



Voraussetzung und verwandte Themen

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

www.versuchsmethoden.de/Multiple_Regression.pdf

Einführung

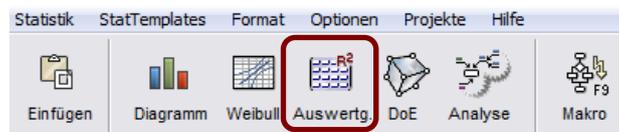
Bei der Auswertung von historischen oder nicht geplanten Daten kann es vorkommen, dass eine Regressionsauswertung Probleme macht.

Ziel und Nutzen

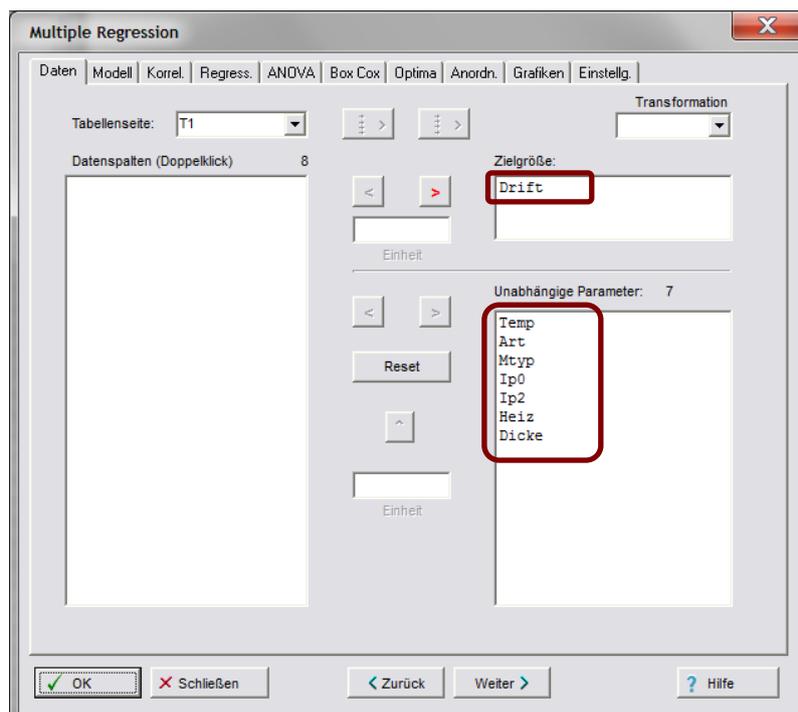
Das Ziel dieser Beschreibung ist es mögliche Auswerteschritte aufzuzeigen.

Beispiel

Im folgendem Beispiel eines Sensors wurde eine multiple Regression durchgeführt. Die Auswertung wird über die Ikone *Auswertung* gestartet.

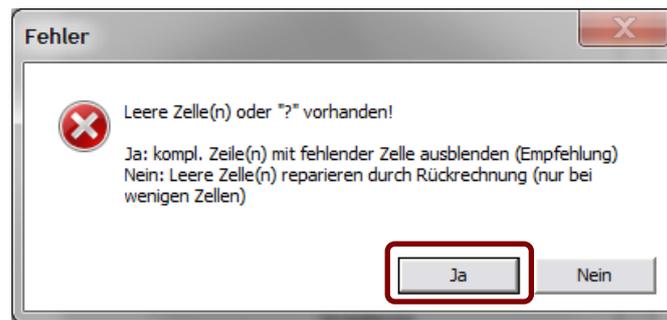


Die weitere Auswahl der Auswertung ist quantitativ – metrisch, sowie mehrere Parameter. Die verwendeten Parameter und die Zielgröße sind:



Auswertung kritischer Daten

Bei der Auswahl des Modells, bzw. bei Taste *Weiter* werden leere Felder erkannt:

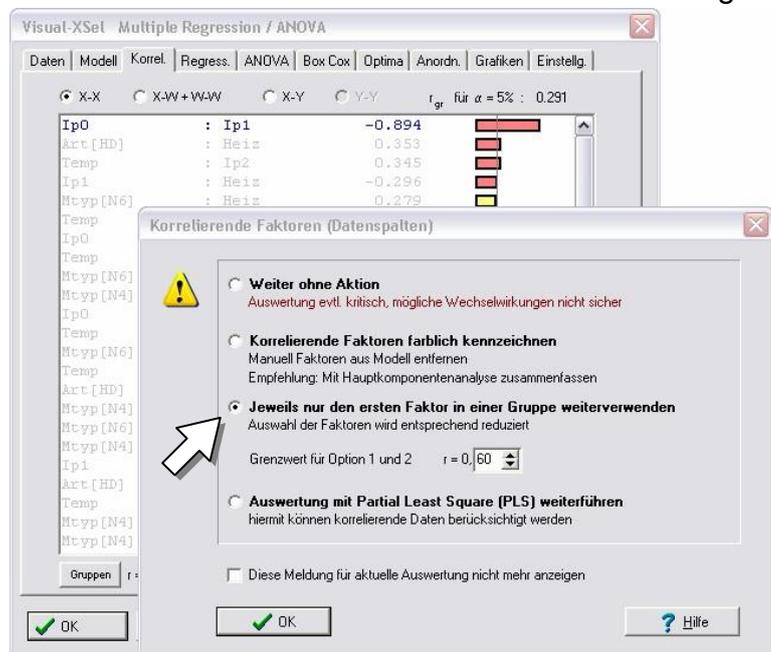


Bei zu vielen leeren Feldern oder unklaren Verhältnissen sind diese Zeilen komplett auszulassen. Dies ist mit *Ja* zu bestätigen. Nur dann, wenn ein oder maximal zwei Lücken bestehen, lassen sich diese optional „reparieren“. Bei leeren Feldern der Zielgröße werden diese automatisch ausgelassen.

Als erstes Modell, soweit sinnvoll, **Quadratisch mit WW** auswählen. Gibt es Einstellungen nur auf 2 Stufen, so wird dies später unter Regression angezeigt. Zunächst ist aber zu prüfen, ob die unabhängigen Parameter zu sehr korrelieren.

Korrelationen

In der Rubrik **Korrelation** wird bei kritischen Daten ein Dialogfenster mit möglichen Optionen geöffnet. Bei hoch korrelierenden Parametern -> **Jeweils nur den ersten Parameter in einer Gruppe weiterverwenden**. In diesem Beispiel korreliert *Ip0* mit *Ip1* zu stark. Rote Balken bedeuten, dass nach statistischem Test die Korrelation eindeutig wird. Dies gilt auch für Parameter *Art* und *Heiz*. Für die Regression sind allerdings Korrelationen erst über 0,6 von Belang. Nach Bestätigung der gewählten Option und Schließen des Dialogfensters, wird automatisch auf die Rubrik Daten zurückgesprungen. *Ip1* erscheint nun wieder in der linken Seite. Sollte stattdessen lieber *Ip0* herausgenommen werden, so ist im Dialogfenster die Option „*Weiter ohne Aktion*“ zu wählen und manuell unter Daten *Ip0* aus der Liste der unabhängigen Parameter gelöscht werden. Danach kann man direkt auf die Rubrik Regression gehen.



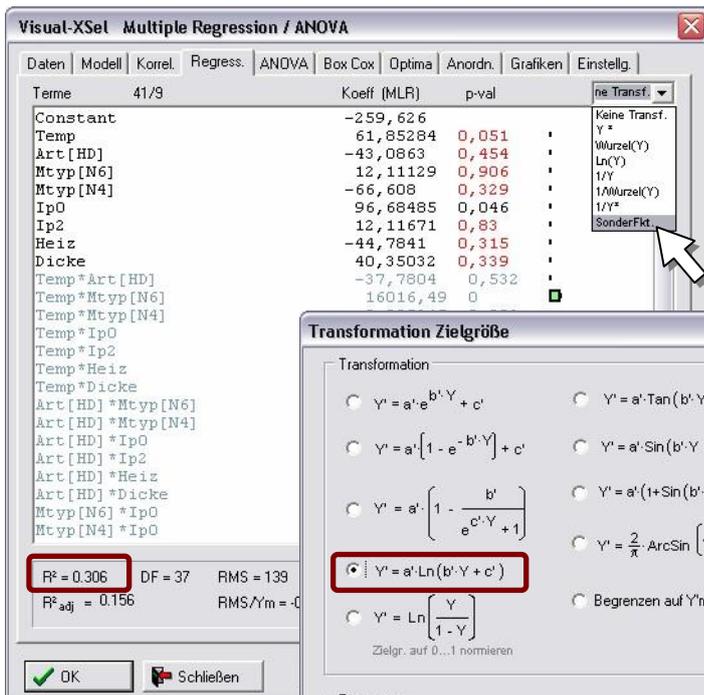
Hinweis: Sind alle Daten untereinander komplett korrelierend, ist keine Auswertung möglich. Ist die maximale Korrelation jedoch kleiner 1, so kann hier mit der Methode Partial Least Square die weitere Auswertung alternativ zur Multiplen Regression erfolgen.

Regression und Modellbildung

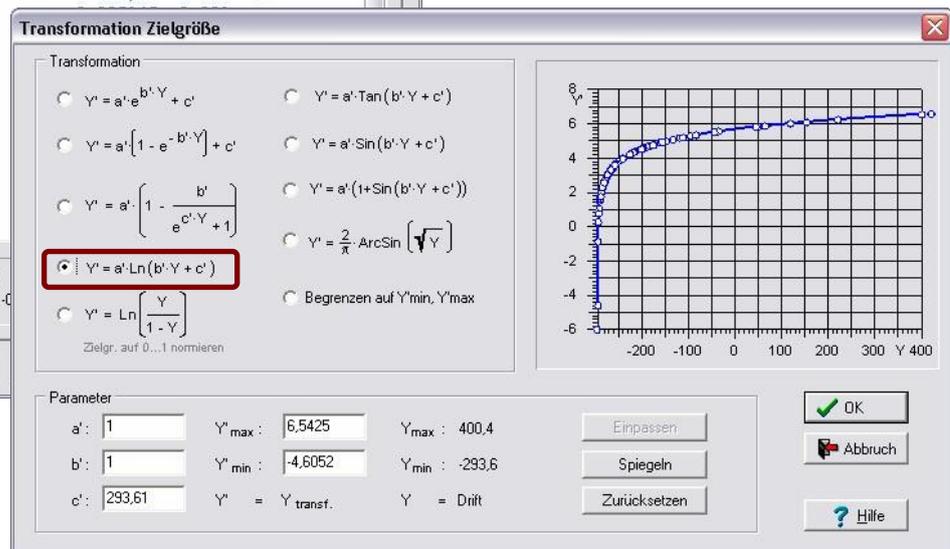
Beim ersten Durchlauf sind alle Terme (Faktoren, Wechselwirkungen, quadratische Terme usw.) im Modell. Es ist die Taste *Auto* zu wählen, um durch eine schrittweise Regression nicht signifikante Terme zu entfernen. Der Durchlauf ist hier nicht erfolgreich, *Auto stepwise* wurde gestoppt. Auch bei sehr vielen nicht signifikanten Termen und bei Termen mit „?“ (singulär), oder bei zu wenig Freiheitsgraden zunächst die eigentlichen Faktoren im Modell lassen, alle höherwertigen Terme manuell herausnehmen. Zunächst sollen nur die Hauptterme im Modell belassen werden (markieren Sie die höherwertigen Terme und entfernen diese aus dem Modell mit der Taste ).



Das ermittelte Modell ist dabei vom Erklärungsgrad noch sehr schlecht, das Bestimmtheitsmaß beträgt nur 0,306. Eine Verbesserung ist über eine Transformation und durch Herausnehmen von Ausreißern möglich (soweit vorhanden).



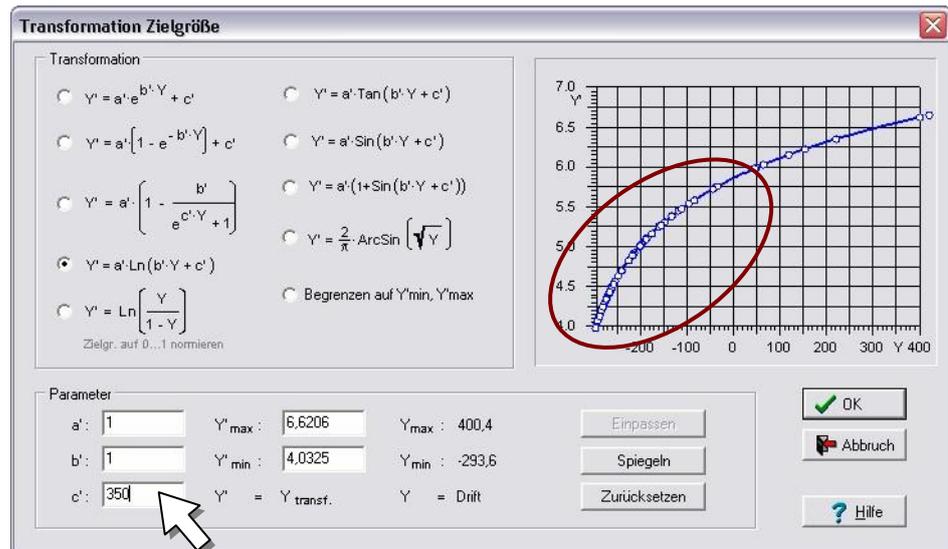
Das Überprüfen einer möglichst sinnvollen Transformation über Box-Cox ist wegen der negativen Werte der Zielgröße nicht möglich. Deshalb ist hier die manuelle Transformation über die *Sonderfunktion* zu wählen und hierin die dargestellte Funktion \ln



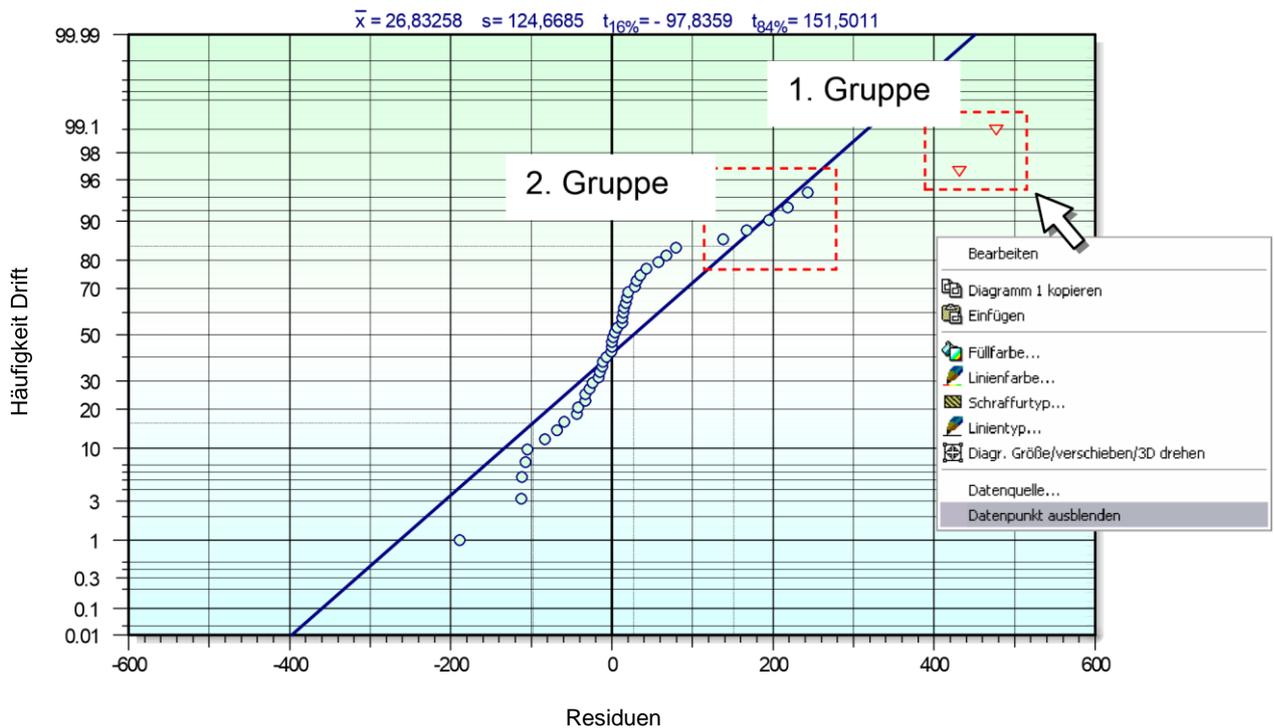
Der Offset c' sorgt dafür, dass im Argument des \ln keine negativen Werte auftreten. Der Vorgabewert ist gerade so groß, damit nur positive Argumente entstehen. Es empfiehlt sich hier einen etwas höheren Wert zu verwenden, um den Verlauf der Transformation zu verbessern. Hier wird vorgeschlagen $c' = 350$ zu wählen. Nach Bestätigung mit der Taste *Ok* ergibt sich ein Bestimmtheitsmaß von $R^2 = 0,402$, was eine deutliche Verbesserung ist, wenn auch noch nicht zufriedenstellend.

Auswertung kritischer Daten

Bei der hier relativ großen Streuung ist zu empfehlen für die folgende Modellbildung das Signifikanzniveau von 5% auf 10% zu vergrößern. Dies wird unter der Rubrik Einstellungen /Regression geändert.

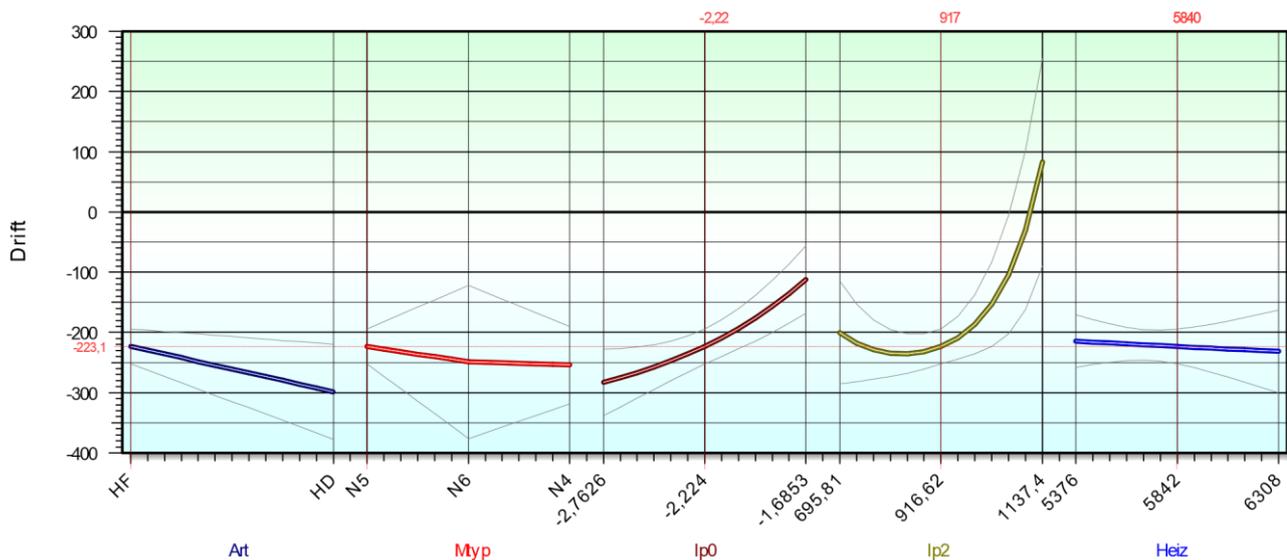


Nun ist erneut mit der Taste Auto unter der Rubrik Regression eine schrittweise Regression zu versuchen. Auch hier läuft diese jedoch nicht durch und das Modell ist nicht vollständig. Die nächsten Schritte behandeln nun die Ausreißer. Hierzu ist das Diagramm **Residuenverteilung** darzustellen. Wählen Sie dieses unter der Rubrik Diagramme aus und bestätigen mit **Ok**. Es entsteht folgendes Bild:



Es werden zwei Ausreißer in rot ganz rechts angezeigt. Ziehen Sie mit der Maus über diese Punkte (1. Gruppe) und entfernen diese mit dem letzten Menüpunkt (rechte Maustaste). Das Regressionsfenster wird automatisch geöffnet (ansonsten Ikone **Auswertung** wählen). Hier kann überprüft werden, ob sich das Modell geändert hat. Wählen Sie die Taste **Auto stepwise**. Auch hier wird der Durchgang gestoppt. Bei der 2. Gruppe von Ausreißern (nur der äußerste Punkt der 5 nicht zur Gesamtverteilung passenden Linie ist rot) stellt sich die Frage, war hierfür der Grund ist. Offensichtlich gibt es noch weitere Einflussparameter, die hier nicht betrachtet wurden. Evtl. lässt sich durch nachträgliche Überprüfung der Messungen feststellen, was hier noch für

„Änderungen“ stattfanden. Zur weiteren Verbesserung des Regressionsmodells sollen die Punkte der 2. Gruppe auch entfernt werden. Es ergeben sich immer wieder neue Ausreißer. Der Vorgang ist solange zu wiederholen, bis keine Ausreißer mehr gemeldet werden. Die Darstellung des Kurvendiagramms (Rubrik Grafiken) zeigt das letzte Modell:



Das Bestimmtheitsmaß von $R^2=0,57$ ist immer noch sehr schlecht. Etwa 43% der Streuung können nicht erklärt werden. Die hier gezeigte Vorgehensweise der schrittweisen Regression ist in der Fachwelt umstritten. Letztlich ist die Modellbildung im Zusammenhang mit dem Hintergrundwissen der Fachleute zu erstellen und nicht nur aus den statistischen Kennwerten.

Die Darstellung der Kurvenverläufe in diesem Beispiel kann als Trend die Wirkungen erkennen lassen. Auch die im Modell enthaltenen Wechselwirkungen zwischen $Ip0$ und $Ip2$, sowie zwischen $Ip2$ und $Heiz$ konnten technisch begründet werden. Häufig gibt es aber eine Menge von Wechselwirkungen, die nicht nachvollziehbar sind. Im Zweifelsfall sollten diese nicht berücksichtigt werden.

Was sind mögliche weitere Schritte:

- Überprüfung der Messwerte, die als Ausreißer erkannt wurden. Gibt es Ursachen in der **Messmethode** in der **Aufzeichnung** oder gibt es systematische Veränderungen eines **nicht beobachteten Parameters** (meist partielle Temperaturunterschiede an Bauteilen).
- Die Zusammenhänge des Modells durch die **Fachleute** begutachten lassen. Vergleich mit Erkenntnissen aus früheren Untersuchungen durchführen, falls vorhanden.
- **Wiederholungsversuche** bei den Extremwerten und bei fehlenden Eckpunkten unter Ansicht Anordnung
- Erstellung eines **D-Optimalen Versuchsplanes** unter Einbeziehung der bisherigen Daten (Inclusions), Messung der hinzugekommenen Einstellungen und erneute Auswertung.

Hinweis:

Je nach Reihenfolge und entsprechender Modellbildung kann es zu unterschiedlichen Ergebnissen kommen. Zum Vergleich kann die gleiche Auswertung zusätzlich über die Methode Partial Least Square durchgeführt werden. Dies ist, wie bereits beschrieben, bei höher korrelierenden Daten grundsätzlich zu empfehlen. Zu beachten ist hierbei jedoch, dass das R^2 in der Regel schlechter ist und die Effekte meist geringer ausfallen, als bei der Multiplen Regression.

Literatur

Taschenbuch der statistischen Qualitäts- und Zuverlässigkeitsmethoden

Die wichtigsten Methoden und Verfahren für die Praxis.

Beinhaltet statistische Methoden für Versuchsplanung & Datenanalyse, sowie Zuverlässigkeit & Weibull.

- Statistische Verteilungen und Tests & Mischverteilungen
- Six Sigma Einführung und Zyklen
- Systemanalysen Wirkdiagramm, FMEA, FTA,
- Matrizen-Methoden
- Shainin- und Taguchi-Methoden
- Versuchsplanung DoE, D-Optimal
- Korrelations- und Regressionsverfahren
- Multivariate Datenauswertungen
- Prozessfähigkeit – Messmittelfähigkeit
MSA 4 und VDA 5
- Regelkarten
- Toleranzrechnung und Monte-Carlo-Simulation
- Statistische Hypothesentests
- Weibull und Lebensdaueranalysen
- Stichprobengröße

190 Seiten, Ringbuch

ISBN: 978-3-00-043678-9

