Erstellung von multiplen linearen Regressiongleichungen

Mit einfachen Regressionsgleichungen können wir in einem definierten Teilmarkt wertbeeinflussende Abhängigkeiten des Mietwerts von bestimmten Merkmalen erkennen und grafisch aufbereitet für den Gutachtenleser anschaulich sichtbar machen. Das Ziel einfacher Regressionsgleichungen ist die Ableitung von Anpassungsparametern aus Marktdaten, zur Berücksichtigung der individuellen Situation des Bewertungsobjekts. Wenn mehrere wertbeeinflussende Merkmale identifizierbar sind, bedient man sich der multiplen linearen Regression. Das hat den Vorteil, dass Überlappungen aus den Einzelbetrachtungen ausgeschlossen werden können. Der Nachteil ist die fehlende Visualisierbarkeit dieser mehrdimensionalen Berechnung mit den Möglichkeiten von Excel.

Ausgangsbasis ist eine aufbereitete und validierte Mietpreissammlung.

In dieser Anleitung verwende ich eine Auswertung der marktüblichen Nettokaltmieten für Wohnungen in Köln-Longerich, einem Stadtteil im Stadtbezirk Nippes im Jahr 2013. Der Bewertungsstichtag soll der 11.01.2014 sein. Das zu bewertende Objekt hat 68 m² Wohnfläche, in mittlerer Wohnlage aus dem Ursprungsbaujahr 1987. Die Ausgangsdatenbank verfügt über 237 Datensätze

WHG-Mieten Köln -Nippes, OT Longerich						© Rolf Schubert + Kollegen, 50769 Köln, www.ImmoWert-RS.com						± 1,5 Intq.	± 2 Sigma
Jahr 2013	BJ	REN	MBJ	WFL	EnKW	Angeb.€/M	Faktor	NKM €/Mo	NKM €/m²	NK €/Mo	WARM €/m	Int.Quart	Konfid.
Anzahl	148	70	169	237	24	237	237	237	237	222	237	223	215
Mindestwert	1902	1975	1960	31 m²	77	226,00 €	0,994	224,64 €	4,66 €/m²	25,00 €	6,06 €/m²	6,06 €/m²	6,54 €/m²
25% Quartil	1961	2010	1983	55 m²	99	456,00 €	0,994	453,26 €	7,54 €/m²	95,00 €	9,40 €/m²	7,65 €/m²	7,68 €/m²
Mittelwert	1975	2010	1989	71 m²	134	595,62 €	0,994	592,07 €	8,45 €/m²	136,32 €	10,27 €/m²	8,43 €/m²	8,42 €/m²
Median	1970	2012	1985	65 m²	133	550,00 €	0,994	546,70 €	8,50 €/m²	120,00 €	10,28 €/m²	8,50 €/m²	8,50 €/m²
75% Quartil	1996	2013	2004	82 m²	158	685,00 €	0,994	680,89 €	9,14 €/m²	169,50 €	11,17 €/m²	9,11 €/m²	9,03 €/m²
Maximum	2010	2013	2013	190 m²	204	1.258,00 €	1,000	1.250,45 €	14,79 €/m²	400,00 €	17,03 €/m²	10,65 €/m²	10,40 €/m²
Streuungsmaße													
Median/Arithm.Mittel	-0,3%	0,1%	-0,2%	-8,7%	-0,6%	-7,7%	0,0%	-7,7%	0,6%	-12,0%	0,1%	0,8%	1,0%
Interquartilabstand	35	3	21	27 m²	59	229,00 €	0,00000	227,63 €	1,60 €/m²	74,50 €	1,77 €/m²	1,46 €/m²	1,36 €/m²
Standardabweichung	21	6	16	26 m²	38	208,52 €	0,00055	207,27 €	1,26 €/m²	63,36 €	1,48 €/m²	0,99 €/m²	0,92 €/m²
Variationskoeffizient	1,0%	0,3%	0,8%	36,3%	28,7%	35,0%	0,1%	35,0%	14,9%	46,5%	14,4%	11,7%	10,9%
Zusammenhangsmaße		Datum	MBJ	WFL				Schwellenwerte unten	6,05 €/m²	Interquartil		6,46 €/m²	Stabw.
Bestimmtheitsmaß (SP:NKM €/m²)		0,0177	0,0699	0,0842				Schwellenwerte oben	10,85 €/m²	Interquartil		10,41 €/m²	Stabw.
Korrelationskoeffizient (SP:NKM €/m²)		0,1332	0,2645	-0,2902									
Kovarianz(s) (SP:NKM €/m²)		17,7859	5,2318	-9,4452									
Ausreißerprüfung bei 5 Stabw.													
unterer Grenzwert				-58 m²	-58	-446,97 €	0,99130	-444,27 €	2,14 €/m²	-180,50 €	2,86 €/m²		
oberer Grenzwert				200 m²	325	1.638,20 €	0,99680	1.628,42 €	14,75 €/m²	453,14 €	17,67 €/m²		

Die in der statistischen Auswertung berechneten Zusammenhangsmaße belegen eine Kovarianz zwischen Datum und relativen Nettokaltmieten von 17,7869, zwischen dem mietrechtlichen Baujahr und den relativen Nettokaltmieten von 5,2318 und für die Wohnfläche zu den relativen Nettokaltmieten von – 9,4452.

Über eine Ausreißerprüfung bei ± 5 Standardabweichungen wurden alle Datensätze, die die oberen Grenzwerte bei einzelnen Merkmalen übersteigen, ausgeschieden. Damit sind in dieser Analyse nur aus statistischer Sicht valide Datensätze enthalten. Der Mittelwert der relativen Nettokaltmieten liegt bei 8,45 €/m².

Die von Excel bereitgestellte Funktion lautet RGP().

Zur Vorbereitung kopieren wir aus der Datentabelle die Daten der Spalten, die die wertbeeinflussenden Merkmale enthalten.

Die Daten der Spalte „A" mit dem Datum, die Daten der Spalte „O" mit den Daten des mietrechtlichen Baujahrs (MBJ), die Daten der Spalte „P" mit den Wohnflächen und die Daten der

Spalte „U“ mit den Nettomieten in €/Mo. Achtung: Beim Kopieren kann es sein, dass Berechnungsformeln, die sich auf nicht mitkopierte Zellen der Ausgangstabelle beziehen, in der neuen Tabelle zu Fehlern führen. Deshalb muss man beim Kopieren die Unterfunktion <Inhalte einfügen> verwenden. Wir verwenden hier nicht die relativen Nettokaltmieten, weil diese bereits den Einfluss der Wohnfläche beinhalten. Das würde eine verbotene Autokorrelation verursachen.

	M	N	O	P	Q	R	S
1	x1	x2	x3				
2	Datum	MBJ	WFL	NKM €/Mo			
3	04.01.2013	1994	56 m²	407,54 €			
4	07.01.2013	1984	80 m²	596,40 €			
5	07.01.2013	1968	123 m²	916,47 €			
6	07.01.2013	1985	75 m²	626,22 €			
7	07.01.2013		48 m²	347,90 €			
8	07.01.2013		92 m²	680,89 €			
9	14.01.2013	2013	63 m²	546,70 €			
10	14.01.2013	1998	60 m²	546,70 €			
11	14.01.2013		55 m²	483,08 €			
12	21.01.2013	1960	65 m²	463,20 €			
13	21.01.2013	1960	65 m²	463,20 €			
14	21.01.2013	1967	77 m²	574,53 €			
15	21.01.2013	1993	48 m²	407,54 €			
16	28.01.2013		80 m²	616,28 €			
17	04.02.2013		76 m²	566,58 €			
18	04.02.2013		76 m²	566,58 €			
19	04.02.2013		31 m²	347,90 €			
20	11.02.2013	1963	99 m²	864,78 €			
227	09.12.2013	2007	81 m²	765,38 €			
228	10.12.2013	2013	63 m²	560,62 €			
229	16.12.2013	1983	60 m²	497,00 €			
230	16.12.2013	2000	55 m²	492,03 €			
231	16.12.2013	1961	60 m²	360,82 €			
232	17.12.2013	1965	55 m²	559,62 €			
233	17.12.2013	1994	49 m²	357,84 €			
234	23.12.2013	1983	58 m²	447,30 €			
235	30.12.2013	1983	55 m²	452,27 €			
236	30.12.2013	1983	67 m²	564,59 €			
237	30.12.2013	1965	60 m²	626,22 €			
238	30.12.2013	1983	52 m²	542,72 €			
239	30.12.2013	1983	93 m²	790,23 €			
240							
241							
242							
243							
244							
245							
246							
247							
248							
249							
250							
251							
252							
253							
254							

MietWHG2013(o.NB).STBZ 5 | MietWHG2013(o.NB).Longerich | Regressionsanalysen

	A	B	C	D
4	21.01.2013	1960	65 m²	463,20 €
5	22.04.2013	1960	64 m²	556,64 €
6	27.05.2013	1960	54 m²	526,82 €
7	07.06.2013	1960	103 m²	690,83 €
8	17.06.2013	1960	65 m²	463,20 €
9	05.08.2013	1960	66 m²	508,93 €
10	19.08.2013	1960	65 m²	497,00 €
11	22.04.2013	1961	63 m²	488,05 €
12	27.05.2013	1961	65 m²	447,30 €
13	16.12.2013	1961	60 m²	360,82 €
14	09.12.2013	1962	76 m²	675,92 €
15	11.02.2013	1963	99 m²	864,78 €
16	09.12.2013	1963	74 m²	656,04 €
17	27.05.2013	1964	73 m²	656,04 €
18	22.04.2013	1965	57 m²	385,67 €
19	10.06.2013	1965	80 m²	596,40 €
20	23.09.2013	1965	91 m²	709,72 €
21	17.12.2013	1965	55 m²	559,62 €
22	30.12.2013	1965	60 m²	626,22 €
23	21.01.2013	1967	77 m²	574,53 €
24	01.04.2013	1967	69 m²	517,87 €
25	01.04.2013	1967	77 m²	577,51 €
26	15.04.2013	1967	76 m²	565,59 €
160	14.01.2013	2013	63 m²	546,70 €
161	18.02.2013	2013	82 m²	795,20 €
162	11.03.2013	2013	51 m²	453,26 €
163	26.08.2013	2013	54 m²	509,92 €
164	23.09.2013	2013	80 m²	652,06 €
165	25.09.2013	2013	67 m²	546,70 €
166	18.11.2013	2013	146 m²	1.043,70 €
167	06.12.2013	2013	82 m²	566,58 €
168	06.12.2013	2013	64 m²	566,58 €
169	06.12.2013	2013	65 m²	582,48 €
170	09.12.2013	2013	56 m²	526,82 €
171	10.12.2013	2013	63 m²	560,62 €
172				

Regressionsanalysen | M

Bereit

Man erkennt, dass ein mietrechtliches Baujahr (MBJ) nicht bei sämtlichen Datensätzen existiert. Die Anwendung der RGP()-Funktion setzt voraus, dass bei allen Datensätzen auch alle Merkmale positiv besetzt sind. Deshalb muss man diese 4 Datenspalten nach Spalte MBJ sortieren und alle Datensätze, die unvollständig sind, löschen. Nach dieser Maßnahme enthält das Beispiel noch 169 Datensätze (von Zeile 3 bis 171).

Beschreibung der Parameter der Funktion RGP()

Zunächst beschreibe ich die RGP()-Funktion für 3 wertbeeinflussende Merkmale
Die Regressionsformel RGP berechnet die Koeffizienten m_n und b der Gleichung y = mx+b

RGP(Y;X;KONST;STATS)

Die **Variable Y**. Y-Werte, das sind im Beispielfall die abhängigen Werte der Marktdaten, d.h. die Nettokaltmieten in € pro Monat. Die Daten stehen in Spalte D von Zeile 3 bis 171. Es sind 169 Datensätze, d.h.

n = 169.

Der gesuchte Wert dieser Funktion ist die Nettokaltmiete in €/Monat für die vom Durchschnitt der Datenbasis abweichende Wohnfläche, bezogen auf einen vom Durchschnitt abweichenden Stichtag und mit einem vom Durchschnitt abweichenden mietrechtlichen Baujahr. Also ebenfalls ein Y-Wert.

Die **Variable X**. X-Werte umfassen den Bereich der unabhängigen Variablen. Bei 3 Merkmalen in unserem Beispiel die Daten in den in den Spalten A3 bis C171.

Bei 3 Merkmalen [x] lautet die Ausgleichs-Gleichung der Geraden

$$y = m_1x_1+m_2x_2+m_3x_3+b$$

Die **Konstante** der RGP()-Funktion kennt die Wahrheitswerte WAHR und FALSCH.
Sofern hier FALSCH eingegeben wird, werden die Koeffizienten so angepasst, dass der Wert b auf 0 gesetzt wird.

Die Ausgleichs-Gleichung lautet dann

$$y = m_1x_1+m_2x_2+m_3x_3$$

STATS ist ebenfalls ein Wahrheitswert, der bei FALSCH nur die Koeffizienten ohne die statistischen Prüfwerte liefert.

Beschreibung der Rückgabewerte der RGP() Funktion:

Spalte Zeile	A	B	C	D
1	m_3	m_2	m_1	b
2	se_3	se_2	se_1	se_b
3	r^2	se_y		
4	F	d_f		
5	ss_{reg}	ss_{resid}		

m_n Mit m werden die Koeffizienten bezeichnet, die innerhalb der Funktionsgleichung mit dem x-Wert multipliziert werden. In diesem Beispiel steht x_1 für das Datum, m_1 für den Koeffizienten des Datums, x_2 steht für das MBJ, das mietrechtliche Baujahr, m_2 ist der Koeffizient des MBJ, x_3 ist die WFL, m_3 der Koeffizient der WFL.

b Die berechnete Konstante der Regressionsgleichung.

se_n Mit se bezeichnet Excel die statistischen Standardfehler der Koeffizienten
n ist dabei der Zähler (1 … 3 … n)

se_b Der Standardfehler der Konstanten b

r^2 Das Bestimmtheitsmaß der Regressionsgleichung. Das Bestimmtheitsmaß vergleicht die erwarteten mit den tatsächlichen y-Werten und kann Werte von 0 bis 1 annehmen. Besitzt es den Wert 1, besteht für die Stichprobe eine vollkommene Korrelation: Ein erwarteter y-Wert und der entsprechende tatsächliche y-Wert unterscheiden sich nicht. Im anderen Extremfall, wenn das Bestimmtheitsmaß 0 ist, ist die Regressionsgerade nicht dazu geeignet, einen y-Wert vorherzusagen.

se_y Der Standardfehler des Schätzwerts y (d.h. des Prognosewerts)

F Der berechnete F-Wert. Anhand der F-Statistik kann man entscheiden, ob die zwischen der abhängigen und der unabhängigen Variablen beobachtete Beziehung zufällig ist oder nicht.

d_f Der Freiheitsgrad. Mit diesem Freiheitsgrad können Sie den jeweiligen kritischen F-Wert (Quantil F) aus einer entsprechenden statistischen Tabelle entnehmen oder berechnen. Vergleichen Sie den jeweils auf diese Weise ermittelten kritischen F-Wert mit der von **RGP** zurückgegebenen F-Statistik, um das Konfidenzniveau Ihres Modells zu beurteilen.

ss_{reg} Die Regressions-Quadratsumme.

ss_{resid} Die Residual-Quadratsumme (Summe der Abweichungsquadrate)

Vorbereitung der Funktion RGP()

Da es sich um eine Arrayformel handelt, muss man die Formel schreiben, man kann die darin enthaltenen Bezüge nicht mit der Maus bestimmen. Es hat sich als hilfreich erwiesen, die Formel zunächst ohne das führende Gleich-Zeichen als Text in eine darüber liegende Zelle zu schreiben und anschließend in die linke obere Zelle des Array-Bereichs zu kopieren. Wenn man dann ein „=“ vor die Formel setzt, kann die eigentliche Erzeugung des Arrays gestartet werden.

G4 | fx =RGP(D3:D171;A3:C171;WAHR;WAHR)

	A	B	C	D	E	F	G	H	I	J
1	x1	x2	x3							
2	Datum	MBJ	WFL	NKM €/Mo			RGP(D3:D171;A3:C171;WAHR;WAHR)			
3	21.01.2013	1960	65 m²	463,20 €		RGP()	m3	m2	m1	b
4	21.01.2013	1960	65 m²	463,20 €		m	7,46907392			
5	22.04.2013	1960	64 m²	556,64 €		se				
6	27.05.2013	1960	54 m²	526,82 €		r				
7	07.06.2013	1960	103 m²	690,83 €		f				
8	17.06.2013	1960	65 m²	463,20 €		ss_e				
9	05.08.2013	1960	66 m²	508,93 €						

Sie sehen in grüner Schrift die vorbereitete Formel. Beginnend mit der Zelle „G4“ finden Sie den vorbereiteten Arraybereich, den Sie für eine RGP()-Funktion mit 3 Merkmalen benötigen. Für jedes weitere Merkmal müssen Sie links eine weitere Spalte ansetzen.

Erzeugung des Arrays der Funktion RGP() bei 3 Merkmalen

Setzen Sie den Cursor in die Zelle G4, Markieren Sie mit der Maus den Bereich des Arrays von Zelle G4 bis J8. Drücken Sie die Taste <F2>. Drücken Sie anschließend gleichzeitig die Tastenkombination <STRG – UMSCHALT + EINGABETASTE>.

	A	B	C	D	E	F	G	H	I	J
1	x1	x2	x3							
2	Datum	MBJ	WFL	NKM €/Mo			RGP(D3:D171;A3:C171;WAHR;WAHR)			
3	21.01.2013	1960	65 m²	463,20 €		RGP()	m3	m2	m1	b
4	21.01.2013	1960	65 m²	463,20 €		m	7,46907392	1,32498712	0,09589013	-6549,49754
5	22.04.2013	1960	64 m²	556,64 €		se	0,25789138	0,39509178	0,05948947	2551,17336
6	27.05.2013	1960	54 m²	526,82 €		r	0,83692257	83,8350123		
7	07.06.2013	1960	103 m²	690,83 €		f	282,263105	165		
8	17.06.2013	1960	65 m²	463,20 €		ss_e	5951497,21	1159671,03		
9	05.08.2013	1960	66 m²	508,93 €						

Da das Array nicht auf Anhieb selbsterklärend ist, habe ich es für das Beispiel mit 3 Markmalen mit erklärenden Zusätzen oberhalb und links versehen.

In der ersten Zeile stehen die Koeffizienten für die Regressionsgleichung. Und zwar immer von rechts nach links aufsteigend. In der Spalte J steht die Gleichungs-Konstante b.

Spalte Zeile	A	B	C	D
1	m_3	m_2	m_1	b
2	se_3	se_2	se_1	se_b
3	r^2	se_y		
4	F	d_f		
5	ss_{reg}	ss_{resid}		

In der Beschreibung der Formel in Excel wird dazu folgendes Schema mitgeteilt:

	A	B	C	D	E	F
1	m_n	m_{n-1}	...	m_2	m_1	b
2	se_n	se_{n-1}	...	se_2	se_1	se_b
3	r_2	se_y				
4	F	d_f				
5	ss_{reg}	ss_{resid}				

Da man erfahrungsgemäß nach kurzer Zeit die genaue Lage der Kenngrößen und ihre Bedeutung leicht vergisst bevorzuge ich Gutachten eine mit begleitenden Erklärungen versehene Darstellung, auch um bei nach Monaten zu erwartende Gerichtsverhandlung die notwendigen Erklärungen im mündlichen Termin liefern zu können.

Anwenden der multiplen Regressionsformel für 3 Merkmale

Dazu habe ich mit Pfeilen die Zuordnung der einzelnen Koeffizienten eingetragen.

Die Merkmale des Bewertungsobjekts ergeben sich aus dem Auftrag

Wenn man die einzelnen Teilmerkmale mit dem Koeffizienten multipliziert und anschließend addiert erhält man den gesuchten Wert – die Nettokaltmiete des Bewertungsobjekts.

RGP()	m3	m2	m1	b	
m	7,469073924	1,32498712	0,095890128	-6549,5	
se	0,257891383	0,39509178	0,059489466	2551,17	
r	0,836922571	83,8350123			
f	282,2631051	165			
ss_e	5951497,207	1159671,03			
	Merkmale d. Bew.Obj		**Koeffizienten der RGP()-Funktion**		$m_n \cdot x_n$
x1	11.01.2014	m1	0,096		3.993,82 €
x2	1987	m2	1,325		2.632,75 €
x3	68 m²	m3	7,469		507,90 €
		b	-6549,498		-6.549,50 €
	Das Ergebnis:		**Miete/Monat**		**584,97 €**
			Miete/m²		**8,60 €/m²**

	Merkmale d. Bew.Obj		Koeffizienten der RGP()-Funktion		$m_n \cdot x_n$
x1	11.01.2014	m1	0,096		3.993,82 €
x2	1987	m2	1,325		2.632,75 €
x3	68 m²	m3	7,469		507,90 €
		b	-6549,498		-6.549,50 €
	Das Ergebnis:		Miete/Monat		584,97 €
			Miete/m²		8,60 €/m²

Es ist durchaus möglich – wie das Beispiel zeigt, mit formatierten Eingabewerten zu arbeiten. Bei der Würdigung dieses Arrays stellt man fest, dass einzelne Koeffizienten absolut geringe Werte sind und andere hohe Werte aufweisen. Ganz wichtig ist, dass man die Zuordnung der einzelnen Koeffizienten nicht durcheinander bringt. Die Reihenfolge der Merkmale ist umgekehrt zur Reihenfolge der Ausgangswerte der Datenbasis. In Spalte „A“ steht das Datum. Für die Statistik werden die Daten in dieser Spalte als „x_1“ bezeichnet. Diese werden multipliziert mit dem Koeffizienten „m_1“, der in dem Array an dritter Stelle steht, links neben der Konstanten „b“.

Erläuterung der Kennzahlen der RGP()-Funktion

In der dritten Zeile und der zweiten Spalte steht der „Standardfehler der Prognose“ se_y mit 83,84 €. Dividiert man den Wert y der Gleichung mit dem statistischen Standardfehler se_y, dann kann man im Rahmen einer Sensitivitätsanalyse eine Aussage zur **Genauigkeit** des gefundenen Mietwerts machen. Die Genauigkeit der berechneten Miete liegt in diesem Beispiel bei ± 14,3%.

RGP() liefert das **Bestimmtheitsmaß R²** des Werts y in der ersten Spalte der dritten Zeile. In diesem Beispiel mit 0,8369.

Zur Erinnerung: Das R^2 liegt immer zwischen 1 und 0. Je näher das R^2 bei 1 ist, desto vollkommener ist die Korrelation dieser Merkmale mit dem Zielwert. Bei einem R^2 von 0 ist die Regressionsgerade nicht geeignet einen y-Wert abzuleiten. Das R^2 ist jedoch nur ein Kennwert von mehreren und sagt für sich genommen noch nichts aus über die Zuverlässigkeit der statistischen Aussage. Insbesondere, wenn eine Autokorrelation vorliegt, kann sich ein hohes R^2 ergeben, obwohl der berechnete Wert ungültig ist.

Aus dem Bestimmtheitsmaß lässt sich ganz einfach der Korrelationskoeffizient r berechnen.

$$r = \text{Wurzel}(R^2)$$

Die F-Statistik bei 3 Merkmalen

Bei einem hohen R^2 geht man davon aus, dass es einen engen Zusammenhang zwischen der unabhängigen Variablen und dem abgeleiteten Mietwert gibt. Mit Hilfe der F-Statistik können Sie prüfen, ob dieses Ergebnis mit einem derart großen Bestimmtheitsmaß zufällig ist oder nicht.

Dazu stellen Sie die Hypothese auf, das zwischen den Variablen eigentlich kein Zusammenhang besteht, sondern dass Sie nur zufällig Daten ausgewertet haben, für die die statistische Analyse einen starken Zusammenhang anzeigt. Um die Wahrscheinlichkeit zu beschreiben, mit der irrtümlich ein Zusammenhang ermittelt wird, wird die Irrtumswahrscheinlichkeit „Alpha“ verwendet.

Außerdem benötigen Sie die von der RGP()-Funktion mitgelieferten Werte für F und d_f. Damit können Sie die Wahrscheinlichkeit eines zufällig auftretenden höheren F-Werts bewerten. In der F-Statistik wird der berechnete F-Wert mit einem situationsabhängigen kritischen f-VerteilungsWert verglichen. Dieser ist in den einschlägigen Fachbüchern tabelliert, lässt sich aber auch problemlos mit der Excel-Funktion FVERT() berechnen. Die Funktion FVERT() arbeitet mit einem Alpha von 95 %.

Der kritische Wert P ist umso größer, desto kleiner die Stichprobe ist. Im vorliegenden Beispiel haben wir 169 Datensätze und 3 Merkmale, d.h. wir haben für immobilienwirtschaftliche Auswertungen verhältnismäßig sehr viele Datensätze und damit eine hohe statistische Sicherheit (viele Freiheitsgrade d_f).

Die FVERT()-Funktion (die Funktion zur Berechnung der F-Verteilung) benötigt 3 Parameter. Den mit der RPG()-Funktion berechneten F-Wert in der ersten Spalte und 4. Zeile des Arrays. In diesem Beispiel ist es 282 und die Freiheitsgrade v1 und v2. Für den Fall, dass Sie die RGP()-Funktion mit der Konstanten WAHR berechnet haben, gilt
für v1 = n – d_f – 1 (bei Verwendung der Funktion mit der Konstanten FALSCH gilt für v1 n – d_f) und für v2 = d_f. (Im Beispiel mit n=169 und d_f = 165 ergibt sich für v1 ein Wert von 3 und für v2 ein Wert von 165.) Der damit berechnete kritische Wert beträgt demnach 9,93093E-65.
Eine sehr kleine Dezimalzahl mit 0, 60 Nullen und dann 99309.

Angenommen, als Alpha-Quantil wird 0,05 verwendet, v1 = 169 – 165 – 1 = 3 und v2 = 165, dann liegt das kritische Niveau von F bei 0,00000000 … 9. Da F mit 282,… viel größer ist als 0,00000 …9, ist es sehr unwahrscheinlich, dass zufällig ein so hoher F-Wert auftritt. (Wird Alpha = 0,05 verwendet, ist die Hypothese, dass kein Zusammenhang zwischen *Y_Werten* und *X_Werten* besteht, ungültig, wenn F das kritische Niveau (0,000000…9) überschreitet. Mithilfe der FVERT-Funktion in Excel können Sie die Wahrscheinlichkeit berechnen, dass zufällig so ein hoher Wert auftritt. Bei FVERT(282;3;165) = 9,93093E-65 ist diese Wahrscheinlichkeit beispielsweise sehr gering.

Da der absolute F-Wert im positiven 3-stelligen Bereich liegt, ist er definitiv größer als der kritische Wert und damit ist der F-Test positiv bestanden.

n	Anzahl Daten	169	
F-Wert	der Gleichung		282,263
d_f	Freiheitsgrade	165	
v1	n - d_f - 1	3	
v2	d_f	165	
FVERT(F;v1;v2)	F-Verteilungskontrollwert		9,93093E-65

FVERT() gibt die Wahrscheinlichkeit eines zufällig auftretenden höheren F-Wertes zurück. Das ist hier sehr unwahrscheinlich.

Die t-Statistik bei 3 Merkmalen

Mithilfe einer anderen Hypothese kann festgestellt werden, ob die einzelnen Steigungskoeffizienten geeignet sind, den gesuchten Mietwert zu berechnen.

Mit dem t-Test lassen sich für jeden Koeffizienten berechnen, ob die einzelnen Steigungskoeffizienten geeignet sind, den Schätzwert z.B. der Miete zu berechnen. Man will z.B. die statistische Wahrscheinlichkeit (Sicherheit) prüfen. Dazu werden die berechneten Koeffizienten durch den Standardfehler dieses Koeffizienten geteilt.

Der kritische Prüfwert P ist abhängig von einem Alpha α und den Freiheitsgraden d_f.
Das Alpha ist meist 95 %, d.h. man will den t-Test mit einer 95 %igen Genauigkeit durchführen.

Man kann den kritischen Wert P aus Tabellen ablesen oder mit der Funktion TINV(α,d_f) berechnen.
Der kritische Wert P hat in diesem Beispiel die Größe P = TINV(0,05;165) = 1,97444563.

Um zum Beispiel den Koeffizienten für den Stichtag bezüglich der statistischen Wahrscheinlichkeit (Sicherheit) zu prüfen, dividieren Sie 0,09589 (Steigungskoeffizient für das Datum) durch 0,059 (den in der 2. Zeile in der 3. Spalte stehenden Standardfehler se_1 des Datumkoeffizienten). Daraus ergibt sich der folgende t-Wert:

$$t = m_1 \div se_1 = 0{,}9589 \div 0{,}0595 = 1{,}61188$$

Dieser t-Wert ist kleiner als der kritische Wert P. Damit trägt dieser Koeffizient nicht dazu bei, die Miete des Bewertungsobjekts mit der angenommenen Prüfgenauigkeit von 95 % zu schätzen. Andernfalls, d.h. wenn der Absolut-Wert von t hoch genug ist, kann geschlussfolgert werden, dass der Steigungskoeffizient für die Berechnung des Mietwertes hilfreich ist. In der folgenden Tabelle sind die Absolut-Werte der drei berechneten t-Werte dargestellt.

				m_n / se_n	t > P
m_1	0,095890128	se_1	0,059	1,61188	FALSCH
m_2	1,324987118	se_2	0,395	3,35362	WAHR
m_3	7,469073924	se_3	0,258	28,9621	WAHR

Mit diesem Ergebnis ist das Datum für die Miete bei diesen Vorgaben ohne Bedeutung.

Der kritische Wert P ist aber auch abhängig vom angenommenen Alpha. Alpha bezeichnet die in der Hypothese angenommene Wahrscheinlichkeit, dass es sich tatsächlich um zufällige Ergebnisse handelt. Üblich wird in der Literatur ein Alpha von 95 % angenommen. D.h. wir berechnen, ob es eine Wahrscheinlichkeit von 5 % gibt, dass der gefundene Wert zufällig aufgetreten ist. Tabelliert sind neben 95 % auch 97 %, 98,5 % 99%; man kann aber in der immobilienwirtschaftlichen Statistik (insbesondere in kaufpreisarmen Märkten mit sehr wenigen Datensätzen) die Annahme treffen, dass diese Wahrscheinlichkeit mit 90 % oder sogar 85 % angenommen wird. Entsprechend niedriger wird der kritische Wert P. Und das Prüfergebnis, das alle Korrelationskoeffizienten den Zielwert zutreffend beschreiben, ist leichter zu erbringen.

Der kritische Wert P liegt bei anderen Alpha wie folgt:		α	P
für 165 Freiheitsgrade	0,150	85,0%	1,44626423
	0,100	90,0%	1,654140976
	0,050	95,0%	1,97444563
	0,025	97,5%	2,262041183
	0,020	98,0%	2,349159919
	0,010	99,0%	2,605953791

Wenn man eine 15%-ige Unsicherheit (85%ige Sicherheit) beim t-Test zulassen möchte, dann sind alle Koeffizienten geeignet, für die Schätzung der Miete, herangezogen zu werden. Wenn man dies im Gutachten nicht mit entsprechender Begründung offen legt, dann grenzt dies schon an den Tatbestand der Manipulation.

Wenn ein Koeffizient im t-Test durchfällt sollten wir die Regressionsrechnung ein weiteres Mal – nun ohne das durchgefallene Merkmal durchführen.

Erzeugung der multiplen linearen Regressionsfunktion bei 2 Merkmalen

Diesmal haben wir nur die beiden Merkmale x_1 aus Spalte B (MBJ) und x_2 aus Spalte C (WFL).

In der RGP-Funktion wird für die X-Werte entsprechend B3:C171 eingetragen.

	RGP(D3:D171;B3:C171;WAHR;WAHR)		
RGP()	m2	m1	b
m	7,443287875	1,35828916	-2636,977049
se	0,2586308	0,39644595	789,0852894
r	0,834354674	84,2376078	
f	418,0705839	166	
ss_e	5933236,46	1177931,78	

Anwenden der multiplen linearen Regressionsformel

			Koeffizienten der	
	Merkmale d. Bew.Obj		RGP()-Funktion	$m_n \cdot x_n$
x1	1987	m1	1,358	2.698,92 €
x2	68 m²	m2	7,443	506,14 €
		b	-2636,977	-2.636,98 €
	Das Ergebnis:		**Miete/Monat**	**568,09 €**
			Miete/m²	**8,35 €/m²**

Dann werden die Prüf-Kennziffern der RPG()-Funktion strukturiert aus dem Array entnommen:

Standardfehler der Prognose se_y 84,24 €

Damit liegt die Genauigkeit der berechneten Miete (Miete/se_y) bei ± 14,8%.

Im Vergleich zur vorherigen Berechnung mit 3 Merkmalen hat die Ungenauigkeit von 14,3 auf 14,8 % zugenommen. Die relative Miete liegt nun bei 8,35 €/m² statt 8,60 €/m². Das sind 3 % weniger als vorher. Der Standardfehler des Zielwertes hat sich von 83,84 € auf 84,24 € erhöht. Das sind Belege für die Unbedeutsamkeit des Datums in diesem Beispiel.

Das Bestimmtheitsmaß R^2 lautet 0,8344

Der Korrelationskoeffizient r lautet 0,9134

Die F-Statistik			
n	Anzahl Daten	169	
F-Wert	der Gleichung		418,071
d_f	Freiheitsgrade	166	
v1	n - d_f - 1	2	
v2	d_f	166	
FVERT(F;v1;v2)	F-Verteilungskontrollwert		1,55551E-65

Solange der F-Wert absolut größer, als der FVERT-Wert ist, hat die Formel den F-Test bestanden. Hier ist der berechnete F-Wert deutlich höher als im ersten Beispiel.

Der t-Test bei 2 Merkmalen

Da wir unsere Auswertung mit einer 95%igen Sicherheit im t-Test anwenden möchten, beträgt das Alpha-Quantil bei einer gewünschten 95%igen Aussage 0,05.

Der kritische Wert P hat in diesem Beispiel die Größe TINV(0,05;166) = 1,974357764

				m_n / se_n	t-Wert > P
m_1	1,358289158	se_1	0,396	3,42616	WAHR
m_2	7,443287875	se_2	0,259	28,7796	WAHR

Der t-Test ist positiv bestanden. Ergebnis: Die beiden Koeffizienten sind geeignet, den Mietwert zu berechnen!

Das Merkmal MBJ ist dabei deutlich weniger bestimmend; wie die WFL.

Überlegungen zur Beurteilung der Zusammenhangsmaße

Zusammenhangsmaße	Datum	MBJ	WFL
Bestimmtheitsmaß (SP:NKM €/m²)	0,0177	0,0699	0,0842
Korrelationskoeffizient (SP:NKM €/m²)	0,1332	0,2645	-0,2902
Kovarianz(s) (SP:NKM €/m²)	17,7859	5,2318	-9,4452

Der Absolut-Wert der Kovarianz für das Merkmal MBJ (Siehe 1 dieser Anleitung) war mit 5,… deutlich kleiner als der Wert der Kovarianz für das Merkmal WFL und der Wert der Kovarianz für das Merkmal Datum. Es gibt eine Regel in der Statistik, dass man mit dem Merkmal beginnt, welches den höchsten Zusammenhang für den Zielwert liefert. Die Frage stellt sich, woran mache ich das fest?

Beim Bestimmtheitsmaß ergibt sich die Rangfolge WFL, MBJ, Datum, wobei diese bei allen 3 Merkmalen besonders niedrig sind und gar keinen Zusammenhang vermuten lassen.

Beim Korrelationskoeffizienten, einem relativen Wert, ergibt sich die Rangfolge der absoluten Werte ebenfalls mit WFL, MBJ, Datum, wobei diese Werte bereits deutlich höher liegen als beim Bestimmtheitsmaß.

Bei der Kovarianz lautet die Rangfolge der absoluten Werte Datum, WFL MBJ.

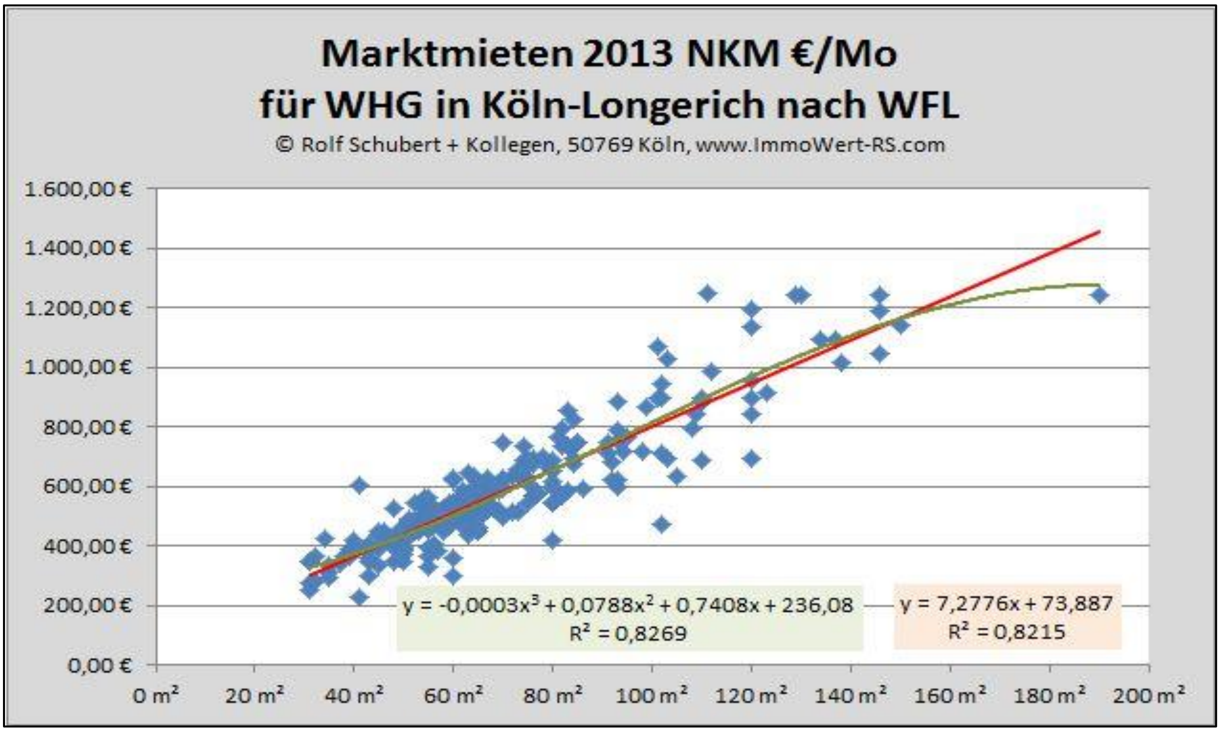
Im t-Test bei den 3 Merkmalen ist das Datum durchgefallen. Der Prüfwert für das Merkmal MBJ liegt nur knapp über dem kritischen Wert und die WFL weicht davon deutlich positiv ab. Dasselbe gilt für den t-Test mit 2 Merkmalen, auch hier liegt der Prüfwert für das MBJ nur knapp über dem kritischen Wert, während das Merkmal WFL davon deutlich positiv abweicht.

Da WFL auch bei den Zusammenhangsmaßen zweimal an erster Rangstelle steht, sollte man testen, ob es ggf. reicht, die Anpassung nur mit diesem einen Merkmal vorzunehmen.

Deshalb wird die Abhängigkeit des Mietwerts von der Wohnfläche nun im Rahmen der einfachen linearen Regression geprüft.

Gegenüberstellung der einfachen linearen Regression

Wir erstellen aus den beiden Datenreihen der WFL und der Nettokaltmiete in €/Monat eine Punktwolke und legen in diese zunächst eine lineare Trendkurve und lassen uns die Regressionsformel und das Bestimmtheitsmaß anzeigen:



Wir stellen fest, dass die lineare Gerade eine deutliche Steigung der absoluten Miete pro Monat in Abhängigkeit von der Wohnfläche zeigt. Das muss nach einschlägiger Erfahrung auch so sein, denn je größer die Wohnung, desto höher ist die Miete.

Um den Einfluss der Wohnfläche auf die relative Miete zu eliminieren müssen wir deshalb eine weitere Regressionsrechnung für das Verhältnis Wohnfläche auf die relative Miete durchführen. Dazu verwenden wir die Wohnfläche als bestimmendes Merkmal und die relative Miete in €/m² als gesuchte abhängige Größe.

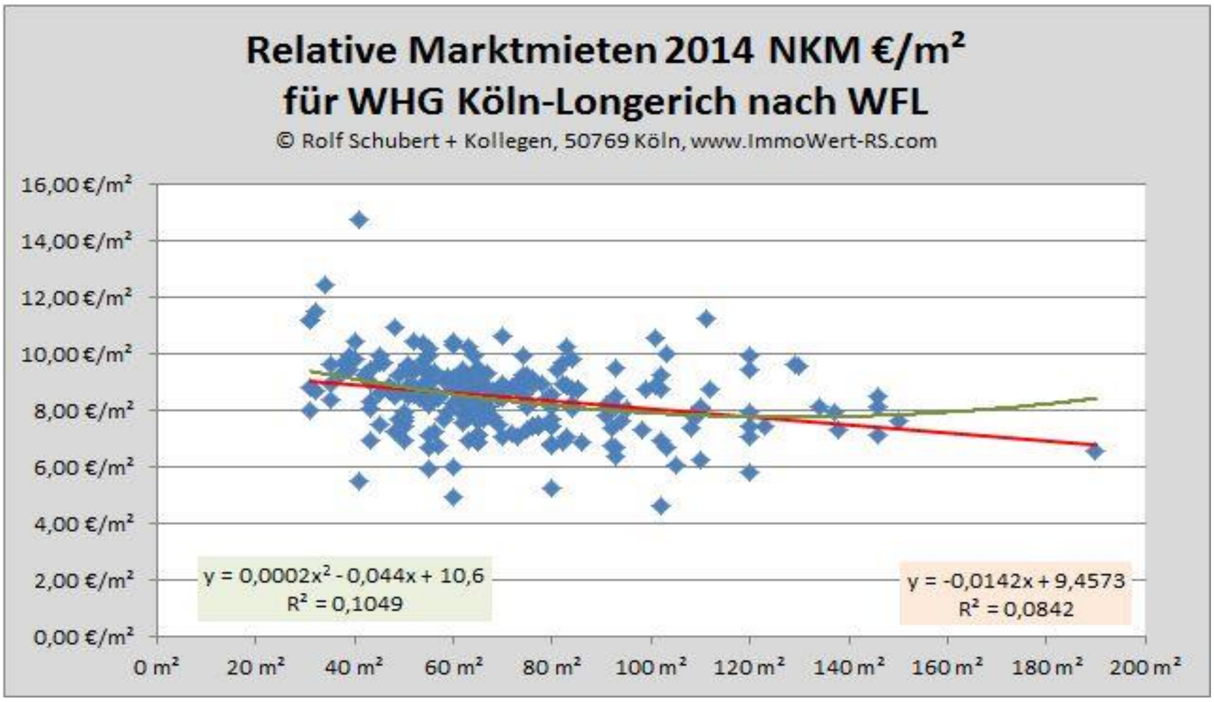
Wir stellen fest, dass die relative Miete wohnflächenbereinigt im Bereich zwischen 30 m² und 120 m² fast linear abnimmt, um danach wieder anzusteigen. Das kann mit dem Einfluss des Mietspiegels auf die Mieten zusammenhängen. Größere Wohnung werden durch den Mietspiegel nicht gedämpft und wegen der mangelnden Transparenz auf dem Markt sowie der besseren Ausstattungsqualität größerer Wohnungen mit mehreren WC und oft einem Zweitbad steigen die relativen Mieten wieder an.

Die daraus abgeleitete relative Miete für das Bewertungsobjekt liegt nach dieser Auswertung bei

Anpassungsparameter		Mietwert
MW	71 m²	8,45 €
WFL	68 m²	8,49 €
Differenz	-3 m²	0,04 €
in Prozent	-4,23%	0,50%

rund 8,50 €/m².

In der multiplen Regressionsrechnung ergaben sich 8,60 €/m², bzw. 8,35 €/m².



Als Fazit kann man mit ziemlicher Sicherheit einen Mietwert zwischen 570 €/Monat und 585 €/Monat zum Stichtag 11.01.2014 feststellen. Damit sind rund 8,50 €/m² oder 578 €/Monat ein statistisch gut gesicherter Wert.

Durch die Prüfkriterien der multiplen Regressionsanalyse hat man feststellen können, dass die anderen Merkmale sich kaum oder gar nicht auf den Mietwert auswirken. Diese Auswertung dient zunächst nur der Anleitung, wie man mit den zur Verfügung stehenden Mitteln zu einem statistisch gesicherten Mietwert kommen kann. Diese Erkenntnisse gelten für das untersuchte Jahr und den untersuchten Teilmarkt. Sie sind auf andere Märkte nicht übertragbar.

Normalerweise hat der zeitliche Aspekt insbesondere in dynamischen Marktphasen einen wesentlich deutlicheren Einfluss auf den Mietwert. Aber mit der grafischen Auswertung der einfachen Regressionsgleichung können wir einen stagnierenden Markt erkennen, was zwangsläufig zu dem gefundenen Ergebnis im ersten Anlauf führen musste.