(Pearson-)Korrelationskoeffizienten höherer Grade

Dipl.-Ing. Björnstjerne Zindler, M.Sc.

www.Zenithpoint.de

Erstellt: 13. März 2014 - Letzte Revision: 7. August 2024

Inhaltsverzeichnis

Literatur

[001] Keine für vorliegenden Text.

---

1 Einleitung

Besteht die Notwendigkeit Datenpaare einer Messwertreihe auszuwerten, dann werden mit hoher Sicherheit während dieses Prozesses ebenfalls Regressionen durchgeführt, von linear über quadratisch, kubisch bis zu biquadratisch. Gleichfalls ist unter Umständen eine Elliptische Regression notwendig. Ein sichtbares Ergebnis ist dann der Lineare Korrelationskoeffizient $\rho^{(1)}$ unter der Berechnungsgrundlage:

Einleitung

[001]ff.

$$\rho^{(1)} = a \cdot \frac{\sigma_x}{\sigma_y}$$

Wobei a den Anstieg der Hauptachse der regressierten Ellipse darstellt und σ_x bzw. σ_y die Standardabweichungen der Datenwerte X und Y.

Nutzt man die Grundlage nach Pearson zur Ermittlung des Linearen Korrelationskoeffizienten, dann lässt sich $\rho^{(1)}$ berechnen über:

$$\rho^{(1)} = \frac{Cov\,(X;Y)}{\sigma_x \cdot \sigma_y}$$

Der Wert Cov ist hier die Kovarianz zwischen X und Y.

Beide Gleichungen für $\rho^{(1)}$ zusammen gefasst, zeigen folgenden Zusammenhang:

$$a \cdot \sigma_x^2 = Cov\,(X;Y)$$

$\Rightarrow$

$$a \cdot Var\,(X) = Cov\,(X;Y)$$

$Var\,(X)$ ist die Varianz von X.

$$a = \frac{Cov\,(X;Y)}{Var\,(X)}$$

$\Rightarrow$

$$a = \frac{E\left[(X - E\,(X)) \cdot (Y - E\,(Y))\right]}{E\,(X^2) - E^2\,(X)}$$

Wobei $E\,(\bullet)$ der Erwartungswert ist.

Der Lineare Korrelationskoeffizient $\rho^{(1)}$ ist ein Repräsentant für den Grad des linearen Zusammenhangs zwischen zwei Merkmalen. Er nimmt Werte zwischen -1 und +1 an. Bei ± 1 besteht ein vollständiger linearer Zusammenhang. Gilt $\rho^{(1)} = 0$ liegt keine Abhängigkeit voneinander vor. Dies gilt jedoch nur für lineare Abhängigkeiten, so kann Merkmal 1 und Merkmal 2 durchaus nichtlinear zusammen hängen, obwohl $\rho^{(1)} = 0$. Daher ist der Lineare Korrelationskoeffizient nicht geeignet zur Untersuchung für vollständig stochastische Abhängigkeiten.

Die Nutzung von $\rho^{(1)}$ verlangt einige Voraussetzungen, welche hier als erfüllt gelten.

Der Anstieg a liegt linear vor. Im weiteren Verlauf dieses Arbeitsblattes wird eine Möglichkeit beschrieben um polynomiale Funktionen für die Hauptachse der Elliptischen Regression nutzen zu können und somit auch Korrelationskoeffizienten höherer Grade zu berechnen.

2 Der Pearson-Korrelationskoeffizient $\rho_P^{(1)}$

Lineare Korrelation

Mit den Datenpaaren X und Y ist eine Lineare Regression durchgeführt worden. Damit liegt eine Berechnungsgrundlage folgender Form vor.

$$y = B \cdot x + A$$

$\Rightarrow$

$$Y_i = y_i \qquad X_i = x_i$$

Ebenso wurde ein Linearer Korrelationskoeffizient berechnet. Die Voraussetzungen für diese Berechnung sind gegeben. Aus der Elliptischen Regression ist der Lineare Korrelationskoeffizient vorab schon bekannt.

$$\rho_P^{(1)} = 0,866$$

Die Berechnung von $\rho^{(1)}$ ist einfach durchführbar mit den bekannten elementaren Mitteln

n	x_i	y_i	X_i	Y_i	X_i - X_M	Y_i - Y_M
1	128	100	+128	+100	-567	-348, 75
2	256	250	+256	+250	-439	-198, 75
3	440	510	+440	+510	-255	+61, 25
4	640	160	+640	+160	-55	-288, 75
5	768	400	+768	+400	+73	-48, 75
6	896	520	+896	+520	+201	+71, 25
7	1152	750	+1 152	+750	+457	+301, 25
8	1280	900	+1 280	+900	+585	+451, 25
-	-	-	+5 560	+3 590	0	0

$(X_i - X_M)^2$	$(Y_i - Y_M)^2$	$(X_i - X_M) \cdot (Y_i - Y_M)$
+321 489	+121 627	+197 741
+192 721	+39 502	+87 251
+65 025	+3 752	-15 619
+3 025	+83 377	+15 881
+5 329	+2 377	-3 559
+40 401	+5 077	+14 321
+208 849	+90 752	+137 671
+342 225	+203 627	+263 981
+1 179 064	+550 091	+697 668

- **Mittelwerte**

$$X_M = \frac{5560}{8} = 695 \qquad Y_M = \frac{3590}{8} = 448,75$$

- **Standardabweichungen**

$$\sigma_X = \sqrt{\frac{1179064}{8}} = 383,905 \qquad \sigma_Y = \sqrt{\frac{550091}{8}} = 262,224$$

- **Kovarianz**

$$Cov(X,Y) = \frac{697668}{8} = 87208,5$$

- **Linearer Korrelationskoeffizient**

$$\rho_P^{(1)} = \frac{Cov\,(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{87208,5}{383,905 \cdot 262,224} = 0,866$$

3 Die erweiterten Korrelationskoeffizenten

3.1 Der Lineare Korrelationskoeffizient $\rho^{(1)}$

Lineare Korrelation

Mit den Datenpaaren X und Y ist eine Lineare Regression durchgeführt worden. Damit liegt eine Berechnungsgrundlage folgender Form vor.

$$y = B \cdot x + A$$

$\Rightarrow$

$$Y_i = y_i \qquad X_i = B \cdot x_i + A$$

Die Voraussetzungen für die Berechnung des Korrelationskoeffizienten sind gegeben. Das Ergebnis der Linearen Regression:

$$y = 0,593 \cdot x + 37,508$$

$\Rightarrow$

$$B = +0,593 \qquad A = +37,508$$

n	x_i	y_i	X_i	Y_i	$X_i - X_M$	$Y_i - Y_M$
1	128	100	+113, 386	+100	-336, 118	-348, 75
2	256	250	+189, 265	+250	-260, 239	-198, 75
3	440	510	+298, 340	+510	-151, 164	+61, 25
4	640	160	+416, 900	+160	-32, 604	-288, 75
5	768	400	+492, 778	+400	+43, 274	-48, 75
6	896	520	+568, 657	+520	+119, 153	+71, 25
7	1152	750	+720, 414	+750	+270, 910	+301, 25
8	1280	900	+796, 292	+900	+346, 788	+451, 25
-	-	-	+3 596	+3 590	0	0

$(X_i - X_M)^2$	$(Y_i - Y_M)^2$	$(X_i - X_M) \cdot (Y_i - Y_M)$
+112 975	+121 627	+117 221
+67 724	+39 502	+51 723
+22 851	+3 752	-9 259
+1 063	+83 377	+9 240
+1 873	+2 377	-2 110
+14 197	+5 077	+8 490
+73 392	+90 752	+81 612
+120 262	+203 627	+156 488
+414 337	+550 091	+413 405

- **Mittelwerte**

$$X_M = \frac{3596}{8} = 449,50 \qquad Y_M = \frac{3590}{8} = 448,75$$

- **Standardabweichungen**

$$\sigma_X = \sqrt{\frac{414337}{8}} = 227,579 \qquad \sigma_Y = \sqrt{\frac{550091}{8}} = 262,224$$

- **Kovarianz**

$$Cov(X,Y) = \frac{413405}{8} = 51675,625$$

- **Linearer Korrelationskoeffizient**

$$\rho^{(1)} = \frac{Cov\,(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{51675,625}{227,579 \cdot 262,224} = 0,866$$

Quadratische Korrelation

3.2 Der Quadratische Korrelationskoeffizient $\rho^{(2)}$

Mit den Datenpaaren X und Y ist eine Quadratische Regression durchgeführt worden. Damit liegt eine Berechnungsgrundlage folgender Form vor.

$$y = C \cdot x^2 + B \cdot x + A$$

$\Rightarrow$

$$y - C \cdot x^2 = B \cdot x + A$$

$\Rightarrow$

$$Y_i = y_i - C \cdot x_i \qquad X_i = B \cdot x_i + A$$

Die Voraussetzungen für die Berechnung des Korrelationskoeffizienten sind gegeben. Das Ergebnis der Quadratischen Regression:

$$y = 454,96 \cdot 10^{-6} \cdot x^2 - 0,0486 \cdot x + 195,71$$

$\Rightarrow$

$$C = +454,96 \cdot 10^{-6} \qquad B = -0,0486 \qquad A = +195,71$$

n	x_i	y_i	X_i	Y_i	X_i - X_M	Y_i - Y_M
1	128	100	+189, 489	+92, 546	+27, 556	-69, 394
2	256	250	+183, 268	+220, 184	+21, 335	+58, 244
3	440	510	+174, 326	+421, 920	+12, 393	+259, 980
4	640	160	+164, 606	-26, 352	+2, 673	-188, 292
5	768	400	+158, 385	+131, 653	-3, 548	-30, 287
6	896	520	+152, 164	+154, 751	-9, 769	-7, 189
7	1152	750	+139, 723	+146, 221	-22, 210	-15, 719
8	1280	900	+133, 502	+154, 594	-28, 431	-7, 346
-	-	-	+1 295, 463	+1 295, 517	0	0

$(X_i - X_M)^2$	$(Y_i - Y_M)^2$	$(X_i - X_M) \cdot (Y_i - Y_M)$
+759, 333	+4 815, 527	-1 912, 220
+455, 182	+3 392, 264	+1 242, 636
+153, 586	+67 589, 600	+3 221, 932
+7, 145	+35 453, 877	-503, 305
+12, 588	+917, 302	+107, 458
+95, 433	+51, 682	+70, 229
+493, 284	+247, 087	+349, 119
+808, 322	+53, 964	+208, 854
+2 784, 873	+112 521, 303	+2 784, 703

- **Mittelwerte**

$$X_M = \frac{1295,463}{8} = 161,933 \qquad Y_M = \frac{1295,517}{8} = 161,940$$

- **Standardabweichungen**

$$\sigma_X = \sqrt{\frac{2784,873}{8}} = 18,658 \qquad \sigma_Y = \sqrt{\frac{112521,303}{8}} = 118,597$$

- **Kovarianz**

$$Cov(X,Y) = \frac{2784,703}{8} = 348,088$$

- **Quadratischer Korrelationskoeffizient**

$$\rho^{(2)} = \frac{Cov\,(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{348,088}{18,658 \cdot 118,597} = 0,157$$

3.3 Der Kubische Korrelationskoeffizient $\rho^{(3)}$

Kubische Korrelation

Mit den Datenpaaren X und Y ist eine Kubische Regression durchgeführt worden. Damit liegt eine Berechnungsgrundlage folgender Form vor.

$$y = D \cdot x^3 + C \cdot x^2 + B \cdot x + A$$

$\Rightarrow$

$$y - D \cdot x^3 - C \cdot x^2 = B \cdot x + A$$

$\Rightarrow$

$$Y_i = y_i - D \cdot x_i^3 - C \cdot x_i^2 \qquad X_i = B \cdot x_i + A$$

Die Voraussetzungen für die Berechnung des Korrelationskoeffizienten sind gegeben. Das Ergebnis der Kubischen Regression:

$$y = 1,578 \cdot 10^{-6} \cdot x^3 - 0,00289 \cdot x^2 + 1,901 \cdot x - 70,611$$

$\Rightarrow$

$$D = 1,578 \cdot 10^{-6} \qquad C = -0,00289 \qquad B = 1,901 \qquad A = -70,611$$

n	x_i	y_i	X_i	Y_i	$X_i - X_M$	$Y_i - Y_M$
1	128	100	+173	+144	-1077, 5	-1106
2	256	250	+416	+413	-834, 5	-837
3	440	510	+766	+935	-484, 5	-315
4	640	160	+1 146	+929	-104, 5	-321
5	768	400	+1 389	+1 388	+138, 5	+135
6	896	520	+1 632	+1 703	+381, 5	+453
7	1152	750	+2 119	+2 169	+868, 5	+919
8	1280	900	+2 363	+2 322	+1112, 5	+1072
-	-	-	+10 004	+10 003	0	0

$(X_i - X_M)^2$	$(Y_i - Y_M)^2$	$(X_i - X_M) \cdot (Y_i - Y_M)$
$+1,161 \cdot 10^{+06}$	$+1,223 \cdot 10^{+06}$	$+1,192 \cdot 10^{+06}$
$+0,696 \cdot 10^{+06}$	$+0,701 \cdot 10^{+06}$	$+0,698 \cdot 10^{+06}$
$+0,235 \cdot 10^{+06}$	$+0,099 \cdot 10^{+06}$	$+0,153 \cdot 10^{+06}$
$+0,011 \cdot 10^{+06}$	$+0,103 \cdot 10^{+06}$	$+0,034 \cdot 10^{+06}$
$+0,019 \cdot 10^{+06}$	$+0,018 \cdot 10^{+06}$	$+0,019 \cdot 10^{+06}$
$+0,146 \cdot 10^{+06}$	$+0,205 \cdot 10^{+06}$	$+0,173 \cdot 10^{+06}$
$+0,754 \cdot 10^{+06}$	$+0,845 \cdot 10^{+06}$	$+0,798 \cdot 10^{+06}$
$+1,238 \cdot 10^{+06}$	$+1,149 \cdot 10^{+06}$	$+1,193 \cdot 10^{+06}$
$+4,260 \cdot 10^{+06}$	$+4,343 \cdot 10^{+06}$	$+4,260 \cdot 10^{+06}$

- **Mittelwerte**

$$X_M = \frac{10004}{8} = 1250,5 \qquad Y_M = \frac{10003}{8} = 1250,375$$

- **Standardabweichungen**

$$\sigma_X = \sqrt{\frac{4,26 \cdot 10^{+6}}{8}} = 729,726 \qquad \sigma_Y = \sqrt{\frac{4,343 \cdot 10^{+6}}{8}} = 736,801$$

- **Kovarianz**

$$Cov(X,Y) = \frac{4,260 \cdot 10^{+6}}{8} = 0,533 \cdot 10^{+6}$$

- **Kubischer Korrelationskoeffizient**

$$\rho^{(3)} = \frac{Cov\,(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{0,533 \cdot 10^{+6}}{729,726 \cdot 736,801} = 0,990$$

Biquadratische Korrelation

3.4 Der Biquadratische Korrelationskoeffizient $\rho^{(4)}$

Mit den Datenpaaren X und Y ist eine Biquadratische Regression durchgeführt worden. Damit liegt eine Berechnungsgrundlage folgender Form vor.

$$y = E \cdot x^4 + D \cdot x^3 + C \cdot x^2 + B \cdot x + A$$

$\Rightarrow$

$$y - E \cdot x^4 - D \cdot x^3 - C \cdot x^2 = B \cdot x + A$$

$\Rightarrow$

$$Y_i = y_i - E \cdot x_i^4 - D \cdot x_i^3 - C \cdot x_i^2 \qquad X_i = B \cdot x_i + A$$

Die Voraussetzungen für die Berechnung des Korrelationskoeffizienten sind gegeben. Das Ergebnis der Biquadratischen Regression:

$$y = -5,168 \cdot 10^{-9} \cdot x^4 + 15,930 \cdot 10^{-6} \cdot x^3 - 0,0161 \cdot x^2 + 6,450 \cdot x - 514,281$$

$\Rightarrow$

$E = -5,168 \cdot 10^{-9}$ $D = +15,930 \cdot 10^{-6}$ $C = -0,0161$ $B = +6,450$ $A = -514,281$

n	x_i	y_i	X_i	Y_i	$X_i - X_M$	$Y_i - Y_M$
1	128	100	+311	+312	-3 657	-3 654
2	256	250	+1 137	+1 061	-2 831	-2 905
3	440	510	+2 324	+2 465	-1 644	-1 501
4	640	160	+3 613	+3 449	-357	-517
5	768	400	+4 439	+4 483	+471	+520
6	896	520	+5 265	+5 324	+1 297	+1 358
7	1152	750	+6 916	+6 875	+2 948	+2 909
8	1280	900	+7 741	+7 756	+3 773	+3 790
-	-	-	+31 746	+31 725	0	0

$(X_i - X_M)^2$	$(Y_i - Y_M)^2$	$(X_i - X_M) \cdot (Y_i - Y_M)$
$+13,374 \cdot 10^{+06}$	$+13,352 \cdot 10^{+06}$	$+13,363 \cdot 10^{+06}$
$+8,015 \cdot 10^{+06}$	$+8,439 \cdot 10^{+06}$	$+8,224 \cdot 10^{+06}$
$+2,703 \cdot 10^{+06}$	$+2,253 \cdot 10^{+06}$	$+2,468 \cdot 10^{+06}$
$+0,127 \cdot 10^{+06}$	$+0,267 \cdot 10^{+06}$	$+0,185 \cdot 10^{+06}$
$+0,222 \cdot 10^{+06}$	$+0,270 \cdot 10^{+06}$	$+0,245 \cdot 10^{+06}$
$+1,682 \cdot 10^{+06}$	$+1,844 \cdot 10^{+06}$	$+1,761 \cdot 10^{+06}$
$+8,691 \cdot 10^{+06}$	$+8,462 \cdot 10^{+06}$	$+8,576 \cdot 10^{+06}$
$+14,236 \cdot 10^{+06}$	$+14,364 \cdot 10^{+06}$	$+14,300 \cdot 10^{+06}$
$+49,05 \cdot 10^{+06}$	$+49,251 \cdot 10^{+06}$	$+49,122 \cdot 10^{+06}$

- **Mittelwerte**

$$X_M = \frac{31746}{8} = 3968,25 \qquad Y_M = \frac{31725}{8} = 3965,625$$

- **Standardabweichungen**

$$\sigma_X = \sqrt{\frac{49,05 \cdot 10^{+6}}{8}} = 2476,103 \qquad \sigma_Y = \sqrt{\frac{48,251 \cdot 10^{+6}}{8}} = 2481,229$$

- **Kovarianz**

$$Cov(X,Y) = \frac{49,122 \cdot 10^{+6}}{8} = 6,140 \cdot 10^{+6}$$

- **Biquadratischer Korrelationskoeffizient**

$$\rho^{(4)} = \frac{Cov\,(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{6,140 \cdot 10^{+6}}{2476,103 \cdot 22481,229} = 0,999$$

4 Zusammenfassung und Erwartungen

Zusammenfassung

- Da eine Polynomregression höherer Grade die vorhandenen Daten x_i und y_i immer besser widerspiegeln kann, ist zu erwarten das gilt:

$$\lim_{n \to +\infty} p^{(n)} = \pm 1$$

Es sei denn, dass die vorhandenen Daten schon beim Linearen Korrelationskoeffizenten eine völlige Unabhängigkeit voneinander anzeigen.

$$\rho^{(1)} = 0$$

- Als Kontrolle der Richtigkeit der einzelnen Werte $\rho^{(1)}, \sigma_X$ und σ_Y kann die Berechnungsgrundlage von $\rho^{(1)}$ aus der Elliptischen Regression heran gezogen werden. So gilt dort:

$$\rho^{(1)} = a \cdot \frac{\sigma_X}{\sigma_Y}$$

Damit für den Anstieg a der Hauptachse:

$$a = \rho^{(1)} \cdot \frac{\sigma_Y}{\sigma_X}$$

Zu erwarten ist ein Übereinstimmen von a und $a^{(1)}$ beim Linearen Korrelationskoeffizienten mit dem Anstieg aus der Elliptischen Regression und durch die Transformation der Datenwerte x_i und y_i zu $X_i; Y_i$ bei den Korrelationskoeffizienten höherer Grade in den linearen Raum ein $a^{(n>1)} = \pm 1$.

Die einzelnen Werte:

Lineare Regression:

$$a^{(1)} = 0,866 \cdot \frac{262,224}{383,905} = 0,592$$

Quadratische Regression:

$$a^{(2)} = 0,157 \cdot \frac{118,597}{18,658} = 0,998$$

Kubische Regression:

$$a^{(3)} = 0,990 \cdot \frac{736,801}{729,726} = 1,000$$

Biquadratische Regression:

$$a^{(4)} = 0,999 \cdot \frac{2481,229}{2476,103} = 1,000$$

- Für die **Lineare Exzentrizität** ε_L einer Ellipse ist bekannt:

$$\varepsilon_L^2 = \left| e^2 - f^2 \right|$$

Weiterhin ist gegeben:

$$f^2 \equiv \sigma_X^2 = \frac{\{f^2\}}{n} \qquad e^2 \equiv \sigma_Y^2 = \frac{\{e^2\}}{n}$$

$\Rightarrow$

$$f^2 \cdot n \equiv \sigma_X^2 \cdot n = \{f^2\} \qquad e^2 \cdot n \equiv \sigma_Y^2 \cdot n = \{e^2\}$$

$\Rightarrow$

$$f^2 \equiv \sigma_X^2 \qquad e^2 \equiv \sigma_Y^2$$

$\Rightarrow$

$$\varepsilon_L^2 = \left| \sigma_Y^2 - \sigma_X^2 \right|$$

Für vorhandene Werte gilt:

Grad	ε_L	σ_X^2	σ_Y^2
Linear	280,400	147 383, 000	68 761, 375
Quadratisch	117,120	348,109	14 065, 163
Kubisch	101,858	532 500	542 875
Biquadratisch	316,030	6 131 250	6 031 375

- Für die **numerische Exzentrizität** ε_N gilt analog:

$$\varepsilon_N^2 = \frac{\varepsilon_L^2}{MAX\,(\sigma_X^2; \sigma_Y^2)}$$

$\Rightarrow$

Grad	ε_N	σ_X^2	σ_Y^2
Linear	0,730	**147 383, 000**	68 761, 375
Quadratisch	0,988	348,109	**14 065, 163**
Kubisch	0,138	532 500	**542 875**
Biquadratisch	0,128	**6 131 250**	6 031 375

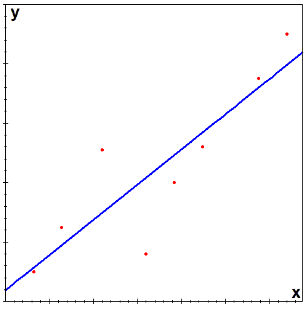
5 Grafische Darstellungen

5.1 Regressionen

Grafisch dargestellt sind in den folgenden Abbildungen die Punktmenge $P\left(x_i; y_i\right)$ der Urliste und den dazugehörigen Regressionsgrafen. **Anhang**

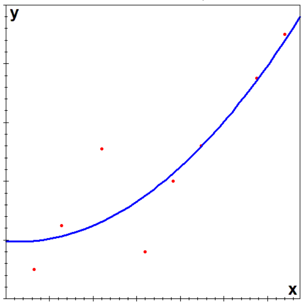
- **Lineare Regression**

$$y_1 = 0,593 \cdot x + 37,508$$



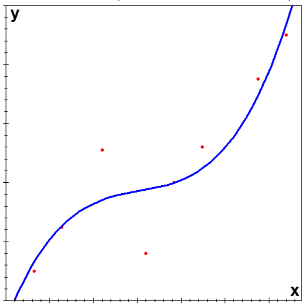
- **Quadratische Regression**

$$y_2 = 454,96 \cdot 10^{-6} \cdot x^2 - 0,0486 \cdot x + 195,71$$



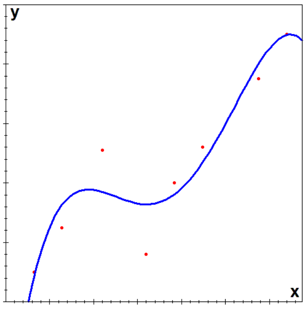
- **Kubische Regression**

$$y_3 = 1,578 \cdot 10^{-6} \cdot x^3 - 0,00289 \cdot x^2 + 1,901 \cdot x - 70,611$$



- **Biquadratische Regression**

$$y_4 = -5,168 \cdot 10^{-9} \cdot x^4 + 15,930 \cdot 10^{-6} \cdot x^3 - 0,0161 \cdot x^2 + 6,450 \cdot x - 514,281$$



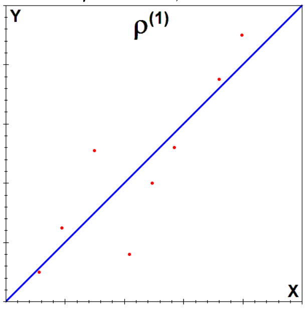
5.2 Korrelationen

Anhang

Grafisch dargestellt sind in den folgenden Abbildungen die Punktmenge $P\left(X_i; Y_i\right)$ und der Funktionsgraf $Y = f\left(X\right) = X$.

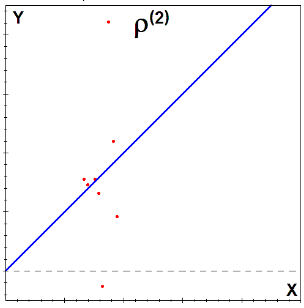
- **Lineare Korrelation**

$$\rho^{(1)} = 0,866$$



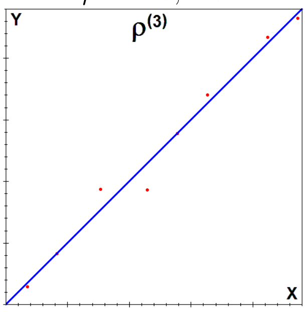
- **Quadratische Korrelation**

$$\rho^{(2)} = 0,157$$



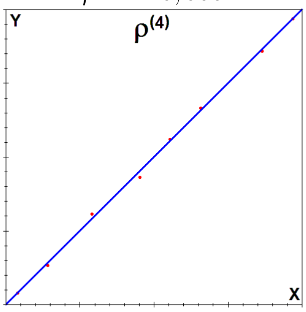
- **Kubische Korrelation**

$$\rho^{(3)} = 0,990$$



- **Biquadratische Korrelation**

$$\rho^{(4)} = 0,999$$



$\LaTeX\,2_\varepsilon$