



# Einzelgrafiken versus Multiple Regression

## Grafische Datendarstellung gegen Statistik

### Inhalt

Voraussetzung und verwandte Themen.....	1
Keywords.....	1
Einführung.....	2
Darstellung von Einzelgrafiken in Visual-XSel.....	3
Matrixplot.....	5
Die Methode Multiple Regression in Visual-XSel .....	7
Fazit.....	10
Literatur - Weiterführende Beschreibungen.....	11
Consulting & Schulungen .....	11

### Voraussetzung und verwandte Themen

Für diese Beschreibungen sind Grundlagen der Statistik vorteilhaft. Weiterführende und verwandte Themen sind:

[www.crgraph.de/Literatur](http://www.crgraph.de/Literatur)

[www.versuchsmethoden.de/Versuchsplanung.pdf](http://www.versuchsmethoden.de/Versuchsplanung.pdf)

[www.versuchsmethoden.de/Multiple\\_Regression.pdf](http://www.versuchsmethoden.de/Multiple_Regression.pdf)

[www.versuchsmethoden.de/PLS.pdf](http://www.versuchsmethoden.de/PLS.pdf)

### Keywords:

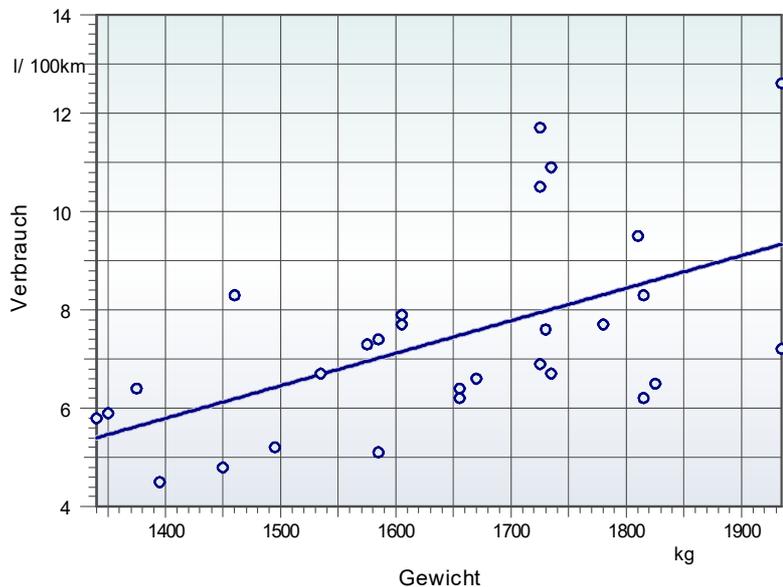
Einzelgrafiken, Streudiagramme, XY-Diagramme, Multiple-Regression, Regressionsanalyse, Datenanalyse

## Einführung

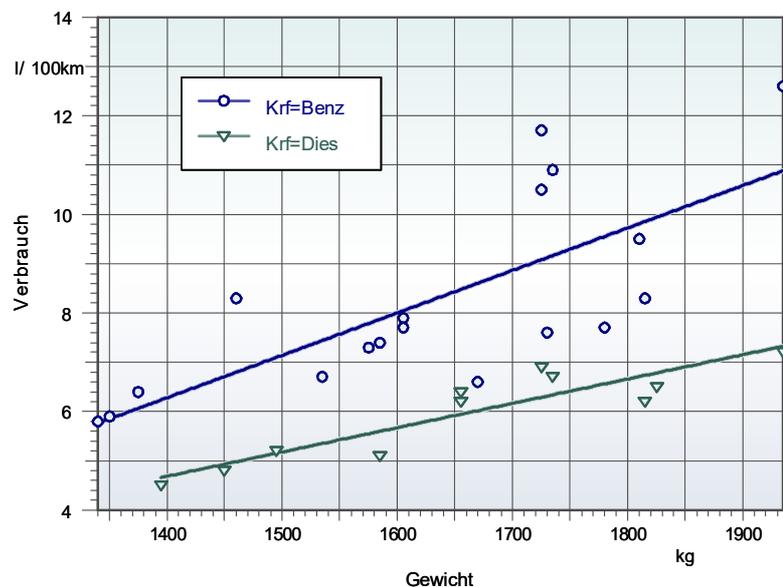
Bei der Darstellung von Zusammenhängen werden oft XY-Diagramme oder sogenannte Streudiagramme verwendet. Diese liefern schnell einen Überblick wie groß ein Einfluss X auf eine abhängige Größe Y ist.

Beispiel Fahrzeugverbrauch  
 ⇒ Y von Fahrzeugen gegenüber Gewicht ⇒ X.

Das Ganze funktioniert soweit gut, wenn die Daten z.B. aus einem voll [faktoriellen Versuchsplan](#) stammen, da die gegenseitigen Einflussgrößen in der Tabelle bezüglich ihrer Werte (hohe und niedrige) ausgeglichen oft vorkommen. Wenn die Daten aber nicht „ausbalanciert“ sind, wie in dem obigen Beispiel, sind diese nicht ausgeglichen und korrelieren meist stark. Das führt dazu, dass die Kurvenverläufe in Einzelgrafiken nicht die vollständigen Beziehungen wiedergeben, da benachbarte Einflussgrößen mit hineinwirken, aber nicht gleichzeitig im Diagramm dargestellt werden können.



Da das Fahrzeuggewicht bekanntlich nicht der einzige Einfluss auf den Fahrzeugverbrauch ist, kann im Diagramm eine Aufteilung nach Gruppen weiteren Aufschluss bringen. So sind im Diagramm rechts die Daten in Benzin- und Dieselfahrzeuge aufgeteilt. Die nächste Gruppenbildung wäre die Zylinderanzahl, die auch einen wesentlichen Einfluss hat.



Für diese Darstellung sollen die weiteren Schritte direkt mit der Umsetzung in Visual-

XSel aufgezeigt. Informationen zur Software sind zu finden unter

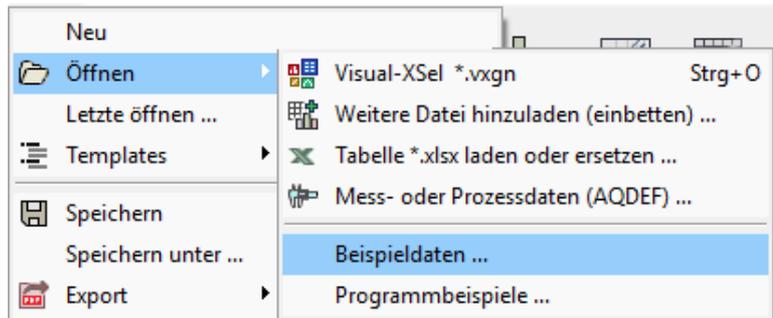
<https://crgraph.de/visual-xsel-software/>

## Darstellung von Einzelgrafiken in Visual-XSel

Visual-XSel besitzt neben den XY-Diagrammen eine sehr leistungsfähige Datenanalyse in Form einer Multiplen Regression oder mit Partial-Least-Square für hoch korrelierende Daten.

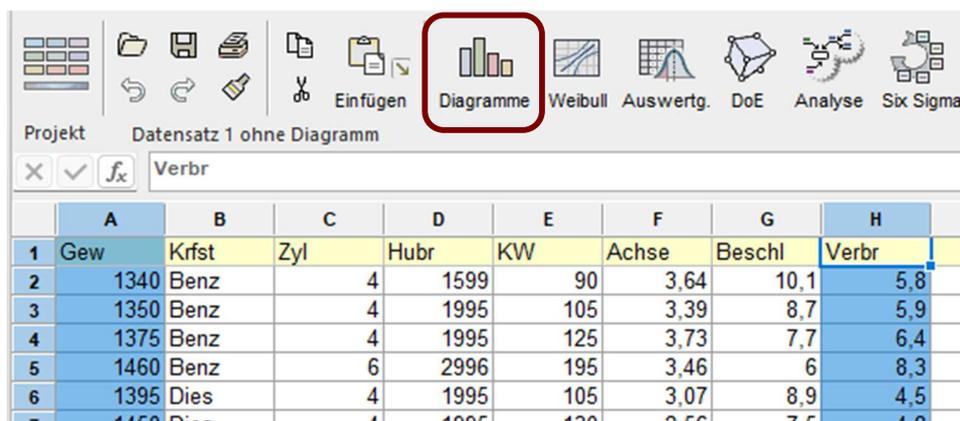
Die vorher gezeigten Diagramme stammten aus dem Beispieldatensatz *Beispiel\_Verbrauch.xls*, der hoch korrelierende Zusammenhänge aufweist und im Folgenden verwendet wird:

Insgesamt gibt es 7 Einflussparameter auf den Verbrauch in der letzten Spalte. Das Beispiel ist technisch leicht nachzuvollziehen, man erwartet u.a. gegenseitige Abhängigkeiten zwischen Hubraum, Leistung in KW und der Beschleunigung.

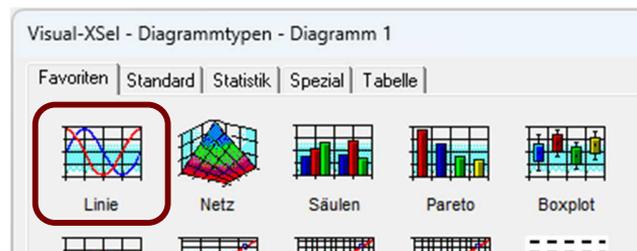


	1	2	3	4	5	6	7	8	
1	Gew	Kfst	Zyl	Hubr	KW	Achse	Beschl	Verbr	
2	1340	Benz		4	1599	90	3,64	10,1	5,8
3	1350	Benz		4	1995	105	3,39	8,7	5,9
4	1375	Benz		4	1995	125	3,73	7,7	6,4
5	1460	Benz		6	2996	195	3,46	6	8,3

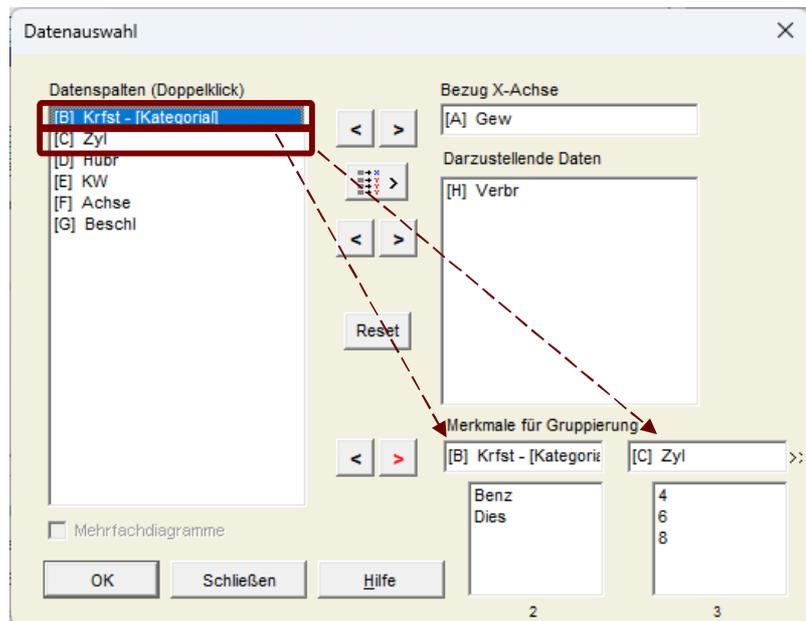
Wie unter der Einführung beschrieben, soll als nächste die Gruppe um die Zylinderzahl erweitert werden. Zum Aufbau dieses Diagrammes ist die Ikone Diagramm zu wählen:



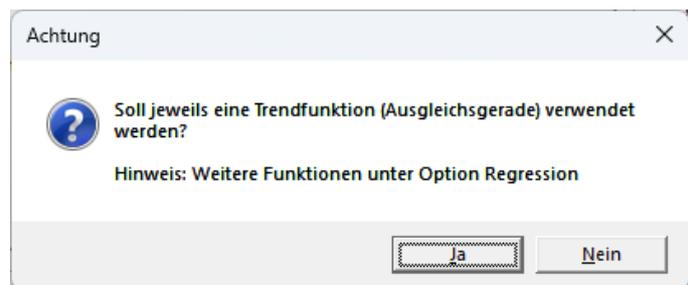
Falls die Spalte *Gew* und *Verbr* bereits markiert wurden, ist im übernächsten Dialog der X-Bezug und die darzustellende Zielgröße bereits gewählt. Zunächst soll aber der Diagrammtyp Linie ausgewählt werden:



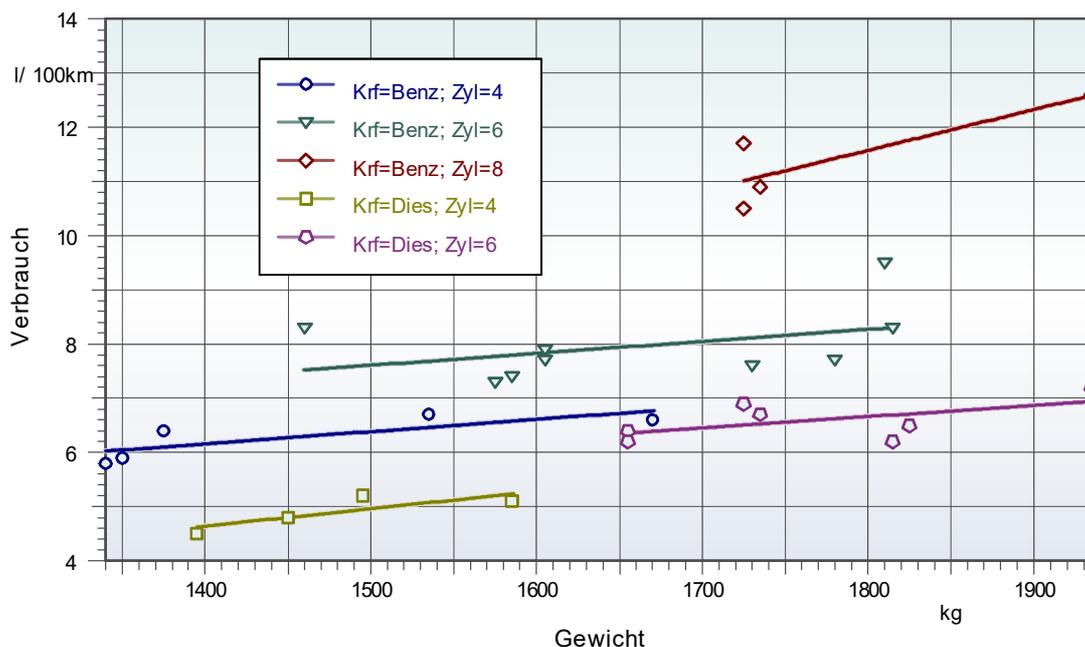
Nach gewähltem Bezug der X-Achse oben rechts und der „Darzustellenden Daten“ ist als Merkmale für Gruppierungen die Kraftstoffart und die Zylinderzahl zu wählen. Die anderen Datenspaten machen hier keinen Sinn für Gruppen, da sich hier keine Gruppen bilden lassen (jeder Wert kommt nur einmal oder sehr wenig vor).



Damit die Daten in der Darstellung sortiert und als Trendfunktion zu sehen sind, ist nach Ok der folgende Hinweis zu bestätigen (ab Version 20.0013)



Nach weiterer Ergänzung der Überschriften und Achsen, erscheint folgende Grafik:



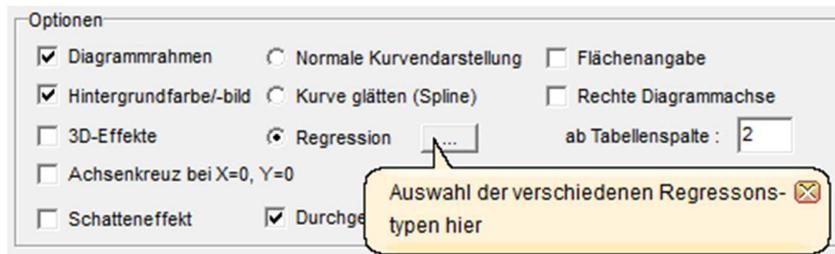
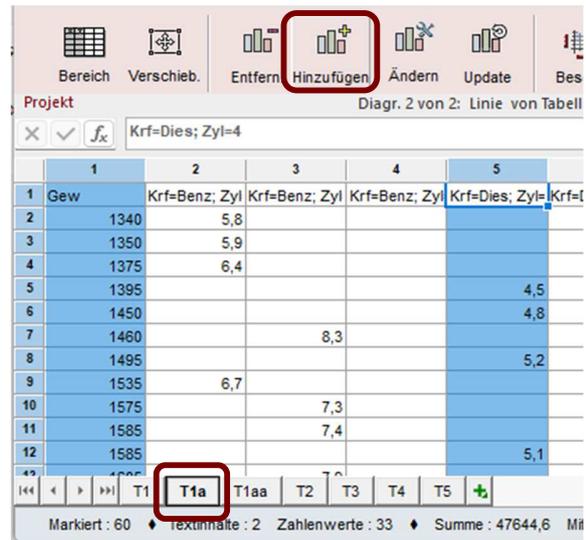
Deutlich zu sehen ist der hohe Verbrauch des Achtzylinders und der größere Einfluss

des Gewichtes aufgrund der steileren Steigung im Diagramm.

Hinweis zum Programm: Für diese Grafik wird durch den Dialog der darzustellenden Daten mit der Gruppenbildung eine neue Tabellenseite T1a angelegt, wo die Gruppen einzusehen sind.

Möchte man anstelle einer einzigen Grafik eine oder weitere nur für eine Gruppe, so sind diese zu markieren und die Ikone Diagramm/Hinzufügen zu wählen.

Soll anstelle einer einfachen Ausgleichsgerade eine andere Funktion gewünscht sein, so kann unter der Ikone *Diagramm/Ändern* im Dialog der Diagrammtyp unter Regression eine beliebige Funktion gewählt werden:

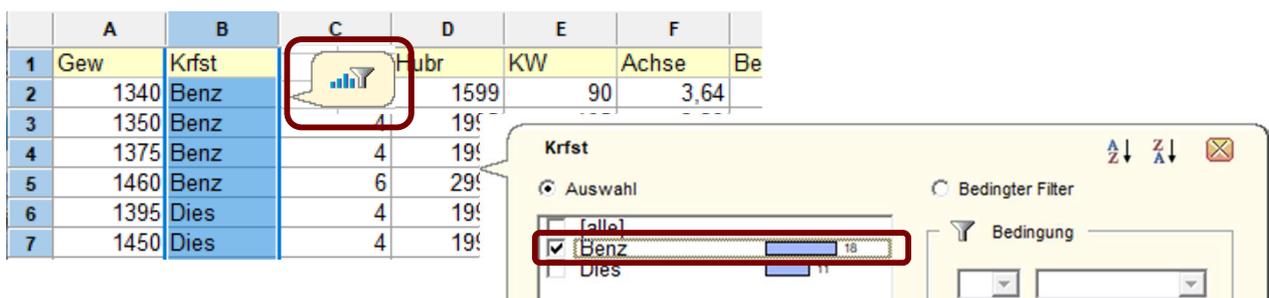


## Matrixplot

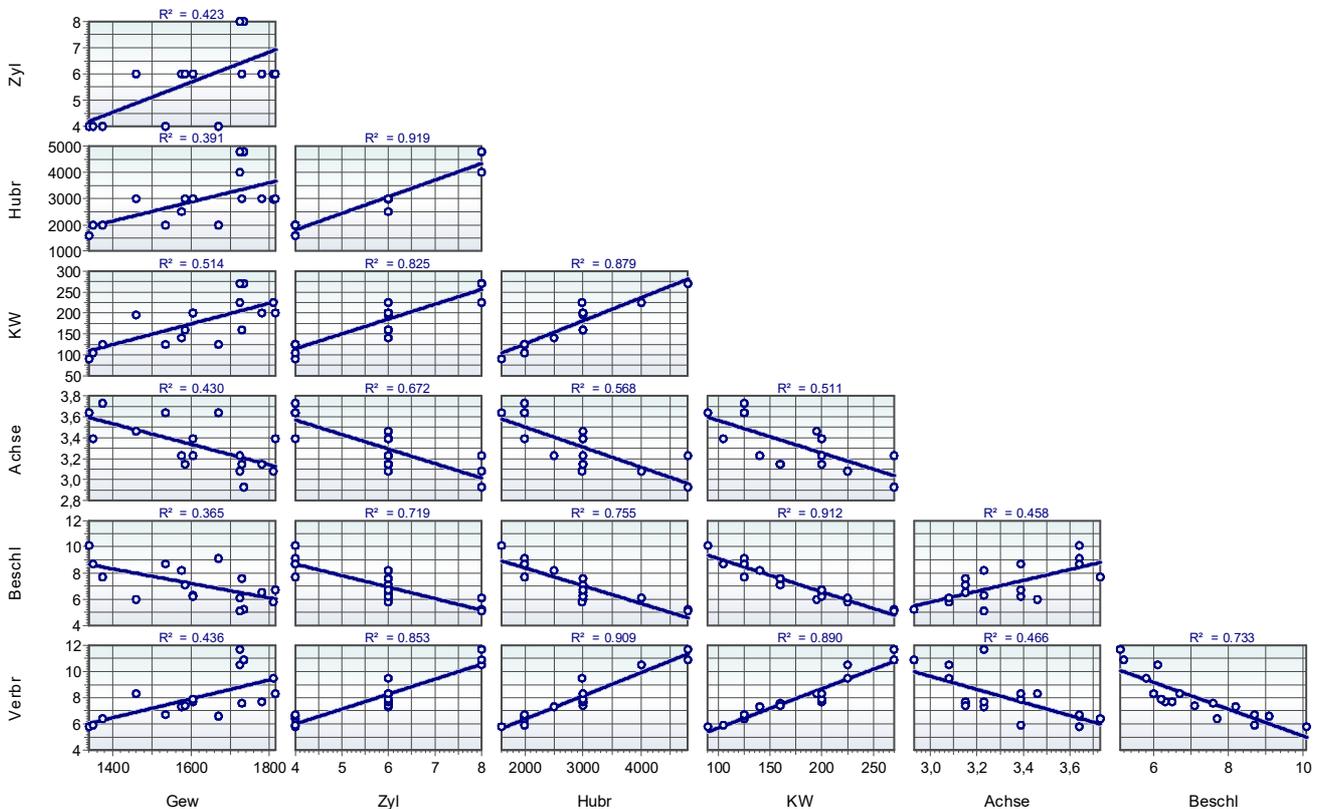
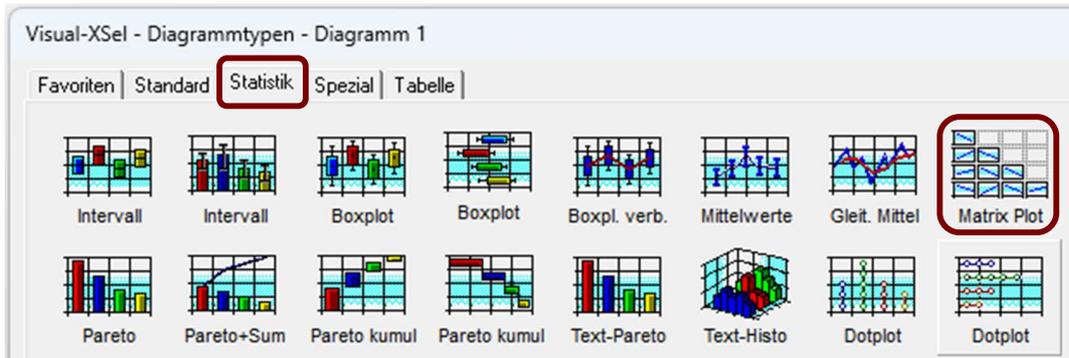
Dieses Kapitel dient nur dazu aufzuzeigen, welche Grafikformen noch verfügbar sind. Es trägt jedoch nicht zur Beantwortung bei, ob Einzelgrafiken ausreichend sind und kann übersprungen werden.

Eine andere Möglichkeit der Einzelgrafiken ist der sogenannte Matrix-Plot. Der Vorteil ist hier, dass auf einen Schlag alle Kombinationen dargestellt werden können.

Hier sind jedoch nur numerische Daten möglich. Man muss sich hier für eine Kraftstoffart entscheiden, z.B. für den Benziner. Hierzu kann der Filter verwendet werden. Klicken Sie oben auf die gewünschte Spalte und dann auf das Filtersymbol:



Wählen Sie dann die Diagramm-Ikone und den Matrix Plot unter der Rubrik Statistik



Das Gleich kann auch für den Diesel gemacht werden. Hierzu muss unter der Tabelle T1 der Filter zurückgesetzt werden und dieser erneut für den Diesel gesetzt werden.

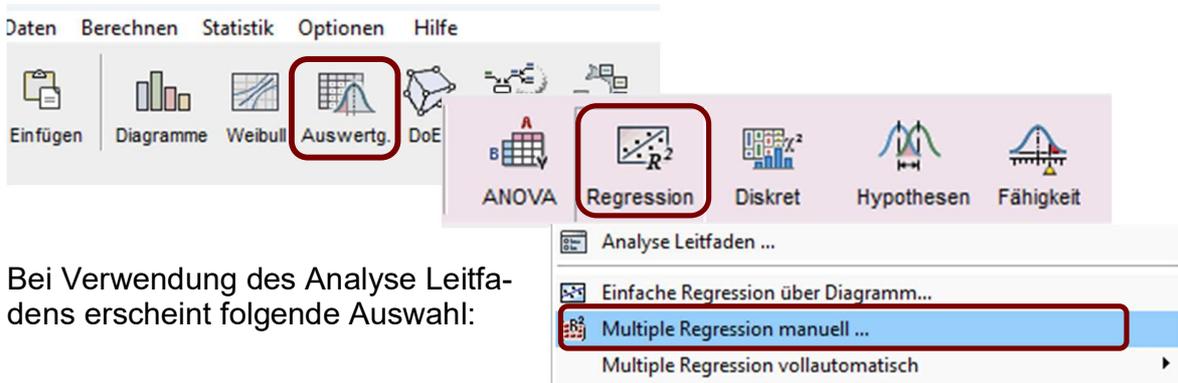
Damit das vorhergehende Diagramm nicht überschrieben wird, muss die Tabelle markiert werden und unter der Diagramm-Ikone Hinzufügen gewählt werden



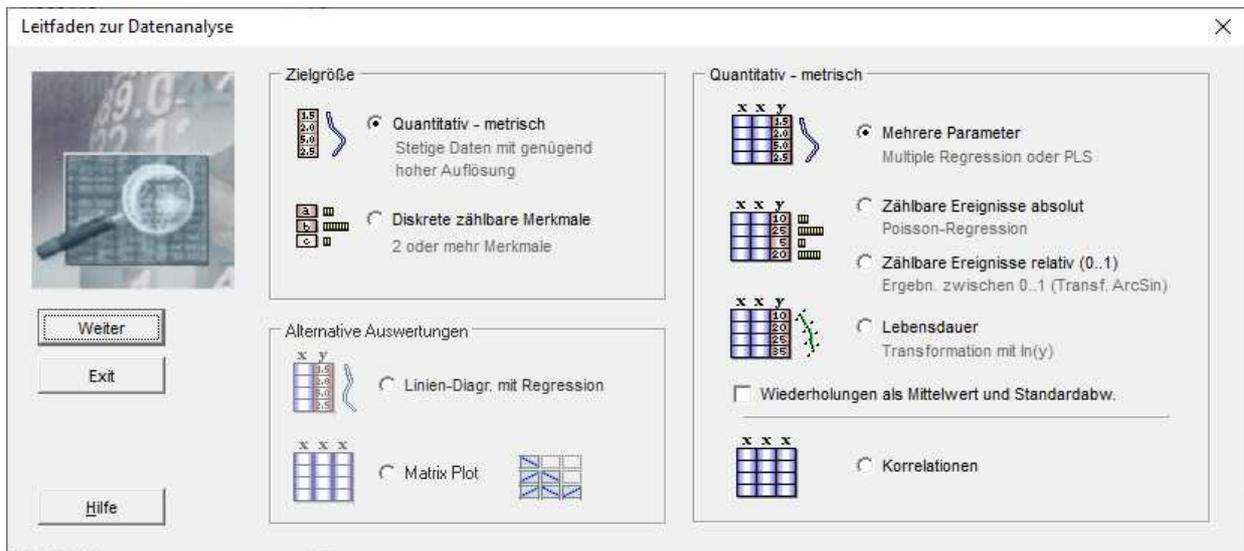
## Die Methode Multiple Regression in Visual-XSel

Wie Eingangs beschrieben, haben die bisher gezeigten Diagramme den großen Nachteil, dass immer nur eine Einflussgröße unter eventueller Beeinflussung anderer dargestellt wird und dass dadurch der „Effekt“ in Abhängigkeit der Steigung der Grafiken verfälscht sein kann. Hier wird deshalb die folgende Methode empfohlen.

Es sollte die Beispieldatei. ...[\Beispieldaten\Beispiel\\_Radaufhängung\\_MulReg.xls](#), erneut geladen werden, um zu verhindern, dass sich Tabellenseiten der bisherigen Auswertungen überschneiden.



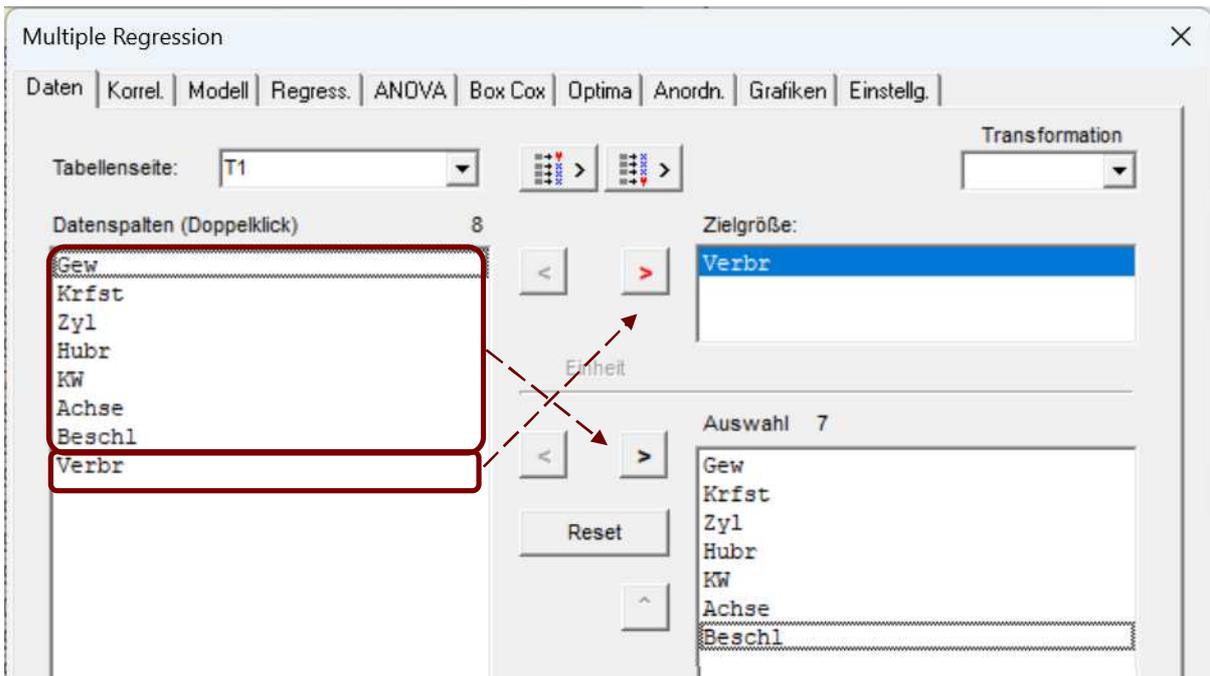
Bei Verwendung des Analyse Leitfadens erscheint folgende Auswahl:



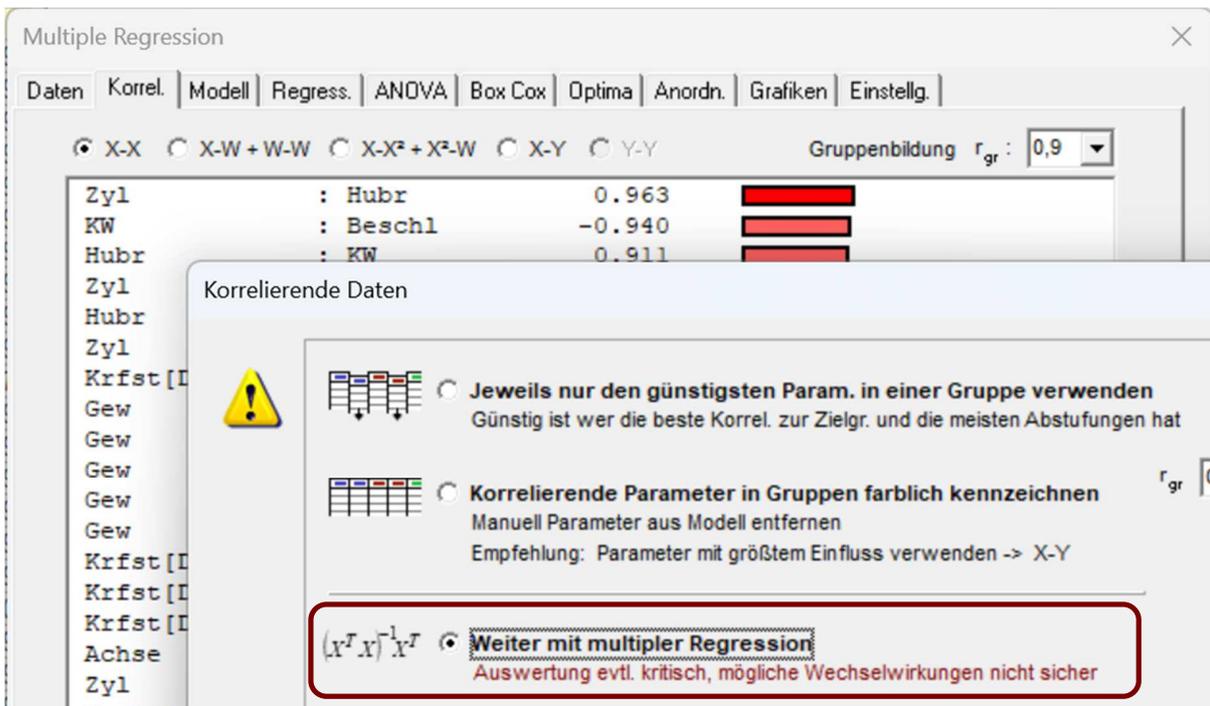
Eine ausführliche Beschreibung der statistischen Grundlagen hierfür ist unter [www.versuchsmethoden.de/Multiple\\_Regression.pdf](http://www.versuchsmethoden.de/Multiple_Regression.pdf)

beschrieben. Für die hier gezeigte Thematik soll nur eine vereinfachte Vorgehensweise zur Darstellung der Grafiken gezeigt werden.

Nach der Auswahl der Multiplen Regression erscheint eine Auswahl der Einflussgrößen und der Zielgröße. Übernehmen Sie die gezeigten Titel nach rechts:

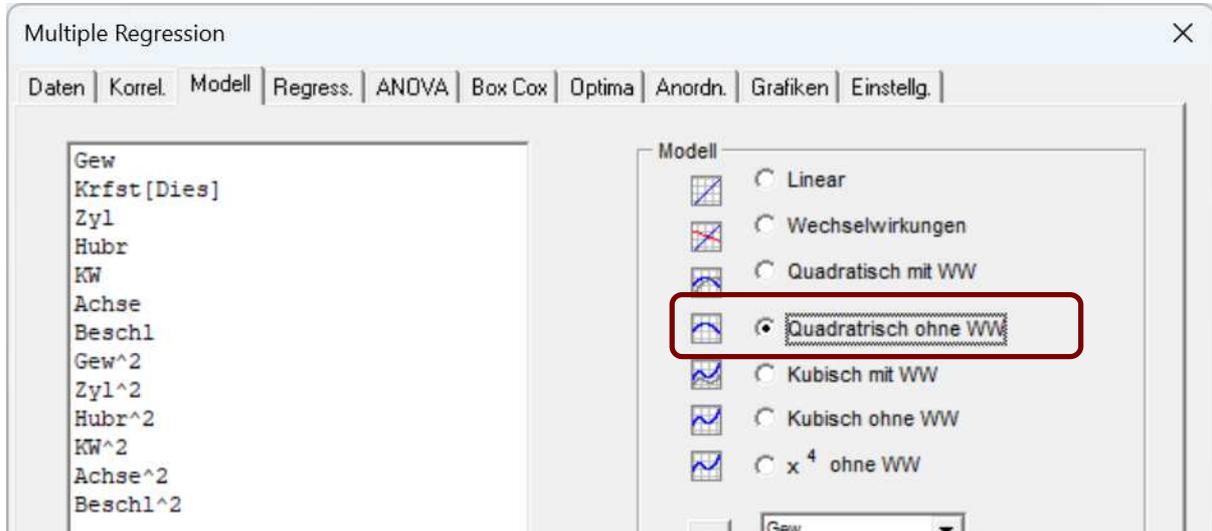


Im nächsten Reiter Korrelation wird überprüft, ob die Daten hinsichtlich ihrer Abhängigkeiten kritisch sind, was hier der Fall ist. Es erscheint eine Warnmeldung mit verschiedenen Optionen weiterzugehen:

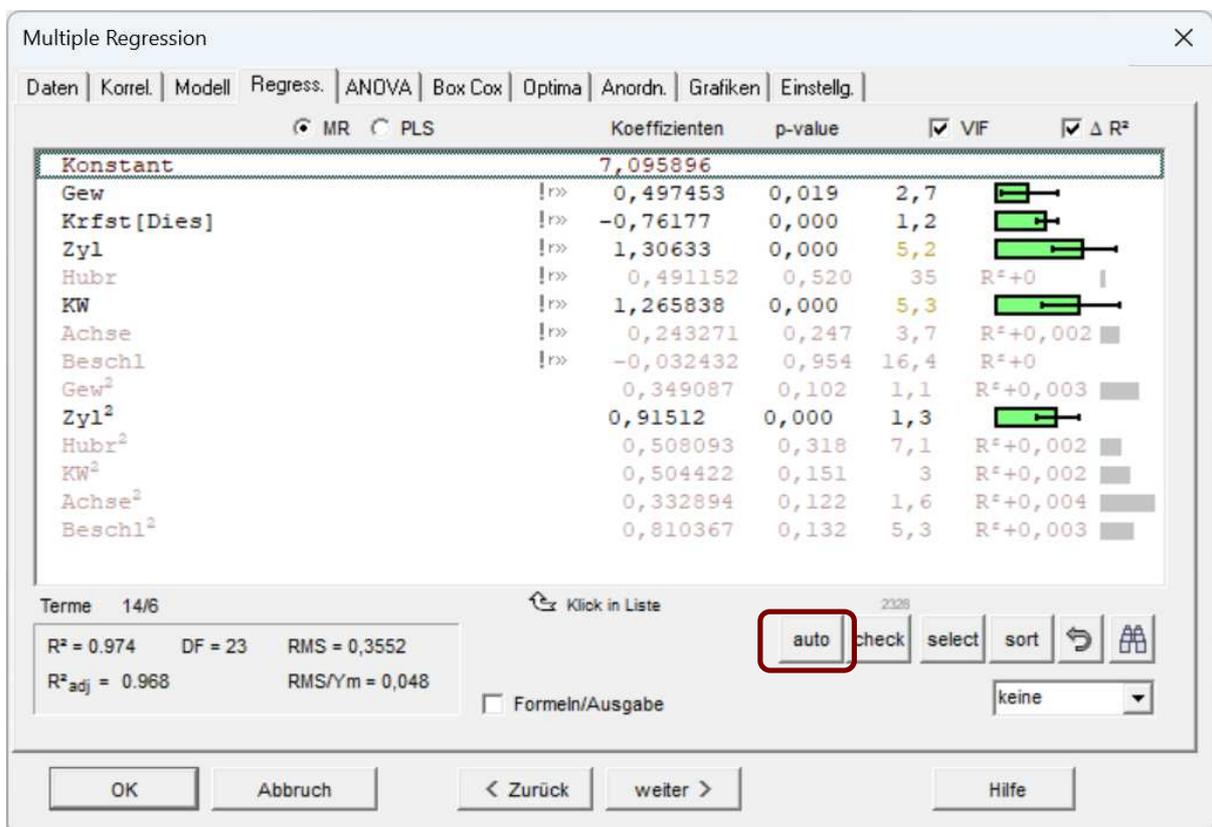


Bei der Auswahl der obersten Standard-Option werden nicht die gewünschten Einflussgrößen verwendet. Die Literatur empfiehlt hier nicht die Multiple Regression, stattdessen das sogenannte Partial-Least-Square Verfahren. An dieser Stelle soll trotzdem die häufig angewendete Multiple Regression genutzt werden. Visual-XSel bietet einige Funktionen, die verhindern, dass hier kritische Situationen entstehen.

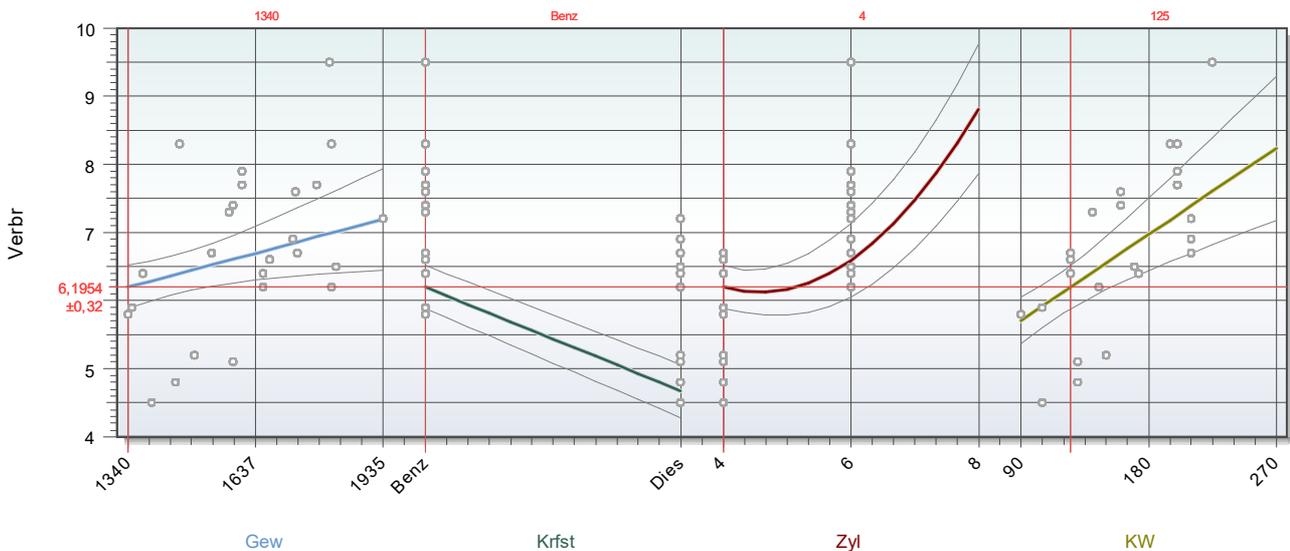
Beim Modell auf dem nächsten Reiter sollte *Quadratisch ohne Wechselwirkungen* gewählt werden, da diese bei hochkorrelierenden Daten auf jeden Fall nicht empfehlenswert sind.



Abschließen ist unter der Regression das Modell um die nicht signifikanten Terme über die Taste Auto zu „bereinigen“. Die Streuung dieser Terme ist größer als ihr Einfluss.



Nach Abschluss über die Taste OK erscheint das sogenannte Kurvendiagramm:



Der Effekt für das Gewicht auf den Verbrauch errechnet sich aus der Differenz des Verbrauches nach Einstellung der vertikalen roten Linie nach ganz links und ganz rechts.

Es ergibt sich ein Mehrverbrauch von

$$7,2 - 6,2 \text{ l/100km} = 1,0 \text{ l/100km}$$

für eine Gewichts Differenz von

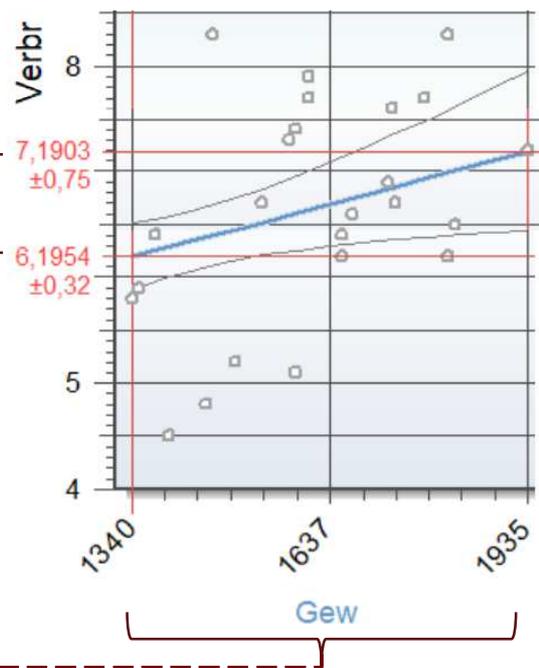
$$1935 \text{ kg} - 1340 \text{ kg} = 595 \text{ kg}$$

Im zweiten Diagramm auf Seite 2 kam ein Mehrverbrauch für Benziner von

$$10,9 \text{ l/100km} - 5,8 \text{ l/100km} = 5,1 \text{ l/100km}$$

heraus, ein eklatanter Unterschied.

Der Mehrverbrauch, ermittelt über die Multiple Regression stellt den wahren Mehrverbrauch, bzw. die wahre Steigung im Diagramm dar, denn nur hier werden die Effekte der anderen Einflussgrößen herausgerechnet. Das Ergebnis liegt hier im Bereich dessen, was man in der Literatur findet. Nur wenn die Daten ausgeglichene Einstellungen in Form eines Versuchsplanes haben, ergeben sich keine Unterschiede.



## Fazit

Dieses Beispiel zeigt eindrucksvoll, dass die Methode der Multiplen Regression bei mehreren Einflussgrößen der Darstellung von Einzelgrafiken vorzuziehen ist.



## Literatur - Weiterführende Beschreibungen

Ausführliche softwareunabhängige Beschreibungen zum Thema DoE und der dazugehörigen Auswertungen gibt es im

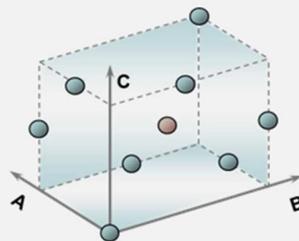
### Taschenbuch der statistischen Qualitäts- und Zuverlässigkeitsmethoden



#### Definitive Screening Designs DSD

Sogenannte Definitive Screening Designs sind sehr neu von Jones und Nachtshiem entwickelte Versuchspläne mit sehr geringem Versuchsumfang.

Sie ermöglichen die Auswertung von quadratischen Modellen und basieren deshalb auf 3 Stufen. Zwischen den Hauptfaktoren untereinander und den quadratischen Termen gibt es keine Vermengung (orthogonal). Die Wechselwirkungen sind nicht zu 100% vermengt.



Nr	A	B	C	D
1	0	1	-1	-1
2	0	-1	1	1
3	-1	0	-1	1
4	1	0	1	-1
5	-1	-1	0	-1
6	1	1	0	1
7	-1	1	1	0
8	1	-1	-1	0
9	0	0	0	0

In der generischen Erzeugung dieser Versuchspläne (iterativ mit Hilfe der Determinante) ergibt sich regulär die Anzahl Versuche mit  $n = 2^p + 2$ . Manche Pläne, z.B. für  $p=5$  sind dann allerdings teilweise zwischen den Hauptfaktoren vermengt. Hier müssen bis zu 3 Versuchszeilen ergänzt werden. Der Gesamtumfang ergibt sich somit zu:

$$n = 2^p + 2 + (1..3)$$

Alle Faktoren müssen durchgehend auf 3 Stufen sein und es lassen sich keine kategorialen Faktoren darstellen. Nachteilig ist auch, dass keine Auswertung aller möglichen

Weitere Informationen und Leseproben:

[crgraph.de/Literatur](http://crgraph.de/Literatur)



## Consulting & Schulungen

Bei unseren Inhouse- oder Online-Schulungen wird die praxisnahe Anwendung von statistischen Methoden vermittelt. Wir haben über 25 Jahre Erfahrung, insbesondere in der Automobilindustrie und unterstützen Sie bei Ihren Problemstellungen, führen Auswertungen für Sie durch, oder erstellen firmenspezifische Auswertevorlagen.



Weitere Informationen finden Sie unter:

[crgraph.de/schulungen](http://crgraph.de/schulungen)

Sie haben ein konkretes Qualitätsproblem, oder wollen ein Produkt effizient und zuverlässig entwickeln? Sie wollen keine Statistik-Software anschaffen, weil diese voraussichtlich zu selten gebraucht wird, oder weil zu wenig Zeit zur Einarbeitung vorhanden ist? Dann sind unsere Q-Support Pakete genau das Richtige:

[crgraph.de/consulting](http://crgraph.de/consulting)