

Übersicht Teil 1 - deskriptive Statistik

- $x = (x_1, x_2, \dots, x_n)$ Stichprobe vom Umfang n

Die x_i heißen Stichprobenwerte.

Alle unterschiedlichen Werte, die unter den x_i vorkommen, nennt man Merkmalswerte und bezeichnet sie mit $x_1^*, x_2^*, \dots, x_k^*$.

Wenn alle Werte, die man betrachtet, Zahlen sind, ist es sinnvoll, die x_k^* aufsteigend anzuordnen, etwa Klausurnoten von $x_1^* = 1$ bis $x_5^* = 5$.

Beispiel 01:

Man sieht sich die Farben von fünf vorbeifahrenden Autos an: Der Reihe nach rot, blau, rot, rot, blau.

Das ist dann eine Stichprobe vom Umfang 5 mit $x_1 = x_3 = x_4 =$ „rot“ und $x_2 = x_5 =$ „blau“.

Es gibt zwei verschiedene (Merkmals-)Werte, $x_1^* =$ „rot“ und $x_2^* =$ „blau“.

- Die Anzahl, wie oft ein bestimmtes Merkmal x_k^* (Note, Farbe, Zahl o.ä.) in der Stichprobe vorkommt, nennt man absolute Häufigkeit von x_k^* und schreibt dafür kurz h_k („h“ für Häufigkeit).

Der Anteil des Merkmals x_k^* am Ganzen heißt relative Häufigkeit von x_k^* , wird mit f_k bezeichnet und durch $f_k = \frac{h_k}{n}$ berechnet.

Beispiel 02:

Bei dem Autobeispiel von oben gab es drei rote und zwei blaue Wagen. „Rot“ war x_1^* , „blau“ war x_2^* ; damit ist dann $h_1 = 3$ und $h_2 = 2$.

Wenn drei von fünf Autos rot sind - wieviel Prozent sind das dann ?
 \Rightarrow rel. Häufigkeit

$$f_1 = \frac{3}{5} = 0,6 = 60\% \text{ rote und } f_2 = \frac{2}{5} = 0,4 = 40\% \text{ blaue Autos.}$$

Kurz: Absolute Häufigkeiten sind immer ganze Zahlen; relative Häufigkeiten sind Prozent- oder Kommazahlen mit Werten zwischen 0% und 100% bzw. zwischen 0 und 1.

- Hinter den relativen Häufigkeiten stecken „genau“-Aussagen in Prozent (wieviel Prozent haben (genau) eine 3 geschrieben ?). Wenn man nach „höchstens“ fragt (etwa „wieviel Prozent haben bestanden ?“ = „wieviel Prozent haben höchstens eine 4 geschrieben ?“), braucht man die sogenannte relative Summenhäufigkeit F_i .

Das ist, wenn man die Merkmale x_k^* aufsteigend sortiert hat, einfach die Summe der ersten i relativen Häufigkeiten, also

$$\begin{aligned} F_1 &= f_1, \\ F_2 &= f_1 + f_2 &= F_1 + f_2, \\ F_3 &= f_1 + f_2 + f_3 &= F_2 + f_3, \\ F_4 &= f_1 + f_2 + f_3 + f_4 = F_3 + f_4 \\ &\text{etc.} \end{aligned}$$

Merke: Die Einzelwerte werden mit einem kleinen f , die Summe wird mit einem großen F bezeichnet.

- Die empirische Verteilungsfunktion braucht man im Prinzip nur, um F_i grafisch darstellen zu können.

Der Graph dieser Verteilungsfunktion war eine Treppenfunktion (sah wie Treppenstufen aus); auf der x -Achse wurden die x_k^* und auf der y -Achse die entsprechenden Wahrscheinlichkeiten (alle zwischen 0 und 1) eingetragen.

Links von x_1^* war der y -Wert 0 (und blieb es auch, Strich nach links), rechts vom größten x^* -Wert war der y -Wert 1 (und blieb es auch, Strich nach rechts). Dazwischen sprang man an jeder x^* -Stelle immer eine Stufe hoch: Bei x_1^* auf den Wert F_1 , bei x_2^* auf den Wert F_2 usw.

Am besten sieht man sich dazu mal die Seiten mit den Kindergrößen in der Materialiensammlung dazu an.

- Lageparameter einer Stichprobe: Mittelwert \bar{x} und Median \tilde{x} .

Den Mittelwert (umgangssprachlich: Durchschnitt) berechnet man am besten mit der einfachen Faustformel: Summe durch Anzahl (also Summe der x_i -Werte durch n).

Wenn ich nicht die Stichprobe selber kenne, aber die absoluten Häufigkeiten der verschiedenen Merkmale (etwa 2 Einsen, 8 Zweien, 0 Dreien, 15 Vieren in der Klausur), dann brauche ich nicht mehr jeden Wert einzeln zu addieren, sondern kann einfach absolute Häufigkeit mal Merkmal rechnen (*viel* weniger Arbeit); nicht vergessen: Anschließend durch n teilen !

Beispiel 03:

Die Notensumme kann auch berechnet werden als

$$h_1 * x_1^* + h_2 * x_2^* + h_3 * x_3^* + h_4 * x_4^* = 2 * 1 + 8 * 2 + 0 * 3 + 15 * 4 = 78, \\ \text{insgesamt } 2 + 8 + 0 + 15 = 25 \text{ Klausuren, also ist der Mittelwert der Noten} \\ \text{(oder auch Notendurchschnitt) } \frac{78}{25} = 3,12.$$

Der Median ist der mittlere oder auch Zentralwert der Stichprobe.

Bei einer ungeraden Anzahl von Meßwerten x_i (d.h. n ungerade) ist der Median der Wert, der in der Mitte steht, nachdem ich die x_i aufsteigend sortiert habe; also etwa der 3. Wert bei 5, oder der 5. bei 9 - das müssen aber wegen der Sortierung NICHT ZWINGEND x_3 und x_5 sein ! (Könnte natürlich zufällig so sein, muß aber nicht.)

Bei einer geraden Anzahl von Meßwerten x_i (d.h. n gerade) ist der Median der Mittelwert der beiden mittleren Werte (die vorher wieder aufsteigend sortiert wurden), also etwa $\tilde{x} = \frac{1}{2}$ (dritter Wert + vierter Wert) bei $n = 6$ Werten.

Beispiel 04:

Den Median praktisch bestimmen (vgl. Ende erste Vorlesung).

Möglichkeit 1: Grafisch. Geht gut, wenn man den Graph der Verteilungsfunktion hat; dann sucht man sich auf der y -Achse den Wert 0,5, geht nach rechts, bis man auf den Graph trifft, und da dann senkrecht nach unten; der x -Wert ist dann der Median. Kommt man auf ein gerades Stück, wenn man nach rechts geht, nimmt man den Mittelwert zwischen den beiden x -Werten, die zu diesem geraden Stück gehören (der, wo das Stück anfängt, und der, wo es aufhört).

Möglichkeit 2: Durch Ablesen in einer Tabelle, in der F_k , die rel. Summenhäufigkeit, drinsteht. Der Wert x_k^* , zu dem der erste Wert F_k gehört, der größer ist als 0,5, ist dann der Median; wenn F_k genau den Wert 0,5 annimmt, muß man wieder das arithmetische Mittel zwischen diesem x_k^* und dem nächsten x^* -Wert bilden.

Man sieht sich besser mal MatSam I.3 an.

- Streuungsmaße einer Stichprobe: Maße für die Spannbreite einer Stichprobe, namentlich Varianz s^2 und Standardabweichung s .

Empirische Varianz einer Stichprobe: Ist die mittlere quadratische Abweichung der einzelnen Meßwerte vom Mittelwert, in Formeln:

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

wenn man nur die Varianz der Stichprobe wissen will (theoretischer, sprich: unbrauchbarer Wert)
bzw.

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

wenn man die Varianz des Ganzen (sog. Grundgesamtheit) sucht (praktische Anwendungen brauchen s_{n-1}).

Zu der jeweiligen Varianz gehört die empirische Standardabweichung s , was einfach die Wurzel aus der Varianz ist.

Der Unterschied: Stichprobe von meinetwegen 1000 Autos hinsichtlich des Spritverbrauchs. Bestimme die durchschnittliche Abweichung der Stichprobe $\Rightarrow s_n$ - im Gegensatz zu: Schätze die durchschnittliche Abweichung der Menge aller Autos $\Rightarrow s_{n-1}$. Der Unterschied wird im dritten Teil der Vorlesung nochmal deutlich gemacht.

Beispiel 05:

Varianz s_n^2 und damit auch die Standardabweichung s_n praktisch bestimmen (vgl. Anfang zweite Vorlesung).

$s_n^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$: Umständlich, weil n mal Differenz bilden, dann quadrieren, dann aufaddieren.

Einfacher: $s_n^2(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$. Statt jedesmal die Differenz zu bilden, muß man nur einen einzigen Wert abziehen.

Dazu sollte man folgende Arbeitstabelle benutzen:

x_i	x_i^2	Entscheidend sind die Werte in der letzten Zeile.
\vdots	\vdots	
\vdots	\vdots	
$\sum x_i = n \cdot \bar{x}$	$\sum x_i^2$	

Damit gilt: $s_n^2(x) = \frac{1}{n} \cdot \underbrace{\sum_{i=1}^n x_i^2}_{\text{rechter Wert}} - \left(\frac{n \cdot \bar{x}}{n}\right)^2 \leftarrow \text{linker Wert: erst durch } n, \text{ dann quadrieren}$

und $s_n = \sqrt{s_n^2}$.

WICHTIG: Diese Arbeitstabelle kann man nur für s_n , nicht für s_{n-1} benutzen !

- Datentransformation: Macht man, um krumme und/oder große Werte x_k^* glatt und/oder klein zu kriegen. Aus den alten Werten x macht man transformierte, neue Werte y durch folgende Formel:

$$y = \frac{x - (\text{kleinster } x\text{-Wert})}{\text{größter gemeinsamer Teiler der Abstände zwischen den } x\text{-Werten}}$$

Allerdings hat die mathematisch korrekte Transformation die Form

$$y = ax + b,$$

also muß man (vgl. Beispiel in der Vorlesung 2) etwa $y = \frac{x-800}{100}$ noch umformen zu $y = \frac{x-800}{100} = \underbrace{\frac{1}{100}}_{=a} x + \underbrace{(-8)}_{=b}$.

Wenn der alte Mittelwert \bar{x} bekannt ist, dann kann man den transformierten Mittelwert mit $\bar{y} = a \cdot \bar{x} + b$ berechnen.

Ist umgekehrt \bar{y} bekannt (=aus den neuen, einfacheren y -Werten berechnet worden), kann man den Mittelwert der ursprünglichen Daten mit $\bar{x} = \frac{\bar{y}-b}{a}$ berechnen.

Wenn die alte Varianz $s_n^2(x)$ bekannt ist, dann kann man die transformierte Varianz mit $s_n^2(y) = a^2 \cdot s_n^2(x)$ berechnen.

Ist umgekehrt $s_n^2(y)$ bekannt (=aus den neuen, einfacheren y -Werten berechnet worden), kann man die Varianz der ursprünglichen Daten mit $s_n^2(x) = \frac{1}{a^2} s_n^2(y)$ berechnen.

- Datenzusammensetzung: Braucht man, wenn man zu alten Daten schon Mittelwert und Varianz kennt und jetzt neue Daten mit berücksichtigen will (also gesucht: Mittelwert und Varianz des ganzen Datensatzes, alte und neue Daten).

Superhäßliche Formeln; stehen in der Formelsammlung. Ich empfehle DRINGEND, sich dazu das gemachte Beispiel nochmal anzusehen, dann versteht man hoffentlich, wie man die Formeln anwendet.

- Zweidimensionale Stichproben: Stichproben von Wertepaaren (x_i^*, y_k^*) .

Wenn es etwa drei verschiedene x^* - und fünf verschiedene y^* -Werte gibt, dann gibt es $3 * 5 = 15$ verschiedene Paare. Die absoluten Häufigkeiten dieser Paare schreibt man in eine 3×5 -Tabelle.

In die Spalte rechts davon schreibt man die Summe aller Zeileneinträge der jeweiligen Zeile und in die Zeile darunter die jeweilige Spaltensumme. Alle

diese Einträge heißen Randverteilungen der Stichprobe.

Ganz unten rechts schreibt man die Summe der Einträge der allerletzten (Summen-)Zeile, was auch die Summe der Einträge der allerletzten (Summen-)Spalte sein soll - beides muß n ergeben !

Die ganze Tabelle mit allen Summeneinträgen heißt Kontingenztafel der Stichprobe.

Wichtige Frage: Stehen die beiden Werte in einem Zusammenhang zueinander ? (etwa Größe und Gewicht, Preis und Leistung, Lebenserwartung und Raucher/Nichtraucher).

- Maß für den Zusammenhang der Merkmale x und y :

$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ heißt empirische Kovarianz; kennt man die, kann man den empirischen Korrelationskoeffizienten $r(x, y)$ berechnen durch

$$r(x, y) = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2\right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2\right)}} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} (n \bar{x})(n \bar{y})}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{1}{n} (n \bar{x})^2\right) \left(\sum_{i=1}^n y_i^2 - \frac{1}{n} (n \bar{y})^2\right)}}$$

Man berechnet diesen Koeffizienten mit folgender Arbeitstabelle:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
\vdots	\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots	\vdots
<hr/>				
$\sum = n \cdot \bar{x}$	$\sum = n \cdot \bar{y}$	$\sum x_i^2$	$\sum y_i^2$	$\sum x_i y_i$

Für dieses r gilt: $-1 \leq r \leq 1$; je näher r bei 1 oder -1 liegt, desto besser ist ein linearer Zusammenhang erkennbar; je näher r bei 0 liegt, desto weniger Zusammenhang besteht zwischen den beiden Merkmalen.

Beispiel 06:

Sei $(x, y) = ((1, 6), (2, 7), (3, 8), (4, 9), (5, 10))$ gegeben; wie gut hängen x und y zusammen ?

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	6	1	36	6
2	7	4	49	14
3	8	9	64	24
4	9	16	81	36
5	10	25	100	50
<hr/>				
$5 \cdot \bar{x} = 15$	$5 \cdot \bar{y} = 40$	$\sum x_i^2 = 55$	$\sum y_i^2 = 330$	$\sum x_i y_i = 130$

$$\Rightarrow \bar{x} = 3, \bar{y} = 8,$$

$$r(x, y) = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2\right)}} = \frac{130 - 5 \cdot 3 \cdot 8}{\sqrt{(55 - 5 \cdot 3^2)(330 - 5 \cdot 8^2)}} = \frac{10}{\sqrt{10 \cdot 10}} = 1$$

\Rightarrow perfekter linearer Zusammenhang, die Punkte liegen alle auf einer Geraden (nämlich auf $y = x + 5$)

WICHTIG: Kombination aus Datentransformation und Korrelationskoeffizient, vgl. Beispiel mit Ottomotor aus der Vorlesung! (Kurz: Der Korr.koeff. der transformierten Daten unterscheidet sich höchstens im Vorzeichen von dem der nicht transformierten, ursprünglichen Daten.)

Beispiel 07:

Werte aus Beispiel 6 mit transformiertem y :

x_i	y_i	$\hat{y}_i = y_i - 6$	x_i^2	\hat{y}_i^2	$x_i \hat{y}_i$
1	6	0	1	0	0
2	7	1	4	1	2
3	8	2	9	4	6
4	9	3	16	9	12
5	10	4	25	16	20
$5 \cdot \bar{x} = 15$		$5 \cdot \bar{\hat{y}} = 10$	$\sum x_i^2 = 55$	$\sum \hat{y}_i^2 = 30$	$\sum x_i \hat{y}_i = 40$

$$\Rightarrow \bar{x} = 3, \bar{\hat{y}} = 2,$$

$$r(x, \hat{y}) = \frac{\sum_{i=1}^n x_i \hat{y}_i - n\bar{x}\bar{\hat{y}}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^n \hat{y}_i^2 - n\bar{\hat{y}}^2\right)}} = \frac{40 - 5 \cdot 3 \cdot 2}{\sqrt{(55 - 5 \cdot 3^2)(30 - 5 \cdot 2^2)}} = \frac{10}{\sqrt{10 \cdot 10}} = 1 = +r(x, y)$$

- Regressionsrechnung: Gutes Schätzen von Zwischenwerten.

Beispiel 06:

Wir verpacken Reiseproviant für Wanderer in Plastiktüten. Für eine Person reicht eine Tüte, für 10 Personen brauche ich 4 Tüten, für 20 sogar 6 Tüten. ($\Rightarrow x$: Anzahl Personen, y : Anzahl Tüten, Wertepaare (1; 1), (10; 4), (20; 6))

Ich hab eine Gruppe von 18 Leuten, wieviele Tüten brauche ich ? (Und wer muß sie tragen ?)

Ich hab nur noch 3 Tüten; wer nix zu essen kriegt, will nicht mit. Wieviele Leute können loswandern ?

Trick: Wir basteln uns eine Gerade (oder eine Parabel), die möglichst gut durch die gegebenen Punkte (= Werte) durch und/oder möglichst nahe an ihnen vorbeigeht; dann können wir fehlende Punkte schätzen, indem wir einfach behaupten, sie würden genau auf der Gerade (oder der Parabel) draufliegen.

Erst die Gerade: $\hat{y} = a + bx$

Man berechnet zunächst b , den sogenannten empirischen Regressions-

koeffizienten, durch $b = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n}(n\bar{x})(n\bar{y})}{\sum_{i=1}^n x_i^2 - \frac{1}{n}(n\bar{x})^2}$ mit einer Arbeitstabelle wie

beim emp. Korrelationskoeffizienten (man braucht die y_i^2 -Spalte aber nicht).

Dann berechnet man a durch $a = \bar{y} - b\bar{x}$. Ergebnis: Regressionsgerade fertig.

Im Beispiel 6 gilt:

$$\sum_{i=1}^n x_i y_i = 161, \quad n\bar{x} = 31, \quad n\bar{y} = 11, \quad \sum_{i=1}^n x_i^2 = 501$$

$$\Rightarrow b = \frac{161 - \frac{1}{3} \cdot 31 \cdot 11}{501 - \frac{1}{3}(31)^2} = \frac{47,333}{180,667} = 0,262,$$

$$n\bar{x} = 31, \quad n = 3 \Rightarrow \bar{x} = \frac{31}{3} = 10,333, \quad n\bar{y} = 11, \quad n = 3 \Rightarrow \bar{y} = \frac{11}{3} = 3,667$$

$$\Rightarrow a = \bar{y} - b\bar{x} = 3,667 - 0,262 \cdot 10,333 = 0,960$$

$$\Rightarrow \hat{y} = 0,96 + 0,262x$$

Die Antworten lauten also: $x = 18 \Rightarrow y = 0,96 + 0,262 \cdot 18 = 5,676$, ich brauche also mindestens 5,676, also 6 Tüten für 18 Leute.

$y = 3 \Rightarrow 3 = 0,96 + 0,262x \Rightarrow x = \frac{3-0,96}{0,262} = 7,786$, ich kann also höchstens 7,786 Leute ernähren, also kommen nur 7 Leute mit.

Jetzt die Parabel: $\hat{y} = a + bx + cx^2$

Um a, b, c zu berechnen, muß man folgendes 3×3 - Gleichungssystem lösen, am besten per Gauß:

$$\begin{aligned} a \sum_{i=1}^n x_i^0 + b \sum_{i=1}^n x_i^1 + c \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i^0 \\ a \sum_{i=1}^n x_i^1 + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i^3 &= \sum_{i=1}^n y_i x_i^1 \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^4 &= \sum_{i=1}^n y_i x_i^2 \end{aligned}$$

Wenn man es etwas umformt, sieht es dann so aus:

$$\begin{aligned} an + b \sum_{i=1}^n x_i + c \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i^3 &= \sum_{i=1}^n y_i x_i \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^4 &= \sum_{i=1}^n y_i x_i^2 \end{aligned}$$

Und die zugehörige Arbeitstabelle ist:

x_i	x_i^2	x_i^3	x_i^4	y_i	$y_i x_i$	$y_i x_i^2$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\sum x_i$	$\sum x_i^2$	$\sum x_i^3$	$\sum x_i^4$	$\sum y_i$	$\sum y_i x_i$	$\sum y_i x_i^2$

Im Beispiel 6 gilt:

$$\begin{aligned} n = 3, \quad \sum_{i=1}^n x_i = 31, \quad \sum_{i=1}^n x_i^2 = 501, \quad \sum_{i=1}^n x_i^3 = 9001, \\ \sum_{i=1}^n x_i^4 = 170001, \quad \sum_{i=1}^n y_i = 11, \quad \sum_{i=1}^n y_i x_i = 161, \quad \sum_{i=1}^n y_i x_i^2 = 2801 \end{aligned}$$

$$\begin{aligned} 3a + 31b + 501c &= 11 \\ \Rightarrow 31a + 501b + 9001c &= 161 \\ 501a + 9001b + 170001c &= 2801 \end{aligned}$$

$$\Rightarrow a = \frac{34}{57} = 0,596, \quad b = \frac{39}{95} = 0,411, \quad c = \frac{-2}{285} = -0,007$$

$$\Rightarrow \hat{y} = a + bx + cx^2 = 0,596 + 0,411x - 0,007x^2$$

Nochmal die Antworten zu Bsp. 6, jetzt bei einer Regressionsparabel:

$x = 18 \Rightarrow y = 0,596 + 0,411 \cdot 18 - 0,007 \cdot 18^2 = 5,726$, also bräuchte ich hier mindestens 5,726, also 6 Tüten für 18 Leute;

$$y = 3 \Rightarrow 3 = 0,596 + 0,411x - 0,007x^2 \Rightarrow x^2 + \frac{0,411}{-0,007}x + \frac{0,596-3}{-0,007} = 0$$

$$\Rightarrow x^2 - 58,714x + 343,429 = 0; \text{ (p, q-Formel)}$$

$$\Rightarrow x_{1,2} = 29,357 \pm \sqrt{29,357^2 - 343,429} = 29,357 \pm 22,768$$

$\Rightarrow x_1 = 52,125 \vee x_2 = 6,589$; unsinnigerweise lautet hier das Ergebnis: Entweder kann ich nur 6 oder mindestens 52 Leute mitnehmen, wenn ich drei Tüten habe.

Die zweite Lösung ist natürlich Blödsinn und kommt dadurch zustande, daß man versucht hat, eine Parabel aus nur drei Werten zu erzeugen - generell: Versucht immer, erstmal ne Gerade draus zu machen; das macht viel Sinn, wenn man nur wenige Werte hat oder der Korrelationskoeffizient nahe ± 1 liegt. Hat man viele Werte und ist der Korr.koeff. zusätzlich beinahe 0, muß man wohl oder übel die Parabel aufstellen.