Veröffentlicht in

Risikomanagement im Unternehmen
Loseblattwerk (Hrsg. Dr. Werner Gleißner)

10. Aktualisierung, 2004

"Rechnen mit Risiken II"

Kapitel 7-3.3, S. 33-48

KOGNOS VERLAG, Augsburg
(www.kognos.de)

Rechnen mit Risiken II

Autor: Marco Wolfrum

Inhalt:

Regressionsanalyse

Zusammenhang zwischen risikobehafteten Größen ermitteln

Risikofaktoren als Ursache für Planabweichungen

Auf ein Unternehmen (bzw. die Zielgrößen eines Unternehmens) wirken in der Regel mehrere risikobehaftete Einflussfaktoren gleichzeitig ein. Teilweise wirken sich diese direkt auf das Unternehmen – genauer gesagt eine Plangröße des Unternehmens – aus (Risikofaktoren) und können so Planabweichungen verursachen. Andere Faktoren wirken sich jedoch nur indirekt auf das Unternehmen aus, indem sie eine Wirkung auf Risikofaktoren entfalten (Indikatoren). Natürlich können sich die Risikofaktoren oder Indikatoren auch gegenseitig beeinflussen.

Aufgabe des Risikomanagementsystems ist es nun, Risikofaktoren und Indikatoren zu identifizieren und die Beziehungen zwischen den identifizierten Variablen zu bestimmen. Die folgenden Ausführungen zeigen mathematische Grundlagen zur Bestimmung der Abhängigkeiten zwischen risikobehafteten Variablen auf.

Beispiel Umsatzwachstumsschwankung

Beispielsweise kann vermutet werden, dass die im ersten Teil des Fachartikels betrachtete Umsatzwachstumsschwankung beeinflusst wird durch eine volkswirtschaftliche Größe wie die Umsatzschwankung innerhalb einer Branche. Ob eine solche Beziehung vorliegt oder nicht, kann mithilfe der Regressionsanalyse ermittelt werden. Sie will die Art der Abhängigkeit bestimmen, d. h. diejenige mathematische Funktion finden, durch die sich die zwischen den Variablen bestehende Abhängigkeit beschreiben lässt.

Einfach- und Mehrfachregression

Unterschieden wird hierbei zwischen der Einfachregression und der Mehrfachregression. Bei der Einfachregression wird der Zusammenhang zwischen zwei Variablen untersucht, während die Mehrfachregression den Zusammenhang zwischen drei oder mehr Variablen klären will. Bei der Einfachregression wird also angenommen, dass eine Variable Y (abhängige oder zu erklärende Variable, Regressand) von einer zweiten Variable X (unabhängige oder erklärende Variable, Regressor) abhängt. Im Beispiel ist somit das Umsatzwachstum eines Unternehmens der Regressand Y und das Branchenumsatzwachstum der Regressor X.

Umsatzwachstum

Wie im ersten Teil dieses Artikels werden folgende Werte für das Umsatzwachstum des Unternehmens für die anschließenden Berechnungen zugrunde gelegt.

Umsatzwachstum (UW) in der Vergangenheit										
Periode	1	2	3	4	5	6	7	8	9	10
UW		2,00%	2,94%	2,86%	2,31%	2,71%	2,20%	2,59%	1,68%	3,31%
Periode	11	12	13	14	15	16	17	18	19	20
UW	4,00%	0,77%	3,05%	2,22%	2,54%	2,47%	3,45%	1,33%	1,97%	3,23%
Periode	21	22	23	24	25	26	27	28	29	30
UW	2,19%	2,75%	2,38%	2,91%	1,69%	2,78%	2,70%	2,63%	2,56%	2,50%

Branchenumsatzwachstum

Die Werte des Branchenumsatzwachstums seien nun wie folgt gegeben.

Branchenumsatzwachstum (BUW) in der Vergangenheit										
Periode	1	2	3	4	5	6	7	8	9	10
BUW		2,45%	3,10%	2,76%	2,57%	2,48%	2,17%	2,26%	1,97%	2,78%
Periode	11	12	13	14	15	16	17	18	19	20
BUW	3,14%	2,68%	2,89%	2,57%	2,48%	2,39%	2,22%	1,96%	2,08%	2,34%
Periode	21	22	23	24	25	26	27	28	29	30
BUW	2,56%	2,79%	2,89%	2,78%	2,68%	2,59%	2,79%	2,65%	2,70%	2,65%

Streuungsdiagramm

Grafische Darstellung

Theoretische Vermutungen über die Beziehung zwischen (ökonomischen) Variablen beinhalten meist keine Information über die mathematische Form dieser Beziehung. Grundlage bilden dann empirische Beobachtungen. Häufig behilft man sich hier zunächst mit einem grafischen Verfahren, beispielsweise einem Streuungsdiagramm (Punktwolke). Hierbei wird im Falle einer Einfachregression jedes der beobachteten Wertepaare (x_i, y_i) in einem xy-Koordinatensystem dargestellt.

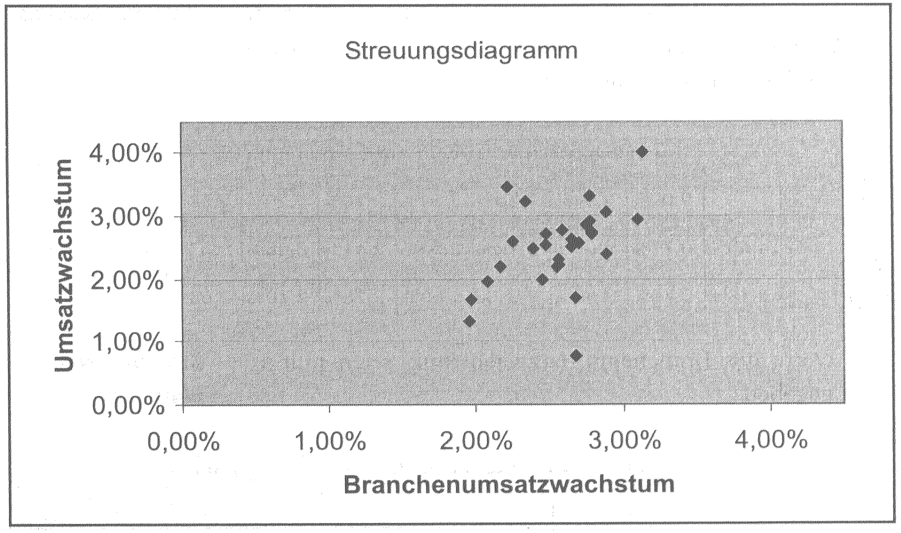

Abb. 1: Streuungsdiagramm

Regressionsgerade

Lineare Regressionsfunktion als ökonomisch bedeutsamster Zusammenhang

Lineare Einfachregression

Zumindest näherungsweise lassen sich viele Zusammenhänge zwischen ökonomischen Variablen durch lineare Funktionen charakterisieren. Häufig ist es auch möglich, nicht lineare Zusammenhänge durch eine geeignete Transformation zu linearisieren. Daher spielen die linearen Zusammenhänge bei der Regressionsanalyse eine bedeutende Rolle. In den folgenden Ausführungen wird sich somit auf die lineare Einfachregression beschränkt.

Regressionskoeffizienten

Regressionsfunktion beschreibt Zusammenhänge

Erster Schritt ist also häufig die Erstellung eines Streuungsdiagramms. Die darin gegebenenfalls erkennbare Tendenz soll durch eine lineare Regressionsfunktion (Regressionsgerade) beschrieben werden.

$$\hat{y} = b_1 + b_2 x$$
$$\hat{y}_i = b_1 + b_2 x_i$$

Dabei werden b_1 und b_2 als Regressionskoeffizienten bezeichnet.

Obige Punktwolke lässt einen linearen Zusammenhang zwischen den beiden Variablen vermuten.

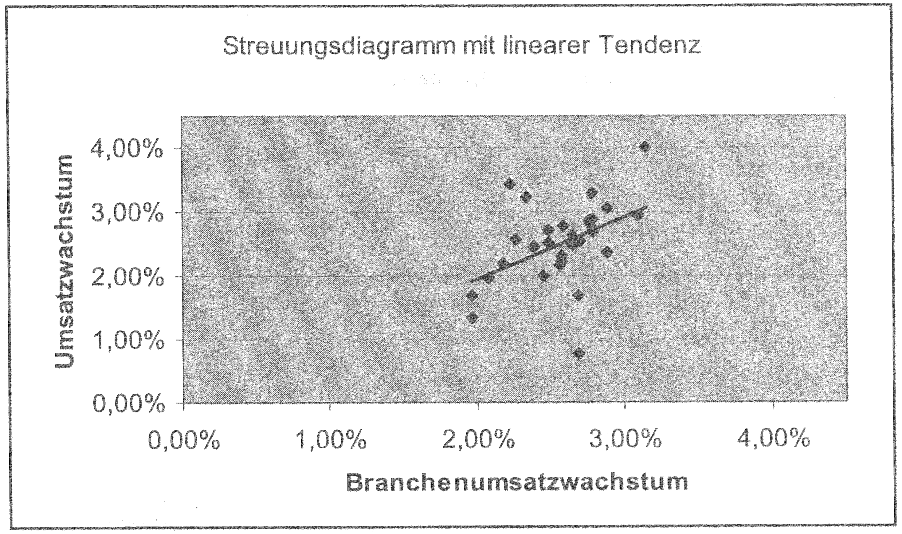

Abb. 2: Streuungsdiagramm mit linearer Tendenz

Tendenz des Zusammenhangs

Eine Regressionsfunktion beschreibt die Tendenz des Zusammenhangs oder den durchschnittlichen Zusammenhang. Sie gibt an, um wie viel Einheiten sich der Wert der Variable Y durchschnittlich ändert, wenn man den Wert von X um eine Einheit ändert.

Residuen

Normalerweise liegen nicht alle (oder oft sogar gar keine) Beobachtungswerte auf der Regressionsfunktion. Die Abweichungen zwischen den tatsächlich beobachteten y-Werten und den geschätzten Werten werden als Residuen bezeichnet.

$$e_i = y_i - \hat{y}_i$$

Gesucht ist unter den theoretisch unendlich vielen Regressionsgeraden nun diejenige, die den Zusammenhang zwischen den Variablen „möglichst gut“ beschreibt.

Methode der kleinsten Quadrate

Verfahren zur eindeutigen Festlegung einer Regressionsgeraden

Kleinste-Quadrat-Methode

Ein Verfahren zur eindeutigen Festlegung einer Regressionsgeraden ist die Methode der kleinsten Quadrate. Bei der Methode der kleinsten Quadrate wird die Summe der Abweichungsquadrate minimiert.

Gesucht sind im Falle einer Regressionsgeraden diejenigen Regressionskoeffizienten, bei denen die folgende Funktion ein Minimum annimmt.

$$SAQ(b_1, b_2) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)^2$$

Die gesuchten Regressionskoeffizienten werden bestimmt durch:

$$b_1 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b_2 = \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Durch Umformung erhält man:

$$b_1 = \frac{\frac{\sum_{i=1}^{n} x_i^2}{n} \bar{y} - \bar{x} \frac{\sum_{i=1}^{n} x_i y_i}{n}}{s_x^2}$$

$$b_2 = \frac{\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \overline{xy}}{s_x^2} = \frac{\mathrm{cov}(X, Y)}{s_x^2}$$

Fortsetzung Fallbeispiel

Für unser Beispiel errechnen sich die Koeffizienten wie folgt:

n = 29

$\bar{y} = 2{,}56\ \%$

$\bar{x} = 2{,}51\ \%$

$$\sum_{i=1}^{n} x_i^2 = 1{,}94\ \%$$

$$\sum_{i=1}^{n} x_i y_i = 1{,}89\ \%$$

$s_x^2 = 0{,}0041\ \%$

Und damit ist

$b_1 = 2{,}04\ \%$

$b_2 = 21{,}00\ \%$

Die Regressionsgerade hat also die Form

$\hat{y} = b_1 + b_2 x = 2{,}04\ \% + 21{,}00\ \%\ x$

Steigt x also um 1 % an, wächst y damit um 0,21 %.

Die geschätzten y-Werte $\hat{y}_i$ – also die y-Werte, die sich errechnen, wenn die x-Werte in die Regressionsgerade eingegeben werden – ergeben sich damit wie folgt:

geschätzte Werte für Branchenumsatzwachstum (BUW g)										
Periode	1	2	3	4	5	6	7	8	9	10
BUW g		2,46%	2,66%	2,64%	2,52%	2,61%	2,50%	2,58%	2,39%	2,73%
Periode	11	12	13	14	15	16	17	18	19	20
BUW g	2,88%	2,20%	2,68%	2,50%	2,57%	2,56%	2,76%	2,32%	2,45%	2,72%
Periode	21	22	23	24	25	26	27	28	29	30
BUW g	2,50%	2,62%	2,54%	2,65%	2,39%	2,62%	2,61%	2,59%	2,58%	2,56%

Die Residuen – also die Abweichungen der geschätzten von den tatsächlich beobachteten Werten für das Branchenumsatzwachstum – ergeben sich zu

Residuen des Branchenumsatzwachstums (BUW r)										
Periode	1	2	3	4	5	6	7	8	9	10
BUW r		-0,01%	0,44%	0,12%	0,05%	-0,13%	-0,33%	-0,32%	-0,42%	0,05%
Periode	11	12	13	14	15	16	17	18	19	20
BUW r	0,26%	0,48%	0,21%	0,07%	-0,09%	-0,17%	-0,54%	-0,36%	-0,37%	-0,38%
Periode	21	22	23	24	25	26	27	28	29	30
BUW r	0,06%	0,17%	0,35%	0,13%	0,29%	-0,03%	0,18%	0,06%	0,12%	0,09%

Summe der Residuen

Wie leicht nachzurechnen ist, ist die Summe der Residuen gleich null. Die Regressionsgerade ist also eine fehlerausgleichende Gerade.

Messung der Stärke des statistischen Zusammenhangs

Streuung anderer Variablen

In der Regel wirken auf eine Variable verschiedene andere Größen (Risikofaktoren) ein. Die Streuung der Variable hängt also von der Streuung anderer Variablen ab. Mit der Regressionsgeraden hat man den Einfluss eines Faktors untersucht. Es soll nun bestimmt werden, inwieweit durch die untersuchte Regressionsgerade die Abweichung bei der abhängigen Variablen durch die unabhängige Variable erklärt ist.

Die Regressionsfunktion an sich erlaubt keine Aussage über die Stärke oder Ausgeprägtheit des statistischen Zusammenhangs. Hierzu kann die Zerlegung der Abweichungsquadratsumme benutzt werden.

Zerlegung der Abweichungsquadratsumme

Abweichungsanalyse

Mit Hilfe der Zerlegung der Abweichungsquadratsumme soll also die Variation (Streuung, Variabilität) der abhängigen Variable aus der Variation der unabhängigen Variable erklärt werden. Als zu erklärende Variation an einer Stelle x_i wird die einfache Abweichung des beobachteten Wertes vom Mittelwert der abhängigen Variablen angesehen.

$$\boxed{y_i - \bar{y}}$$

Als Erklärungswert für y_i wird der geschätzte Wert geliefert. Dementsprechend kann als die durch die Regressionsfunktion erklärte Variation die Abweichung von geschätztem Wert und Mittelwert angesehen werden.

$$\boxed{\hat{y}_i - \bar{y}}$$

Als nicht erklärte Abweichung kann somit das Residuum herangezogen werden.

$$\boxed{y_i - \hat{y}_i}$$

Somit erhält man

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

$$\begin{pmatrix}\text{zu erklärende}\\ \text{Abweichung}\end{pmatrix} = \begin{pmatrix}\text{nicht erklärte}\\ \text{Abweichung}\end{pmatrix} + \begin{pmatrix}\text{erklärte}\\ \text{Abweichung}\end{pmatrix}$$

Auch hier werden für die Messung der Variation die Abweichungsquadrate betrachtet.

$$SQT = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$SQR = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

$$SQE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Abweichungsquadratsummen

Für unser Beispiel ergeben sich die Abweichungsquadratsummen zu

$SQT = 0{,}0255\ \%$

$SQR = 0{,}0203\ \%$

$SQE = 0{,}0052\ \%$

In unserem Beispiel kann die Regressionsfunktion also nur einen geringen Teil der Variation der abhängigen Variable erklären. Es muss noch nach anderen Einflussfaktoren gesucht werden, um ein gut funktionierendes Frühaufklärungssystem aufzubauen.

SQT, SQR, SQE

Die zu erklärende Gesamtabweichungsquadratsumme SQT lässt sich also in die nicht erklärte Abweichungsquadratsumme SQR und in die durch die Regressionsfunktion erklärte Abweichungsquadratsumme SQE zerlegen. Diese einfache Zerlegung gilt für die lineare Kleinste-Quadrate-Regressionsfunktion. Bei anderen linearen Regressionsfunktionen gilt dies nicht unbedingt.

Lineares einfaches Bestimmtheitsmaß

Als Maß für die durch die lineare Regressionsfunktion gelieferte Erklärung der Variation der abhängigen Variablen aus der Variation der unabhängigen Variablen kann folgende Größe herangezogen werden.

$$r^2 = \frac{SQE}{SQT} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Für unser Beispiel ergibt sich r^2 somit zu 0,21.

r^2 wird als lineares einfaches Bestimmtheitsmaß bezeichnet. r^2 stellt den Anteil der durch die Regressionsfunktion erklärten Abweichungsquadratsumme an der zu erklärenden Gesamtabweichungsquadratsumme dar.

Bestimmtheitsmaß

In unserem Beispiel werden also lediglich 21 % der gesamten Abweichung durch die Regressionsfunktion erklärt werden. Auch mit dem Branchenwachstum als erklärendem Risikofaktor bleiben damit 79 % (= 100 % – 21 %) der

Schwankungen des eigenen Umsatzes unseres Beispiel-Unternehmens nicht prognostizierbar. Entsprechend hoch bleibt das Risiko (die Standardabweichung der – nicht erklärbaren – Residuen $(y_1 - \hat{y}_1)$ der Regression). Je höher das Bestimmtheitsmaß einer Regressionsschätzung, desto deutlicher kann durch diese die Zukunft z. B. einer Zielvariablen (wie der Umsatz) vorhergesagt werden und desto stärker reduziert ein solches Prognosesystem das Risiko. Damit führen Regressionsmodelle mit hohem Bestimmtheitsmaß zu hoher Planungssicherheit.

Das Bestimmtheitsmaß r^2 kann nur Werte zwischen den beiden Grenzfällen $r^2 = 0$ (kein Erklärungsbeitrag) und $r^2 = 1$ (vollständige Erklärung) annehmen.

Linearer Einfachkorrelationskoeffizient

Statt dem linearen einfachen Bestimmtheitsmaß wird teilweise der lineare Einfachkorrelationskoeffizient r verwandt. Dieser wird ermittelt als Quadratwurzel aus r^2.

Damit schwankt der absolute Betrag des linearen Einfachkorrelationskoeffizienten ebenfalls zwischen 0 (fehlende Erklärung, fehlende Korrelation) und 1 (vollständige Erklärung, vollständige Korrelation). Definitionsgemäß gibt man dem linearen Einfachkorrelationskoeffizienten das Vorzeichen der Steigung der Regressionsgeraden.

In unserem Beispiel ergibt sich r zu 0,45.

Verwendung einer Kennzahl

Das lineare Bestimmtheitsmaß und damit der lineare Einfachkorrelationskoeffizient lässt sich nur bei einer gegebenen Regressionsfunktion bestimmen. Diese ist in vielen Fällen aber nicht bekannt. Somit gilt es eine Kennzahl zu finden, die den Zusammenhang zwischen zwei Variablen auch ohne Kenntnis einer Regressionsfunktion misst. Ein

solches Maß ist der so genannte (Pearson'sche) Korrelationskoeffizient.

(Pearson'scher) Korrelationskoeffizient

Korrelation als Maß für statistischen Zusammenhang

Der (Pearson'sche) Korrelationskoeffizient ist ein Maß für die Ausgeprägtheit des linearen Zusammenhangs zwischen zwei Zufallsvariablen X und Y und wird mittels der so genannten Kovarianz berechnet. Die Kenntnis einer Regressionsgeraden wird also nicht vorausgesetzt.

Die Kovarianz $\mathrm{cov}(X, Y)$ gibt die Stärke des Zusammenhangs zwischen den (Zufalls-)Variablen X und Y wieder. Sie wird berechnet mittels

$$\boxed{\mathrm{cov}(X, Y) = \mathrm{E}((X - E(X))*(Y - E(Y))) = E(XY) - E(X)E(Y)}$$

Im Falle vorliegender Datenreihen – wie in unserem Beispiel – kann dies berechnet werden durch

$$\boxed{\mathrm{cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}}$$

Der Wert der Kovarianz hängt somit von den Einheiten ab, in denen X und Y gemessen werden. Um dies zu vermeiden, betrachtet man meist die in Vielfachen der jeweiligen Standardabweichung gemessenen Abweichungen von den jeweiligen Erwartungswerten. Man bezeichnet die resultierende Größe r_{xy} als Korrelationskoeffizient.

$$\boxed{r_{XY} = \frac{\mathrm{cov}(X, Y)}{s_X s_Y}}$$ [7]

Bei unserem Beispiel berechnet sich r_{XY} zu 0,45.

Es lässt sich zeigen, dass der Korrelationskoeffizient zwischen –1 und 1 liegt. Der absolute Betrag schwankt somit zwischen 0 (fehlende Korrelation) und 1 (vollständige Korrelation).

Unabhängigkeit von X und Y

Bei Unabhängigkeit von X und Y gilt

$\mathrm{cov}(X, Y) = 0$

und damit auch

$$\boxed{r_{XY} = 0}$$

Die umgekehrte Schlussfolgerung ist aber nicht zulässig!

Der Korrelationskoeffizient sollte nur berechnet werden, wenn der untersuchte Zusammenhang sinnvoll durch eine lineare Regressionsfunktion beschrieben werden kann. Diese Regressionsgerade muss aber nicht bekannt sein.

Wie berechnet werden kann (und auch am Zahlenbeispiel zu sehen ist), stimmen für eine lineare Regressionsfunktion der lineare Einfachkorrelationskoeffizient und der (Pearson'sche) Korrelationskoeffizient überein.

[7] Liegen Standardabweichungen aus theoretischen Verteilungen vor, wird s_X bzw. s_Y ersetzt durch σ_X bzw. σ_Y.

Die Formel lautet somit $r_{XY} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$

Scheinkorrelation

Nachteile statistischer Verfahren

Mit Hilfe statistischer Verfahren wird empirisches Datenmaterial verarbeitet. Daraus kann gegebenenfalls ein statistischer Zusammenhang errechnet werden. Dies heißt jedoch nicht, dass auch ein kausaler Zusammenhang (also eine Ursache-Wirkungs-Beziehung) besteht. Es besteht also die Gefahr, sachlich nicht in Zusammenhang stehende Variablen miteinander in Beziehung zu bringen. Man spricht in diesem Fall von Scheinkorrelationen. Es sollte demnach normalerweise zunächst geprüft werden, ob zwischen Variablen ein sachlicher Zusammenhang hergeleitet werden kann oder nicht. Nur wenn dies der Fall ist, ergeben errechenbare statistische Zusammenhänge auch tatsächlich Sinn.