



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

www.crgraph.de/MultipleRegression.pdf

www.crgraph.de/PLS.pdf

Einführung

Der Varianz-Inflations-Faktor (Variance Inflation Factor) beschreibt, welcher Grad der Korrelation r zwischen den Faktoren (Multikollinearität) in einer Regressionsanalyse vorliegt. Der Name Varianzinflationsfaktor resultiert daraus, dass sich mit zunehmender Multikollinearität die Varianzen der Regressionskoeffizienten um diesen Betrag vergrößern.

Ziel und Nutzen

Die Schätzungen der Faktoreinflüsse werden also mit zunehmender Multikollinearität immer schlechter, insbesondere wenn die Streuung der Daten groß ist. Die Regression wird dann instabil und schwierig zu interpretieren. Der *VIF* hilft dabei zu entscheiden, welche Regressions-Terme man im Modell lässt, oder besser nicht. Dies trifft insbesondere für Wechselwirkungen zu.

Grundlagen

Für die Definition des *VIF* geht von folgendem allgemeinen Regressionsmodell aus:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_k \cdot x_k$$

Für $i=1$ für den ersten Faktor wird die Regression auf x_1 bezogen

$$x_1 = b_0 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_k \cdot x_k$$

und hierfür $R^2 = \text{Korrelation}^2$ bestimmt. Somit ergibt sich allgemein für alle Faktoren:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Für die weiteren Faktoren ist die Vorgehensweise die gleiche, es wandert jeweils der Faktor auf die linke Seite, für den der *VIF* berechnet werden soll. Allgemein lässt sich hieraus folgende Interpretation treffen:

$VIF = 1$: $r = 0$, keine Korrelation vorhanden \Leftrightarrow z.B. orthogonaler Versuchsplan

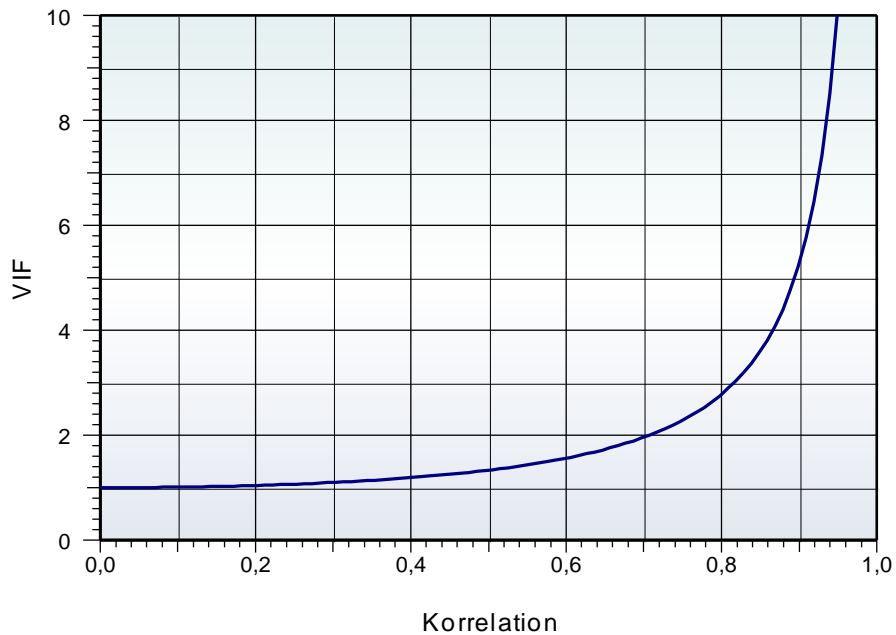
$1 < VIF < 5^*$: $0 < r < 0,9$, mäßige bis mittlere Korrelation, Auswertung möglich

$5 < VIF < 10$: $0,9 < r < 0,95$, hohe Korrelation, Auswertung kann nicht mehr eindeutig werden, insbesondere für Wechselwirkungen

$VIF > 10$: $r > 0,95$ Korrelation kritisch, Auswertung nicht mehr vertrauenswürdig.

* wird in einigen Literaturquellen bereits bei der Grenze von 4 als kritisch betrachtet

Den Zusammenhang zwischen *VIF* und der Korrelation zeigt folgende Grafik:



Je höher der *VIF*-Wert, desto größer werden die Vertrauensbereiche.

Bei zu hohen *VIF*-Werten ist der Ausweg, entweder diesen Faktor aus dem Modell zu entfernen, da der Zusammenhang durch die anderen Faktoren bereits beschrieben wird, oder das PLS-Verfahren anzuwenden.

Zu beachten ist, dass *VIF* nicht die paarweise Betrachtung der Korrelation jeweils zweier Faktoren ersetzt. Es kann sein, dass es nur eine hohe Korrelation zwischen 2 Faktoren in einem Modell gibt, obwohl *VIF* nicht kritisch erscheint.

Anwendung in Visual-XSel 14.0

Ab Version 14 gibt es optional die Ausgabe der *VIF*-Kennzahl. Weitere Informationen zeigt die Sprechblase, für jeden Term, wenn man in der Spalte *VIF* auf den jeweiligen Term klickt.

Partial Least Square						
Daten Modell Korrel. Regress. ANOVA Box Cox Optima Anordn. Grafiken Einstellg.						
Terme	14/14	MR	PLS	Koeff (PLS 5C)	VIP	VIF
Constant				6,981447		
Gew				0,306916	0,83	15,7
Krfst [Dies]				-0,55139	0,93	5,1
Zyl				0,474596	1,26	57,4
Hubr				0,426629	1,30	96,9
KW				0,28026	1,26	>100,0
Achse				0,31584	0,59	8,1

Die gesamte Datenauswertung ist beschrieben in:

[www.crgraph.de/
Visual XSel DoE 14.pdf](http://www.crgraph.de/Visual_XSel_DoE_14.pdf)

VIF
 Varianz Invlations Faktor
[Weitere Infos \(Hilfe-Taste\)](#)

i Der VIF ist ein Maß für Korrelation des Terms mit allen anderen und beschreibt die sogenannte Multikollinearität. (unter "Korrelation" sieht man nur die paarweisen)

🎓 Der aktuelle VIF=57,4 bedeutet, dass der Term mit $r=0,991$ extrem mit anderen korreliert, Streuungen verfälschen die Auswertung evtl. stark. Gibt es kritische Werte bei WW oder x^2 , so sind diese zuerst herauszunehmen.