

An Approach to Process Improvement

Multiple Linear Regression

Dr. rer. nat. Thomas Pfeilsticker, TEQ® Training & Consulting GmbH

This essay explains how to create and evaluate a regression study based on process data. The aim is to find an empirical model $y=f(x_1, x_2, x_3\dots)$ for our process data explaining their impact on the response.

The injection molding process of a thermoplastic produced components causing problems during the assembly due to high shrinkage. In order to find out about any method to reduce the percentage shrinkage, data about the **injection temperature, injection speed and holding pressure** were collected. The general model equation leads to **percentage shrinkage = f (injection temperature, injection speed, holding pressure)**.

Which one of the three influencing factors has the major impact on the percentage shrinkage of the injection molding process? And how do you have to adjust the influencing factors in order to minimize the percentage shrinkage of the injection molded parts as far as possible?

The following reliable data about the process have been gained:

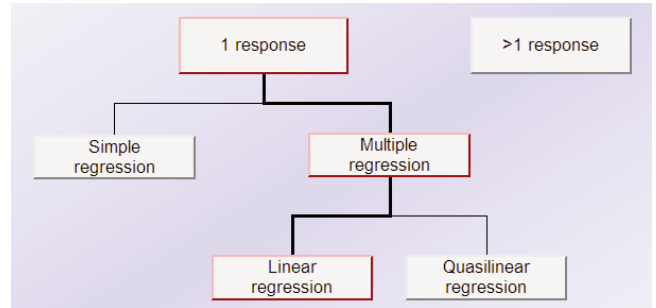
	injection temperature	injection speed	holding pressure	percentage shrinkage
1	295.0	205.2	37.31	0.87
2	296.8	205.2	35.48	0.93
3	293.7	204.3	36.01	0.87
4	295.4	205.5	37.51	0.92
5	293.2	204.4	37.49	0.82
6	295.9	205.2	37.21	0.89
7	294.7	205.1	36.64	0.89
8	296.3	205.9	37.21	0.90
9	294.2	204.2	37.21	0.83
10	296.1	204.9	35.87	0.87
11	295.0	205.8	34.52	0.86
12	294.9	204.9	36.66	0.91
13	293.8	203.3	37.23	0.84
14	295.3	205.3	37.36	0.88
15	294.4	205.7	36.45	0.86

The average percentage shrinkage amounts to 0,886%.

1 Selection of Regression Method – Multiple Linear Regression and Model Structure

The software *destra*® supports users in creating and evaluating process data, field data and experimental data obtained through systematic planning. *destra*® helps to answer questions about the main factors affecting the process (their absolute value) and their effective direction (their algebraic sign), about interactions and about how to adjust factors optimally.

The collected process data are inspected in a multiple linear regression analysis.



There are two different approaches available. You may select and compare the multiple linear regression and the multiple quasilinear regression.

Multiple linear regression								
Selection		Data						
Char.No	Characteristic	n	No. Lev	Input qb	Response	Constr.	E lo	E up
1	injection temperature	15	14	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	293.2	296.8
2	injection speed	15	12	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	203.3	205.9
3	holding pressure	15	15	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	34.52	37.51
4	percentage shrinkage	15	12	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		

Click the “next” icon to start the evaluation.

2 Evaluation of a Multiple Linear Regression

Correlation coefficient	=	r	0.8814
Coefficient of determination	=	R ²	77.683%
Adjusted coefficient of determination	=	R ² *	71.597%
Residual variance	=	s ²	0.000421
Residual standard deviation	=	s	0.0205
y = 0.895815 + 0.0431122*x1 + 0.00647798*x2 - 0.0259212*x3			

The evaluation result provides an overview of important statistics describing the quality of the regression model.

The **coefficient of determination (R²)** = 77,68% is quite high. (R²) indicates the degree to which the variance of the response (percentage shrinkage) can be explained by the three influencing factors or how strongly the three selected influencing factors affect the variation of the shrinkage values.

The **adjusted coefficient of determination (R²***) is a fair instrument to compare alternative models, e.g. the linear and the quadratic model. It considers the number of regression parameters and the number of measured values.

The variation of the percentage shrinkage that cannot be explained by the selected influencing factors is given in the **residual variance (s²)** or the **residual standard deviation (s)**.

3 Estimation of Coefficients (Estimation of Parameters)

In order to calculate the regression model you estimate the minimum sum of square deviations (squared residuals). There is no regression model including a smaller **sum of square deviations**.

Char Descr	xi	bi	sb _i	t _{stat}	1% [*]	5%	P	VIF
percentage shrinkage	Const.	-6.565	1.872	3.507**			0.00491	---
injection temperature	X1	0.0240	0.00709	3.370**			0.00618	1.752
injection speed	X2	0.0050	0.0101	0.493			0.632	1.774
holding pressure	X3	-0.0172	0.00059	2.928*			0.0234	1.979

$$y = -6.56489 + 0.0239512 \times X1 + 0.00498306 \times X2 - 0.0173386 \times X3$$

After describing the responses and factors, you specify the **regression coefficients (bi)** and their **confidence intervals** (95%, confidence level) **bi** (...). In order to evaluate the regression coefficients the software provides the respective **standard deviations sb_i (bi)** and the **t-test statistic** calculated by dividing the regression coefficient by its standard deviation. In addition, there are three options available to evaluate the significance. The program highlights the t-test statistic with a superscript star, offers a bar chart with the three usual significance levels marked by red lines and provides the **P-value** as an alternative.

In case the $\alpha = 5\%$ limit is exceeded, the respective influencing factor is considered to be significant and the corresponding bar is displayed in yellow. If even the $\alpha = 1\%$ limit is exceeded, the corresponding bar exceeds the second red line, the software will show a red bar. The P-value must be less than 5% (0,05). This value indicates the residual risk to detect a significant value that is actually insignificant. Injection temperature and holding pressure prove to be significant but the injection speed is not.

The **VIF values** (variance inflation factor) show whether there is interdependence between the single influencing factors.

Regression models are often reduced, i.e. they are adjusted for insignificant influencing factors and their interactions accordingly. In this case the software recalculates the model and evaluates it again.

Injection temperature and **holding pressure** have a particularly **strong impact** on the percentage shrinkage. The reduced model shows this fact even more clearly.

Char Descr	xi	bi	sb _i	t _{stat}	1% [*]	5%	P	VIF
percentage shrinkage	Const.	-6.176	1.843	3.359**			0.00273	---
injection temperature	X1	0.0262	0.00533	4.902**			0.000364	1.067
holding pressure	X3	-0.0178	0.00632	2.818*			0.0195	1.067

$$y = -6.17574 + 0.0261521 \times X1 - 0.0178006 \times X3$$

By raising the injection temperature by 1°C the percentage shrinkage increases by about 0,026% and raising the holding pressure by 1MPa means reducing the percentage shrinkage by about 0,018%.

4 Model Evaluation

4.1 Model Significance

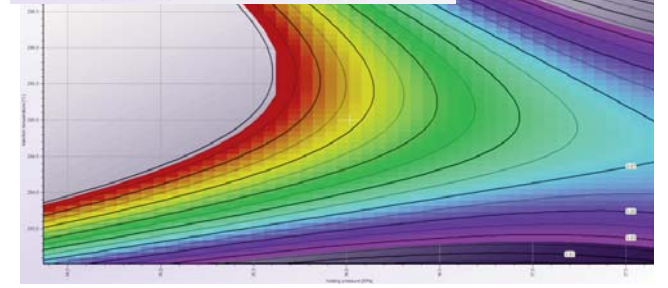
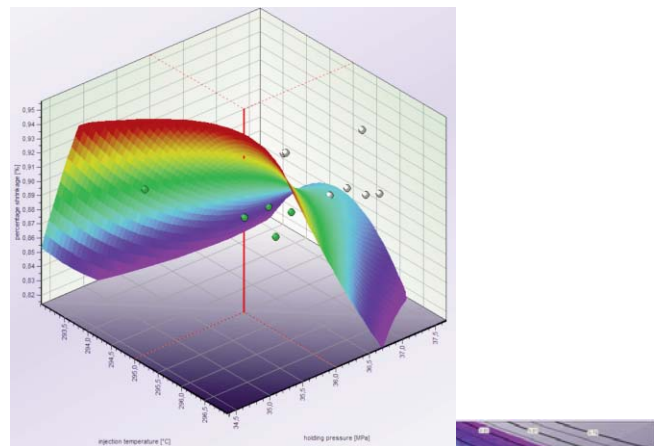
The F-test analyzes whether the influencing factors (coefficients) have a joint significant impact on the response and do not affect it individually as explained in section 3 about the t-test of coefficients. In case the null hypothesis is rejected (red), at least one factor is significant.

Test name	Stat	Critical values	Result	Test status
F=4.7%	---	---	3.09	---
F=1.1%	---	---	6.02	---
F=0.1%	---	---	11.56	---

4.2 Model Fit – Linearity F-test (LoF)

In order to evaluate a model, you use a linearity F-test based on the analysis of variance in addition to the coefficient of determination and the significance t-test. The linearity F-test is also referred to as lack-of-fit test (LoF) and helps to analyze the fit of the model. However, the test requires several observations of identical process adjustments and thus it is particularly important for planned experimental data.

5 Overview through Response Surface Plot and Contour Plot



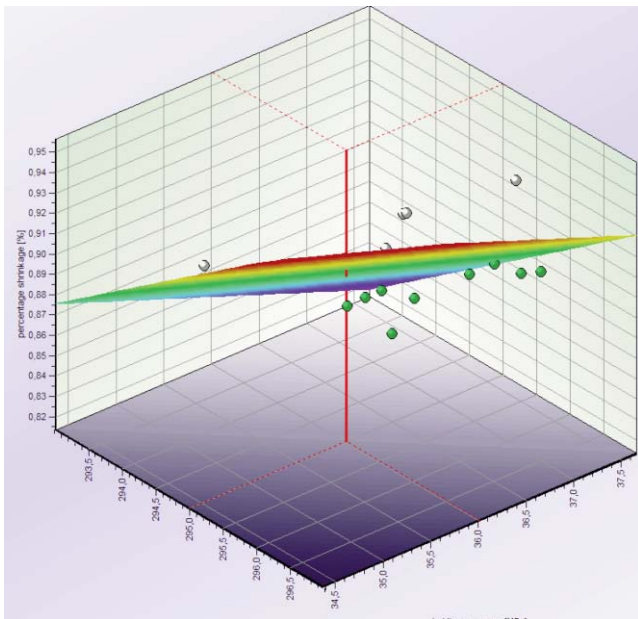
$$y = -485.085 + 9.23625 \times X1 - 7.90165 \times X2 - 3.76411 \times X3 - 0.022722 \times X1^2 +$$

$$0.0242253 \times X1 \times X2 - 0.0212555 \times X1 \times X3 - 0.00212439 \times X2^2 + 0.0444841 \times X2 \times X3 + 0.0121351 \times X3^2$$

The graphics show the impact of two influencing factors on the response in case the conditions of the third factor remain constant.

One of these graphics shows the 3-dimensional display of the variable before reducing the model for the quasi-linear approach (interactions included). The other one projects the percentage shrinkage onto the plane spanned by the influencing factors.

Reducing the linear model leads to the following graphic:

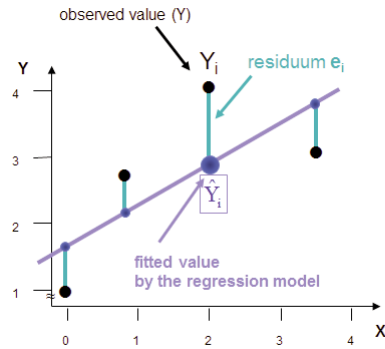


$$y = -6.17574 + 0.0261521 \times X1 - 0.0178006 \times X3$$

6 Single-factor Plots Overview:

The single-factor plots overview shows the effect of each selected influencing factor and displays the prediction value and prediction interval to be expected in case of the currently specified settings. The software forecasts a percentage shrinkage of 0,85% for the adjustment values highlighted by red lines.

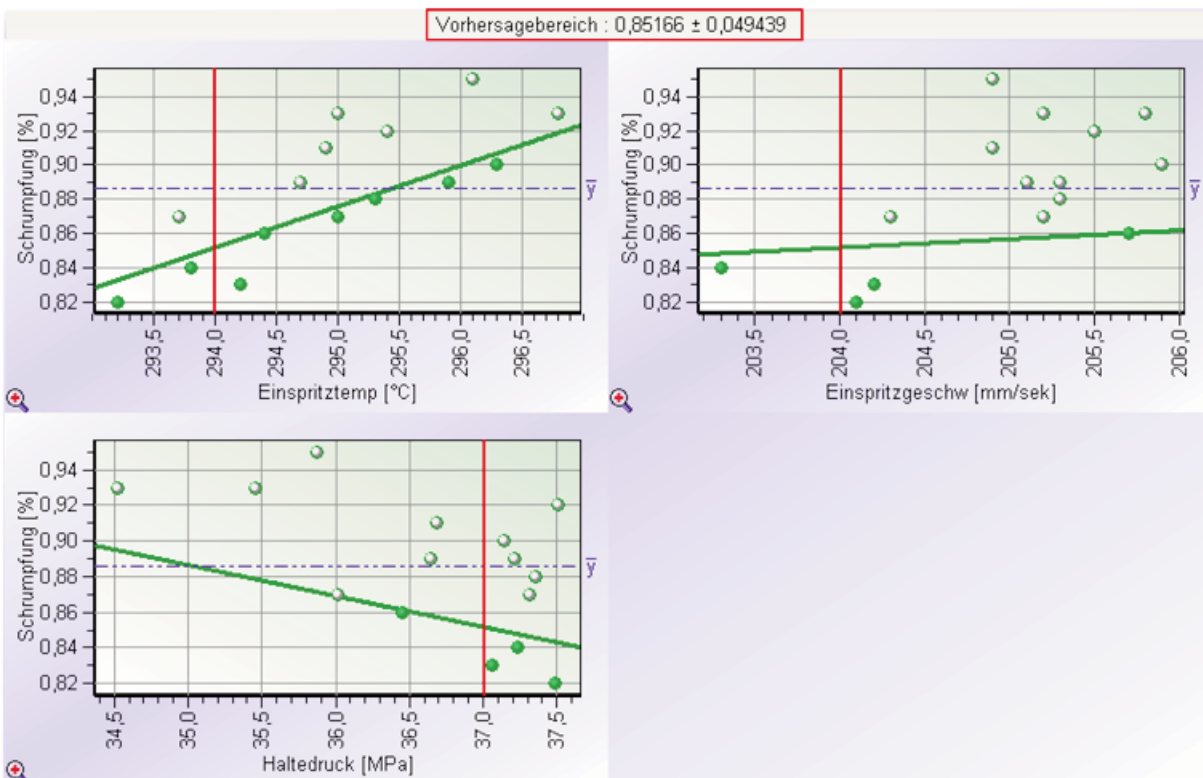
7 Analysis of Residuals



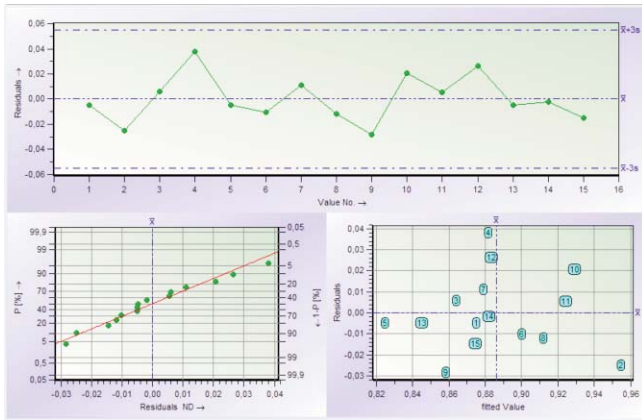
Residuals (lat. rest, singular residuum) are the differences between the observed values (y_i) and the values of the response estimated by the model (\hat{y}_i). As a result, residuals are the deviations that are not explained by the model.

The graphical analysis of residuals gives information about whether the model assumptions are fulfilled.

The upper graphic plots the residuals on the value number. Ideally, the residuals show a random distribution without any specific structure.

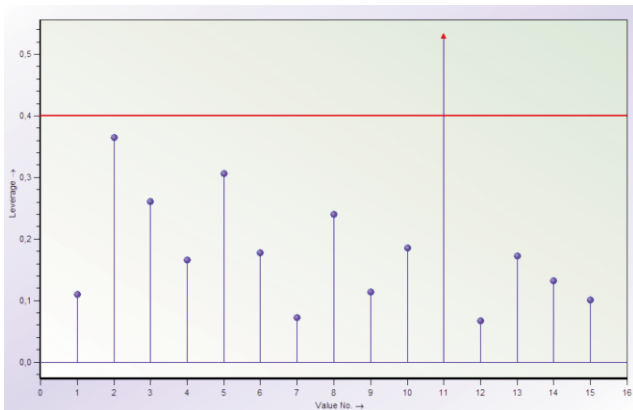


The lower left-hand graphic displays residuals on a probability plot. They are tested for normality. The assumption that residuals are normally distributed is a requirement for some tests within the scope of regression analysis.



The lower right-hand graphic compares the estimated values (fitted values) to the residuals. The values should be randomly distributed around the zero line from the left to the right in this graphic.

8 Leverage, Cook:



These diagnostics help to check whether single values (datasets) have a supererogatory impact on the entire regression model.

The “Leverage” graphic analyzes whether single sets of values of the influencing factors can be interpreted as outliers in x-direction.

The Cook’s distance graphic helps to recognize an influential value that also affects the entire model (its own estimate in the model) considerably.

9 VIF and Red%

The variance inflation factor (VIF) indicates the mutual dependence between the single influencing factors. The VIF should not exceed 10. In case there are no interactions, the VIF amounts, not exceeds, to VIF = 1. If the VIF does not rise, the estimations of the coefficients become more imprecise. This might affect the significance test of the coefficients.

Red% indicates the percentage reduction of the coefficient of determination in case a specific influencing factor is removed from the model (reduced model).

Red% is a variation complement corresponding to the coefficients t-statistic analyzing the expectation of the response, i.e. how much it can be raised.

Compared to the t-statistic, Red% inspects the impact of single influencing factors on the variation of the response.

23,13% means that the coefficient of determination amounts to 77,68% - 23,13% = 54,55% without injection temperature.

Even after reducing the variation, the injection temperature is still more important than the holding pressure.

		VIF	Red%
		---	---
injection temperature	X1	1.752	23.13
injection speed	X2	1.774	0.492
holding pressure	X3	1.079	14.03

10 Optimization

Before optimizing the process, the average percentage shrinkage amounted to 0,886%.

After finding the optimal adjustment of parameters, the percentage shrinkage is reduced to 0,825%. The control test proved that the process has improved.

Find Optima

Maximize

Minimize

Starting points

Number of random starting points: 5

Resolution of grid points: 2

Index	Index	
Factors	injection temperature	293.2
	holding pressure	37.51
percentage shrinkage	Predicted value	0,82437
	Standard dev.	0,022717