ,5160 | ,5199 | ,5239 | ,5279 | ,5319 | ,5359 |
| 0,1 | 0,5398 | ,5438 | ,5478 | ,5517 | ,5557 | ,5596 | ,5636 | ,5675 | ,5714 | ,5753 |
| 0,2 | 0,5793 | ,5832 | ,5871 | ,5910 | ,5948 | ,5987 | ,6026 | ,6064 | ,6103 | ,6141 |
| 0,3 | 0,6179 | ,6217 | ,6255 | ,6293 | ,6331 | ,6368 | ,6404 | ,6443 | ,6480 | ,6517 |
| 0,4 | 0,6554 | ,6591 | ,6628 | ,6664 | ,6700 | ,6736 | ,6772 | ,6808 | ,6844 | ,6879 |
| 0,5 | 0,6915 | ,6950 | ,6985 | ,7019 | ,7054 | ,7088 | ,7123 | ,7157 | ,7190 | ,7224 |
| 0,6 | 0,7257 | ,7291 | ,7324 | ,7357 | ,7389 | ,7422 | ,7454 | ,7486 | ,7517 | ,7549 |
| 0,7 | 0,7580 | ,7611 | ,7642 | ,7673 | ,7703 | ,7734 | ,7764 | ,7794 | ,7823 | ,7852 |
| 0,8 | 0,7881 | ,7910 | ,7939 | ,7967 | ,7995 | ,8023 | ,8051 | ,8078 | ,8106 | ,8133 |
| 0,9 | 0,8159 | ,8186 | ,8212 | ,8238 | ,8264 | ,8289 | ,8315 | ,8340 | ,8365 | ,8389 |
| 1,0 | 0,8413 | ,8438 | ,8461 | ,8485 | ,8508 | ,8531 | ,8554 | ,8577 | ,8599 | ,8621 |
| 1,1 | 0,8643 | ,8665 | ,8686 | ,8708 | ,8729 | ,8749 | ,8770 | ,8790 | ,8810 | ,8830 |
| 1,2 | 0,8849 | ,8869 | ,8888 | ,8907 | ,8925 | ,8944 | ,8962 | ,8980 | ,8997 | ,9015 |
| 1,3 | 0,9032 | ,9049 | ,9066 | ,9082 | ,9099 | ,9115 | ,9131 | ,9147 | ,9162 | ,9177 |
| 1,4 | 0,9192 | ,9207 | ,9222 | ,9236 | ,9251 | ,9265 | ,9279 | ,9292 | ,9306 | ,9319 |
| 1,5 | 0,9332 | ,9345 | ,9357 | ,9370 | ,9382 | ,9394 | ,9406 | ,9418 | ,9429 | ,9441 |
| 1,6 | 0,9452 | ,9463 | ,9474 | ,9484 | ,9495 | ,9505 | ,9515 | ,9525 | ,9535 | ,9545 |
| 1,7 | 0,9554 | ,9564 | ,9573 | ,9582 | ,9591 | ,9599 | ,9608 | ,9616 | ,9625 | ,9633 |
| 1,8 | 0,9641 | ,9649 | ,9656 | ,9664 | ,9671 | ,9678 | ,9686 | ,9693 | ,9699 | ,9706 |
| 1,9 | 0,9713 | ,9719 | ,9726 | ,9732 | ,9738 | ,9744 | ,9750 | ,9756 | ,9761 | ,9767 |
| 2,0 | 0,97725 | ,97778 | ,97831 | ,97882 | ,97932 | ,97982 | ,98030 | ,98077 | ,98124 | ,98169 |
| 2,1 | 0,98214 | ,98257 | ,98300 | ,98341 | ,98382 | ,98422 | ,98461 | ,98500 | ,98537 | ,98574 |
| 2,2 | 0,98610 | ,98645 | ,98679 | ,98713 | ,98745 | ,98778 | ,98809 | ,98840 | ,98870 | ,98899 |
| 2,3 | 0,98928 | ,98956 | ,98983 | ,99010 | ,99036 | ,99061 | ,99086 | ,99111 | ,99134 | ,99158 |
| 2,4 | 0,99180 | ,99202 | ,99224 | ,99245 | ,99266 | ,99286 | ,99305 | ,99324 | ,99343 | ,99361 |
| 2,5 | 0,99379 | ,99396 | ,99413 | ,99430 | ,99446 | ,99461 | ,99477 | ,99492 | ,99506 | ,99520 |
| 2,6 | 0,99534 | ,99547 | ,99560 | ,99573 | ,99585 | ,99598 | ,99609 | ,99621 | ,99632 | ,99643 |
| 2,7 | 0,99653 | ,99664 | ,99674 | ,99683 | ,99693 | ,99702 | ,99711 | ,99720 | ,99728 | ,99736 |
| 2,8 | 0,99744 | ,99752 | ,99760 | ,99767 | ,99774 | ,99781 | ,99788 | ,99795 | ,99801 | ,99807 |
| 2,9 | 0,99813 | ,99819 | ,99825 | ,99831 | ,99836 | ,99841 | ,99846 | ,99851 | ,99856 | ,99861 |
| 3,0 | 0,99865 | 3,1 | 0,99903 | 3,2 | 0,99931 | 3,3 | 0,99952 | | | |

Die Werte in der Tabelle entsprechen der unten abgebildeten Fläche:

$$z_0 \geq 0$$

$$P(z \leq z_0) = \Phi(z_0)$$



Zwei Zufallsexperimente

Zufallsexperiment "Würfeln"

Ereignisraum: $R = \{\text{Augen: } 1, 2, 3, 4, 5, 6\}$

Wahrscheinlichkeiten der Ereignisse e :

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6 = 0.17$$

Anzahl n von 100: 100

Absolute Häufigkeiten $H(e)$

$$H(1) = 20$$

$$H(2) = 15$$

$$H(3) = 13$$

$$H(4) = 11$$

$$H(5) = 19$$

$$H(6) = 22$$

Relative Häufigkeiten $h(e) = H(e) / n$

$$h(1) = 0.20$$

$$h(2) = 0.15$$

$$h(3) = 0.13$$

$$h(4) = 0.11$$

$$h(5) = 0.19$$

$$h(6) = 0.22$$

Zufallsexperiment "Urne"

(Ziehen MIT Zurücklegen)

Ereignisraum: $R = \{\text{Kugeln: 6 rote, 4 blaue, 2 grüne}\}$

Wahrscheinlichkeiten der Ereignisse e :

$$p(\text{rot}) = 6/12 = 1/2 = 0.50$$

$$p(\text{blau}) = 4/12 = 1/3 = 0.33$$

$$p(\text{grün}) = 2/12 = 1/6 = 0.17$$

Anzahl n von 100: 100

Absolute Häufigkeiten $H(e)$

$$H(\text{rot}) = 58$$

$$H(\text{blau}) = 29$$

$$H(\text{grün}) = 13$$

Relative Häufigkeiten $h(e) = H(e) / n$

$$h(\text{rot}) = 0.58$$

$$h(\text{blau}) = 0.29$$

$$h(\text{grün}) = 0.13$$

Wahrscheinlichkeitstheorie

Wahrscheinlichkeitstheorie, Teil 1

Elementarereignisse und Ereignisraum

Ein Zufallsexperiment ist ein beliebig oft wiederholbarer Vorgang, dessen Ausgang ungewiss (zufällig) ist.

Ein solcher Versuchsausgang heißt Elementarereignis $e(i)$. Die Menge aller n Elementarereignisse $\{e(1), e(2), \dots, e(n)\}$ heißt Ereignisraum (Grundraum) G des Zufallsexperiments.

Beispiel "Münze": Beim Aufwerfen einer Münze gibt es zwei Elementarereignisse $e(1) = \text{"Zahl Z"}$ und $e(2) = \text{"Kopf K"}$. Daher ist $G = \{Z, K\}$.

Beispiel "Würfel": Beim Würfeln mit einem Würfel gibt es sechs Elementarereignisse (die Augenzahlen des Würfels). Daher ist $G = \{1, 2, 3, 4, 5, 6\}$.

Beispiel "Urne": In einer Urne befinden sich 6 rote, 4 blaue und 2 grüne Kugeln. Die drei Elementarereignisse bestehen im Herausgreifen einer Kugelfarbe $e(1) = \text{"r"}$, $e(2) = \text{"b"}$, $e(3) = \text{"g"}$. Daher ist $G = \{r, b, g\}$.

Alle drei Beispiele sind einstufige Zufallsexperimente, weil sich nur ein bestimmtes Experiment zufällig wiederholt.

Bei mehrstufigen Zufallsexperimenten werden verschiedene Experimente hintereinander oder gleichzeitig ausgeführt, beispielsweise das gleichzeitige Würfeln mit zwei Würfeln. Dabei gibt es $6 \cdot 6 = 36$ Elementarereignisse und der Ereignisraum ist $\{(1,1), (1,2), \dots, (2,1), (2,2), \dots, \dots, (6,5), (6,6)\}$. Das sind genau 36 mögliche Augenpaare.

Bei zweistufigen Zufallsexperimenten ist der Ereignisraum die Produktmenge der zwei Ereignisräume $G_1 \times G_2$. Das ist die Menge aller geordneten Paare (x,y) , wo x ein Elementarereignis aus G_1 ist und y ein Elementarereignis aus G_2 ist.

Ereignisse und Ereignisfeld

Ereignisse A, B, C, \dots sind Teilmengen des Ereignisraumes G . Beispielsweise ist A das Ereignis "gerade Augenzahl" beim einfachen Würfeln. Dieses Ereignis besteht aus den Elementarereignissen 2, 4 und 6 und entspricht einer Teilmenge von G .

Alle beim Zufallsexperiment möglichen Ereignisse fasst man zum so genannten Ereignisfeld F zusammen. Ereignisfeld F und Ereignisraum G beschreiben somit ein Zufallsexperiment.

Das Ereignisfeld muss eine so genannte Sigma-Algebra bilden. Zu jeder Teilmenge müssen auch die Komplementärmenge und alle Vereinigungsmengen im Ereignisfeld liegen.

$(A * B)$ = Durchschnittsmenge
 $(A + B)$ = Vereinigungsmenge
 $(A \setminus B)$ = Differenzmenge
 $\neg A$ = Komplementärmenge zu A
 $\{ \}$ = leere Menge (unmögliches Ereignis)
 G = der ganze Ereignisraum (sicheres Ereignis)

Wenn der Durchschnitt zweier Ereignisse A und B leer ist, d.h. $(A * B) = \{ \}$, dann heißen die Ereignisse unvereinbar oder disjunkt, d.h. sie treten niemals gemeinsam auf.

Einige Ereignisse beim Würfeln mit einem einfachen Würfel:

$G = \{1, 2, 3, 4, 5, 6\}$ = "sicheres Ereignis"

$A = \{2, 4, 6\}$ = "gerade Augenzahl"

$B = \{3, 6\}$ = "durch 3 teilbare Augenzahl"

$A * B = \{6\}$

$A \setminus B = \{2, 4\}$

$\neg A = \{1, 3, 5\}$

$A + B = \{2, 3, 4, 6\} = \{2, 4\} + \{6\} + \{3\} =$

$(A \setminus B) + (A * B) + (B \setminus A).$

Das letzte Ereignis liefert eine disjunkte Zerlegung von $(A + B)$.

Häufigkeit und Wahrscheinlichkeit

Ein Zufallsexperiment wird n Mal durchgeführt. Die absolute Häufigkeit $H(A)$ gibt an, wie oft Ereignis A eingetreten ist. $h(a) = H(A) / n$ ist die relative Häufigkeit von Ereignis A und $100 * h(a)$ ist die prozentuelle Häufigkeit von Ereignis A . Durch einfaches Abzählen des Zutreffens der entsprechenden Ereignisse können folgende wichtige Eigenschaften für die relative Häufigkeit festgestellt werden:

[E1] $0 \leq h(A) \leq 1$

[E2] $h(G) = 1$ und $h(\{ \}) = 0$

[E3] Wenn $(A * B) = \{ \}$, dann $h(A + B) = h(A) + h(B)$

Wenn nicht $(A * B) = \{ \}$, dann $h(A + B) = h(A) + h(B) - h(A * B)$, weil sonst auf der rechten Seite doppelt gezählt wird!

Erhöht man die Anzahl n der Wiederholungen eines Zufallsexperiments, so stabilisieren sich die relativen Häufigkeiten der Versuchsausgänge. Unter der Wahrscheinlichkeit $p(A)$ eines Ereignisses A (probability) versteht man den Grenzwert der relativen Häufigkeit $h(A)$, wenn die Versuchsanzahl gegen Unendlich strebt: $p(A) = \lim h(A)$ für n gegen Unendlich. Dieser Sachverhalt wird auch das "Gesetz der großen Zahlen" genannt.

Nachdem wir Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit definiert haben, können wir die Eigenschaften [E1], [E2] und [E3] auch für Wahrscheinlichkeiten aussagen. Gegeben ist von einem Zufallsexperiment der Ereignisraum G und das Ereignisfeld F . Dann kann jedem Ereignis A aus F eine reelle Zahl $p(A)$ eindeutig zugeordnet werden. Für diese Zuordnung gelten folgende 3 Grundgesetze (Kolmogorov-Axiome):

$$[K1] \quad 0 \leq p(A) \leq 1 \quad \text{für alle } A \text{ aus } F$$

$$[K2] \quad p(G) = 1 \quad \text{und} \quad p(\{\}) = 0$$

$$[K3] \quad p(A + B) = p(A) + p(B), \quad \text{nur wenn } (A * B) = \{\}$$

[K3] ist das "Additionsgesetz der Wahrscheinlichkeiten". Es bestimmt, wie oft das Ereignis A oder das Ereignis B eintritt, und es gilt für beliebige Vereinigungsmengen $(A + B)$ von F (Sigma-Additivität).

Das ganze System $\{G, F, p\}$ heißt Wahrscheinlichkeitsraum. Aus [K1], [K2] und [K3] werden weitere Gesetze hergeleitet:

$$[W1] \quad p(\neg A) = 1 - p(A)$$

Beweis: $(A * \neg A) = \{\}$ und $(A + \neg A) = G$.

Daraus folgt $1 = p(G) = p(A + \neg A) = p(A) + p(\neg A)$.

Daraus folgt $p(\neg A) = 1 - p(A)$.

$$[W2] \quad p(A \setminus B) = p(A) - p(A * B)$$

Beweis: A wird disjunkt zerlegt in $(A \setminus B)$ und $(A * B)$.

Daraus folgt $p(A) = p(A \setminus B) + p(A * B)$.

Daraus folgt $p(A \setminus B) = p(A) - p(A * B)$.

$$[W3] \quad p(A + B) = p(A) + p(B) - p(A * B)$$

Beweis: $(A + B)$ wird disjunkt zerlegt in

$$(A \setminus B) + (A * B) + (B \setminus A).$$

Daraus folgt

$$p(A + B) = p(A) - p(A * B) + p(A * B) + p(B) - p(A * B).$$

Daraus folgt

$$p(A + B) = p(A) + p(B) - p(A * B).$$

Zufallsexperimente, wo jeder elementare Versuchsausgang gleich wahrscheinlich ist, nennt man Laplace-Experimente. Das sind beispielsweise der Münzwurf oder das Würfeln. Ein Ereignis A ist ja eine Zusammenfassung von Elementarereignissen. Um die Wahrscheinlichkeit von A zu ermitteln, müssen nur alle Elementarereignisse, welche im Ereignis A enthalten sind zusammengezählt (g) und durch die Anzahl (n) der Versuche dividiert werden: $p(A) = g/n$. So gilt die Regel:

$$p(A) = \frac{\text{Anzahl der günstigen Fälle } (g)}{\text{Anzahl der möglichen Fälle } (n)}$$

Beispiel "Urne": In einer Urne befinden sich 6 rote, 4 blaue und 2 grüne Kugeln. Die drei Elementarereignisse bestehen im Herausgreifen einer Kugelfarbe $e(1) = "r"$, $e(2) = "b"$, $e(3) = "g"$. Ereignisraum $G = \{r, b, g\}$, Ereignisfeld $F = \text{Potenzmenge } P(G)$. Diese drei Ereignisse sind nicht gleich wahrscheinlich. Um trotzdem ein Laplace-Experiment zu erhalten, denkt man sich die Kugeln von 1 bis 12 durchnummeriert. Die roten Kugeln tragen dann die Nummern von 1 bis 6, die grünen sind 7 bis 10 und die blauen sind 11 und 12. Durch einfaches Abzählen gilt:

$$\begin{aligned} A &= \text{"rote Kugel ziehen"}, & p(A) &= 6/12 = 1/2. \\ B &= \text{"blaue Kugel ziehen"}, & p(B) &= 4/12 = 1/3. \\ C &= \text{"grüne Kugel ziehen"}, & p(C) &= 2/12 = 1/6. \end{aligned}$$

Diese drei Ereignisse sind paarweise unvereinbar (disjunkt). Also gilt $p(A + B + C) = p(A) + p(B) + p(C) = 1 = p(G)$
 $D = \text{"blau oder grün ziehen"}, \quad p(D) = p(B + C) = p(B) + p(C) = 1/2.$
 $E = \text{"nicht rot ziehen"}, \quad p(E) = p(\neg A) = 6/12 = 1/2.$

Damit ist der erste Teil der Wahrscheinlichkeitstheorie beendet.

Wahrscheinlichkeitstheorie, Teil 2

Bedingte Wahrscheinlichkeit

Ein Zufallsexperiment mit Wahrscheinlichkeitsraum $\{G, F, p\}$ hat n Elementarereignisse. A und B sind beliebige Ereignisse mit den absoluten Häufigkeiten $H(A)$ und $H(B)$. Dann sind die Ereigniswahrscheinlichkeiten die Grenzwerte $p(A) = \lim H(A)/n$ und $p(B) = \lim H(B)/n$ für n gegen Unendlich.

Wir fragen nun nach der Wahrscheinlichkeit von A unter der Voraussetzung, dass B eingetreten ist. Diese Wahrscheinlichkeit heißt bedingte Wahrscheinlichkeit von A und man schreibt dafür $p(A|B)$. Um sie zu ermitteln, zählt man einfach, wie oft A eintritt, wenn B eingetreten ist. Das ist die absolute Häufigkeit der Durchschnittsmenge $(A*B)$, also $H(A*B)$. Relativiert man diese Häufigkeit auf die Häufigkeit von B, dann erhält man die relative Häufigkeit $h(A|B) = H(A*B)/H(B)$ und es gilt:

$$h(A|B) = H(A*B)/H(B) = (H(A*B)/n)/(H(B)/n) = h(A*B)/h(B).$$

$$p(A|B) = \lim h(A*B)/h(B) \text{ für } n \text{ gegen Unendlich.}$$

$$p(A|B) = p(A*B)/p(B).$$

Beispiel "Würfel": $G = \{1,2,3,4,5,6\}$, $p(x) = 1/6$ für x aus G .
 $B = \text{"gerade } x\text{"} = \{2,4,6\}$, $A = \text{"durch 3 teilbare } x\text{"} = \{3,6\}$.
 $p(A|B) = p(A*B)/p(B) = p(\{6\})/p(\{2,4,6\}) = (1/6)/(1/2) = 1/3$.

Man kann nun zeigen, dass die bedingte Wahrscheinlichkeit $p(A|B)$ die drei Axiome von Kolmogorov [K1], [K2], [K3] erfüllt und somit in diesem Sinne eine echte Wahrscheinlichkeit ist, mit der dann auch gerechnet werden kann.

Es sei $\{G, F, p\}$ ein Wahrscheinlichkeitsraum mit A, B aus F , $p(B) > 0$ und $(A*B)$ in F . Dann gilt für $p(A|B) = p(A*B)/p(B)$:

$$[K1] \quad p(A|B) = p(A*B)/p(B) \geq 0.$$

$$[K2] \quad p(G|B) = p(G*B)/p(B) = p(B)/p(B) = 1.$$

$$[K3] \quad A \text{ und } C \text{ sind zwei disjunkte Ereignisse, d.h. } (A*C) = \{ \}.$$

Es gilt $p((A+C)|B) = p((A+C)*B)/p(B) = p((A*B)+(C*B))/p(B)$.
 Weil A und C disjunkt sind, sind es auch $(A*B)$ und $(C*B)$.
 $p((A+C)|B) = p(A*B)/p(B) + p(C*B)/p(B) = p(A/B) + p(C/B)$.

Also gilt das Additionsgesetz: $p((A+C)|B) = p(A/B) + p(C/B)$.

Beispiel: Ein Ehepaar hat zwei Kinder. Wie groß ist die Wahrscheinlichkeit, dass es zwei Jungen sind, wenn wir bereits wissen, dass eines der Kinder ein Junge ist?

$$G = \{JJ, JM, MJ, MM\}, \quad p(MM) = p(JM) = p(MJ) = p(JJ) = 1/4.$$

$$A = \{JJ\} \text{ und } B = \text{"ein Kind ist ein Junge"} = \{JJ, JM, MJ\}.$$

$$p(A|B) = p(A*B)/p(B) = p(\{JJ\})/p(\{JJ, JM, MJ\}). \text{ Also gilt}$$

$$p(A|B) = (1/4)/(3/4) = 1/3.$$

Mit Hilfe der bedingten Wahrscheinlichkeit kann ausgesagt werden, wann zwei Ereignisse voneinander unabhängig sind. Ereignis A ist von Ereignis B offensichtlich genau dann unabhängig, wenn sein Zutreffen das Ereignis B nicht voraussetzt, d.h. $p(A|B) = p(A)$. Wegen $p(A|B) = p(A*B)/p(B)$ gilt dann $p(A*B) = p(A|B) * p(B) = p(A) * p(B)$.

Zwei Ereignisse A und B heißen voneinander unabhängig, wenn für sie gilt: $p(A \cdot B) = p(A) \cdot p(B)$, ("Multiplikationssatz"). Die Wahrscheinlichkeit, dass zwei unabhängige Ereignisse zugleich eintreten ist gleich dem Produkt aus ihren Wahrscheinlichkeiten.

Mehrstufige Zufallsexperimente

Bei mehrstufigen Zufallsexperimenten werden verschiedene Experimente hintereinander oder gleichzeitig ausgeführt, beispielsweise das gleichzeitige Würfeln mit zwei Würfeln. Bei zweistufigen Zufallsexperimenten ist der Ereignisraum die Produktmenge der zwei Ereignisräume $G_1 \times G_2$. Das ist die Menge aller geordneten Paare $(x;y)$, wo x ein Elementarereignis aus G_1 ist und y ein Elementarereignis aus G_2 ist. Dabei ist die Anzahl der Elemente der Produktmenge gleich dem Produkt der Elementanzahlen der beiden Mengen.

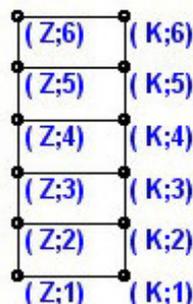
Beispiel "Münze und Würfel"

$$G_1 = \{Z, K\}, p(Z) = p(K) = 1/2.$$

$$G_2 = \{1,2,3,4,5,6\}, p(1) = p(2) = \dots = p(6) = 1/6.$$

$$G_1 \times G_2 = \{(Z;1), (Z;2), \dots, (Z;6), (K;1), (K;2), \dots, (K;6)\}$$

$$p((x;y)) = 1/(2 \cdot 6) = 1/12 \text{ für alle } (x;y) \text{ aus } G_1 \times G_2.$$



Der Ereignisraum des zweistufigen Experiments besteht aus 12 gleich wahrscheinlichen Ereignispaaren $(x;y)$. Sie sind grafisch als Gitterpunkte in der Ebene aufgezählt.

Aus diesen Elementarereignissen setzen sich alle anderen Ereignisse zusammen. Für einige Ereignisse A, B sollen die Wahrscheinlichkeiten ermittelt werden.

$$A = \text{"durch 3 teilbare Augenzahl"} = \{(K;3), (K;6), (Z;3), (Z;6)\}.$$

$$B = \text{"Kopf K geworfen"} = \{(K;1), (K;2), (K;3), (K;4), (K;5), (K;6)\}.$$

$$p(A) = 4/12 = 1/3,$$

$$p(B) = 6/12 = 1/2.$$

$$A \cdot B = \{(K;3), (K;6)\}$$

$$p(A \cdot B) = 2/12 = 1/6.$$

$$p(A|B) = p(A \cdot B)/p(B) = (1/6)/(1/2) = 2/6 = 1/3.$$

$$A+B = \{(K;1), (K;2), (K;3), (K;4), (K;5), (K;6), (Z;3), (Z;6)\}.$$

$$p(A+B) = 8/12 = 2/3.$$

$$p(A+B) = p(A) + p(B) - p(A \cdot B) = 1/3 + 1/2 - 1/6 = 4/6 = 2/3.$$

Im letzten Beispiel wurden gesuchte Wahrscheinlichkeiten einfach durch Abzählen der "günstigen" Elementarereignisse ermittelt. Ist die Anzahl der Elementarereignisse groß, dann ist dieses Verfahren unübersichtlich. Als alternative Methode sollen so genannte "Baumdiagramme" besprochen werden.

Beispiel "Einstufiges Ziehen aus einer Urne"

In einer Urne befinden sich 6 rote Kugeln, 4 blaue Kugeln und 2 grüne Kugeln. Die drei Elementarereignisse bestehen im Herausgreifen einer Kugelfarbe $e(1) = "r"$, $e(2) = "b"$, $e(3) = "g"$. Ereignisraum $G = \{r, b, g\}$, Ereignisfeld $F = \text{Potenzmenge } P(G)$. Diese drei Ereignisse sind nicht gleich wahrscheinlich. Um trotzdem ein Laplace-Experiment zu erhalten, denkt man sich die Kugeln von 1 bis 12 durchnummeriert. Die roten Kugeln tragen dann die Nummern von 1 bis 6, die grünen sind 7 bis 10 und die blauen sind 11 und 12. Durch einfaches Abzählen gilt:

$A = \text{"rote Kugel ziehen"}, \quad p(A) = 6/12 = 1/2.$

$B = \text{"blaue Kugel ziehen"}, \quad p(B) = 4/12 = 1/3.$

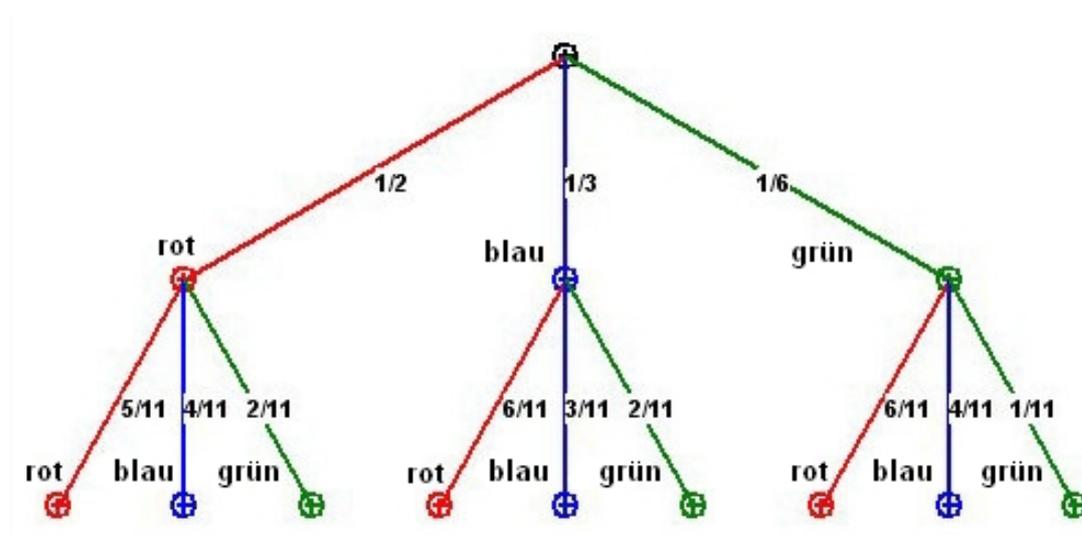
$C = \text{"grüne Kugel ziehen"}, \quad p(C) = 2/12 = 1/6.$

Diese drei Ereignisse sind paarweise unvereinbar (disjunkt), d.h. $p(A + B + C) = p(A) + p(B) + p(C) = 1 = p(G)$.

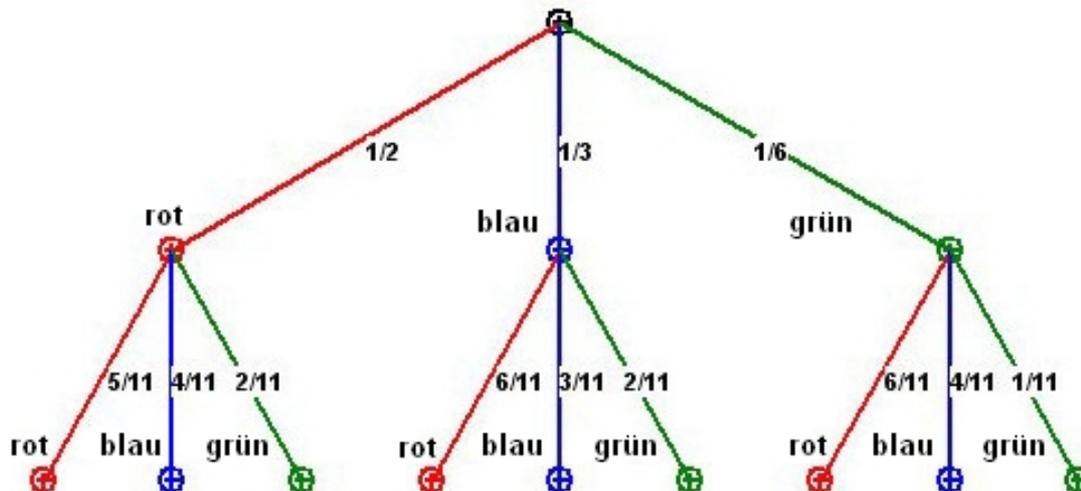
Beispiel "Mehrstufiges Ziehen aus einer Urne"

In einer Urne liegen 6 rote, 4 blaue und 2 grüne Kugeln. Hier sollen hintereinander zwei Kugeln aus der Urne gezogen werden, OHNE dass die erste Kugel zurückgelegt wird.

Das Ereignis A sei "Es wird eine rote (r) und eine blaue (b) Kugel gezogen, egal in welcher Reihenfolge". Zur Ermittlung der Wahrscheinlichkeit von A wird ein Baumdiagramm verwendet.



Ein Baum beginnt im Urknoten oben und verzweigt nach unten. Jede Pfadebene entspricht einem unabhängigen Experiment, welches mit seinen Wahrscheinlichkeiten an den Knoten der Pfade beginnt. Der gesamte Versuch verläuft entlang eines Pfades vom Urknoten oben bis zum letzten Endknoten unten



In einer Urne liegen 6 rote, 4 blaue und 2 grüne Kugeln.

Bei der ersten Ziehung gibt es für jede Kugelfarbe einen Pfad mit den Wahrscheinlichkeiten $p(r) = 1/2$, $p(b) = 1/3$, $p(g) = 1/6$. Weil die gezogene Kugel nicht zurückgelegt wird, sind bei der zweiten Ziehung die Wahrscheinlichkeiten verändert.

Das Ereignis $A =$ "Ziehen einer roten und blauen Kugel, egal in welcher Reihenfolge" zerfällt in zwei disjunkte Ereignisse $B =$ "zuerst rot, dann blau" und $C =$ "zuerst blau, dann rot". $p(A) = p(B) + p(C)$. Die Ereignisse B und C bestehen jeweils aus zwei unabhängigen Ziehungen, deren Wahrscheinlichkeiten multipliziert werden: $p(B) = (1/2) \cdot (4/11) = 4/22 = 2/11$, und $p(C) = (1/3) \cdot (6/11) = 2/11$. $p(A) = 2/11 + 2/11 = 4/11$.

Zusammenfassung der Spielregeln für Baumdiagramme:

- (1) Jene Pfade bestimmen, die zu einem Ereignis A gehören.
- (2) Die Wahrscheinlichkeiten entlang der Pfade multiplizieren.
- (3) Die dabei erhaltenen End-Wahrscheinlichkeiten addieren.

Ein letztes Übungsbeispiel mit der Urne:

$A =$ "die zweite gezogene Kugel ist grün".

$$p(A) = (1/2) \cdot (2/11) + (1/3) \cdot (2/11) + (1/6) \cdot (1/11) = 11/66 = 1/6.$$

Damit ist Teil II der Wahrscheinlichkeitstheorie beendet.

Wahrscheinlichkeitstheorie, Teil 3

In diesem dritten und letzten Teil der Wahrscheinlichkeitstheorie sollen nur die wichtigen Begriffe "Zufallsvariable", "Wahrscheinlichkeitsdichte" und "Verteilungsfunktion" erklärt werden. Für weiterführende Untersuchungen sei auf die einschlägige Fachliteratur verwiesen.

Zufallsvariable

Bei Experimenten mit zufälligem Ausgang interessiert man sich häufig nicht für die zufälligen Ausgänge, sondern für Zahlen oder mathematische Größen, die durch den zufälligen Ausgang des Experiments bestimmt werden. Solche Größen nennt man Zufallsvariable.

Eine reelle Zufallsvariable X ist eine Funktion, die einem Elementarereignis e aus einem Ereignisraum G eine reelle Zahl x zuordnet: $x = X(e)$. Das große X steht für Funktion und das kleine x für ihren Funktionswert (ihre Realisation).

Diese Funktion muss zusätzlich eine messbare Funktion sein, d.h. die Menge aller Elementarereignisse e , bei denen der Wert der zugeordneten Zufallsvariablen $X(e)$ kleiner als eine beliebige reelle Zahl x ist, muss ein Ereignis aus G sein. $X: G \rightarrow \mathbb{R}$. Für alle x aus \mathbb{R} gilt, dass $\{e \text{ mit } X(e) \leq x\}$ in G liegt (Messbarkeits-Axiom).

Ermittelt man zu den Elementarereignissen e eines Ereignisraumes die relativen Häufigkeiten bzw. Wahrscheinlichkeiten, dann erhält man eine Häufigkeits- bzw. Wahrscheinlichkeits-Verteilung. Durch die eindeutige Zuordnung von Zahlen x zu den Elementarereignissen wird die Verteilung der Ereignisse auf eine Verteilung von Zahlen übertragen. Man spricht dann von einer Verteilung der Zufallsvariablen X . Dadurch wird man unabhängig von der tatsächlichen Bedeutung der Ereignisse.

Diskrete und stetige Zufallsvariable.

Betrachten wir das Zufallsexperiment des "Münzwurfes" mit dem Ereignisraum $G = \{\text{Zahl (Z)}, \text{Kopf (K)}\}$. Diesen beiden Elementarereignissen kann eine Zufallsvariable X zwei Werte zuordnen, beispielsweise $X(Z) = 1$ und $X(K) = 0$. Wenn die Zufallsvariable solche abzählbare (d.h. nummerierbare) reelle Zahlenwerte (Realisationen) liefert, dann heißt sie "diskret". Wollen wir hingegen die Körpergröße von Menschen messen, dann entspricht dem Merkmal Körpergröße offensichtlich eine kontinuierliche (stetige) Zufallsvariable.

Errichtet man auf der Zufallsvariablen X noch eine weitere Funktion $Y = g(X)$, dann ist auch Y eine Zufallsvariable.

Wahrscheinlichkeitsdichte und Verteilungsfunktion

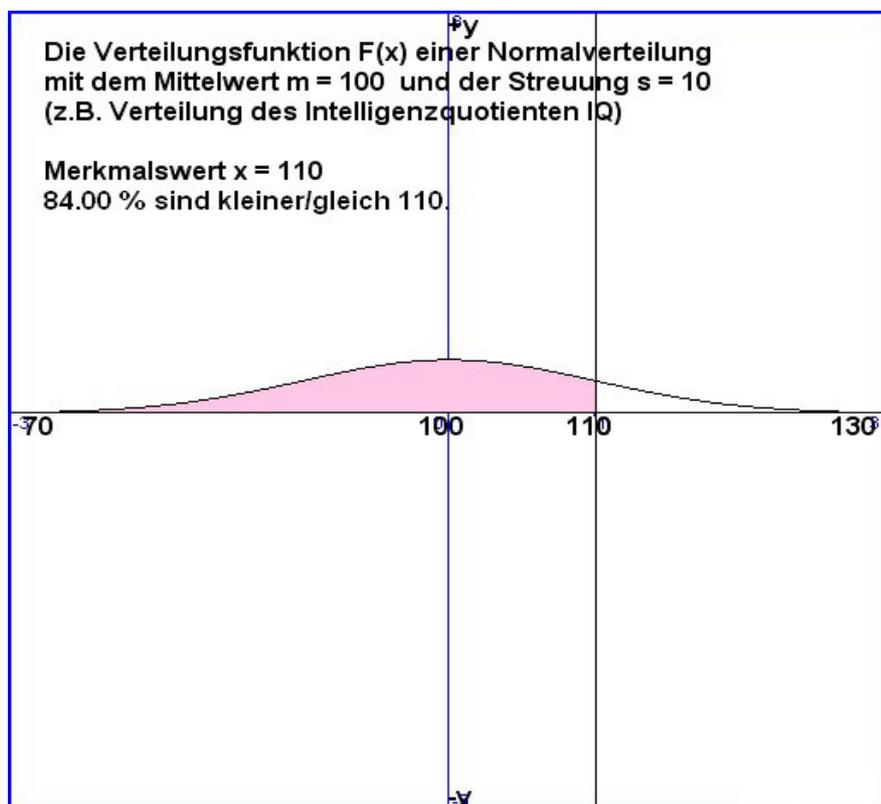
Es sei $p(e)$ die Wahrscheinlichkeit eines Elementarereignisses e aus einem Ereignisraum G . Es sei weiters $X: G \rightarrow \mathbb{R}$ eine Zufallsvariable mit $x = X(e)$, so heißt $p(e)$ Wahrscheinlichkeitsdichte $f(x)$ der Verteilung an der Stelle $X = x$. Oft wird sie auch mit Wahrscheinlichkeitsfunktion $w(x)$ bezeichnet.

Die kumulative Häufigkeit einer Realisation x eines Merkmals gibt an, wie viele Individuen einer Stichprobe kleiner oder gleich dem Wert x sind. Man kann sie einfach durch Abzählen und Summieren der Häufigkeiten ermitteln, und dadurch die Position von x in der Stichprobe beschreiben.

Überträgt man diese Definition auf eine Zufallsvariable X mit ihren Wahrscheinlichkeiten, dann erhält man die so genannte Verteilungsfunktion F mit $F(x) = p(X \leq x)$. Wenn X stetig ist, dann wird die Summierung durch eine Integration ersetzt:

$$\int_{-\infty}^x f(t) dt = F(x) \text{ mit } x \text{ aus } \mathbb{R}.$$

Integriert (summiert) wird über der Wahrscheinlichkeitsdichte $f(t)$, welche die Wahrscheinlichkeit $p(X = t)$ angibt.



Die Verteilungsfunktion $F(x) = p(X \leq x)$ der Zufallsvariablen X ist der zentrale Ausgangspunkt für weitere statistische Untersuchungen. Direkt ersichtlich sind folgende Eigenschaften:

$$0 \leq F(x) \leq 1$$

$F(x)$ ist monoton wachsend

$$p(X > x) = 1 - F(x)$$

$$p(a < x \leq b) = p(X \leq b) - p(X \leq a) = F(b) - F(a)$$

Überträgt man die Definition von Mittelwert und Varianz einer Merkmals-Verteilung von einer endlich großen Stichprobe auf eine reelle stetige Zufallsvariable X , dann erhält man, indem man die Summierung durch eine Integration ersetzt, folgende Formeln:

$$\int_{-\infty}^{+\infty} x \cdot f(x) dx = M, \text{ Mittelwert, } \sum x_i \cdot h(x_i) \text{ mit } i = 1 \text{ bis } n$$

$$\int_{-\infty}^{+\infty} (x-m)^2 \cdot f(x) dx = V, \text{ Varianz, } \sum (x_i - m)^2 \cdot h(x_i)$$

In den obigen Integralformeln ist $f(x)$ die Wahrscheinlichkeitsdichte einer reellen stetigen Zufallsvariablen X . Das entspricht der relativen Häufigkeit $h(x_i)$.

Lineare Skalentransformationen

Es sei X eine reelle stetige Zufallsvariable mit der Wahrscheinlichkeitsdichte $f(x)$. Von der Verteilung kennt man den Mittelwert M_x und die Varianz V_x . Die lineare Transformation $Y = k \cdot X + d$ erzeugt eine neue Zufallsvariable Y . Dabei wird die Dichte nicht verändert, d.h. $f(y) = f(x)$. Aber Mittelwert und Varianz ändern sich. Man kann zeigen, dass gilt: $M_y = k \cdot M_x + d$ und $V_y = k^2 \cdot V_x$.

$$M_y = \int_{-\infty}^{+\infty} y \cdot f(y) dy = \int_{-\infty}^{+\infty} (k \cdot x + d) \cdot f(x) dx$$

$$M_y = k \cdot \int_{-\infty}^{+\infty} x \cdot f(x) dx + d \cdot \int_{-\infty}^{+\infty} f(x) dx = k \cdot M_x + d \cdot 1 = k \cdot M_x + d$$

$$V_y = \int_{-\infty}^{+\infty} (Y - M_y)^2 \cdot f(y) dy = k^2 \cdot \int_{-\infty}^{+\infty} (X - M_x)^2 \cdot f(x) dx = k^2 \cdot V_x$$

Standardisierung

Es sei X eine reelle stetige Zufallsvariable mit der Wahrscheinlichkeitsdichte $f(x)$. Von der Verteilung kennt man Mittelwert M_x und Varianz V_x bzw. Streuung $S_x = \sqrt{V_x}$.

Satz: Die lineare Transformation $Z = (X - M_x)/S_x$ führt den Mittelwert in $M_z = 0$ und die Varianz in $V_z = S_z^2 = 1$ über. Eine solche Skalentransformation heißt "Standardisierung".

Beweis: $Z = (X - M_x)/S_x = (1/S_x) \cdot X - M_x/S_x = k \cdot X + d$
 mit $k = (1/S_x)$ und $d = -M_x/S_x$
 Damit gilt dann:
 $M_z = k \cdot M_x + d = (1/S_x) \cdot M_x - M_x/S_x = 0$
 $V_z = k^2 \cdot V_x = 1/S_x^2 \cdot V_x = 1$

Der Sinn dieser Standardisierung besteht in Folgendem:

Wenn beispielsweise zwei Realisationen $x_1 < x_2$ von einer normal verteilten Zufallsvariablen X vorliegen, dann will man diese zwei Werte oft miteinander vergleichen. Dazu werden sie standardisiert zu z_1 und z_2 . $F(z_2) - F(z_1)$ gibt den Prozentsatz der Population an, der zwischen z_1 und z_2 liegt. $F(z)$ ist die Verteilungsfunktion der standardisierten Normalverteilung, welche Tabellen entnommen werden kann.

Beispiel:

Ein Schüler beendet die Hauptschule mit der Note 2 in Mathematik und mit der Zeit von 14.00 Sekunden im 100-Meter-Lauf. In welcher der beiden Disziplinen ist er besser?

Lösung:

Aus Untersuchungen weiß man:

- (1) Die Mathematiknoten X der Hauptschulabsolventen sind annähernd normalverteilt mit Mittelwert $M_x = 2.80$ und Streuung $S_x = 0.75$
- (2) Die Laufzeiten Y der Hauptschulabsolventen sind auch annähernd normalverteilt mit $M_y = 14.20$ sec und $S_y = 1.50$ sec.

Zuerst werden die Standardwerte bestimmt:

$$Z_x = (2 - 2.80)/0.75 = -1.0667$$

$$Z_y = (14.00 - 14.20)/1.50 = -0.1333$$

Nachschlagen in der Tabelle der standardisierten Normalverteilung ergibt folgende prozentuellen Abweichungen vom Durchschnitt:

$$P_x = F(1.0667) - 0.5 = 0.8570 - 0.5 = 0.3570 = +36\%$$

$$P_y = F(0.1333) - 0.5 = 0.5530 - 0.5 = 0.0530 = +5\%$$

Ergebnis: Der Schüler ist in beiden Disziplinen besser als der Durchschnitt, in Mathematik um 36%, im Laufen nur um 5%.

Damit ist der letzte Teil der Wahrscheinlichkeitstheorie beendet.

23 Übungsaufgaben

(A) Sechs Beispiele [096]

(B) Siebzehn Beispiele [100]

Hinweise zu den Mengenoperationen

A, B, \dots sind Mengen in einer Grundmenge G
bzw. Ereignisse in einem Ereignisraum G

$(A * B)$ = Durchschnittsmenge

$(A + B)$ = Vereinigungsmenge

$(A \setminus B)$ = Differenzmenge

$\neg A$ = Komplementärmenge zu A

$\{ \}$ = leere Menge (unmögliches Ereignis)

G = Grundmenge (Ereignisraum)

$p(A), p(B), \dots$ sind die Wahrscheinlichkeiten
für das Eintreffen der Ereignisse

(A) Sechs Beispiele

Beispiel 01

Aus einem sechsköpfigen Ausschuss werden genau drei Mitglieder in den Vereinsvorstand gewählt. Jede Wahl ist dabei gleichwahrscheinlich.

[a] Gib den Ereignisraum G an, wenn die Personen nach ihrem Alter gereiht mit 1, 2, 3, 4, 5, 6 bezeichnet werden!

[b] Wie groß ist die Wahrscheinlichkeit, dass die älteste Person gewählt wird?

[c] Wie groß ist die Wahrscheinlichkeit, dass die zwei ältesten Personen gewählt werden?

[d] Wie groß ist die Wahrscheinlichkeit, dass mindestens eine der beiden ältesten Personen gewählt wird?

[e] Wie groß ist die Wahrscheinlichkeit, dass die jüngste und die älteste Person gewählt werden?

[f] Wie groß ist die Wahrscheinlichkeit, dass die drei ältesten Personen gewählt werden?

Lösung 01

[a] Der Ereignisraum $G = \{(1,2,3), (1,2,4), (1,2,5), \dots, (4,5,6)\}$ besteht aus $K(5,3) = 5 \cdot 4 \cdot 3 / (1 \cdot 2 \cdot 3) = 20$ Elementarereignissen, welche alle möglichen Wahlergebnisse darstellen. Jedes Ergebnis $e = (x,y,z)$ hat die Wahrscheinlichkeit $p(e) = 1/20$.

[b] Das Ereignis $A_6 = \text{"Person 6 wird gewählt"} = \{(6,x,y)\}$ mit x und y von 1 bis 5, kommt unter den Wahlergebnissen $K(5,2) = 10$ mal vor und es gilt daher $p(A_6) = 10/20 = 1/2$.

[c] Für Ereignis $B = \text{"Die Personen 6 und 5 werden gewählt"} gilt: $B = A_6 * A_5 = \{(6,5,x)\}$ mit x von 1 bis 4. $p(B) = 4/20 = 1/5$.$

[d] Für das Ereignis $C = \text{"Person 6 oder 5 wird gewählt"} gilt: $C = A_6 + A_5$. $p(C) = p(A_6 + A_5) = p(A_6) + p(A_5) - p(A_6 * A_5)$.
 $p(C) = 1/2 + 1/2 - 1/5 = 4/5$.$

[e] Für Ereignis $D = \text{"Die Personen 1 und 6 werden gewählt"} gilt: $D = A_1 * A_6 = \{(1,6,x)\}$ mit x von 2 bis 5. $p(D) = 4/20 = 1/5$.$

[f] Für Ereignis $E = \text{"Die Personen 6,5,4 werden gewählt"} gilt: $E = A_6 * A_5 * A_4 = \{(6,5,4)\}$. $p(E) = 1/20$.$

Beispiel 02

Von der Bevölkerung einer Stadt sind 40% mit einer Krankheit infiziert. Aus repräsentativen Studien weiß man, dass 30% der Infizierten auch erkranken und 5% der Erkrankten sterben. Wie groß ist die Wahrscheinlichkeit, dass ein mit dieser Krankheit infizierter Bewohner der Stadt stirbt?

Lösung 02

Das Zufallsexperiment ergibt für jeden Stadtbewohner x drei Ereignisse: "infiziert", "erkrankt" und "gestorben".
 $A = "x \text{ ist infiziert mit der Krankheit}"$, $p(A) = 40/100 = 4/10$.
 $B = "x \text{ ist erkrankt}"$ und $C = "x \text{ stirbt}"$.
 $D = "x \text{ ist infiziert und erkrankt}" = A * B$.
 Die bedingte Wahrscheinlichkeit $p(B|A) = 30/100 = 3/10$.
 $p(D) = p(A*B) = p(A) * p(B|A) = 4/10 * 3/10 = 12/100$.
 $E = "x \text{ ist infiziert und erkrankt und stirbt}" = A * B * C$.
 Die bedingte Wahrscheinlichkeit $p(C|D) = 5/100$.
 $p(E) = p(A*B*C) = p(D*C) = p(D) * p(C|D) = 12/100 * 5/100$.
 $p(E) = 60/10000 = 6/1000 = 0.006$

Ergebnis: Während der Epidemie sterben 6 von 1000 Bewohner der Stadt an dieser Krankheit.

Beispiel 03

In einem Fundbüro befinden sich genau 8 Regenschirme. Drei Personen melden den Verlust. Der Beamte im Fundbüro gibt willkürlich die Schirme heraus.

[a] Wie groß ist die Wahrscheinlichkeit, dass jede der drei Personen ihren Schirm erhält?

[b] Wie groß ist die Wahrscheinlichkeit, dass mindestens eine der drei Personen ihren Schirm erhält?

Lösung 03

Das Experiment ist die dreimalige Ausgabe eines Schirmes.

[a] Ereignis X ist die Herausgabe des richtigen Schirmes für die Person X . $p(A) = 1/8$. $p(B|A) = 1/7$. $p(A*B) = 1/8 * 1/7 = 1/56$.
 $p(C|A*B) = 1/6$. $p(A*B*C) = 1/56 * 1/6 = 1/8 * 1/7 * 1/6 = 0.003$

[b] $\neg X$ ist das Gegenereignis zu X mit $p(\neg X) = 7/8$.
 $D = "dreimal kein richtiger Schirm" = (\neg A * \neg B * \neg C)$. $p(D) = (7/8)^3$.
 $E = \neg D = "mindestens ein richtiger Schirm"$. $p(E) = 1 - p(D)$.
 $p(E) = 1 - (7/8)^3 = 0.33$

Beispiel 04

Aus vielen Untersuchungen weiß man, dass eine Nachricht bei der Weitergabe durch einen Mittelsmann nur mehr zu 90% richtig ist. Nach wie vielen Berichterstattem ist die Wahrscheinlichkeit der Informationsrichtigkeit auf weniger als 50% gesunken?

Lösung 04

Das Zufallsexperiment ist die Weitergabe einer Nachricht. Zwei Elementarereignisse $e = \text{"richtig"}$ mit $p(e) = 9/10$ und $\neg e = \text{"nicht richtig"}$ mit $p(\neg e) = 1/10$. $G = \{e, \neg e\}$.

Das Experiment wird n Mal wiederholt: $G_n = G \times G \times \dots \times G$. Ereignis $A = \{e, e, \dots, e\}$ mit $p(A) = (9/10)^n = 5/10$. $0.9^n = 0.5$, $n \cdot \lg(0.9) = \lg(0.5)$, $n = \lg(0.5)/\lg(0.9) = 6.58$

Ergebnis: Bei der 7-ten Weitergabe der Nachricht beträgt ihre inhaltliche Richtigkeit nur mehr weniger als 50%.

Beispiel 05

Bei der maschinellen Erzeugung eines Produktes sind 8% der erzeugten Stücke leicht fehlerhaft (zweite Qualität). Wie groß ist die Wahrscheinlichkeit, dass

- [a] bei 8 Stücken nur erste Qualität vorliegt?
- [b] bei 16 Stücken mindestens 1 Stück von zweiter Qualität ist?
- [c] bei 8 Stücken höchstens 2 Stücke von zweiter Qualität sind?

Lösung 05

Es liegt eine Binomialverteilung vor mit $e = \text{"zweite Qualität"}$ und $\neg e = \text{"erste Qualität"}$. $p(e) = 0.08$, $p(\neg e) = q = 0.92$

- [a] Ereignis A mit $p(A) = p(k=0) = K(8,0) \cdot p^0 \cdot q^8 = 0.5132$
- [b] Ereignis B mit $p(B) = p(1 \leq k \leq 16) = 1 - p(k=0)$.
 $p(B) = \sum (K(16,k) \cdot p^k \cdot q^{16-k})$ für k von 1 bis 16
 $p(B) = 1 - K(16,0) \cdot p^0 \cdot q^{16} = 1 - q^{16} = 0.7366$
- [c] Ereignis C mit $p(C) = p(0 \leq k \leq 2) = \sum (K(8,k) \cdot p^k \cdot q^{8-k})$
für k von 0 bis 2. $p(C) = 0.9789$

Beispiel 06

Glühbirnen werden maschinell erzeugt. Die Brenndauer ist normal verteilt. Repräsentative Stichproben ergeben dabei einen Mittelwert von $m = 1200$ Stunden und eine Streuung von $s = 200$ Stunden.

[a] Wie groß ist die Wahrscheinlichkeit, dass die Brenndauer einer Glühbirne zwischen 1300 und 1400 Stunden liegt?

[b] Welche Toleranzgrenzen muss man setzen, wenn nur 5% der Produktion als Ausschuss deklariert werden?

Lösung 06

[a] Normalverteilte Zufallsvariable X . $p(1300 \leq x \leq 1400) = ?$
Standardisierung $Z = (X - m)/s$. $Z(1300) = 100/200 = 0.5$ und
 $Z(1400) = 200/200 = 1$. Die Verteilungsfunktion $F(z) = p(0 \leq z)$
kann für die z -Werte aus Tabellen entnommen oder mittels
numerischer Integration berechnet werden.

$$p(0.5 \leq z \leq 1.0) = F(1.0) - F(0.5) = 0.8413 - 0.6915 = 0.15$$

[b] Ausschuss 5%, d.h. $F(g) = 1 - 0.0250 = 0.9750$. Der Wert g
wird den Tabellen entnommen: $g = 1.96$, $x = s * g + m = 1592$.
Toleranzbereich in Stunden: $[m - 392; m + 392] = [808; 1592]$.

(B) Siebzehn Beispiele

Hinweis: Bei mehrstufigen Zufallsexperimenten ist es vorteilhaft die Lösungen mit Baumdiagrammen zu ermitteln.

Beispiel 1:

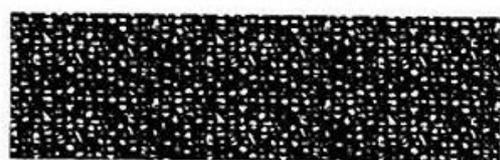
Beiliegende Kopie zeigt Vorder- und Rückseite eines glücklosen Briefloses. Errechnen Sie aus den vorliegenden Angaben folgende Wahrscheinlichkeiten (für den ersten Käufer) und geben Sie die Gewinnchancen jeweils in Prozent an.

Wie groß ist die Wahrscheinlichkeit, beim Kauf eines Loses

- mindestens € 100,- zu gewinnen?
- 100,- € oder 500,- € zu gewinnen?
- 10 000,- € zu gewinnen, wenn man bereits weiß, dass das Los keine Niete ist?
- einen Zusatz-Hauptgewinn zu machen?
- eine Niete zu haben?
- irgendetwas zu gewinnen?

Wie groß ist die Wahrscheinlichkeit, beim Kauf von 2 Losen

- jedes Mal ein Gewinnlos zu haben?
- zwei Hauptgewinne zu machen?



Das wertvollste Brieflos aller Zeiten.

► Das Goldene Brieflos sorgt für Aufregung bei allen AufreißerInnen. Sage und schreibe 8 x 100.000 Euro und weitere 1.312.160 Sofortgewinne warten darauf, gewonnen zu werden. ► Exklusiv nur in Losserie 790. ► Viel Glück!

| So viel können Sie gewinnen. | |
|--------------------------------------|--------------|
| 3 Hauptgewinne | zu € 100.000 |
| 5 Zusatz-Hauptgewinne | zu € 100.000 |
| exklusiv im Goldenen Brieflos | |
| 10 Gewinne | zu € 10.000 |
| 50 Gewinne | zu € 1.000 |
| 100 Gewinne | zu € 500 |
| 2.000 Gewinne | zu € 100 |
| 20.000 Gewinne | zu € 10 |
| 390.000 Gewinne | zu € 2 |
| 900.000 Gewinne | zu € 1 |

So kommen Sie zu Ihrem Gewinn.

Diese Serie besteht aus 6.600.000 Losen. Ihr Gewinn ist im weißen Feld des Gewinnabschnittes aufgedruckt. Gewinne bis einschließlich EUR 1.000,- erhalten Sie in jeder Vertriebsstelle. Gewinne von EUR 1.000,10 bis EUR 80.000,- erhalten Sie in jeder Großgewinnauszahlungsstelle oder werden, ebenso wie Gewinne über EUR 80.000,-, nach Übermittlung einer Gewinnanforderung und des Gewinnloses, direkt vom Kunden-Service-Center der Österreichischen Lotterien G. m. b. H., 1038 Wien, Rennweg 44, Tel.: 0810/100 200 (zum Ortstarif) angewiesen. Die Gewinnanforderung liegt in jeder Lotto-Annahmestelle auf. Alle Gewinne können nur gegen Abgabe des Gewinnabschnittes dem Inhaber ausbezahlt werden und können nur bis zu dem am Gewinnabschnitt aufgedruckten Termin („Auszahlung“) geltend gemacht werden. Bei Verlust des Briefloses oder verspäteter Einlösung erlischt der Gewinnanspruch. Mit Ihrer Spielteilnahme erkennen Sie die Spielbedingungen an. Ihre Vertriebsstelle hält diese gerne für Sie bereit.

Lösungen: (a) 0.0328%, (b) 0.0318%, (c) 0.000762%, (d) 0.0000758%, (e) 80.12%, (f) 19.88%, (g) 3.95%, (h) $1.38 \cdot 10^{-11}$

Beispiel 2:

Eine Forscherin macht Experimente mit Ratten:

Sie lässt 20 Ratten durch ein Labyrinth laufen und stoppt folgende Zeiten (in Sekunden):

364, 322, 280, 302, 383, 420, 205, 250, 375, 445, 310, 256, 340, 470, 317, 220, 427, 264, 406, 355

- a) Teilen Sie die Zeiten in Klassen ein (0 – 1 Minute, 1 – 2 Minuten,), ermitteln Sie für jede Klasse die absolute und relative Häufigkeit und zeichnen Sie dazu ein Histogramm. Berechnen Sie zusätzlich den Mittelwert m und die Streuung s der Daten.

Die Ratte „Rudi“ läuft 10 Tage hintereinander durch das Labyrinth und erreicht dabei folgende Zeiten:

| Tag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Zeit | 364 | 350 | 281 | 239 | 192 | 220 | 208 | 175 | 144 | 150 |

- b) Ermitteln Sie die Gleichung der Regressionsgeraden.
Wie hoch ist die durchschnittliche Verbesserung pro Tag?
- c) Geben Sie den Korrelationskoeffizienten an und interpretieren Sie ihn.
- d) Machen Sie eine Prognose für die Laufzeit am 12. Tag.
- e) Wann wird die Ratte das Labyrinth in 1 Minute bewältigen können?

Die Forscherin untersucht außerdem, ob die Tiere bestimmte Farben bevorzugen. Daher hat sie ein Labyrinth mit einem blauen und einem roten Gang gebaut. Die Wahrscheinlichkeit, dass die Ratte „Adolf“ den roten Gang wählt, liegt bei 45%; die Wahrscheinlichkeit, dass die Ratte „Walter“ diesen Gang wählt, liegt bei 55%. Nun werden beide Ratten losgeschickt.

Wie groß ist die Wahrscheinlichkeit, dass

- f) beide den roten Gang wählen?
- g) nur „Adolf“ den roten Gang wählt?
- h) mindestens einer von beiden den roten Gang wählt?
- i) höchstens einer den roten Gang wählt?
- j) genau einer den roten Gang wählt?

Lösungen:

- (a) $m = 335.55$ Sek, $s = 74.19$ Sek
- (b) Regressionsgerade: $y = -24.0182 \cdot x + 364.4000$
Tägliche Verbesserung = 24 Sekunden.
- (c) Korrelation $r = -0.9402$, d.h. starker negativer Zusammenhang
- (d) $y(12) = 76$ Sekunden
- (e) am 12. Tag
- (f) 24.75%, (g) 20.25%, (j) 75.25%, (i) 75.25%, (j) 50,5%

- 3) Beim einmaligen Würfeln mit einem Würfel sollen die Wahrscheinlichkeiten für folgende Ereignisse berechnet werden:
- a) es wird eine ungerade Augenzahl gewürfelt
 - b) die geworfene Augenzahl ist kleiner als 5
 - c) die geworfenen Augenzahl ist 6

Lösungen: (a) $\frac{1}{2}$, (b) $\frac{2}{3}$, (c) $\frac{1}{6}$

- 4) In einer Urne befinden sich 8 rote, 5 weiße und 3 schwarze Kugeln. Wie groß ist die Wahrscheinlichkeit, beim einmaligen Herausnehmen einer Kugel eine rote oder eine weiße Kugel zu ziehen?

Lösung: $\frac{13}{16}$

- 5) Die Zusammensetzung einer 5A und einer 5B Klasse, aufgeteilt nach Klasse und Geschlecht, ist aus nachstehender Tabelle ersichtlich:

| | Knaben | Mädchen | Σ |
|----------|--------|---------|----------|
| 5 A | 18 | 7 | 25 |
| 5 B | 11 | 9 | 20 |
| Σ | 29 | 16 | 45 |

Aus allen Schülern wird eine Person für ein Referat vor dem Schulforum ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass die ausgewählte Person

- a) aus der 5A stammt?
- b) männlich ist?
- c) ein Knabe aus der 5A ist?

Lösungen: (a) $\frac{25}{45}$, (b) $\frac{29}{45}$, (c) $\frac{18}{29}$

- 6) In einer Urne sind Lose, die mit den Nummern 1-20 versehen sind. Ein Los wird willkürlich gezogen. Wie groß ist die Wahrscheinlichkeit, dass das gezogene Los
- a) eine ungerade Nummer trägt, wenn die Nummer des Loses durch 3 teilbar ist?
 - b) eine durch 5 teilbare Nummer trägt, wenn die Nummer des Loses eine gerade Zahl ist?

Lösungen: (a) $\frac{1}{2}$, (b) $\frac{1}{5}$

- 7) In einem Karton befinden sich 50 Glühlampen, von denen 4 defekt sind. Es werden zwei Lampen herausgenommen,
- a) mit Zurücklegen
 - b) ohne Zurücklegen
- Wie groß ist die Wahrscheinlichkeit, dass man das erste Mal eine defekte und das zweite Mal eine nicht defekte Lampe herausgenommen hat?

Lösungen: (a) $\frac{46}{625}$, (b) $\frac{92}{1225}$

- 8) Wie groß ist die Wahrscheinlichkeit, in zwei Würfeln mit einem Würfel
- a) zuerst 4 und danach eine ungerade Zahl zu werfen?
 - b) zuerst eine gerade und danach eine durch 3 teilbare Zahl zu würfeln?
 - c) bei beiden Würfeln gleiche Augenzahlen zu werfen?
 - d) bei beiden Würfeln nicht die gleichen Augenzahlen zu werfen?

Lösungen: (a) $\frac{1}{12}$, (b) $\frac{1}{6}$, (c) $\frac{1}{6}$, (d) $\frac{5}{6}$

- 9) Wie groß ist die Wahrscheinlichkeit, mit einem Würfel
- viermal hintereinander die Augenzahl 6 zu werfen?
 - viermal hintereinander nicht 5 zu würfeln?
 - bei drei Versuchen mindestens ein Mal 4 zu werfen?

Lösungen: (a) $1/1296$, (b) $625/1296$, (c) $91/216$

- 10) In einer Urne befinden sich 6 weiße und 10 schwarze Kugeln. Wie groß ist die Wahrscheinlichkeit, beim zweimaligen Ziehen
- mit Zurücklegen
 - eine weiße und eine schwarze Kugel zu ziehen?
 - zwei weiße Kugeln zu ziehen?
 - zwei schwarze Kugeln zu ziehen?
 - ohne Zurücklegen
 - zwei schwarze Kugeln zu ziehen?
 - mindestens eine weiße Kugel zu ziehen?

Lösungen: (a1) $5/64$, (a2) $9/64$, (a3) $25/64$, (b1) $3/8$, (b2) $5/8$

- 11) Aus einer Urne mit 5 roten und 8 gelben Kugeln werden gleichzeitig 3 Kugeln gezogen. Wie groß ist die Wahrscheinlichkeit, dass
- alle 3 Kugeln rot sind?
 - alle 3 Kugeln gelb sind?
 - 2 Kugeln rot sind?
 - mindestens eine Kugel gelb ist?
 - höchstens 2 Kugeln gelb sind?

Lösungen: (a) $5/143$, (b) $28/143$, (c) $40/143$, (d) $138/143$, (e) $115/143$

- 12) In einer Urne befinden sich 5 rote, 3 gelbe und 2 schwarze Kugeln.
- Es wird zweimal ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass
 - beim zweiten Zug eine rote Kugel gezogen wird, wenn beim ersten Zug eine gelbe gezogen wurde?
 - beim zweiten Zug eine schwarze gezogen wird, wenn beim ersten Zug keine gelbe gezogen wurde?
 - beim zweiten Zug keine schwarze gezogen wird, wenn beim ersten Zug eine gelbe gezogen wurde?
 - wie a), aber mit Zurücklegen.

Lösungen: (a1) $1/6$, (a2) $2/15$, (a3) $7/30$, (b1) $3/20$, (b2) $7/50$, (b3) $6/25$

- 13) In einer Warensendung befinden sich 25 (gleichartige) Taschenrechner, von denen laut Auskunft der Lieferfirma vier Stück fehlerhaft sind. Es werden zufällig 3 Rechner ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass
- kein Taschenrechner
 - genau ein Taschenrechner
 - mindestens ein Taschenrechner einen Schaden aufweist?

Lösungen: (a) $133/230$, (b) $42/115$, (c) $97/230$

- 14) In einem Warteraum befinden sich 5 Väter mit je 1 Tochter. 2 Personen werden zufällig aufgerufen. Wie groß ist die Wahrscheinlichkeit, dass
- a) es sich um Vater mit Tochter handelt?
 - b) sie unterschiedlichen Geschlechtes sind?

Lösungen: (a) $\frac{1}{9}$, (b) $\frac{5}{9}$

- 15) Die Schützen A und B schießen gleichzeitig auf ein Ziel. Die Trefferwahrscheinlichkeit von A wird mit 80%, die von B mit 45% eingeschätzt. Wie groß ist die Wahrscheinlichkeit, dass das Ziel
- a) nur von A
 - b) von beiden
 - c) von genau einem der beiden Schützen
 - d) überhaupt getroffen wird?

Lösungen: (a) 0,44, (b) 0,36, (c) 0,53, (d) 0,89

- 16) Die Wahrscheinlichkeit eines Verkehrsunfalls innerhalb einer Woche an einer bestimmten Baustelle wird von Experten mit 0,35 eingeschätzt. Die Wahrscheinlichkeit, dass dabei ein Lenker unter 20 Jahren beteiligt ist, beträgt laut Statistik 78%. Wie groß ist die Wahrscheinlichkeit, dass sich im Laufe von einer Woche an dieser Stelle ein Verkehrsunfall ereignet, an dem ein Lenker unter 20 Jahren beteiligt ist?

Lösung: 27.3%

- 17) In einer Schachtel befinden sich 60 Lose, davon sind $\frac{1}{3}$ blau gefärbt und der Rest rot. Von den blau gefärbten sind 5 Lose mit N(Niete), von der roten 12 mit N gekennzeichnet. Es wird ein Los gezogen. Wie groß ist die Wahrscheinlichkeit
- a) für einen Gewinn?
 - b) für einen Gewinn, wenn bekannt ist, dass das gezogene Los ein rotes ist?

Lösungen: (a) $\frac{43}{60}$, (b) $\frac{7}{10}$

ENDE von MATHE 6