ZESZYTY NAUKOWE POLITECHNIKI ŁÓDZKIEJ
SCIENTIFIC BULLETIN OF ŁÓDŹ TECHNICAL UNIVERSITY
Nr 845 WŁÓKIENNICTWO, z. 58 2000

DATA ANALYSIS FOR QUALITY CONTROL IN THE TEXTILE BRANCH

Jiří Militký, Milan Meloun
Dept. of Textile Materials, Technical University of Liberec,
Hálkova 6, 461 17 Liberec, Czech Republic
Dept. of Analytical Chemistry, Uiversity of Pardubice
Pardubice, Czech Republic

Karel Kupka
TriloByte, Ltd., Pardubice, Czech Republic

The Box Cox transformation are given for construction of asymmetric quantiles and control limits for control charts. It is shown on simulated and real data that the transformation may significantly improve applicability of control charts.

1. INTRODUCTION

In statistical analysis of experimental data it is very often assumed that the data have normal (Gaussian) distribution. For example, when using arithmetic average, least squares regression, symmetric confidence intervals (t-tests), estimates of quantiles of the data for assessing non-conformities, and constructing of control limits in control charts. In textile branch however, many measurements of mechanical parameters such as fiber strength or low concentrations of pollutants usually do not have or physically cannot have normal distribution. They have asymmetric non-normal distributions and the above-mentioned methods generally fail. In case of Shewhart charts, this problem is partially avoided by taking subgroup means instead of individual data, which improves normality. It fails however in cases of stronger asymmetry and/or small subgroups generating many false alarms at the longer tail of the distribution and ignoring excessive values at the shorter one. For unimodal asymmetrically distributed homogeneous data a non-linear transformation can be found which improves symmetry or normality of the measured data. All computations, simulations and graphs were made with S-Plus [6].

2. THEORY

2.1. Asymmetric distribution

Construction of classical Shewhart control charts (see [2, 4]) as well as estimation of many other statistical values is closely tied to an assumption of normality. A non-parametric kernel estimate of probability density (Gaussian kernel, fixed width) was used to find important quantiles of asymmetrically distributed data.

Example 1. Simulated data with positive skewness generated from Weibull distribution are in the table 1.

Table 1

Simulated data

1.249	0.597	1.328	0.439	0.916	0.529	0.951	1.293
0.944	0.431	1.031	1.676	1.539	2.274	0.382	0.614
1.476	0.359	0.727	1.956	1.221	0.609	0.758	1.15
0.931	2.375	0.880	0.667	0.857	2.091	0.933	0.825
1.339	0.958	0.595	1.867	1.679	1.875	0.852	0.711

Basic characteristics of these data are in the table 2.

Table 2

Basic characteristics of data

Mean:	1.097
Skewness:	1.494
Median:	0.938
Mode:	0.816

	Normal	Nonpar. estimate
LCL (2.5%)	0.039	0.223
UCL (97.5%)	2.155	2.356

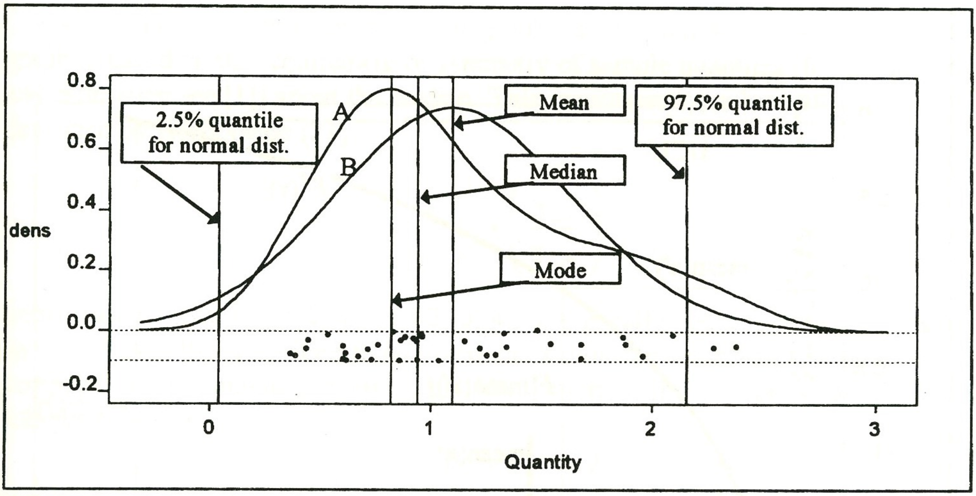

Fig. 1 Simulated data modeled by normal (B) and non-parametric kernel type (A) density function

Data in Example 1 was analyzed in order to determine 2.5% upper and lower control limits. Classical normal quantiles gave the following values: LCL = 0.039, UCL = 2.155 with average value 1.097. Fig. 1 shows how normal (Gaussian) model of probability density (curve B) differs from a non-parametric density curve (A), which describes the data better. Theoretically, the frequency of data x>UCL and x<LCL should be roughly the same, 2.5%. As a result of asymmetry of the data distribution however, the data exceeding UCL will be much more frequent than data x<LCL. And the control limits cannot be used efficiently to control the process.

2.2. Data transformation

Homogeneous data x with asymmetric unimodal distribution can be thought of as some original data y with normal distribution, which was skewed by some monotonous non-linear transformation F (like logarithmic, exponential). If an inverse of F, F^{-1} could be found, we could obtain a normally distributed data $y = F^{-1}(x)$. Since y has normal distribution, the classical methods could be used to determine parameters of the distribution, confidence intervals of the mean, and all necessary quantiles like control limits, etc. [3,4]. F can be then used to retransform these quantiles back to the original scale and unit of x. Fig. 2 shows original data x, transformed data $y = F^{-1}(x)$, and re-transformed mean, and asymmetric control limits $F(\mathrm{LCL}')$ and $F(\mathrm{UCL}')$.

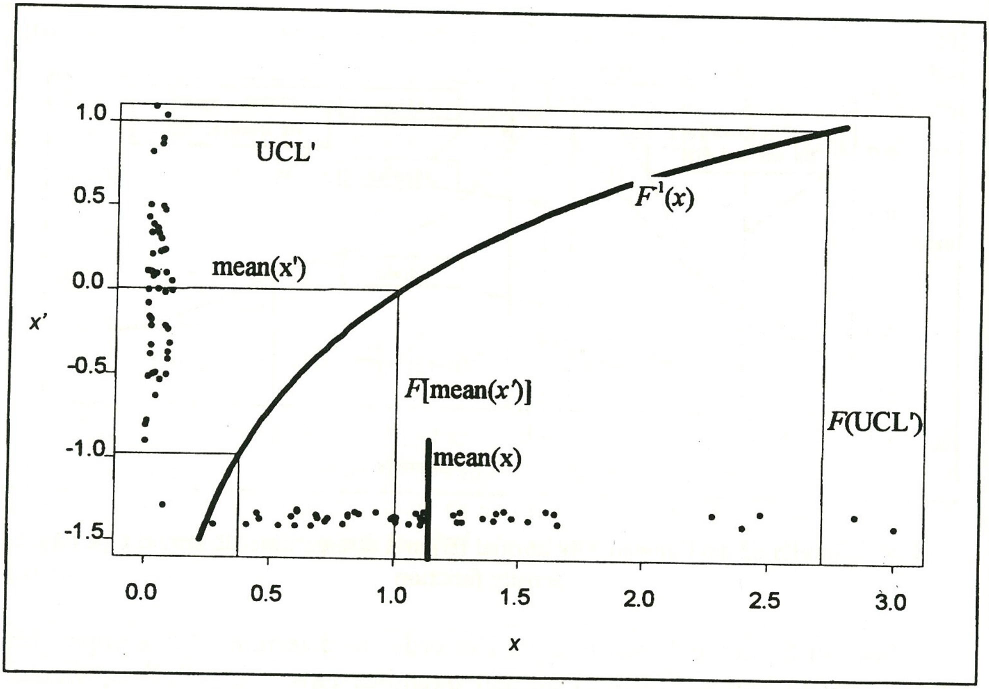

Fig. 2. Symmetrization of data distribution by non-linear transformation

Power transformation is used as a tool for simplifying of data distribution. Suitable power-law transformations may result in a distribution that is nearly symmetrical and perhaps more nearly normal.

Power transformation enables to select of suitable location estimators for skewed distribution and construction of corresponding asymmetrical confidence bands. In many cases the using of simple power transformation

$$\begin{aligned} y = g(x) &= x^{\lambda} && \text{for } \lambda > 0 \\ y = g(x) &= \ln(x) && \text{for } \lambda = 0 \\ y = g(x) &= x^{-\lambda} && \text{for } \lambda < 0 \end{aligned} \tag{1}$$

leads to the improving of the data distribution. This transformation is not scale invariant and is not continuous function of lambda. It requires the positive data only. Optimal transformation can be selected by minimizing of some robust measures of skewness

$$g_R = \frac{(y_{0.75} - y_{0.5}) - (y_{0.5} - y_{0.25})}{((y_{0.75} - y_{0.25})}$$

As a diagnostic tool the selection graph can be simply constructed. This graph is based on the requirement of symmetry of sample quantiles (for definition and estimation see [1]) about the median. This requirement can be mathematically described by relationship [21]

$$\left(\frac{x_{P_i}}{x_{0.5}}\right)^{\lambda}+\left(\frac{x_{0.5}}{x_{1-P_i}}\right)^{-\lambda}=2 \tag{2}$$

Letter values, where $P_i = 2^{-i}$ for i = 2, 3, 4,...are usually chosen. Selection graph has on y-axis the quantities $x_{Pi}/x_{0.5}$ and on x-axis the quantities $x_{0.5}/x_{1-Pi}$. For comparison of computed points with ideal courses for constant lambda the solution of equation

$$y^{\lambda}+x^{-\lambda}=2$$

is superimposed to graph.

Another exploratory technique for graphical estimation of optimal power is described by Emerson and Stoto [10]. After selection of optimal power the location parameter can be estimated from relation

$$x_{MR}=\left[\frac{\sum_{i=1}^{N}x_i^{\lambda}}{N}\right]^{1/\lambda} \quad \text{for } \lambda\neq 0$$

$$x_{MR}=\exp\left(\frac{\sum_{i=1}^{N}\ln(x_i)}{N}\right) \quad \text{for } \lambda=0 \tag{3}$$

Corresponding confidence interval is described in [1]. The Box-Cox power transformation family, which is continuous function of power lambda, can be defined by equation

$$y=g(x)=\frac{x^{\lambda}-1}{\lambda} \quad \text{for } \lambda\neq 0$$

$$y=g(x)=\ln(x) \quad \text{for } \lambda=0 \tag{4}$$

This transformation is limited for positive data only. After slight modification the range of applicability can be arbitrarily extended.

Properties of this transformation family are studied in the [7]. Based on the assumption that for some power lambda the y variable is normally distributed

$N(\mu_y, s_y^2)$ the likelihood function can be constructed. Logarithm of likelihood function has the form

$$\ln L(\lambda) = (N/2)\ln(s_y^2) + (\lambda - 1)\sum_{i=1}^{N}\ln(x_i)$$

The s_y^2 is sample variance of transformed data. The likelihood function can be plotted against lambda in suitable range (standard range is from-3 to 3).To this plot the 100(1-a)%th confidence interval of power

$$2\left[\ln(L(\lambda^*) - \ln(L(\lambda)\right] < \chi^2(1) \tag{5}$$

The maximum likelihood estimator of power is here indicated by star. If this interval does not contain 1, then the transformation is statistically significant, see Fig. 3

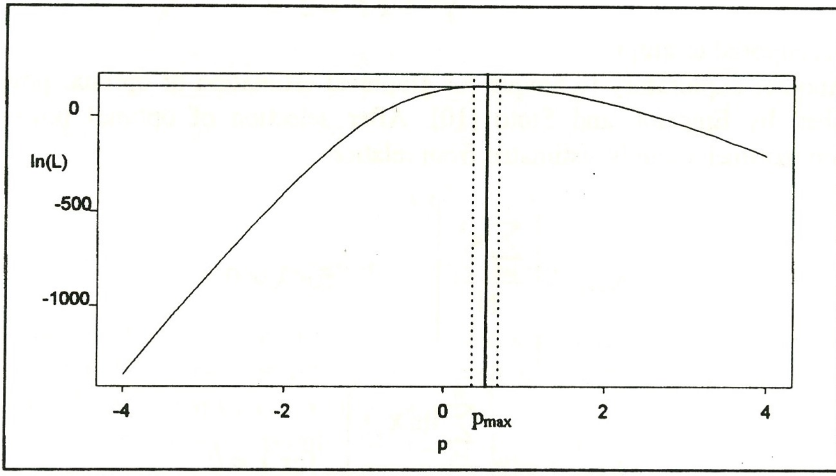

Fig. 3. Confidence interval for λ^* according to (5); λ^*=0.525, confidence interval = (0.363, 0.686)

3. EXPERIMENTAL PART

The aim is to construct the control charts for strength [GPa] distribution of high modulus carbon fiber. The 50 data points were experimentally determined. For construction of control charts the subgroup size = 3 was selected. Because the data are highly skewed the simple power transformation with skewness criterion and Box-Cox transformation has been used. Results are summarized in the tab.3.
Table 3 Results of power transformation

	No transform	Simple power	Box-Cox transform
Opt. power λ	1	-0.35	-0.45
Mean	2.278	2.249	2.247
LCL	1.859	2.066	2.070
UCL	2.697	3.065	3.174

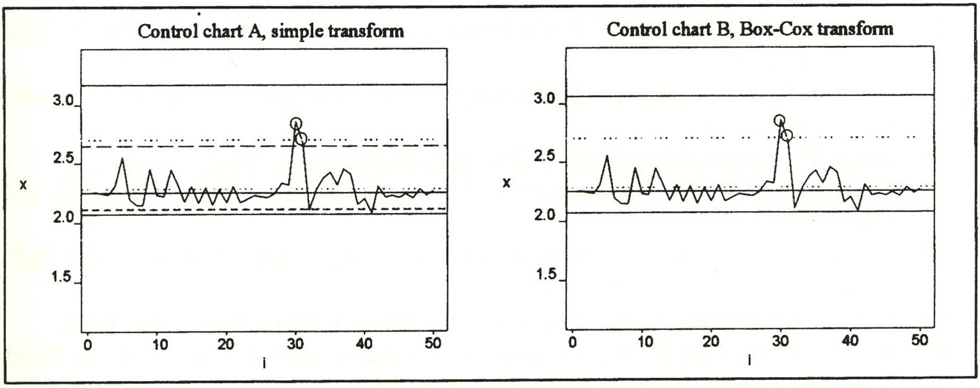

Fig. 4. Fiber strength, with warning and control limits, simple power and Box-Cox transformation. Dotted lines-target and control limits without transform, solid-retransformed target and control limits

Fig, 4 shows two points exceeding dotted classical UCL at i=30 and 31, while LCL falls under 2 GPa, which is to low for the data. After both simple power transform and Box-Cox transform, UCL and LCL increase, points 30 and 31 are inside re-transformed control levels, but above upper warning limits $\pm 2\sigma$ (see Chart A, dashed line). The transformation revealed one point at i=41 that exceeds re-transformed LCL and i=32 that exceeds lower warning limit, which is in contrast with the non-transformed chart.

4. CONCLUSION

The above examples show that even for grouped data, x-bar control charts may lead to false conclusions if the measured data have asymmetric distribution. A diagnosis of normality is always necessary before constructing quantiles, t-tests or confidence intervals based on normal distribution. Generally, both simple and Box-Cox transformation give similar results. Using maximal likelihood and Box-Cox transform give usually more stable values of λ and allow for statistical testing of suitability of the transformation.

Acknowledgment: This work was supported by the Czech Grant Agency; grant GACR No. 106/99/1184 and Czech Ministry of Education Grant VS 97084.

REFERENCES

[1] **M. Meloun, J. Militký,** Chemometrics for Analytical Chemistry: Part 1, Ellis Horwood, 1992.

[2] **Shewhart, W.A.,** Statistical Method from the Viewpoint of Quality Control, Dover Pubns, 1987.

[3] **Ryan, P.,** Statistical Methods for Quality Improvement, J. Wiley, 1994.

[4] **Montgomery D. C.,** Introduction to Statistical Quality Control, Chapman and Hall, 1990.

[5] **Mittag, Rinne,** Statistical Methods for Quality Assurance, Chapman and Hall, 1993.

[6] **Becker, Chambers, Wilks,** The New S Language, Chapman and Hall, 1996.

[7] **Box G. E. P., Cox D. R.,** An analysis of transformations. Journal of the Royal Statistical Society, Series B 26(2) 1964: 211-243.

[8] **Landwehr J.M., Pregibon D., Shoemaker A.C.,** J. Amer. Statist. Assoc., 79, 61 (1984).

[9] **Hines W.G.S., Hines R.J.H.,** Amer. Statist., 41, 21 (1987).

[10] **Emerson J. D., Stoto M. A.,** J. Amer. Statist. Assoc., 77, 103 (1982).

POLITECHNIKA ŁÓDZKA

ZESZYTY NAUKOWE

WŁÓKIENNICTWO

Nr 58

VI INTERNATIONAL TEXTILE CONFERENCE IMTEX" 2000

ŁÓDŹ 2000