

# Hochschule Karlsruhe - Technik und Wirtschaft Fakultät für Elektro- und Informationstechnik



Prof. Dr.-Ing. M. Strohrmann

Doolgii i oi oix oigiila Ei i iv	Design	For	Six	Sigma	<b>EITM</b>
----------------------------------	--------	-----	-----	-------	-------------

Klausur SS 2018

Name:	•
Matrikel-Nr.:	

#### Hinweise:

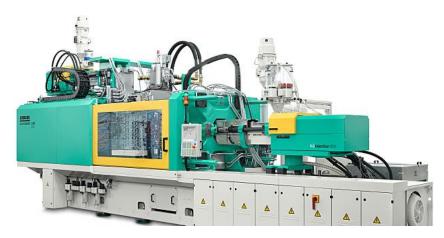
- Diese Klausur enthält 1 Aufgabe.
- Die Unterlagen bestehen aus 11 Seiten inkl. Deckblatt.
- Die Dauer der Klausur beträgt 180 Minuten.
- Erlaubte Hilfsmittel:
  - Vorlesungsskript, Vorlesungsfolien
  - Schriftliche Unterlagen
  - Zugewiesener Laborrechner
  - Taschenrechner
- Notieren Sie auf jedem Blatt Ihren Namen und Matrikelnummer.
- Stellen Sie Ihren Lösungsweg ausführlich dar und begründen Sie Ihr Vorgehen.
- Geben Sie dieses Deckblatt, die Aufgabenstellung und alle bearbeiteten Blätter mit Ihrer Klausur ab.

max. Punktzahl	Ist-Punktzahl
100	

Note	
------	--

## Beschreibung der Aufgabe

In der Industrie werden viele Kunststoffteile über Spritzgussprozesse hergestellt. Dabei wird Kunststoffgranulat aufgeschmolzen und über eine Spindel in eine Negativform gepresst. Während des Vorgangs werden unterschiedliche Drücke, Temperaturen und Zeiten eingestellt. Nach Abkühlen der Bauteile werden sie aus der Spritzgussform genommen.



Quelle: www.arburg.com

Es gehen unterschiedliche Eingangsgrößen in die Masse m des Spritzgussteils ein. Sie sind in folgender Tabelle zusammengefasst.

Eingangsgröße	Variable	Einheit	
Nachdruckzeit	t <sub>N</sub>	s	
Einspritzgeschwindigkeit	VE	mm/s	
Staudruck	рѕт	bar	
Werkzeugkühlmitteltemperatur	Тк	°C	
Zylindertemperaturen	T <sub>Z</sub>	°C	
Nachdruck	ри	bar	
Drehzahl	n	U/min	

In den folgenden Aufgaben wird der Zusammenhang zwischen der Masse des gefertigten Spritzgussteils und den aufgeführten Eingangsgrößen analysiert.

## Signifikanzbewertung von Eingangsgrößen

Um den Versuchsraum einzugrenzen, werden Voruntersuchungen durchgeführt. In Diskussionen mit den Fertigungsverantwortlichen wird die Hypothese aufgestellt, dass die Drehzahl n keinen signifikanten Einfluss auf die Masse des hergestellten Bauteils besitzt. Deshalb wird ein Versuch durchgeführt, bei dem die Masse m bei unterschiedlichen Drehzahlen bewertet wird. Die Daten sind im Datensatz *Signifikanz.mat* gespeichert.

n <sub>1</sub> = 75 U/min		n <sub>2</sub> = 125 U/min		
108.2232	108.2184	108.2209	108.2170	
108.2220	108.2180	108.2100	108.2149	
108.2139	108.2177	108.2128	108.2136	
108.2151	108.2103	108.2138	108.2054	

- a) Ist die Eingangsgröße Drehzahl n signifikant für die Masse des Spritzgussteils? Führen Sie mit den angegebenen Daten einen Hypothesentest auf gleichen Mittelwert durch. Wie lautet die Nullhypothese, wie die Alternative? Welche Zufallsvariable wählen Sie für den Hypothesentest?
- b) Bewerten Sie die Signifikanz der Eingangsgröße Drehzahl n auf Basis der angegebenen Daten mit Hilfe einer ANOVA-Tabelle. Welche Zufallsvariable liegt der ANOVA-Tabelle zugrunde?
- c) Führen Sie eine lineare Regression der Masse m als Funktion der Drehzahl n durch. Bewerten Sie die Signifikanz des linearen Terms. Welche Zufallsvariable liegt der Signifikanzbewertung zugrunde?
- d) Vergleichen Sie die Signifikanzbewertungen miteinander. Führen sie zu demselben Ergebnis? Begründen Sie Ihre Antwort.
- e) Plausibilisieren Sie das Ergebnis des Hypothesentests mit einer geeigneten grafischen Darstellung.

## Statistische Versuchsplanung und Regressionsrechnung

Über einen Versuch soll der Zusammenhang zwischen den verbleibenden Eingangsgrößen und der Masse des gefertigten Kunststoffteils ermittelt werden. Die Eingangsgrößen und ihre Einstellbereiche sind in der folgenden Tabelle zusammengefasst. Die Drehzahl wird auf den Wert n=100~U/min~eingestellt.

Eingangsgröße	Variable	Einheit	Minimum	Zentralpunkt	Maximum
Nachdruckzeit	t <sub>N</sub>	S	7	8	9
Einspritzgeschwindigkeit	VE	mm/s	70	80	90
Staudruck	рѕт	bar	8	10	12
Werkzeugkühlmitteltemperatur	Tĸ	°C	35	40	45
Zylindertemperaturen	Tz	°C	220	230	240
Nachdruck	ри	bar	50	55	60

Zur Ermittlung der Versuchsergebnisse wird die MATLAB-Funktion Masse.p verwendet. Sie ermittelt die Masse des gefertigten Spritzgussteils als Funktion an der Stelle X, die sich aus den Variablen  $X = [t_N \ v \ p_{ST} \ T_K \ T_Z \ p_N \ n]$  zusammensetzt. Alle Variablen können auch als Spaltenvektor eingesetzt werden.

- f) Stellen Sie einen teilfaktoriellen 2<sup>(K-x)</sup> Versuchsplan auf, der es erlaubt, alle Haupteffekte und Zweifach-Wechselwirkungen eindeutig zu bestimmen und eine minimale Anzahl von Versuchen aufweist. Überprüfen Sie für den vorliegenden Versuchsplan, ob alle Haupteffekte und Zweifach-Wechselwirkungen eindeutig bestimmt werden können.
- g) Bestimmen Sie die Versuchsergebnisse und speichern Sie das Versuchsergebnis ab. Verwenden Sie für die folgenden Aufgabenteile diesen Datensatz weiter.
- h) Stellen Sie eine Regressionsfunktion auf und reduzieren Sie die Regressionsfunktion auf die signifikanten Terme. Wie ändert sich das adjungierte Bestimmtheitsmaß?
- i) Geben Sie den Prognosebereich mit  $\gamma = 99$  % für den Zentralpunkt des Versuchsraums an, der sich auf Basis der Regressionsrechnung ergibt.
- j) Erzeugen Sie unabhängig von dem Versuchsplan für den Zentralpunkt 32 Versuchsergebnisse. Berechnen Sie aus den Versuchsergebnissen den Prognosebereich mit  $\gamma = 99 \%$ .
- k) Vergleichen Sie die beiden Prognosebereiche miteinander. Führen sie zu demselben Ergebnis? Begründen Sie Ihre Antwort.
- Für den Fall, dass die signifikanten Regressionskoeffizienten bekannt sind, lässt sich der Versuchsaufwand stark reduzieren, indem ein D-Optimaler Versuchsplan eingesetzt wird. Erstellen Sie einen entsprechenden Versuchsplan.

## **Statistische Tolerierung**

Es zeigt sich, dass Zylindertemperatur T<sub>Z</sub> einen dominanten Einfluss auf die Masse m des gefertigten Spritzgussteils hat. Deshalb soll der Einfluss genauer analysiert und ein vollquadratisches Modell bestimmt werden. Dazu wird folgender Datensatz verwendet, der als *Tolerierung.mat* gespeichert ist. Alle übrigen Größen werden auf ihrem Zentralpunkt gehalten.

T <sub>z</sub> / °C	220	230	240
m / g	108.0583	108.1943	108.2765
m / g	108.0560	108.1958	108.2733
m / g	108.0624	108.1878	108.2704
m / g	108.0589	108.1877	108.2656

- m) Stellen Sie eine vollquadratische Regressionsfunktion auf und prüfen Sie die einzelnen Terme auf Signifikanz.
- n) Bestimmen Sie die Kovarianzmatrix der signifikanten Regressionsparameter. Sind die Regressionsparameter voneinander unabhängig? Begründen Sie Ihre Antwort.

Für die in der folgenden Tabelle aufgeführte Toleranzverteilung wird für den Zentralpunkt eine statistische Tolerierung durchgeführt. Dabei wird die unter m) bestimmte Regressionsfunktion verwendet. Die Eingangsgröße  $T_Z$  weist die in der folgenden Tabelle ausgewiesenen Toleranzverteilung auf.

Variable	Einheit	Zentral-	Toleranz		Bemerkungen	
Variable	Limen	punkt	Minimum	Maximum		
Tz	°C	230	225	235	Normalverteilung, $ \text{Toleranz entspricht} \pm 3 \cdot \sigma $	

Auch die Regressionskoeffizienten streuen. Gehen Sie für die Toleranzrechnung davon aus, dass die geschätzte Kovarianzmatrix S der Kovarianz der Grundgesamtheit  $\Sigma$  entspricht.

- o) Linearisieren Sie die Maßkette und führen Sie eine statistische Tolerierung nach dem Grenzwertsatz der Wahrscheinlichkeit durch. Geben Sie die ± 3·σ Toleranzspanne der Masse m an.
- p) Wodurch wird die Varianz der Masse m in dieser Toleranzrechnung maßgeblich bestimmt, die Streuung der Regressionskoeffizienten oder die Streuung der Zylindertemperatur?

#### Musterlösung

a) Der t-Test zum Vergleich zweier normalverteilter Stichproben mit identischem Stichprobenumfang N auf gleichen Mittelwert wird mit der Zufallsvariable

$$t = \frac{\overline{x} - \mu}{\sqrt{\frac{2}{N}} \cdot s}$$

durchgeführt. Die Nullhypothese H0 lautet  $\mu = 0$ , damit ergibt sich

$$t = \frac{\overline{x}}{\sqrt{\frac{2}{N}} \cdot s}$$

Die Alternative lautet  $\mu \neq 0$ . Der Verwerfungsbereich ist beidseitig. Praktische Durchführung mit der MATLAB-Funktion ttest2. Der p-Value beträgt PT = 0.1088. Er ist größer als 5 %, damit ist die Drehzahl für die Masse des gefertigten Produktes nicht signifikant.

```
% Signifikanzprüfung der Drehzahl
load Signifikanz.mat;

% Bewertung über t-Test zeigt, dass die Drehzahl nicht signifikant
% ist, p-value muss kleiner als 5 % sein, PT = 0.1088
[H,PT] = ttest2(m1,m2)
```

b) Der Varianzanalyse liegt die Grundidee zugrunde, die Varianz von Gruppe zu Gruppe in Relation zur Varianz innerhalb einer Gruppe zu bewerten. Das Verhältnis der Varianzen wird mit der f-verteilten Zufallsvariable

$$v_0 = \frac{s_{\alpha}^2}{s_{\epsilon}^2} = \frac{\frac{q_{\alpha}}{J-1}}{\frac{q_{\epsilon}}{J \cdot (N-1)}}$$

beschrieben. Der entsprechende Hypothesentest lautet, dass die Werte  $\alpha_n$  =0 sind und dass die Streuung  $s_{\alpha}{}^2$  = 0 ist. Der p-Value beträgt PA = 0.1088. Er ist größer als 5 %, damit ist die Drehzahl für die Masse des gefertigten Produktes nicht signifikant. Praktische Durchführung mit der MATLAB-Funktion anova1

```
% Bewertung über Anova-Tabelle zeigt, dass die Drehzahl nicht signifikant
% ist, p-value muss kleiner als 5 % sein, PA = 0.1088
[PA] = anova1([m1 m2])
```

c) Der Signifikanzanalyse bei den Termen einer linearen Regression liegt die t-verteilte Zufallsvariable

$$t = \frac{b_1 - \beta_1}{s_{b1}}$$

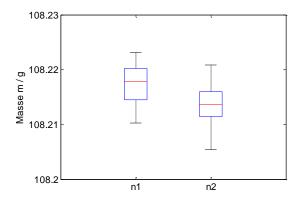
zugrunde. Die Nullhypothese ist, dass  $\beta_1 = 0$  ist.

$$t = \frac{b_1}{s_{b1}}$$

Die Alternative ist  $\beta_1 \neq 0$ . Der Verwerfungsbereich ist beidseitig. Praktische Durchführung mit der MATLAB-Funktion regstats.

```
% Signifikanzbewertung bei linearer Regression zeigt, dass die Drehzahl nicht
signifikant ist, p-value muss kleiner als 5 % sein, PR = 0.1088
n1 = [75 75 75 75 75 75 75]';
n2 = [125 125 125 125 125 125 125]';
stats = regstats([m1;m2],[n1;n2],[0;1],'tstat');
PR = stats.tstat.pval(2)
```

- d) Alle Signifikanzbewertungen führen zu derselben Aussage, nämlich, dass sich die Drehzahl nicht signifikant auf die Masse des produzierten Teils auswirkt. Interessant ist, dass alle p-Werte denselben Wert aufweisen.
- e) Plausibilisierung des Ergebnisses des Hypothesentests mit einem Boxplot zeigt, dass sich die Inter-Quartil-Ranges überlappen, die Größe Drehzahl ist damit nicht signifikant.



f) Es liegen K = 6 Eingangsgrößen vor. Um alle Haupteffekte und Zweifach-Wechselwirkungen eindeutig bestimmen zu können, muss der Versuchsplan eine Auflösung von V aufweisen. Es müssen deshalb  $N = 2^{6-1} = 2^5 = 32$  Versuche durchgeführt werden.

Es wird ein lineares Modell mit Wechselwirkungen aufgestellt. Das Modell ist Ausgangspunkt für die Erzeugung der entsprechenden Generatoren. Mit ihnen wird der teilfaktorielle Versuchsplan aufgestellt. Zur Kontrolle werden die Confoundings angezeigt. Dabei zeigt sich, dass keine unzulässigen Vermischungen existieren.

g) Für den normierten Versuchsplan werden die physikalischen Eingangsgrößen bestimmt und die Masse mit Hilfe der MATLAB-Funktion Masse(X) berechnet.

```
% Berechnen der Einstellwerte für die Versuche und Bestimmen der
% Versuchsergebnisse
X = zeros(32,7);
X(:,1) = 8 + doe(:,1);
X(:,2) = 80 + 10*doe(:,2);
X(:,3) = 10 + 2*doe(:,3);
X(:,4) = 40 + 5*doe(:,4);
X(:,5) = 230 + 10*doe(:,5);
X(:,6) = 55 + 5*doe(:,6);
X(:,7) = 100;
m1 = Masse(X);
save Versuchsplan.mat m3
```

h) Zur Reduktion der Regressionsfunktion auf signifikante Terme wird der p-Wert der t-Statistik verwendet. Es werden nacheinander die Terme mit dem größten p-Wert entfernt, bis alle p-Werte kleiner als 5 % sind. Das adjungierte Bestimmtheitsmaß ist dabei sehr nahe an 1, es ändert sich nicht wesentlich.

i) Der Prognosebereich wird gemäß Herleitung berechnet. Aufwendig ist die Berechnung des Arbeitspunktes  $\underline{x}_0$  und die Matrix X mit allen erforderlichen Termen. Dazu wird die Funktion x2fx verwendet. Der Prognosebereich hat mit dem erzeugten Datensatz die untere Grenze  $m_{Min} = 108.1714$  g und die obere Grenze  $m_{Max} = 108.2048$  g. Diese Werte hängen von dem zugrundeliegenden Datensatz ab.

```
% Bestimmung des Prognosebereichs der Regressionsfunktion
x0 = [8 80 10 40 230 55];
X0 = x2fx(x0,model);
XM = x2fx(X(:,1:6),model);
r = stats.r;
b = stats.beta;
FG = length(r)-length(b);
gamma = 0.99;
mMin = b'*X0' + tinv((1-gamma)/2,FG)*sqrt(r'*r/FG)*sqrt(1+X0*inv(XM'*XM)*X0')
mMax = b'*X0' + tinv((1+gamma)/2,FG)*sqrt(r'*r/FG)*sqrt(1+X0*inv(XM'*XM)*X0')
```

j) Mit der Funktion Masse werden 32 Versuchspunkte generiert. Aus diesen Werte wird der Prognosebereich bestimmt. Weder der Mittelwert, noch die Standardabweichung sind bekannt, deshalb gilt:

$$\overline{x} + c_1 \cdot s \cdot \sqrt{1 + \frac{1}{N}} < x \leq \overline{x} + c_2 \cdot s \cdot \sqrt{1 + \frac{1}{N}}$$

Die Parameter  $c_1$  und  $c_2$  werden aus der inversen t-Verteilung mit N - 1 Freiheitsgradenbestimmt. Der Prognosebereich hat mit dem erzeugten Datensatz die untere Grenze  $m_{Min} = 108.1782$  g und die obere Grenze  $m_{Max} = 108.2071$  g. Diese Werte hängen von dem zugrundeliegenden Datensatz ab.

```
% Bestimmung der Versuchsergebnisse für den Zentralpunkt
% Berechnung Prognosebereich für unbekannten Mittelwert und unbekannte Varianz
X = ones(32,1)*[x0 100];
% m4 = Masse(X);
% save Zentralpunkt.mat m4
load Zentralpunkt.mat;
mquer = mean(m4);
s = std(m4);
mMinPro = mquer + tinv(0.005,32-1)*s*sqrt(1+1/32)
mMaxPro = mquer + tinv(0.995,32-1)*s*sqrt(1+1/32)
```

k) Der Prognosebereich ist bei der Regressionsfunktion mit 0.0334 g größer als bei dem univariaten Fall mit 0.0289 g. Ursache sind die unterschiedlichen Modell die der Prognose zugrunde liegen. Bei dem Prognosebereich der Regressionsfunktion wird die Unsicherheit des Modells

```
\begin{split} m_1 &= b_0 + b_1 \cdot t_N + b_2 \cdot v_E + b_3 \cdot p_{ST} + b_4 \cdot T_K + b_5 \cdot T_Z + b_6 \cdot p_N \\ &+ b_7 \cdot t_N \cdot p_{ST} + b_8 \cdot t_N \cdot T_K + b_9 \cdot t_N \cdot T_Z + b_{10} \cdot t_N \cdot p_N \\ &+ b_{11} \cdot v_E \cdot T_K + b_{12} \cdot v_E \cdot T_Z + b_{13} \cdot T_K \cdot T_Z + b_{14} \cdot T_K \cdot p_N + b_{14} \cdot T_Z \cdot p_N \\ &+ \epsilon \end{split}
```

bewertet, während bei der wiederholten Messung im Zentralpunkt nur die Unsicherheit der Messung  $\epsilon$  relevant ist:

```
m_2 = m_{Z0} + \epsilon
```

l) Es müssen 16 Regressionskoeffizienten bestimmt werden, deshalb werden 24 Versuche durchgeführt. Mit dem bekannten Modell ergibt sich der D-Optimale Versuchsplan.

```
% D-Optimaler Versuchsplan für das bestimmte Modell mit 16 Regressionstermen
nfactors = 6;
nruns = 16*1.5;
dCE = cordexch(nfactors,nruns,model,'tries',10);
```

m) Wenn ausschließlich die Größe  $T_Z$  einen Einfluss hat und ein vollquadratischer Ansatz verwendet werden soll, lautet die Regressionsfunktion

$$m = b_0 + b_1 \cdot T_Z + b_2 \cdot T_Z^2$$

Mit den angegebenen Daten wird eine Regressionsfunktion bestimmt. Die Regressionsanalyse zeigt, dass alle Terme signifikant sind. Das Bestimmtheitsmaß ist mit 99.81 % sehr groß.

```
% Berechnung quadratisches Regressionsmodell, alle Terme sind signifikant,
% Bestimmtheitsmaß ist mit 0.9989 sehr groß
load Tolerierung.mat;
model = [0; 1; 2];
stats = regstats(m5,TZ,model);
pval = stats.tstat.pval
Radj = stats.adjrsquare
```

n) Die Kovarianzmatrix der Regressionskoeffizienten wird nach den Herleitungen im Skript abgeschätzt über

$$\boldsymbol{\Sigma} \approx \boldsymbol{S} = \boldsymbol{C} \cdot \boldsymbol{C}^T \cdot \boldsymbol{s}^2 = \left(\boldsymbol{X}^T \cdot \boldsymbol{X}\right)^{\!-1} \cdot \frac{1}{N\!-\!M\!-\!1} \cdot \left(\underline{\boldsymbol{r}}^T \cdot \underline{\boldsymbol{r}}\right)$$

Die Kovarianzmatrix besitzt außer den Elementen auf der Hauptdiagonalen weitere Elemente, die ungleich null sind. Die einzelnen Regressionskoeffizienten sind also nicht unkorreliert.

$$\boldsymbol{\Sigma} \approx \boldsymbol{S} = \left(\boldsymbol{X}^T \cdot \boldsymbol{X}\right)^{-1} \cdot \frac{1}{N - M - 1} \cdot \left(\underline{r}^T \cdot \underline{r}\right) = \begin{pmatrix} 12.3536 & -0.6451 & 0.0080 \\ -0.6451 & 0.0338 & -0.0004 \\ 0.0080 & -0.0004 & 0.0053 \cdot 10^{-3} \end{pmatrix} \cdot 10^{-3}$$

```
% Berechnung der Kovarianzmatrix nach Skript
b = stats.beta;
r = stats.r;
FG = length(r)-length(b);
X = [ones(size(TZ)) TZ TZ.^2];
kovarbeta = inv(X'*X)*(r'*r/FG)
```

o) Linearisierung der Regressionsgleichung führt unter Berücksichtigung der streuenden Regressionskoeffizienten zu

$$\Delta m = \Delta b_0 + T_{Z0} \cdot \Delta b_1 + b_{10} \cdot \Delta T_Z + T_{Z0}^2 \cdot \Delta b_2 + 2 \cdot b_{20} \cdot T_{Z0} \cdot \Delta T_Z$$

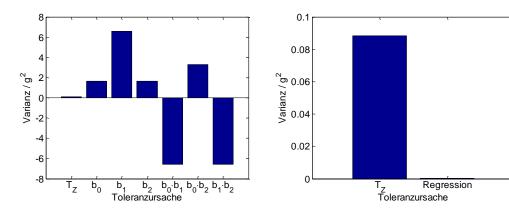
Statistische Tolerierung unter Berücksichtigung der geschätzten Kovarianz

$$\begin{split} \sigma_{m}^{2} &= \sigma_{b0}^{2} + T_{Z0}^{2} \cdot \sigma_{b1}^{2} + b_{10}^{2} \cdot \sigma_{TZ}^{2} + T_{Z0}^{4} \cdot \sigma_{b2}^{2} + 4 \cdot b_{20}^{2} \cdot T_{Z0}^{2} \cdot \sigma_{TZ}^{2} \\ &+ 2 \cdot 1 \cdot T_{Z0} \cdot \sigma_{b0b1} + 2 \cdot 1 \cdot T_{Z0}^{2} \cdot \sigma_{b0b2} + 2 \cdot T_{Z0} \cdot T_{Z0}^{2} \cdot \sigma_{b1b2} \end{split}$$

Die Toleranz T ergibt sich aus der 6-fachen Standardabweichung zu T = 1.7819 g.

```
% Statistische Tolerierung bei streuenden Regressionskoeffizienten
% Standardabweichungen Varianzen der Eingangsgrößen
varTZ = (10/6)^2;
varb0 = kovarbeta(1,1);
varb1 = kovarbeta(2,2);
varb2 = kovarbeta(3,3);
varb0b1 = kovarbeta(2,1);
varb0b2 = kovarbeta(3,1);
varb1b2 = kovarbeta(3,2);
% Definition Arbeitspunkt
TZ0 = 230;
% Umrechnung auf Ausgangsgröße nach Linearisierung
varmTZ = (b(2)^2+4*b(3)^2*TZ0^2)*varTZ;
varmb0 = 1*varb0;
varmb1 = TZ0^2*varb1;
varmb2 = TZ0^4*varb2;
varmb0b1 = 2*TZ0*varb0b1;
varmb0b2 = 2*TZ0^2*varb0b2;
varmb1b2 = 2*TZ0^3*varb1b2;
varm = (varmTZ+varmb0+varmb1+varmb2+varmb0b1+varmb0b2+varmb1b2);
varmReg = (varmb0+varmb1+varmb2+varmb0b1+varmb0b2+varmb1b2);
Tol = 6*sqrt(varm);
```

p) Die Anteile der einzelnen Varianzen werden grafisch dargestellt. Dabei werden zum einen die Auswirkungen aller Teilvarianzen und zum anderen die Toleranzursachen der Regressionsrechnung zusammengefasst.



Die Darstellung zeigt, dass bei der Regressionsrechnung eine Kompensation von Toleranzanteilen entsteht, weil die Kovarianz der Regressionskoeffizienten von null verschieden sind. Die Kovarianz von  $b_0$  und  $b_1$  sowie  $b_1$  und  $b_2$  sind negativ und reduzieren damit die Toleranzen, die von der Regressionsrechnung hervorgerufen werden.

Deshalb ist die Streuung der Zylindertemperatur T<sub>Z</sub> in der Toleranzrechnung gegenüber der Toleranz. die durch die Regression entsteht, dominant.