# INFORMAČNÍ BULLETIN


České statistické společnosti

**Ročník 25, číslo 3, září 2014**

# JEDNODUCHÝ FUZZY REGRESNÍ MODEL
# A SIMPLE FUZZY REGRESSION MODEL

**Zdeněk Půlpán, Jiří Kulička**

*Adresa*: Zdeněk Půlpán, Na Brně 1952/39, 500 09 Hradec Králové 9

Mgr. Jiří Kulička, Ph.D., Univerzita Pardubice, DF JP,
Studentská 95, 532 10 Pardubice

*E-mail*: Zdenek.Pulpan@klikni.cz, jiri.kulicka@upce.cz

**Abstrakt:** Ukážeme řešení problému lineární fuzzy regrese. Využijeme principů teorie fuzzy množin na problematiku odhadu lineární závislosti výstupní proměnné $Y$ na vstupní proměnné $X$. Předpokládáme při tom, že vstupní proměnná $X$ není fuzzy, ale „ostrá“ hodnota, měřená s vyšší přesností než výstupní proměnná $Y$, která bude fuzzifikována. V úvaze o řešení bude problem postupně formálně rozšiřován až k situaci, kdy v modelu lineární závislosti $Y = A + BX$ jsou proměnné $A, B, Y$ považovány za trojúhelníková fuzzy čísla.
**Klíčová slova:** Lineární fuzzy regrese, použití trojúhelníkových fuzzy čísel.

**Abstract:** We will show solutions to the problem of fuzzy linear regression. We will use the principles of fuzzy set theory to problems of estimating the linear dependence of the output variable $Y$ on the input variable $X$. We assume the input variable $X$ as the fuzzy variable, but with the “real” value, measured with greater accuracy than the output variable $Y$, which was fuzzificating. During our work we will formally widen the question we consider to include the situation, where the variables $A, B$ and $Y$ in the model of the linear relationship $Y = A+BX$, are regarded to be triangular fuzzy numbers.
**Keywords:** Linear fuzzy regression, using triangular fuzzy numbers.

## 1. Úvod

V práci [6] jsme uvedli použití lineární regrese a logistické regrese k odhadu závislostí dvou proměnných z empirických zjištění. Zde ukážeme použití teorie fuzzy množin k odhadu lineární závislosti za podmínky, že o charakteru vztahu dvou veličin máme jen málo informací. Jak bude dále vidět, je několik možných přístupů k řešení problému odhadu závislosti jedné proměnné na lineární kombinaci zbývajících proměnných, které využívají teorie fuzzy množin. Jiný přístup, využívající fuzzy inference je ukázán v [7].

Předpokládejme, že vztah spojitých náhodných veličin $X$ a $Y$ máme dokumentován měřením tak, že získané dvojice $(x_i, y_i),\ i = 1, 2, \ldots, N$, jsou

odpovídajícími hodnotami uvedených veličin měřených současně na $N$ statistických jednotkách. Uvažujme situaci, kdy není zaručeno splnění Gauss-Markovových podmínek pro odvození statistických vlastností odhadu předpokládaného vztahu mezi veličinami $X$ a $Y$, např. ve tvaru lineární závislosti $Y = \alpha+\beta x+\varepsilon$, kde $\varepsilon$ je chyba odhadu. V případech, kdy jsou hodnoty veličiny $Y$ spíše odhadovány než měřeny (například v některých psychologických, lékařských nebo pedagogických šetřeních), je možné se opřít jen o představu, že veličina $X$ není náhodná (z hlediska teorie fuzzy množin uvažujeme, že je veličinou ostrou – anglicky crisp), ale veličina $Y$ je neostrá, tedy fuzzy. Jde například o situaci, kdy k výkonu žáka v určitém psychologickém testu je expertně (odhadem učitele) stanovována známka z matematiky nebo k určité naměřené hodnotě dechové frekvence nebo rozdílu objemu vdechnutého a vydechnutého vzduchu je lékařem odhadováno na určité škále stadium astmatu případně míra poruchy dýchání. Odhady zvláště druhé z veličin jsou subjektivní a tedy vágní povahy. Také počet měření je obvykle velmi malý.

Označme proto veličinu $Y$ jako fuzzy náhodnou veličinu ([3], [4 ], [5]) znakem $\underline{Y}$ a konstruujme pro ni vhodnou věrohodnostní funkci $\mu_Y$. Soustředíme se na situaci, kdy hodnoty veličiny $X$ jsou odhadnutelné s větší přesností než hodnoty veličiny $Y$ a tak ke každé naměřené hodnotě $x_i$ můžeme změřit či odhadnout více hodnot veličiny $Y$:

$$x_i \to y_{i1}, y_{i2}, \ldots, y_{in_i}, \quad i = 1, 2, \ldots, N; \tag{1}$$

při tom nejsme schopni předpovědět typ rozdělení náhodné veličiny $Y = Y(x)$. Víra v možnost aspoň přibližného odhadu hodnot veličiny $Y$ z hodnot veličiny $X$ (tedy víra v odhad hodnoty druhé proměnné $Y(x)$ jako fuzzy množiny $\underline{Y}(x)$, která by měla být fuzzy číslem) musí vyplývat z naší zkušenosti. Situaci interpretujeme i tak, že hodnoty veličiny $X$ chápeme jako ostré vstupní, hodnoty veličiny $Y(x)$ jako neostré, fuzzy výstupní. Jen výstupní hodnoty budou v našem modelu charakterizované neurčitostí popsatelnou fuzzy množinami. Pak jsou dvě možnosti. Buď věrohodnostní funkce je s ohledem na zkušenost (a případně i experimentální výsledky) odhadnuta expertem nebo tvar věrohodnostní funkce je předpokládán (například, že je trojúhelníková) a pak se parametry věrohodnostní funkce odhadují z experimentu opět metodami, vycházejícími ze zkušenosti hodnotitele. Nejjednodušší způsob je ten, že odhad parametrů $\alpha, \beta$ předpokládané závislosti se realizuje ostrými (ne-fuzzy) hodnotami $a, b$ z dvojic $(x_i, y_i),\ i = 1, 2, \ldots, N$, metodou nejmenších čtverců tak, aby aproximující funkce $\widehat{y}\,(x) = a + bx$

splňovala podmínku minimalizace výrazu

$$Q(a,b) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(y_{ij} - \widehat{y}\,(x_i)\right)^2 = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(y_{ij} - a - bx_i\right)^2 \tag{2}$$

v intervalu proměnnosti veličiny $X$. Dostaneme tak pro $\widehat{y}\,(x)$ vztah

$$\begin{gathered} \widehat{y}\,(x) = \overline{y} + b(x - \overline{x});\ \text{kde}\ \overline{x} = \frac{1}{n}\sum_{i=1}^{N} x_i n_i;\ \overline{y} = \frac{1}{n}\sum_{i,j} y_{ij}; \\ n = \sum_{i=1}^{N} n_i;\ b = \frac{n\sum_{i,j} x_i y_{ij} - \sum_i n_i x_i \sum_{i,j} y_{ij}}{n\sum_i x_i^2 n_i - \left(\sum_i x_i n_i\right)^2}. \end{gathered} \tag{3}$$

K popisu závislé veličiny $Y$ můžeme pak použít symetrické trojúhelníkové fuzzy číslo $\underline{Y}\,(x)$ s věrohodnostní funkcí $\mu_Y(x,y)$:

$$\mu_Y(x,y) \begin{cases} = 1 - \frac{1}{c} \cdot |y - \widehat{y}\,(x)|,\ \text{když}\ |y - \widehat{y}\,(x)| \le c \\ = 0,\ \text{jinak.} \end{cases} \tag{4}$$

Kladná konstanta $c$ je mírou rozptýlenosti hodnot náhodné veličiny $Y(x)$ pro pevná $x$ z intervalu proměnnosti veličiny $X$. Hodnotu konstanty $c$ můžeme určit buď expertně nebo výpočtem z dvojic naměřených hodnot $(x_i, y_i)$ například takto:

$$\begin{gathered} c_i = \max_j |y_{ij} - \widehat{y}\,(x_i)|, \\ c = \tfrac{1}{N}\textstyle\sum_{i=1}^{N} c_i. \end{gathered} \tag{5}$$

Je-li $c = 0$ (nebo „blízké“ 0), pak naměřené hodnoty veličiny $Y(x)$ jsou bez registrovaného rozptýlení okolo hodnot $\widehat{y}\,(x_i)$; je pak otázkou, zda představa veličiny $Y(x)$ jako fuzzy náhodné $\underline{Y}(x)$ je oprávněná. Předchozí výpočet je podmíněn tím, že variabilita hodnot $y_{ij}$ okolo hodnot $\widehat{y}\,(x_i)$ je v celém rozsahu proměnnosti veličiny $X$ stejná (kladné odchylky od $\widehat{y}\,(x_i)$ jsou stejně možné jako záporné v přibližně stejných hodnotách); takto uvažujeme také když o možných odchylkách nemáme dostatek konkrétnějších informací. Výsledkem úvahy je pak „rozmazaný“ (fuzzy-) odhad lineární regresní funkce. Je zřejmé, že uvedený postup lze zobecnit i na obdobný případ vztahu více proměnných.

*Poznámka 1:* Není-li možné předpokládat rozptýlenost hodnot $y_{ij}$ okolo $\widehat{y}\,(x_i)$, nezávislou na $i$, pak pro každé $i$ konstruujeme předpověď $\widetilde{\underline{Y}}\,(x_i)$, kde ve vztahu pro $\mu_Y$ nahrazujeme univerzální konstantu $c$ hodnotou $c_i$, například podle (5). I tato možnost není na závadu použití dalšího našeho postupu.

*Poznámka 2:* Je-li veličina $Y$ diskrétní, pak určitou předpověď hodnoty $Y(x)$ můžeme chápat také jako fuzzy množinu $\underset{\sim}{Y}(x)$ (která by měla být ovšem fuzzy číslem). Příkladem toho mohou být známky školní klasifikace nebo úsudky lékaře o stupni progrese určité choroby na základě jistých prvotních dat. Ani tato eventualita neohrozí použití navrženého postupu.

*Příklad 1.* Máme k dispozici měření, zaznamenané v tab. 1. Úkolem je stanovit předpověď hodnoty $Y(5)$ (která v tabulce měření uvedena není).

Tabulka 1: Hodnoty měřených veličin $X, Y$ k *Příkladu 1.*

| $\boldsymbol{i}$ | $\boldsymbol{x_i}$ | $\boldsymbol{y_{ij}}$ | $\boldsymbol{\widehat{y}\,(x_i)}$ | $\boldsymbol{\lvert y_{ij} - \widehat{y}\,(x_i)\rvert}$ | $\boldsymbol{\max_j \lvert y_{ij} - \widehat{y}\,(x_i)\rvert}$ |
|---|---|---|---|---|---|
| 1 | 1 | 1; 2; 4 | 2,51 | 1,51; 0,51; 1,49 | 1,51 |
| 2 | 3 | 3; 5 | 4,01 | 1,01; 0,99 | 1,01 |
| 3 | 4 | 4; 7 | 4,76 | 0,76; 2,24 | 2,24 |
| 4 | 6 | 4; 5; 9 | 6,26 | 4,26; 1,26; 2,74 | 4,26 |

Dosazením hodnot z tab. 1 do vztahů (3) dostaneme postupně:

$$\begin{gathered} n = 10;\ N = 4; \\ \overline{x} = 0{,}1 \cdot (3 \cdot 1 + 2 \cdot 3 + 2 \cdot 4 + 3 \cdot 6) = 3{,}5; \\ \overline{y} = 0{,}1 \cdot (1 + 2 + 4 + 3 + 5 + 4 + 7 + 4 + 5 + 9) = 4{,}4; \\ b = \tfrac{10 \cdot 183 - 35 \cdot 44}{10 \cdot 161 - 35^2} = \tfrac{290}{385} = 0{,}753; \\ y(x) = 4{,}4 + 0{,}75(x - 3{,}5) = 1{,}76 + 0{,}75x. \end{gathered} \tag{6}$$

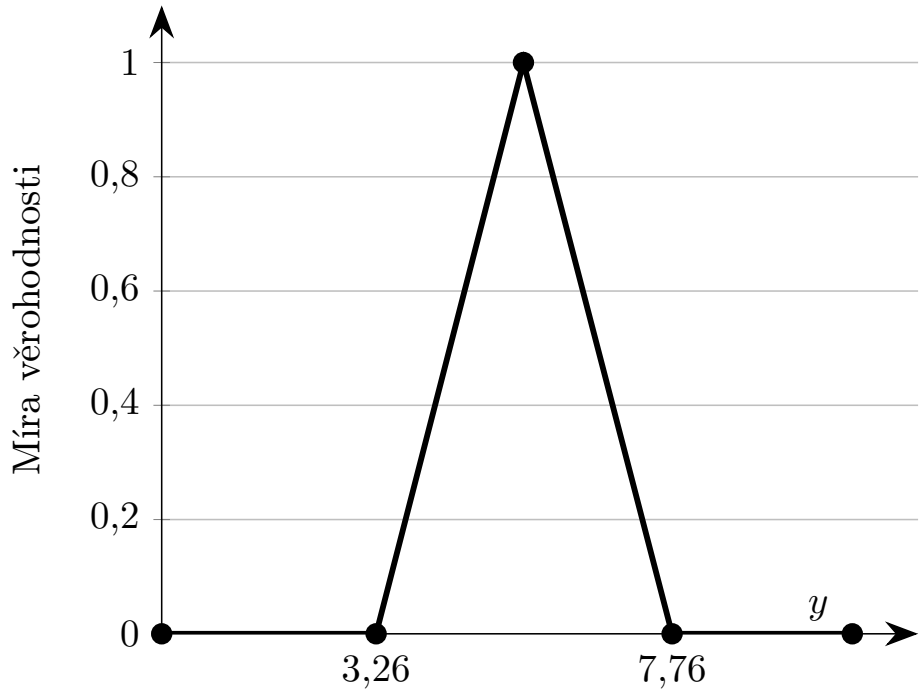
Předpověď ostré hodnoty v bodě $x = 5$ je $\widehat{y}\,(5)$, po vyčíslení je $\widehat{y}\,(5) = 5{,}51$; pro náš případ si ale určeme $c = \frac{1}{4}\sum_{i=1}^{4} c_i = 2{,}25$. Hodnoty veličiny $Y(x)$ budou reprezentovány totiž trojúhelníkovým fuzzy číslem $\underset{\sim}{Y}(x)$; zachycují neurčitost předpovědi prostřednictvím věrohodnostní funkce $\mu_Y(x, y)$:

$$\mu_Y(x,y) \begin{cases} = 1 - \frac{1}{2{,}25} \cdot |y - 1{,}76 - 0{,}75x|, \text{ když } |y - 1{,}76 - 0{,}75x| \le 2{,}25; \\ = 0, \text{ jinak.} \end{cases} \tag{7}$$

Pro $X = 5$ je podle předchozího vztahu $\underset{\sim}{Y}(5)$ dáno věrohodnostní funkcí

$$\mu_Y(5,y) \begin{cases} = 1 - 0{,}44 \cdot |y - 5{,}51|, \text{ když } 3{,}26 \le y \le 7{,}76; \\ = 0, \text{ jinak.} \end{cases} \tag{8}$$

Fuzzy množina $\underline{\widetilde{Y}}(5)$ je zobrazena na obr. 1. Podle vztahu (8) je věrohodnost výroku „Y(5) = 3“ rovna nule, věrohodnost výroku „Y(5) = 5“ je rovna 0,78, viz také obr. 1. Podobně určíme i věrohodnost výroku „Y(2) = 4“ ze vztahu (7) dosazením $x = 2,\ y = 4$; dostaneme přibližně hodnotu $\mu_Y(2,4) = 0{,}67$. Graf jednoduché fuzzy regrese závislosti $Y$ na $X$ je na obr. 2 na další straně.



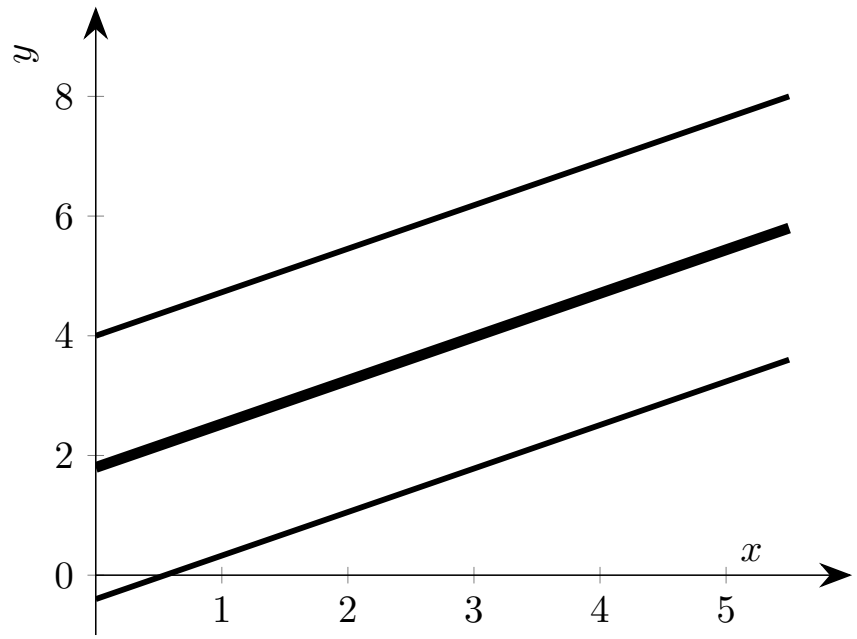

Obrázek 1: Obraz věrohodnostní funkce fuzzy čísla $\underline{\widetilde{Y}}(5)$ k *Příkladu 1.*

## 2. První fuzzy model

Upravme nyní uvedenou metodu tak, že ve vyjádření příslušného regresního odhadu budou parametry $a, b$ trojúhelníkovými fuzzy čísly a veličina $X$ bude „ostrá“. Pak výsledný odhad proměnné $Y$ musí být také trojúhelníkovým fuzzy číslem (lineární kombinace trojúhelníkových fuzzy čísel je opět trojúhelníkové fuzzy číslo).

Předpokládejme tedy, že hledáme takový regresní vztah tvaru $\underline{Y} = \underline{A} + \underline{B}X$, kde $\underline{A}, \underline{B}, \underline{Y}$ jsou trojúhelníková fuzzy čísla a veličina $X$ je ostrá. Přitom volíme pro trojúhelníková fuzzy čísla $\underline{A}, \underline{B}$ za parametry jejich jádra $p_0, p_1$ a $c_0, c_1$ jako poloměry jejich nosičů jež charakterizují jejich „rozmazanost“:

$$\mu_A(x,y) \begin{cases} = 1 - \frac{|p_0 - a|}{c_0},\ \text{když } p_0 - c_0 \le a\ \le p_0 + c_0;\ c_0 > 0; \\ = 0;\ \text{jinak}, \end{cases} \tag{10}$$




Obrázek 2: Jednoduchá fuzzy regrese k *Příkladu 1.*
*Horní přímka představuje horní odhad fuzzy regresní funkce, nejspodnější je grafem dolního odhadu fuzzy regresní funkce; v bodech $(x, y)$ těchto grafů je věrohodnostní funkce $\mu_Y(x, y)$ rovna nule, pro body s vyšším $y$ jsou však její hodnoty nenulové až do bodu horního odhadu – ovšem při stejném $x$, třetí graf (tučnější čára) je grafem fuzzy regresní funkce, je to vlastně graf souřadnic $(x, y)$ bodů pro něž je věrohodnostní funkce $\mu_Y(x, y) = 1$, viz také obr. 1 na předchozí straně.*

$$\mu_B(x,y)\begin{cases} = 1 - \frac{|p_1 - b|}{c_1},\ \text{když } p_1 - c_1 \le b \le p_1 + c_1;\ c_1 > 0; \\ = 0;\ \text{jinak}, \end{cases} \tag{11}$$

pak ale také musí být

$$\mu_Y(x,y)\begin{cases} = 1 - \frac{|p_0 + p_1 x - y|}{c_0 + c_1|x|},\ \text{když} & p_0 + p_1 x - c_0 - c_1|x| \le y \wedge \\ & y \le p_0 + p_1 x + c_0 + c_1|x| \\ = 0;\ \text{jinak}. & \end{cases} \tag{12}$$

Volme $\alpha \in (0; 1)$ a stanovme podmínky pro to, aby $\mu_Y(x, y) \ge \alpha$. Pro fuzzy číslo $\underline{Y}$ volíme jeho $\alpha$-řez jako ostrou množinu $\alpha$-přípustných hodnot $y$. Vyjdeme-li ze vztahu (12) pro $\mu_Y(x, y)$, dostaneme podmínku pro $y$ ve tvaru:

$$1 - \frac{|p_0 + p_1 x - y|}{c_0 + c_1|x|} \ge \alpha. \tag{13}$$

Jednoduchou úpravou a odstraněním absolutní hodnoty získáme dvě základní nerovnosti:

$$y \geq p_0 + p_1 x - (1-\alpha) \cdot (c_0 + c_1|x|), \tag{14a}$$

$$y \leq p_0 + p_1 x + (1-\alpha) \cdot (c_0 + c_1|x|). \tag{14b}$$

Ze všech možných regresních vztahů vybíráme ten, který minimalizuje „rozmazanost“ výstupní fuzzy množiny $\underline{Y}$. „Rozmazanost“ fuzzy množiny $\underline{Y}$ je $c_0 + c_1|x|$ pro každou dvojici měření $(x, y)$. Požadavek minimalizace „rozmazanosti“ vztahujeme pro všechna měření k výrazu $Q$ tvaru (15). Výraz (15) představuje součet „rozmazaností“ všech $\underline{Y}(x)$:

$$Q = N \cdot c_0 + c_1 \sum_{i=1}^{N} |x_i|. \tag{15}$$

Nalezení vhodných parametrů $c_0 > 0, c_1 > 0$ a $p_0, p_1$, které splňují (14a) a (14b) za podmínky minimalizace funkce $Q = Q(c_0, c_1)$ z (15) tak představuje řešení úlohy lineárního programování. Úloha se tak může řešit standardními metodami úloh lineárního programování.

*Příklad 2.* Použijme dat z *Příkladu 1* a sestavme podmínky úlohy lineárního programování pro určení hodnot parametrů $c_0, c_1$ a $p_0, p_1$ z minimalizace funkce $Q = Q(c_0, c_1)$. Volme při tom $\alpha = 0{,}75$. Postupně dostaneme soustavu nerovností nejprve dosazením do (14a) a pak také do (14b), nerovnosti tvoří podmínku pro přípustná řešení:

a) první část nerovnic:

$$\begin{aligned}
x = 1 : \; & p_0 + \phantom{3}p_1 - 0{,}25c_0 - 0{,}25c_1 \leq 1 \\
& p_0 + \phantom{3}p_1 - 0{,}25c_0 - 0{,}25c_1 \leq 2 \\
& p_0 + \phantom{3}p_1 - 0{,}25c_0 - 0{,}25c_1 \leq 4 \\
x = 3 : \; & p_0 + 3p_1 - 0{,}25c_0 - 0{,}75c_1 \leq 3 \\
& p_0 + 3p_1 - 0{,}25c_0 - 0{,}75c_1 \leq 5 \\
x = 4 : \; & p_0 + 4p_1 - 0{,}25c_0 - \phantom{0{,}75}c_1 \leq 4 \\
& p_0 + 4p_1 - 0{,}25c_0 - \phantom{0{,}75}c_1 \leq 7 \\
x = 6 : \; & p_0 + 6p_1 - 0{,}25c_0 - 1{,}50c_1 \leq 4 \\
& p_0 + 6p_1 - 0{,}25c_0 - 1{,}50c_1 \leq 5 \\
& p_0 + 6p_1 - 0{,}25c_0 - 1{,}50c_1 \leq 9
\end{aligned}$$

b) druhá část nerovnic:

$$\begin{array}{rrrrrrr}
x=1: & p_0 + & p_1 + & 0{,}25c_0 + & 0{,}25c_1 \geq 1 \\
 & p_0 + & p_1 + & 0{,}25c_0 + & 0{,}25c_1 \geq 2 \\
 & p_0 + & p_1 + & 0{,}25c_0 + & 0{,}25c_1 \geq 4 \\
x=3: & p_0 + & 3p_1 + & 0{,}25c_0 + & 0{,}75c_1 \geq 3 \\
 & p_0 + & 3p_1 + & 0{,}25c_0 + & 0{,}75c_1 \geq 5 \\
x=4: & p_0 + & 4p_1 + & 0{,}25c_0 + & c_1 \geq 4 \\
 & p_0 + & 4p_1 + & 0{,}25c_0 + & c_1 \geq 7 \\
x=6: & p_0 + & 6p_1 + & 0{,}25c_0 + & 1{,}50c_1 \geq 4 \\
 & p_0 + & 6p_1 + & 0{,}25c_0 + & 1{,}50c_1 \geq 5 \\
 & p_0 + & 6p_1 + & 0{,}25c_0 + & 1{,}50c_1 \geq 9
\end{array}$$

při současné minimalizaci funkce $Q(c_0, c_1) = 4c_0 + 14c_1$. Zjednodušením této soustavy získáme tab. 2, použitelnou již pro řešení některou z metod lineárního programování.

Výsledkem řešení uvedené úlohy minimalizace funkce $Q$ na množině přípustných hodnot pro $p_0, p_1$ a nezáporná $c_0, c_1$ jsou postupně $Q_{\min} = 58, p_0 = p_1 = 1, c_0 = 11, c_1 = 1$.

Tento model předpovídá hodnotu proměnné $Y$ v bodě $x = 5$ jako trojúhelníkové fuzzy číslo $\widetilde{Y}(5)$, viz obr. 3, s věrohodnostní funkcí danou čísly $p_0, p_1, c_0, c_1$ v obecném tvaru

$$\mu_Y(5,y)\begin{cases} = 1 - \frac{|p_0+5p_1-y|}{c_0+5c_1}, \text{ když} & p_0 + 5p_1 - c_0 - 5c_1 \leq y \wedge \\ & y \leq p_0 + 5p_1 + c_0 + 5c_1, \\ = 0; \text{ jinak,} & \end{cases}$$

tedy v našem případě to je

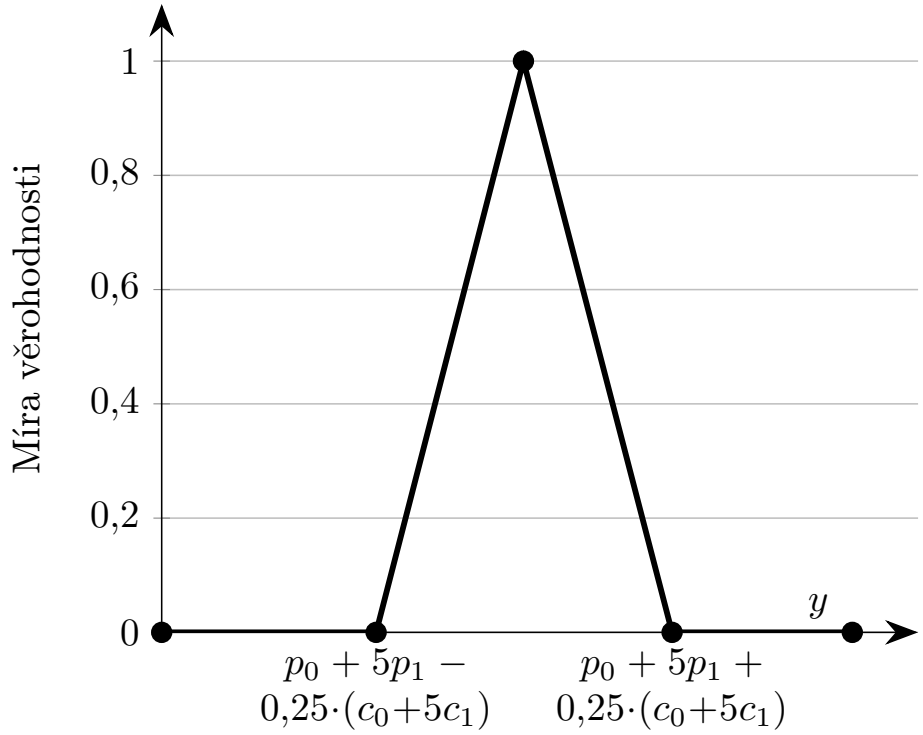
$$\mu_Y(5,y)\begin{cases} = 1 - \frac{|6-y|}{16}; & -10 \leq y \leq 22, \\ = 0; \text{ jinak.} & \end{cases}$$

„Rozmazanost“ výsledného fuzzy čísla je relativně velká a pochopitelně závisí také na volbě hodnoty pro $\alpha$.

## 3. Druhý fuzzy model

Ještě dále zobecníme naše předpoklady. Dosud jsme uvažovali, že měřená veličina $Y$ je ostrá a její odhad je fuzzy číslo $\widetilde{Y}$. Nyní budeme předpokládat, že i naměřené hodnoty $y_{ij}$ jsou realizacemi jistého fuzzy čísla $\underline{Y}_i$ s věrohodnostní

Obrázek 3: Obraz fuzzy čísla $\underline{Y}(5)$ k *Příkladu 2.*




Tabulka 2: Tabulka pro numerické řešení úlohy lineárního programování z *Příkladu 2.*

| $\boldsymbol{X}$ | $\boldsymbol{p_0}$ | $\boldsymbol{p_1}$ | $\boldsymbol{c_0}$ | $\boldsymbol{c_1}$ | **Nerovnost** | **Absolutní člen** |
|---|---|---|---|---|---|---|
| **1** | 1 | 1 | −0,25 | −0,25 | ≤ | 1 |
| **3** | 1 | 3 | −0,25 | −0,25 | ≤ | 3 |
| **4** | 1 | 4 | −0,25 | −0,25 | ≤ | 4 |
| **6** | 1 | 6 | −0,25 | −0,25 | ≤ | 4 |
| **1** | 1 | 1 | 0,25 | 0,25 | ≥ | 4 |
| **3** | 1 | 3 | 0,25 | 0,75 | ≥ | 5 |
| **4** | 1 | 4 | 0,25 | 1,00 | ≥ | 7 |
| **6** | 1 | 6 | 0,25 | 1,50 | ≥ | 9 |
| $\boldsymbol{Q}$ | **0** | **0** | **4** | **14** | **Minimalizace** | — |

funkcí (16):

$$\mu_{Y_i}(x,y)\begin{cases} = 1 - \frac{|\overline{y}_i - y|}{e_i};\ e_i > 0;\ \text{když } \overline{y}_i - e_i \le y \le \overline{y}_i + e_i, \\ = 0;\ \text{jinak.} \end{cases} \tag{16}$$

V (16) je $\overline{y}_i$ aritmetický průměr z hodnot $y_{ij}, j = 1, 2, \ldots, n_i$.

Kladná hodnota $e_i$ definuje „přesnost“ měření veličiny $Y$. V případě, že máme k určité hodnotě $x_i$ naměřeno více hodnot $y_{ij}$ druhé (výstupní) proměnné $Y$, můžeme konstantu $e_i$ určit například podle

$$e_i = \max_j |y_{ij} - \overline{y}_i|, i = 1, 2, \ldots, N. \tag{17}$$

Pro fuzzy výstup $\underline{Y}$ z regresní rovnice $\underline{Y} = \underline{A} + \underline{B}X$ pak máme věrohodnostní funkci (12). Snažíme se, aby fuzzy regresní odhad $\widetilde{\underline{Y}}$ splňoval pro co největší počet $x_i$ podmínku pro inkluzi $\alpha$-řezů fuzzifikovaných výstupních hodnot $\underline{Y}_i = \underline{Y}(x_i)$ a regresního odhadu $\widetilde{Y}_i = \underline{A} + \underline{B}x_i$ ve tvaru

$$\underline{Y}_i^{\alpha_i} \subseteq \widetilde{\underline{Y}}_i^{\alpha}, \text{ tj. } \alpha_i \geq \alpha \text{ pro co největší počet } i; \tag{18}$$

$\alpha_i$ odpovídají $\alpha$-řezům fuzzy čísla $\underline{Y}_i$, $\alpha$ odpovídá $\alpha$-řezu regresního odhadu $\widetilde{\underline{Y}}_i = \widetilde{\underline{Y}}(x_i)$, viz také obr. 4 pro jediné $i$. Mělo by tedy platit

$$\alpha_i = 1 - \frac{|p_0 + p_1 x_i - y_{ij}|}{c_0 + c_1 x_i - e_i} \geq \alpha \text{ pro co největší počet } i.$$

„Nejlepší“ hodnotou $\alpha$ (snažíme se, aby $\alpha$ bylo co největší) pak je

$$\alpha = \min_i \{\alpha_i\}. \tag{19}$$

Pro odhad fuzzy čísel $\underline{A}$, $\underline{B}$ požadujeme, aby „rozmazanost“ regresního odhadu byla co nejmenší, tedy opět jde o minimalizaci funkce (15).

Rozepíšeme-li podmínku (18) pro $\alpha$-řezy, máme pro všechny dvojice naměřených hodnot $(x_i, y_{ij})$ dva druhy nerovnic:

$$\begin{aligned} y_{ij} &\geq p_0 + p_1 x_i - (1-\alpha)(c_0 + c_1 x_i) + (1-\alpha)e_i \\ y_{ij} &\leq p_0 + p_1 x_i + (1-\alpha)(c_0 + c_1 x_i) - (1-\alpha)e_i; \text{ pro všechna } i, j. \end{aligned} \tag{20}$$

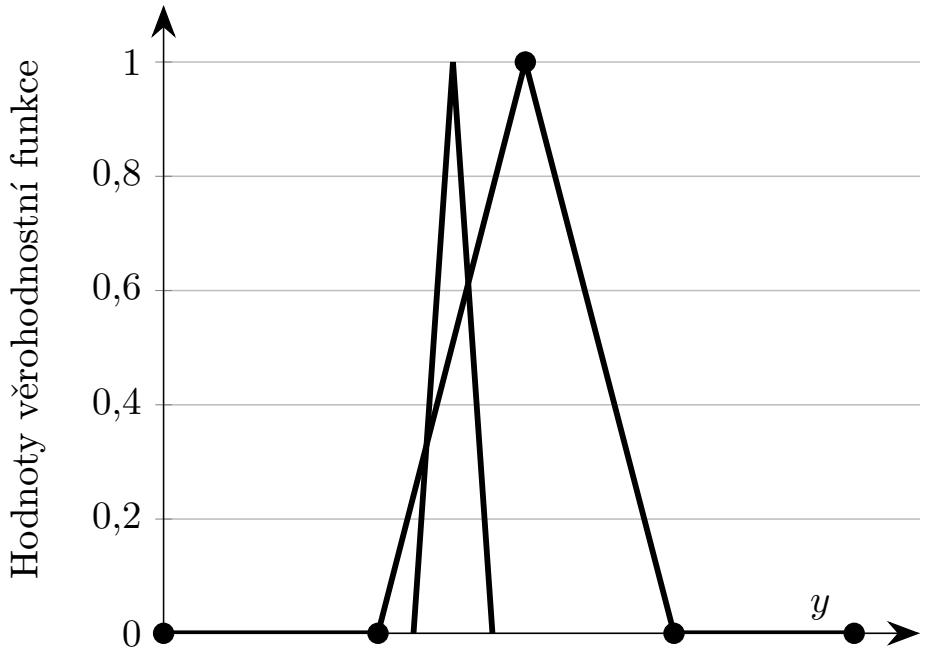
Soustavu nerovnic (20) lze zjednodušit na tvar (21):

$$\begin{aligned} \min_j \{y_{ij}\} &\geq p_0 + p_1 x_i - (1-\alpha)(c_0 + c_1 x_i) + (1-\alpha)e_i \\ \max_j \{y_{ij}\} &\leq p_0 + p_1 x_i + (1-\alpha)(c_0 + c_1 x_i) - (1-\alpha)e_i; \text{ pro všechna } i. \end{aligned} \tag{21}$$

Opět se jedná o úlohu lineárního programování; za podmínek (12) a (21) hledáme minimum funkce $Q = Q(c_0, c_1)$. Ukážeme si to na konkrétním příkladě.

*Příklad 3.* Pokusme se aplikovat předchozí teorii na data z *Příkladu 1*, viz tab. 1.

Obrázek 4: Komentář ke vztahu (18). Vyplněná část zobrazuje věrohodnostní funkci fuzzy množiny $\underline{Y}_i$ a nevyplněná fuzzy množinu, která reprezentuje hodnoty fuzzy regresní funkce $\underline{\widetilde{Y}}(x_i)$.




Protože předpokládáme, že naměřené výstupní hodnoty jsou realizacemi trojúhelníkových fuzzy čísel (16), určíme nejprve hodnoty jejich parametrů $e_i$ podle (17):

$$\begin{aligned} i=1:&\ \overline{y}_1 = 2{,}33;\ e_1 = 1{,}67\\ i=2:&\ \overline{y}_2 = 4{,}00;\ e_2 = 1{,}00\\ i=3:&\ \overline{y}_3 = 5{,}50;\ e_3 = 1{,}50\\ i=4:&\ \overline{y}_4 = 6{,}00;\ e_4 = 3{,}00. \end{aligned}$$

Volíme stejně jako v předcházejících případech $\alpha = 0{,}75$. Dosazením dat z tab. 1 dostáváme soustavu nerovnic ve tvaru:

$$\begin{aligned} 1 &\geq p_0 + p_1 - 0{,}25\cdot(c_0 + c_1) + 0{,}25\cdot 1{,}67\\ 3 &\geq p_0 + 3p_1 - 0{,}25\cdot(c_0 + 3c_1) + 0{,}25\cdot 1{,}00\\ 4 &\geq p_0 + 4p_1 - 0{,}25\cdot(c_0 + 4c_1) + 0{,}25\cdot 1{,}50\\ 4 &\geq p_0 + 6p_1 - 0{,}25\cdot(c_0 + 6c_1) + 0{,}25\cdot 3{,}00\\ \hline 4 &\leq p_0 + p_1 + 0{,}25\cdot(c_0 + c_1) - 0{,}25\cdot 1{,}67\\ 5 &\leq p_0 + 3p_1 + 0{,}25\cdot(c_0 + 3c_1) - 0{,}25\cdot 1{,}00\\ 7 &\leq p_0 + 4p_1 + 0{,}25\cdot(c_0 + 4c_1) - 0{,}25\cdot 1{,}50\\ 9 &\leq p_0 + 6p_1 + 0{,}25\cdot(c_0 + 6c_1) - 0{,}25\cdot 3{,}00 \end{aligned}$$

Řešením této soustavy nerovnic, viz také tab. 3, získáme přípustné hodnoty pro $p_0, p_1, c_0, c_1$. Hodnoty optimální získáme z nich výběrem těch, které minimalizují funkci $Q = 4c_0 + 14c_1$.

Jsou to při $Q_{\min} = 50$ hodnoty $p_0 = p_1 = 1; c_0 = 9; c_1 = 1$, které určují $\widetilde{\underline{Y}}(5)$ věrohodnostní funkcí $\mu_Y(5, y)$ ve tvaru

$$\mu_Y(5,y) \begin{cases} = 1 - \frac{|p_0+5p_1-y|}{c_0+5c_1} = 1 - \frac{|6-y|}{14}, \text{ když } -8 \le y \le 20, \\ = 0; \text{ jinak.} \end{cases} \tag{22}$$

„Rozmazanost“ tohoto fuzzy čísla je o něco menší.

Tabulka 3: Tabulka pro řešení úlohy lineárního programování z *Příkladu 3.*

| $X$ | $p_0$ | $p_1$ | $c_0$ | $c_1$ | Nerovnost | Absolutní člen |
|---|---|---|---|---|---|---|
| **1** | 1 | 1 | −0,25 | −0,25 | ≤ | 0,58 |
| **3** | 1 | 3 | −0,25 | −0,75 | ≤ | 2,75 |
| **4** | 1 | 4 | −0,25 | −1,00 | ≤ | 3,63 |
| **6** | 1 | 6 | −0,25 | −1,50 | ≤ | 3,25 |
| **1** | 1 | 1 | 0,25 | 0,25 | ≥ | 4,42 |
| **3** | 1 | 3 | 0,25 | 0,75 | ≥ | 5,25 |
| **4** | 1 | 4 | 0,25 | 1,00 | ≥ | 7,38 |
| **6** | 1 | 6 | 0,25 | 1,50 | ≥ | 9,75 |
| **Q** | **0** | **0** | **4** | **14** | **Minimalizace** | — |

*Poznámka:* Při řešení této části úlohy jsme mohli každé naměřené $y_{ij}$ nahradit fuzzy číslem $\underline{Y}_{ij}$ s jádrem $y_{ij}$ a přesností danou $e_i$, určenou z naměřených hodnot $y_{ij}$, viz vztahy (20) a (21). Možné bylo uvažovat i jediné fuzzy číslo $\underline{Y}_i$, určené jádrem $\overline{y}_i$ a s přesností $e_i$. Tak jsme to udělali v předchozím příkladě. Bylo to metodologicky asi čistší.

## Závěr

Prezentované fuzzy modely lineární regrese využívají důležité vlastnosti trojúhelníkových fuzzy čísel (jejich lineární kombinace je opět fuzzy číslo). Omezuje se tím však obecnost metody, protože se tak dají řešit jen problémy, kde obstojí předpoklad, že závislá veličina má všude stejnou variabilitu. Výhoda spočívá v možnosti užití standardních algoritmů řešení úlohy lineárního programování. Nevýhodou je často nevyhnutelná vysoká „rozmazanost“

takových fuzzy odhadů: $\alpha$ pro $\alpha$-řezy nemá tak jasnou interpretaci jako je pravděpodobnost, proto se „z bezpečnostních důvodů“ nedoporučuje ho volit např. větší než 0,5.

## Literatura

[1] Ross, T. J.: *Fuzzy logic with engineering applications*, second edition, J. Wiley & Sons, Ltd., The Atrium, Southern Gate, Chichester, West Sussex PO 198SQ, England, June 2005.

[2] Klir, G. J., Yuan, Bo: *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, 1995.

[3] Kwakernaak, H.: Fuzzy Random Variables I and II. *Inf. Sci. (USA),* 15: 1–29, 1979.

[4] Viertl, R.: Univariate Statistical Analysis with Fuzzy Data, *Computational Statistics & Data Analysis*, 51(1): 133–147, 2006. ISSN 0167-9473.

[5] Wang, G., Yhang, Y.: The Theory of Fuzzy Stochastic Processes, *Fuzzy Sets and Systems,* 51: 161–178, 1992. ISSN 0165-0114.

[6] Půlpán, Z.: *K problematice zpracování empirických šetření v humanitních vědách,* Academia, Praha 2004.

[7] Žák, L.: Fuzzy regrese, *Informační bulletin ČStS,* 22(2), 209–220, červen 2011. ISSN 1210-8022.

# VÝPOČET INDEXŮ ZPŮSOBILOSTI PROCESU PŘI NENORMÁLNĚ ROZLOŽENÝCH DATECH

# COMPUTATION OF PROCESS CAPABILITY INDICES FOR NON-NORMALLY DISTRIBUTED DATA

## Martin Kovářík

*Adresa*: Univerzita Tomáše Bati ve Zlíně, Fakulta managementu a ekonomiky, Ústav statistiky a kvantitativních metod, náměstí T. G. Masaryka 5555, 760 01 Zlín, Česká republika

*E-mail*: kovarik.fame@seznam.cz

**Abstrakt:** Pokud charakter rozdělení procesu je nenormální, poté indexy $C_p$ a $C_{pk}$, vypočítané pomocí konvenčních metod, občas vedou k nesprávné interpretaci způsobilosti procesu. Ačkoliv byly k výpočtu náhradních indexů způsobilosti pro nenormální rozložení procesu navrženy různé metody, je neustále nedostatek literatury nabízející komplexní vyhodnocení a srovnání těchto metod. V případech mírné i velké odchylky od normality tyto náhradní indexy způsobilosti procesu adekvátně zachycují spolehlivost procesu. Ale otázka zní, jaká je nejlepší metoda, která nejlépe odráží skutečnou způsobilost procesu za těchto okolností? V následujícím textu bude zhodnoceno a konfrontováno devět metod, které jsou vybrány pro porovnání jejich schopnosti zacházení s nenormalitou při výpočtu indexů způsobilosti. Pro ilustraci porovnání těchto metod budou použita simulovaná data, která pocházejí z Weibullova a lognormálního rozdělení. Výsledky budou prezentovány porovnáním krabicových grafů. Výsledky simulace ukáží, že účinnost metody je závislá na její schopnosti zachytit chování procesu na chvostech případně v krajních částech nosiče rozdělení pravděpodobnosti.

**Klíčová slova:** Indexy způsobilosti procesu (PCI), Asymetrické rozdělení, Metoda pravděpodobnostního grafu, Toleranční intervaly nezávislé na rozdělení, Metoda váženého rozptylu, Clementsova metoda, Burrova percentilová metoda, Box-Coxova mocninná transformace, Johnsonova transformace, Wrightův procesní index způsobilosti $C_s$.

**Abstract:** When the distribution of a process characteristic is nonnormal, the indices $C_p$ and $C_{pk}$, calculated using conventional methods, often lead to erroneous interpretation of the process's capability. Though various methods have been proposed for computing surrogate process capability indices (PCIs) under nonnormality, there is a lack of literature offering a comprehensive evaluation and comparison of these methods. In particular, under mild and severe departures from normality, do these surrogate PCIs adequately

capture process capability, and what is the best method for reflecting the true capability under each of these circumstances? The paper provides a review of nine methods that are chosen for performance comparison in their ability to handle nonnormality in PCIs. For illustration purposes the comparison is carried out by simulating Weibull and lognormal data, and the results are presented using box plots. Simulation results show that the performance of a method is dependent on its capability to capture the tail behavior of the underlying distributions.

**Keywords:** Process capability indices (PCI), Asymetric distribution, Probability plot method, Distribution-free tolerance intervals, Weighted variance method, Clements' method, Burr percentile method, Box-Cox power transformation, Johnson transformation, Wright's process capability index $C_s$.

## Úvod

Indexy způsobilosti procesu (PCI) jsou široce používány k určení, zda je proces schopen produkovat výrobky v dané toleranci. Mezi nejpoužívanější indexy patří $C_p$ a $C_{pk}$, které jsou definovány následovně:

$$C_p = \frac{\text{šířka tolerance}}{\text{šířka procesu}} = \frac{USL - LSL}{6\sigma}, \tag{1}$$

$$C_{pk} = \min\left\{C_{pu}, \mathrm{C}_{pl}\right\} = \min\left\{\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right\}, \tag{2}$$

kde $USL$ a $LSL$ znamenají horní a dolní specifikační limit, v tomto pořadí. Protože střední hodnota procesu $\mu$ a rozptyl $\sigma^2$ nejsou známy, obvykle se odhadují z výběrových statistik $\overline{x}$ a $s^2$.

English a Taylor zkoumali vliv nenormality na indexy způsobilosti procesu a dospěli k závěru, že $C_{pk}$ je více citlivý na odchylky od normality než $C_p$. Kotz a Johnson poskytli výsledky výzkumu ve svojí práci týkající se vlastností indexů způsobilosti procesu a jejich odhadů v případě, že rozdělení není normální. Mezi nimi jsou Clementsova metoda, Johnson-Kotz-Pearnova metoda a indexy způsobilosti procesu „nezávislé na rozdělení".

Nový index $C_s$ navržený Wrightem, navíc včleňuje šikmostní korekční faktor ve jmenovateli $C_{pmk}$ indexu navrženém Johnsonem. Choi a Bai navrhli heuristickou metodu vážených rozptylů k úpravě hodnot PCI podle stupně šikmosti tím, že zvažuje směrodatnou odchylku nad a pod střední hodnotou procesu samostatně.

Někteří autoři navrhli novou generaci PCI založenou na předpokladu o základním souboru. Pearn a Kotz založili jejich nové indexy způsobilosti procesu na Pearsonovu chí-kvadrát rozdělení. Johnson a kol. navrhli nahrazování

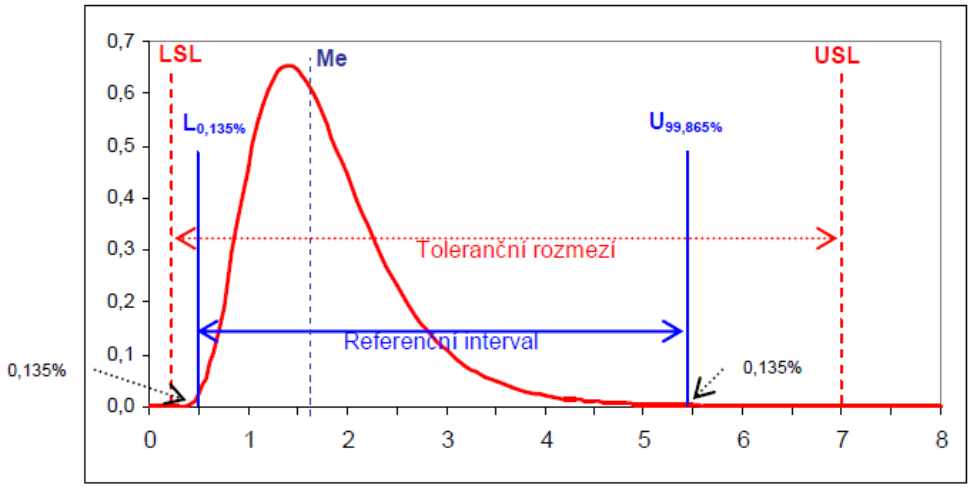
$6\sigma$ ve jmenovateli rovnice (1) výrazem $6\theta$, kde $\theta$ je voleno tak, aby „způsobilost“ nebyla výrazně ovlivněna tvarem rozdělení. Castagliola představil výpočetní metody pro nenormální PCI odhadováním podílu nevyhovujících pomocí Burrova rozdělení. Vännman navrhl novou rodinu indexů $C_p\,(u,\,v)$ parametrizací $(u,\,v)$, které zahrnují mnoho dalších indexů jako speciálních případů. Deleryd zjišťoval vhodné hodnoty pro $u$ a $v$ v případě, kdy rozdělení procesu bylo zešikmené. Index $C_p\,(1,\,1)$, který je ekvivalentní s $C_{pmk}$, se doporučuje jako nejvhodnější k zajištění nenormality v indexech způsobilosti procesu.

I když je dobře známo, že $C_p$ a $C_{pk}$ nejsou určeny pro způsobilost procesu při nenormálním rozdělení a tyto různé metody jsou vhodné pro výpočet některých náhradních PCI, chybí komplexní srovnání těchto metod. Především, odborníky z praxe nejvíce zajímá, které metody jsou navzájem srovnatelné, které jsou citlivé na dodržení předpokladu normality a které jsou nejvhodnější k použití za předpokladu nenormality. V sekci 2 prověříme devět různých metod s ohledem na skutečnou způsobilost procesu srovnáním $C_{pu}$ vypočtenou simulací s cílovou hodnotou $C_{pu}$. Definice indexu je zmíněna na další straně v sekci 1.1. Pokud individuální hodnoty sledovaného znaku kvality v procesu jsou rozděleny nenormálně, potom pro odhady ukazatelů způsobilosti a výkonnosti je třeba odhadnout kvantily $\widehat{L}_{0,00135}$, $\widehat{U}_{0,99865}$ a $\widehat{M}_{0,5}$.

V sekci 1 budeme uvažovat přístupy, jak požadované kvantily odhadnout. Někdy je tvar rozdělení sledovaného znaku kvality znám buď ze zkušenosti, nebo na základě znalosti fyzikálních vlastností procesu. Potom je třeba podrobit napozorovaná data testu dobré shody, ověřit zda předpokládaný model správně vystihuje napozorovaná data, odhadnout parametry rozdělení a vypočítat potřebné kvantily. Nejčastěji se tak setkáváme s rozdělením log-normálním, Weibullovým, exponenciálním apod. V literatuře lze nalézt vzorce pro odhad příslušných parametrů rozdělení a následně pro výpočet požadovaných kvantilů. Ty mohou být rovněž vypočteny s podporou vhodného softwaru. Na obrázku 1 je znázorněna situace, jakým způsobem se odvozuje ukazatel způsobilosti pro nenormálně rozdělený znak jakosti. [1], [2], [3]

## 1. Náhradní indexy způsobilosti procesu pro nenormální data

V následující části představíme devět různých metod výpočtu indexů způsobilosti procesu pro nenormální data.




Obrázek 1: Odvození ukazatele způsobilosti
pro nenormálně rozdělený znak kvality. Zdroj: [4].

## 1.1. Metoda pravděpodobnostního grafu

Široce uznávaný přístup k výpočtu PCI je použití kvantilového grafu funkce normálního rozdělení, tudíž lze zároveň ověřit předpoklad normality dat. Podobně jako graf funkce normálního rozdělení, kde šířka takového procesu se pohybuje mezi 0,135 a 99,865% percentilem, tak i náhradní hodnoty PCI lze získat prostřednictvím vhodné pravděpodobnostní funkce:

$$C_p = \frac{USL - LSL}{\text{horní } 0{,}135\% \ - \ \text{dolní } 0{,}135\% \text{ percentil}} = \frac{USL - LSL}{U_p - L_p}, \qquad (3)$$

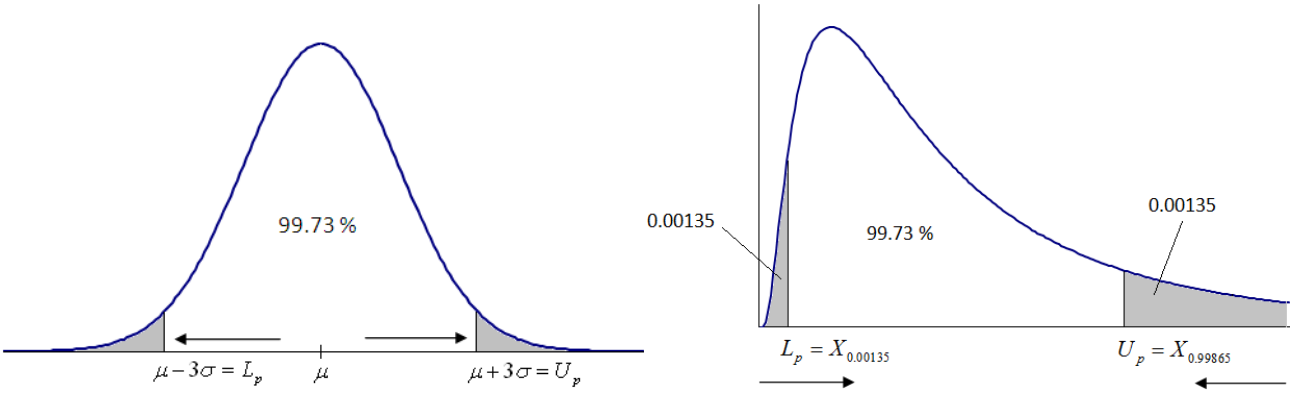
kde $U_p$ a $L_p$ jsou v tomto pořadí 99,865 a 0,135% percentily pozorování. Tyto percentily lze snadno získat pomocí vhodného statistického softwaru, jako například Minitab, NCSS, XLStatistics apod. Vzhledem k tomu, že medián je preferován jako charakteristika polohy asymetrického rozdělení, k tomu ekvivalentní $C_{pu}$ a $C_{pl}$ jsou definovány jako

$$C_{pu} = \frac{USL - \text{ medián}}{x_{0,99865} - \text{ medián}}, \qquad (4)$$

$$C_{pl} = \frac{\text{medián} - LSL}{\text{medián} - x_{0,00135}}. \qquad (5)$$

$C_{pk}$ je potom bráno jako minimum $(C_{pu}, C_{pl})$. Protože se jedná o grafickou metodu, pokud je skutečně použita pouze s pomocí pravděpodobnostního papíru bez softwaru, může být i dosti nepřesná. Tato metoda funguje tak, že do pravděpodobnostního papíru, kde stupnice na ose $x$ je lineární

a stupnice na ose $y$ je pravděpodobnostní, odpovídající určitému rozdělení pravděpodobnosti, se zakreslí $N$ naměřených hodnot, uspořádaných podle velikosti $(x(i))$ s odpovídajícími hodnotami empirické distribuční funkce $(i/N)$. Těmito body proložená přímka je odhadem distribuční funkce předpokládaného rozdělení pravděpodobnosti studovaného znaku kvality. Na této přímce odečteme příslušné kvantily odpovídající zvoleným podílům (0,00135; 0,50; 0,99865) a ty pak dosadíme do výrazů pro odhady ukazatelů výkonnosti. [5], [6]




Obrázek 2: Normální rozdělení: vymezení intervalu, kde se nachází 99,73% hodnot a nenormální rozdělení: vymezení intervalu, kde se nachází 99,73% hodnot.

## 1.2. Toleranční intervaly nezávislé na rozdělení

Chan a kol. poskytli přístup tolerančního intervalu nezávislého na rozdělení pro výpočet $C_p$ a $C_{pk}$ pro proces pocházející z nenormálního rozdělení pomocí

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{USL - LSL}{\frac{3}{2}(4\sigma)} = \frac{USL - LSL}{3(2\sigma)}. \tag{6}$$

To implikuje v

$$C_p = \frac{USL - LSL}{\omega} = \frac{USL - LSL}{\frac{3}{2}\omega_2} = \frac{USL - LSL}{3\omega_3}, \tag{7}$$

$$C_{pk} = \frac{\min\left[(USL - \mu),\ (\mu - LSL)\right]}{\omega/2}, \tag{8}$$

kde $\omega$ je šířka tolerančního intervalu. Všimněme si, že pokud je požadována 99% spolehlivost, interval, který pokrývá celý vzorek dat, zaručuje dosažení pokrytí pouze 75 % populačních hodnot. Toto je jediná zde prezentovaná

metoda, která bude dávat jiný výsledek, i když základní rozdělení je normální. Je zřejmé, že pro případ normálního rozdělení leží v mezích $\pm 3s$ přibližně 99,73 % hodnot parametru jakosti. Hodnota $C_p = 1$ pak ukazuje, že 99,73 % výrobků je v tolerančních mezích. Hodnota 6 ve jmenovateli je obecně závislá na velikosti výběru, z něhož se odhaduje $s$ a na rozdělení znaku jakosti. Omezení problému s nenormalitou rozdělení znaku jakosti lze docílit vhodnou volbou procenta výrobků ležících v tolerančních mezích. Pokud zvolíme tento parametr 99 %, můžeme použít ve jmenovateli hodnotu 5,15 platnou pro řadu rozdělení (se šikmostí od 0 do 3,111 a špičatostí od 1 do 5,997). [7], [8]

## 1.3. Metoda váženého rozptylu

Choi a Bai navrhli heuristickou metodu váženého rozptylu pro úpravu hodnot PCI podle stupně šikmosti základního souboru. Nechť $P_x$ je odhad pravděpodobnosti, že proměnná $X$ je menší nebo rovna své střední hodnotě,

$$P_x = \frac{1}{n} \sum_{i=1}^{n} I\left(\overline{X} - X_i\right), \tag{9}$$

kde $I(x) = 1$, pokud $x > 0$ a $I(x) = 0$, pokud $x < 0$. PCI založený na metodě váženého rozptylu je definován jako [7]

$$C_p = \frac{USL - LSL}{6\sigma W_x}, \tag{10}$$

kde $W_x = \sqrt{1 + |1 - 2P_x|}$ . Tudíž

$$C_{pu} = \frac{USL - \mu}{2\sqrt{2P_x}\sigma}, \tag{11}$$

$$C_{pl} = \frac{\mu - LSL}{3\sqrt{2\,(1 - P_x)}\sigma}. \tag{12}$$

## 1.4. Clementsova metoda

Clements nahradil $6\sigma$ v rovnici (1) výrazem $U_p - L_p$:

$$C_p = \frac{USL - LSL}{U_p - L_p}, \tag{13}$$

kde $U_p$ je 99,865% percentil a $L_p$ je 0,135% percentil. Pro $C_{pk}$ se střední hodnota $\mu$ odhaduje mediánem $M$ a $3\sigma$ jsou odhadovány pomocí $U_p - M$,

resp. $M - L_p$. Dostáváme tedy

$$C_{pk} = \min\left[\frac{USL - M}{U_p - M},\ \frac{M - LSL}{M - L_p}\right]. \tag{14}$$

Clementsův přístup používá klasických odhadů šikmosti a špičatosti, které jsou založeny na třetí a čtvrté momentové charakteristice, ale zároveň mohou být poněkud nespolehlivé pro velmi malý rozsah výběru. Indexy jsou navrženy pro třídu tzv. *Pearsonových rozdělení* (kam patří např. rozdělení *beta*, *gamma* a *Studentovo rozdělení*). [1], [9]

**Postup při použití těchto indexů:**

1. stanovit toleranční meze $USL$, $LSL$,
2. vypočítat charakteristiky $\overline{x}$, $s$, $Sk$ (skewness), $Ku$ (kurtosis),
3. nalézt standardizované kvantily $L'_p$, $U'_p$ pro zvolené pravděpodobnosti $p$ ($p = 0{,}00135$ a $p = 0{,}99875$) a vypočítané $Sk$ a $Ku$ v tabulkách Clements (1989) nebo na PC (např. NCSS, Minitab),
4. nalézt v tabulkách Clements (1989) standardizovaný medián $M'$,
5. vypočítat kvantily $L_p$ a $U_p$,
   $L_p = \overline{x} - s \cdot L'_p$
   $U_p = \overline{x} + s \cdot U'_p$
6. vypočítat medián $M$:
   $M = \overline{x} + s \cdot M'$
7. vypočítat indexy $C_p$, $C_{pk}$ (viz výše).

Pearn a Kotz Clementsovu metodu výpočtu $C_p$ a $C_{pk}$ doplnili o indexy $C_{pm}$, $C_{pm^*}$, $C_{pmk}$. Definiční rovnice dalších tří výše zmíněných indexů jsou dle [1] následující.

$$C'_{pm} = \frac{USL - LSL}{6\sqrt{\left(\frac{U_p - L_p}{6}\right)^2 + (M - T)^2}} \tag{15}$$

$$C'_{pm^*} = \frac{\min\left(USL - T,\ T - LSL\right)}{3\sqrt{\left(\frac{U_p - L_p}{6}\right)^2 + (M - T)^2}} \tag{16}$$

$$C'_{pmk} = \min \left[ \frac{USL - M}{3\sqrt{\left(\frac{U_p - M}{3}\right)^2 + (M - T)^2}}, \frac{M - LSL}{3\sqrt{\left(\frac{M - L_p}{3}\right)^2 + (M - T)^2}} \right] \quad (17)$$

## 1.5. Burrova percentilová metoda

Burr navrhl nové rozdělení, tzv. Burrovo XII rozdělení k výpočtu požadovaných percentilů náhodné veličiny $X$. Hustota pravděpodobnosti Burrova XII rozdělení je definována následujícím způsobem

$$f(y|c,k) = \begin{cases} \dfrac{cky^{c-1}}{(1+y^c)^{k+1}} & \text{pokud } y \geq 0,\ c \geq 1,\ k \geq 1, \\ 0 & \text{pokud } y < 0, \end{cases} \quad (18)$$

kde $c$ a $k$ reprezentují šikmost a špičatost Burrova rozdělení.

Liu a Chen představili modifikaci založenou na Clementsově metodě, kdy místo použití percentilů Pearsonovy křivky nahradili tyto percentily z příslušného Burrova rozdělení. Navržená metoda je popsána v následujících krocích.

1. Odhadne se výběrový průměr $\overline{x}$, výběrová směrodatná odchylka $s$, šikmost $s_3$ a špičatost $s_4$ empirických výběrových dat. Šikmost a špičatost dostaneme následovně:

$$s_3 = \frac{n}{(n-1)(n-2)} \sum \left( \frac{x_j - \overline{x}}{s} \right)^3 \quad (19)$$

a

$$s_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{x_j - \overline{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}. \quad (20)$$

2. Vypočítají se standardizované momenty šikmosti $\alpha_3$ a špičatosti $\alpha_4$ dle následujících vztahů:

$$\alpha_3 = \frac{(n-2)}{\sqrt{n(n-1)}} s_3 \quad (21)$$

a

$$\alpha_4 = \frac{(n-2)(n-3)}{(n^2-1)} s_4 + 3\frac{(n-1)}{(n+1)}. \quad (22)$$

Standardizovaná špičatost je obyčejně známá jako exces a pokud hodnota samotné špičatosti vyjde 3, znamená to, že rozdělení je normální

(mesokurtické). Záporná hodnota celého výrazu špičatosti označuje platykurtické (plošší) rozdělení, kdy směrodatná odchylka je větší. Kladná hodnota značí leptokurtické rozdělení (špičatější), kdy je směrodatná odchylka menší. [10], [11], [12]

3. Dále použijeme hodnoty $\alpha_3$ a $\alpha_4$ pro výběr vhodných parametrů $c$ a $k$. Poté se použije Burrovo rozdělení XII k získání standardizované veličiny

$$Z = \frac{Y - \mu}{\sigma}, \tag{23}$$

kde $Y$ je vybraná Burrova veličina, $\mu$ a $\sigma$ její korespondující střední hodnota a směrodatná odchylka. Všechny tyto momentové charakteristiky můžeme najít v tabulkách (Burr, Liu a Chen). V těchto tabulkách jsou tabelovány standardizované 0,00135, 0,5 a 0,99865 percentily, které odpovídají $Z_{0,00135}$, $Z_{0,5}$ a $Z_{0,99865}$. Dostaneme odpovídající percentily porovnáním dvou standardizovaných hodnot

$$\frac{X - \overline{x}}{s} = \frac{Y - \mu}{\sigma}.$$

4. Z předchozího tedy dostáváme odhady percentilů pro spodní, prostřední a horní percentily, které jsou

$$L_p = \overline{x} + s\ Z_{0,00135},$$

$$M = \overline{x} + s\ Z_{0,5},$$

$$U_p = \overline{x} + s\ Z_{0,99865}.$$

5. Vypočítáme indexy způsobilosti procesu použitím rovnic z Clementsovy metody.

Místo použití momentových charakteristik šikmosti a špičatosti, jak bylo uvedeno výše, jsou i jiné metody, jako je metoda maximální věrohodnosti, pravděpodobnostní metoda vážených momentů a metoda věrohodnostních momentů, které se také používají k odhadu parametrů Burrova rozdělení. Převážně se však používá metoda momentových charakteristik z důvodu její jednoduchosti a rychlosti výpočtu. [11], [12]

## 1.6. Transformační techniky

1. **Stabilizace rozptylu** vyžaduje nalezení transformace $y = g(x)$, ve které je již rozptyl $\sigma^2(y)$ konstantní. Pokud je rozptyl původní proměnné $x$ funkcí typu $\sigma^2(x) = f_1(x)$, lze rozptyl $\sigma^2(y)$ určit

$$\sigma^2(y) \approx \left[\frac{\mathrm{d}\, g(x)}{\mathrm{d}\, x}\right]^2 f_1(x) = C \tag{24}$$

kde $C$ je konstanta. Hledaná transformace $g(x)$ je pak řešením diferenciální rovnice

$$g(x) \approx C \int \frac{\mathrm{d}\, x}{\sqrt{f_1(x)}} \tag{25}$$

U řady instrumentálních metod a přístrojů je zajištěna konstantnost relativní chyby $\delta(x)$. To znamená, že rozptyl $\sigma^2(x)$ je dán funkcí $f_1(x) = \delta^2(x)\, x^2 = \text{konst}\ x^2$. Po dosazení vyjde $g(x) = \ln x$. Optimální je pro tento případ logaritmická transformace původních dat. Z toho vyplývá také vhodnost použití geometrického průměru. Pokud je závislost $\sigma^2(x) = f_1(x)$ mocninná, bude optimální transformace $g(x)$ také mocninná. Jelikož pro normální rozdělení je střední hodnota na rozptylu nezávislá, bude transformace stabilizující rozptyl také zajišťovat přiblížení k normalitě.

2. **Zesymetričtění rozdělení výběru** je možné provést jednoduchou mocninou transformací

$$y = g(x) = \begin{cases} x^{\lambda} & \lambda > 0 \\ \ln x \quad \text{pro} & \lambda = 0 \\ -x^{-\lambda} & \lambda < 0 \end{cases} \tag{26}$$

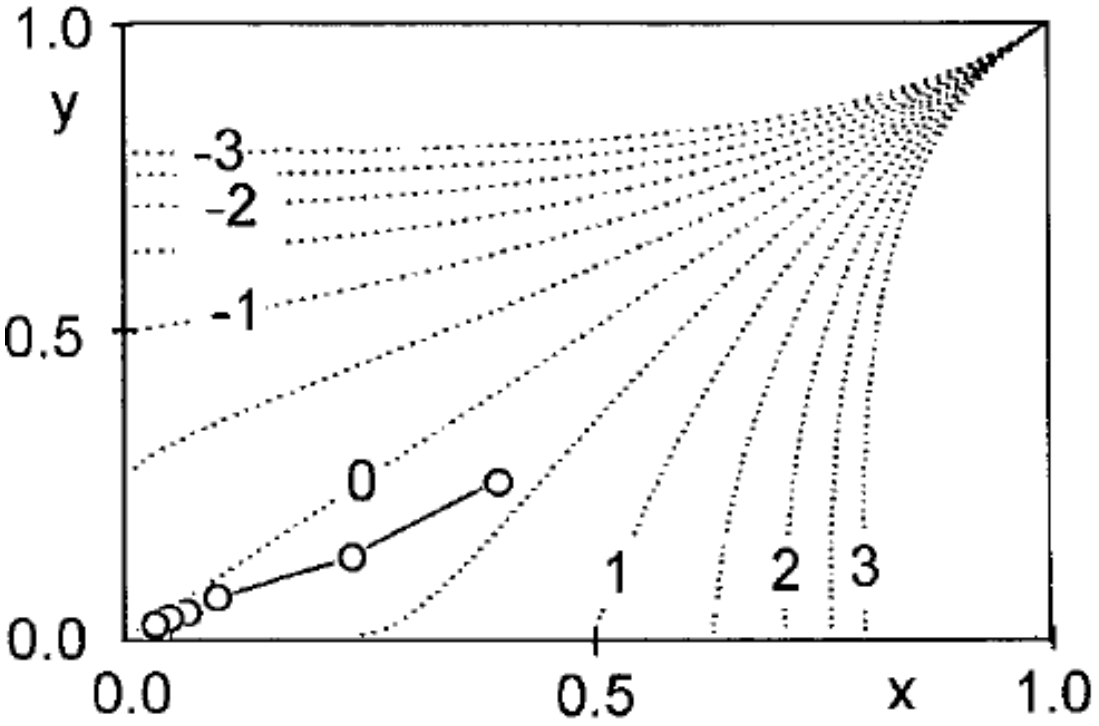
Tato transformace však nezachovává měřítko, není vzhledem k hodnotě $\lambda$ všude spojitá a hodí se pouze pro kladná data. Optimální odhad $\widehat{\lambda}$ se hledá s ohledem na minimalizaci vhodných charakteristik asymetrie. Kromě šikmosti $\widehat{g}_1(y)$ je možné užít i robustní verzi šikmosti definovanou výrazem

$$\widehat{g}_{1R}(y) = \frac{\left(\overline{y}_{0,75} - \overline{y}_{0,50}\right) - \left(\overline{y}_{0,50} - \overline{y}_{0,25}\right)}{\overline{y}_{0,75} - \overline{y}_{0,25}} \tag{27}$$

Pro symetrické rozdělení je $g_1(y) = g_{1R}(y) = 0$. Hodnotu $\widehat{\lambda}$ lze hledat pomocí rankitového grafu. Pro optimální $\widehat{\lambda}$ budou ležet kvantily $y_{(i)}$ přibližně na přímce.

3. **Hines-Hinesův selekční graf** (osa $x$: $\widetilde{x}_{0,5}/x_{1-P_i}$, osa $y$: $\widetilde{x}_{P_i}/\widetilde{x}_{0,5}$)

   Diagnostickou pomůckou pro odhad optimálního parametru $\lambda$ je selekční graf, ukázka je na obrázku 3.




Obrázek 3: Selekční graf pro výběr z lognormálního rozdělení. Zdroj: [13].

Vychází z požadavků symetrie jednotlivých kvantilů kolem mediánu

$$\left(\frac{\widetilde{x}_{P_i}}{\widetilde{x}_{0,5}}\right) + \left(\frac{\widetilde{x}_{0,5}}{\widetilde{x}_{1-P_i}}\right)^{-\lambda} = 2 \tag{28}$$

kde pro pořadové pravděpodobnosti jsou obvykle voleny hodnoty, $P_i = 2^{-i}$, $i = 2, 3$. K porovnání průběhu experimentálního bodu s ideálním (teoretickým) pro zvolené $\lambda$ se do grafu zakreslují i řešení rovnice $y^{\lambda} + x^{-\lambda} = 2$ pro $0 \leq x \leq 1$ a $0 \leq y \leq 1$:

(a) pro $\lambda = 0$ je řešením přímka $y = x$,

(b) pro $\lambda < 0$ je řešením vztah $y = \left(2 - x^{-\lambda}\right)^{1/\lambda}$,

(c) pro $\lambda > 0$ je řešením vztah $x = \left(2 - y^{\lambda}\right)^{-1/\lambda}$.

Podle umístění experimentálních bodů na teoretických křivkách selekčního grafu lze odhadovat velikost $\lambda$ a posuzovat kvalitu transformace v různých vzdálenostech od mediánu.

Pro přiblížení rozdělení výběru k rozdělení normálnímu vzhledem k šikmosti a špičatosti se užívá Boxovy-Coxovy transformace. [13]

**1.6.1. Box-Coxova transformace** Nevýhody prosté mocninné transformace (zejména nespojitost v okolí nuly a nesrovnatelnost měřítek v transformaci) odstraňuje rodina Box-Coxových transformací $X^{(\lambda)}$, která je lineární transformací prosté mocninné transformace $X_p^{(\lambda)}$. Box-Coxova třída polynomických transformací má tvar

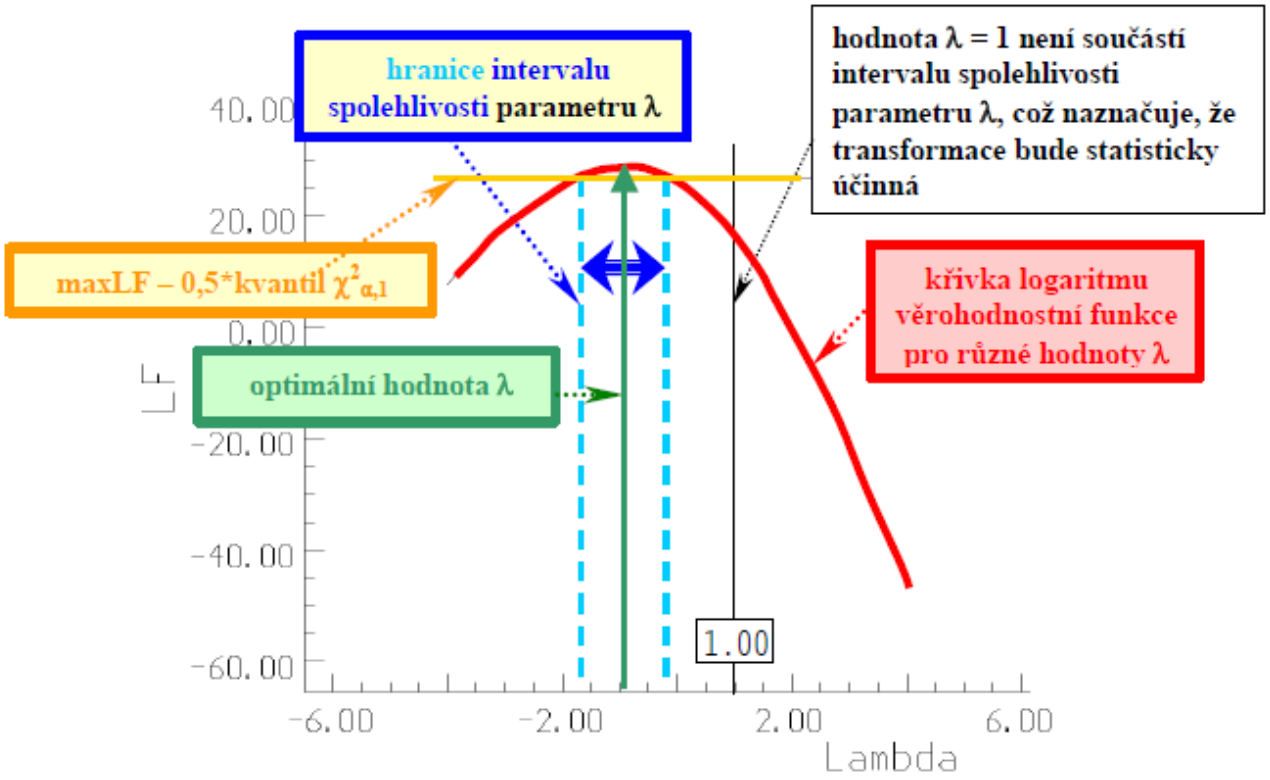
$$X^{(\lambda)} = \begin{cases} \dfrac{X^\lambda - 1}{\lambda} & \text{pro } \lambda \neq 0, \\ \ln X & \text{pro } \lambda = 0. \end{cases} \tag{29}$$

Pro posouzení kvality transformace, resp. nalezení optimálního $\lambda$, je také možno použít grafu rozptýlení s kvantily (GRK), resp. kvantilových grafů (Q-Q grafů). Výhodnější je použití testů normality dat po mocninné transformaci. Známý Shapiro-Wilksův test je úměrný testu významnosti směrnice v Q-Q grafu, takže lze také posuzovat linearitu v Q-Q grafech. Tato rodina Box-Coxových transformací závisí na jediném parametru $\lambda$, který může být odhadnut metodou maximální věrohodnosti (MLE) nebo metodou nejmenších čtverců (LSE). Pro $\lambda = 1$ resultuje aditivní model měření a pro $\lambda = 0$ model multiplikativní. Nejprve se ze zvolené oblasti volí hodnota $\lambda$. Pro zvolené $\lambda$ vypočítáme

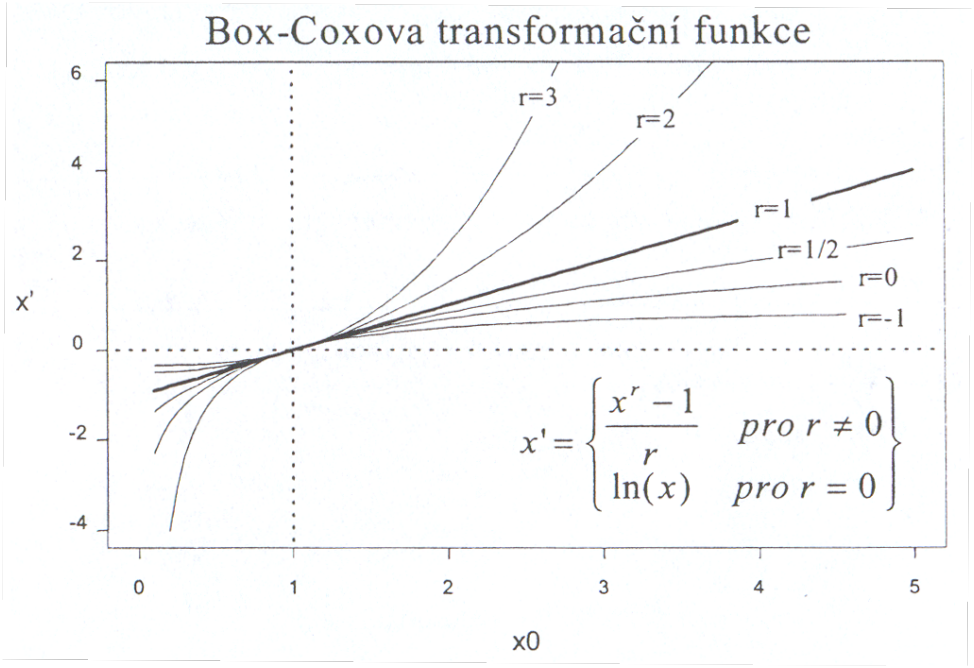
$$L_{\max} = -\frac{1}{2} \ln \widehat{\sigma}^2 + \ln J(\lambda, X) = -\frac{1}{2} \ln \widehat{\sigma}^2 + (\lambda - 1) \sum_{i=1}^{n} \ln X_i, \tag{30}$$

kde $J(\lambda, X) = \prod_{i=1}^{n} \frac{\partial W_i}{\partial X_i} = \prod_{i=1}^{n} X_i^{\lambda-1}$, pro všechna $\lambda$, tak, aby $\ln J(\lambda, X) = (\lambda - 1) \sum_{i=1}^{n} \ln X_i$. Odhad $\widehat{\sigma}^2$ pro pevné $\lambda$ je $\widehat{\sigma}^2 = S(\lambda)/n$, kde $S(\lambda)$ je reziduální součet čtverců v analýze rozptylu $X(\lambda)$. Po vyčíslení $L_{\max}(\lambda)$ pro několik různých hodnot $\lambda$ v rozsahu mohou být hodnoty $L_{\max}(\lambda)$ vyneseny proti $\lambda$. Maximálně věrohodný odhad $\lambda$ je získán z hodnoty $\lambda$, která maximalizuje funkci $L_{\max}(\lambda)$. S optimální hodnotou $\lambda^*$ každá $X$ hodnota specifikačních mezí je transformována na normální veličinu pomocí rovnice (29). [3], [14]

Odpovídající hodnoty PCI jsou počítány z průměru a směrodatné odchylky transformovaných dat pomocí rovnic (1) a (2). Box-Coxova transformace odhaduje hodnotu $\lambda$, která minimalizuje směrodatnou odchylku normalizované transformované proměnné. Software prozkoumá mnoho transformací, z nichž jsou uvedeny jen ty, které jsou pro zaokrouhlené hodnoty $\lambda$. $y$ je transformovaná hodnota proměnné (naměřeného znaku kvality) $x$. [3]




Obrázek 4: Nelineární transformace s optimální hodnotou $\lambda^*$. Zdroj: [15].




Obrázek 5: Tvar Box-Coxovy transformační funkce pro parametr $r = -1,\ 0,\ 0{,}5,\ 1,\ 2,\ 3$. Zdroj: [16].

Parametr $\lambda$ určuje tvar transformační funkce a umožňuje nalezení takového tvaru $F(x)$, který vyhoví podmínce maximální normality, popř. symetrie. Tvary Box-Coxovy transformační funkce pro různá $\lambda$ (v grafu znázorněna jako $r$) jsou zřejmé z obrázku 5 na straně 26. Cílem Box-Coxovy transformace je tedy nalézt takovou hodnotu $r$, která po transformaci zajistí maximální symetrii, nebo (lépe) maximální normalitu dat. Symetrii lze „měřit" šikmostí, normalitu věrohodností. Iterativním postupem lze tedy najít maximálně symetrickou (podmínka nulové šikmosti) nebo maximálně věrohodnou (podmínka maximální věrohodnosti) transformační funkci. [17]

**1.6.2. Zpětná transformace po Box-Coxově a mocninné transformaci** Pokud se podaří nalézt vhodnou transformaci, která vede k přibližné normalitě, lze určit $\overline{y}$ a $s^2(y)$, interval spolehlivosti $\overline{y} \pm t_{1-\alpha/2}\,(n-1)\,s\,(y)\,/\sqrt{n}$ a provádět i statistické testování. Problém však spočívá v tom, že všechny statistické charakteristiky je třeba pro původní proměnné.

1. **Nekorektní přístup** spočívá v prosté zpětné transformaci $\overline{x}_R = g^{-1}\,(\overline{y})$. Pro jednoduchou mocninnou transformaci vede zpětná transformace na obecný průměr definovaný vztahem

$$\overline{x}_R = \overline{x}_\lambda = \left[ \frac{\sum_{i=1}^{n} {x_i}^\lambda}{n} \right]^{1/\lambda} \tag{31}$$

   Pro $\lambda = 0$ se místo $x^\lambda$ používá $\ln x$ a místo $x^{1/\lambda}$ pak $e^x$. Hodnota $\overline{x}_R = \overline{x}_{-1}$ představuje harmonický průměr, $\overline{x}_R = \overline{x}_0$ geometrický průměr, $\overline{x}_R = \overline{x}_1$ aritmetický průměr a $\overline{x}_R = \overline{x}_2$ kvadratický průměr. Tento způsob zpětné transformace vede často ke zkreslujícím výsledkům.

2. **Korektní (přesnější) přístup** zpětné transformace vychází z Taylorova rozvoje funkce $y = g\,(x)$ v okolí $\overline{y}$. Pro netransformovaný průměr $\overline{x}_R$ lze pak odvodit přibližný vztah

$$\overline{x}_R \approx g^{-1}\left[ \overline{y} - \frac{1}{2} \frac{\mathrm{d}^2 g\,(x)}{\mathrm{d}\,x^2} \left( \frac{\mathrm{d}\,g\,(x)}{\mathrm{d}\,x} \right)^{-2} s^2\,(y) \right] \tag{32}$$

   Pro rozptyl vyjde

$$s^2\,(x_R) \approx \left( \frac{\mathrm{d}\,g\,(x)}{\mathrm{d}\,x} \right)^{-2} s^2\,(y)\,. \tag{33}$$

Zde jsou jednotlivé derivace vyčísleny v bodě $x = \overline{x}_R$.

Pro $100\,(1-\alpha/2)\,\%$ interval spolehlivosti střední hodnoty původního souboru dat platí $\overline{x}_R - I_D \leq \mu \leq \overline{x}_R + I_H$, kde

$$I_D = g^{-1}\left(\overline{y} + G - t_{1-\alpha/2}\,(n-1)\,\frac{s\,(y)}{\sqrt{n}}\right) \tag{34}$$

$$I_H = g^{-1}\left(\overline{y} + G + t_{1-\alpha/2}\,(n-1)\,\frac{s\,(y)}{\sqrt{n}}\right) \tag{35}$$

$$G = -0{,}5\frac{\mathrm{d}^2 g\,(x)}{\mathrm{d}\,x^2}\left(\frac{\mathrm{d}\,g\,(x)}{\mathrm{d}\,x}\right)^{-2} s^2\,(y) \tag{36}$$

Symbolem $t_{1-\alpha/2}\,(n-1)$ je označen $100\,(1-\alpha/2)\,\%$ kvantil Studentova rozdělení s $n-1$ stupni volnosti. Při znalosti hodnot konkrétní transformace $y = g\,(x)$ a odhadů $\overline{y}$, $s^2\,(y)$ je snadné vyčíslit hodnoty $\overline{x}_R$ a $s^2\,(x_R)$:

(a) Pro speciální případ $\lambda = 0$, tzn. logaritmickou transformaci typu $g\,(x) = \ln x$, bude $\overline{x}_R \approx \exp\,[\overline{y} + 0{,}5 s_2\,(y)]$.
Rozptyl se určí $s^2\,(x_R) \approx \overline{x}_R^2 s^2\,(y)$.

(b) Pro případ $\lambda \neq 0$ a Boxovy-Coxovy transformace bude $\overline{x}_R$ jedním z kořenů kvadratické rovnice, pro které platí

$$\begin{aligned}\overline{x}_{R,1,2} = \Big[&0{,}5\,(1+\lambda\overline{y}) \pm \\ &\pm 0{,}5\,\sqrt{1 + 2\lambda\big(\overline{y} + s^2\,(y)\big) + \big(\lambda^2 - 2s^2\,(y)\big)}\Big]^{1/\lambda}\end{aligned} \tag{37}$$

Jako odhad $\overline{x}_R$ se pak bere kořen $\overline{x}_{R,i}$, který je nejbližší mediánu $\overline{x}_{0,5} = g^{-1}\left(\overline{y}_{0,5}\right)$. Při znalosti netransformovaného průměru lze vyčíslit i odpovídající rozptyl $s^2\,(x) = {\overline{x}_R}^{-2\lambda+2} s^2\,(y)$. [13]

**1.6.3. Johnsonova transformace** Pokud nejsou hodnoty sledovaného znaku kvality normálně rozdělené, je možno použít Johnsonovu transformaci tak, že nová transformovaná data jsou potom rozdělena normálně $N(0,1)$. Obecná forma transformace je dána vztahem

$$z = \gamma + \eta\tau\,(x;\varepsilon,\lambda)\,,\ \eta > 0,\ -\infty < \gamma < \infty,\ \lambda > 0,\ -\infty < \varepsilon < \infty, \tag{38}$$

kde $z$ je standardizovaná normální náhodná veličina a $x$ je proměnná nafitována pomocí Johnsonova rozdělení. Čtyři parametry $\gamma$, $\eta$, $\varepsilon$ a $\lambda$, mají být

odhadovány, a $\tau$ je libovolná funkce, která může mít jen jednu z následujících tří forem. Johnsonova transformace tedy vybere jeden ze tří typů rovnic v závislosti na tom, zda náhodná veličina je „ohraničená“, „lognormální“ nebo „neohraničená“. [18], [19], [20]

**Lognormální soustava** $(S_L)$

$$\tau_1(x;\varepsilon,\lambda) = \log\left(\frac{x-\varepsilon}{\lambda}\right),\ x \geq \varepsilon. \tag{39}$$

Toto je Johnsonovo $S_L$ rozdělení, které se vztahuje k lognormální soustavě. Požadované odhady parametrů jsou

$$\widehat{\eta} = 1{,}645\left[\log\left(\frac{x_{0,95}-x_{0,50}}{x_{0,50}-x_{0,05}}\right)\right]^{-1}, \tag{40}$$

$$\widehat{\gamma}^* = \widehat{\eta}\log\left(\frac{1-\exp\left(-1{,}645/\widehat{\eta}\right)}{x_{0,50}-x_{0,05}}\right), \tag{41}$$

$$\widehat{\varepsilon} = x_{0,5} = \exp\left(-\widehat{\gamma}^*/\widehat{\eta}\right), \tag{42}$$

kde $100\alpha\%$ percentil je získán jako $\alpha\,(n+1)$ hodnota z $n$ pozorování. Pokud je to nutné, může být lineární interpolace mezi dvěma sousedními hodnotami použita k určení požadovaného percentilu. [19], [20]

**Neohraničená soustava** $(S_U)$

$$\tau_2(x;\varepsilon,\lambda) = \sinh^{-1}\left(\frac{x-\varepsilon}{\lambda}\right),\ -\infty < x < \infty. \tag{43}$$

Křivky v $S_U$ soustavě jsou neomezené. Tato soustava zahrnuje mimo jiné $t$ a normální rozdělení. Hahn a Shapiro poskytli pro fitování tohoto rozdělení tabulku pro stanovení $\widehat{\gamma}$ a $\widehat{\eta}$ na základě daných hodnot šikmosti a špičatosti.

**Ohraničená soustava** $(S_B)$

$$\tau_3(x;\varepsilon,\lambda) = \log\left(\frac{x-\varepsilon}{\lambda+\varepsilon-x}\right),\ \varepsilon \leq x \leq \varepsilon+\lambda. \tag{44}$$

$S_B$ soustava zahrnuje ohraničená rozdělení, která obsahují gamma a beta rozdělení. Vzhledem k tomu, že rozdělení může být ohraničeno buď u dolního konce $\varepsilon$, horního $\varepsilon+\lambda$, nebo obou, vede to k následujícím situacím.

- I. případ. *Rozsah kolísání je znám.* Pro případ, kdy jsou známy hodnoty obou koncových bodů, jsou parametry získané jako

$$\widehat{\eta} = \frac{z_{1-\alpha'} - z_{\alpha}}{\log\left(\frac{\left(x_{1-\alpha'}-\varepsilon\right)\left(\varepsilon+\lambda-x_{\alpha}\right)}{\left(x_{\alpha}-\varepsilon\right)\left(\varepsilon+\lambda-x_{1-\alpha'}\right)}\right)}, \tag{45}$$

$$\widehat{\gamma} = z_{1-\alpha'} - \widehat{\eta}\log\left(\frac{x_{1-\alpha'}-\varepsilon}{\varepsilon+\lambda-x_{1-\alpha'}}\right). \tag{46}$$

- II. případ. *Jeden známý koncový bod.* V tomto případě je dodatečná rovnice získána odpovídajícími mediány dat potřebnými k doplnění rovnice (45) a (46). Tato rovnice je dána vztahem

$$\begin{aligned}\widehat{\lambda} = (x_{0,5}-\varepsilon)\Big\{&(x_{0,5}-\varepsilon)(x_{\alpha}-\varepsilon)+(x_{0,5}-\varepsilon)(x_{1-\alpha}-\varepsilon)-\\ &-2(x_{\alpha}-\varepsilon)(x_{1-\alpha}-\varepsilon)\Big\}\cdot\Big\{(x_{0,5}-\varepsilon)^2-(x_{\alpha}-\varepsilon)(x_{1-\alpha}-\varepsilon)\Big\}^{-1}.\end{aligned} \tag{47}$$

- III. případ. *Žádný známý koncový bod.* V případě, kdy není znám ani jeden koncový bod, čtyři datové percentily musí být srovnány s odpovídajícími percentily normovaného normálního rozdělení. Výsledná rovnice pro $i = 1, 2, 3, 4$,

$$z_i = \widehat{\gamma} + \widehat{\eta}\log\left(\frac{x_i-\widehat{\varepsilon}}{\widehat{\varepsilon}+\widehat{\lambda}-x_i}\right), \tag{48}$$

  je nelineární a musí být řešena pomocí numerických metod.

Algoritmus, který vyvinul Hill a kol., se používá k přiřazení prvních čtyř momentů veličiny $X$ k výše uvedeným typům rozdělení. PCI jsou vypočteny pomocí rovnice (1) a (2). [19], [20]

## 1.7. Wrightův procesní index způsobilosti $C_s$

Wright navrhl index způsobilosti procesu $C_s$, který bere šikmost v úvahu začleněním šikmostního korekčního faktoru ve jmenovateli $C_{pmk}$, a je definován jako

$$C_s = \frac{\min\left(USL-\mu,\ \mu-LSL\right)}{3\sqrt{\sigma^2+\left(\mu-T\right)^2+|\mu_3/\sigma|}} = \frac{\min\left(USL-\mu,\ \mu-LSL\right)}{3\sqrt{\sigma^2+|\mu_3/\sigma|}}, \tag{49}$$

kde $T = \mu$ a $\mu_3$ je třetí centrální moment. [5]

Některé z výše popsaných metod byly a jsou široce používány v průmyslu, jako například pravděpodobnostní grafy a Clementsova metoda, avšak například metoda jako je Box-Coxova transformace je pro odborníky z oblasti kvality relativně neznámý pojem. Je třeba poznamenat, že když je základní rozdělení normální, teoreticky, všechny výše uvedené metody vyjímaje metody nezávislé na rozdělení, by měly dát stejný výsledek jako tradiční $C_p$ a $C_{pk}$ uvedené v rovnicích (1) a (2). Nicméně, stejně, i když se používají různé statistiky anebo různé způsoby odhadu souvisejících statistik, výsledné odhady $C_p$ a $C_{pk}$ projeví určitou variabilitu. Zejména variabilita u $C_p$ by mohla být snížena s rostoucí velikostí vzorku jako v případě jeho rozdělení $\chi^2_{1-\alpha}$ za předpokladu normality. Nicméně, variabilita indexu $C_{pk}$ může být docela významná pro všechny přijatelné rozsahy výběrů, což také závisí na variabilitě procesu způsobené posunem střední hodnoty či jiné změně.

# 2. Případová studie

## 2.1. Metrika pro srovnání metod

Samotná skutečnost, že různé srovnávací metriky mohou vést k odlišným závěrům, zdůrazňuje, že je nutné určit vhodné měřítko ke srovnání výkonnosti devíti výše uvedených metod. V literatuře různí výzkumníci využili rozdílná měřítka při řešení problému nenormality u PCI. English a Taylor použili pevné hodnoty $C_p$ a $C_{pk}$ (roven 1,0) pro všechny jejich simulace při vyšetřování robustnosti indexů způsobilosti procesu v případě nenormality. Základem pro jejich srovnání byl vzájemný poměr $\widehat{C}_p$ a $\widehat{C}_{pk}$ (odhadnuty ze simulace) větší než 1,0 v případě normálního rozdělení. Deleryd se zaměřil na podíl nevyhovujících položek k odvození odpovídající hodnoty $C_p$. Vychýlení a rozptyl odhadovaných hodnot $C_p\,(u, v)$ jsou porovnávány s cílovými hodnotami $C_p$. Jelikož existuje přímá souvislost mezi indexem způsobilosti a zmetkovitostí, je volbou indexu stanoveno i přípustné procento zmetků (nevyhovujících). Rivera a kol. měnili horní specifikační meze základního rozdělení k získání skutečného počtu nevyhovujících položek a odpovídajících hodnot $C_{pk}$. Odhadované $C_{pk}$ hodnoty vypočítané ze simulace transformovaných dat jsou porovnány s těmito cílovými hodnotami $C_{pk}$. V praxi jsou PCI běžně používány k monitorování výkonnosti a srovnání mezi různými procesy. Ačkoli, takovéto používání, aniž by se zkoumalo základní rozdělení, by nemělo být doporučováno, „správný“ náhradní PCI pro nenormální data by měl být kompatibilní s PCI vypočtených na základě normality, kdy příslušné podíly nevyhovujících jsou přibližně stejné. Výše uvedené zdůvodňuje návrh podobný tomu od Rivery a kol., kde podíl nevyhovujících je stanoven a pri-

ori pomocí vhodných specifikačních limitů a PCI vypočtené pomocí různých metod jsou pak porovnány s cílovou hodnotou. To vede k posouzení jednostranné tolerance, kde $C_{pu}$, jednostranný toleranční limit indexu způsobilosti, je používán jako měřítko srovnání v této simulační studii. Pro $C_{pu}$ hodnotu, zaměřenou na podíl nevyhovujících jednotek za předpokladu normality se určuje pomocí

$$\text{Podíl nevyhovujících } = \Phi\left(-3C_{pu}\right) \tag{50}$$

Procento nevyhovujících se vypočítá pro oboustrannou nesymetrickou toleranci (opět za předpokladu normality)

$$\text{Podíl nevyhovujících } = \Phi(-3C_{pl}) + \Phi(-3C_{pu}) \tag{51}$$

V následující simulační studii jsou použity cílové hodnoty $C_{pu} = 1{,}0;\ 1{,}5$ a 1,667, a jsou získány odpovídající hodnoty $USL$ pro lognormální a Weibullovo rozdělení se stejným procentem nevyhovujících. Tyto hodnoty $USL$ jsou potom použity pro odhad indexu $C_{pu}$ vztahující se k různým metodám ze simulovaných dat. Tyto odhadované indexy $C_{pu}$ jsou potom porovnány s cílovými hodnotami $C_{pu}$. Lepší metoda je ta, jejíž výběrový průměr odhadovaného $C_{pu}$ má nejmenší odchylku od cílové hodnoty (referenční) a má co nejmenším rozptyl odhadovaných hodnot indexů $C_{pu}$. Grafické znázornění, které příhodně zobrazuje charakteristiky polohy a variability, je jednoduchý box-and-whisker graf (krabicový graf).

**2.1.1. Testovací rozdělení** K vyšetření účinku nenormálních dat na indexy způsobilosti procesu jsou používány lognormální a Weibullovo rozdělení, neboť je známo, že mají hodnoty parametrů, které mohou představovat nepatrnou, střední a velkou odchylku od normality. Tato rozdělení jsou také známa tím, že mají výrazně odlišné vlastnosti na „chvostu“, které výrazně ovlivňují způsobilost procesu. Hodnoty šikmosti a špičatosti pro obě tato rozdělení jsou v tabulce 2. Distribuční funkce lognormálního a Weibullova rozdělení jsou dle uvedeného pořadí dány vztahy

$$F\left(x;\mu,\sigma\right) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right),\ x \geq 0, \tag{52}$$

$$F\left(x;\eta,\sigma\right) = 1 - \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right],\ x \geq 0. \tag{53}$$

Následující obrázky ukazují dle uvedeného pořadí tvary lognormálního a Weibullova rozdělení, které jsou v simulaci použity.

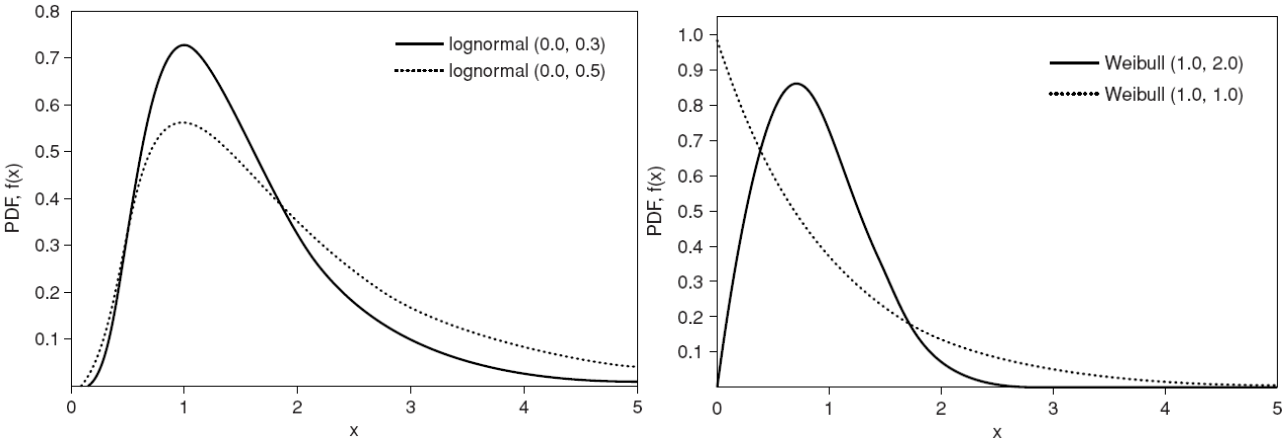
Tabulka 1: Kombinace parametrů používané v simulacích pro porovnávání výkonnosti.

| Lognormální rozdělení | | Weibullovo rozdělení | |
|---|---|---|---|
| Parametr $\mu$ | Parametr $\sigma^2$ | Parametr tvaru $\eta$ | Parametr měřítka $\sigma$ |
| 0,0 | 0,1 | **1,0** | **1,0** |
| **0,0** | **0,3** | **2,0** | **1,0** |
| **0,0** | **0,5** | 4,0 | 1,0 |
| 1,0 | 0,5 | 0,5 | 1,0 |

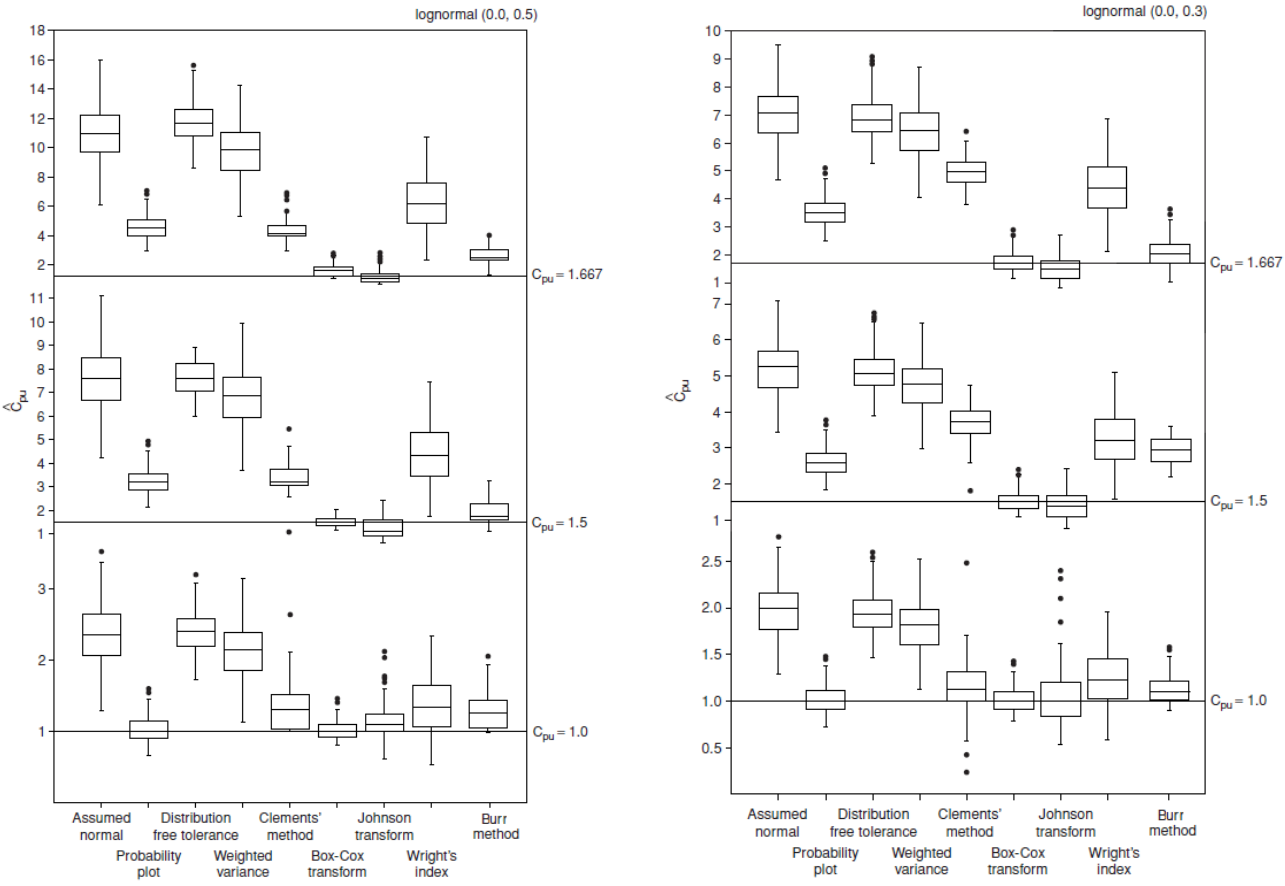
## 2.2. Simulace metodou Monte Carlo

Byla provedena série simulací s rozsahem výběru $n = 50$, 75, 100, 125 a 150 a s cílovými hodnotami $C_{pu} = 0{,}1$, 1,33, 1,5, 1,667 a 1,8 pomocí lognormálního a Weibullova rozdělení. Uvádíme také vypočtené výsledky standardních PCI za předpokladu normality dat. Zde uvádíme pouze výsledky simulace s $n = 100$ a cílovou hodnotou $C_{pu} = 1{,}0$, 1,5 a 1,667. V každém běhu simulace byly z náhodných proměnných generovaných z příslušných rozdělení získány nezbytné statistiky vyžadované v různých metodách, jako je $\overline{x}$, $s$, horní a dolní 0,135% percentil a medián. Pro metody zahrnující transformaci jsou tyto statistiky v podstatě získány z transformovaných hodnot. Odhady pro $C_{pu}$, byly stanoveny pomocí výše uvedených devíti metod, z těchto odhadovaných parametrů pro jednotlivé cílové hodnoty $USL$. [3], [14], [21]

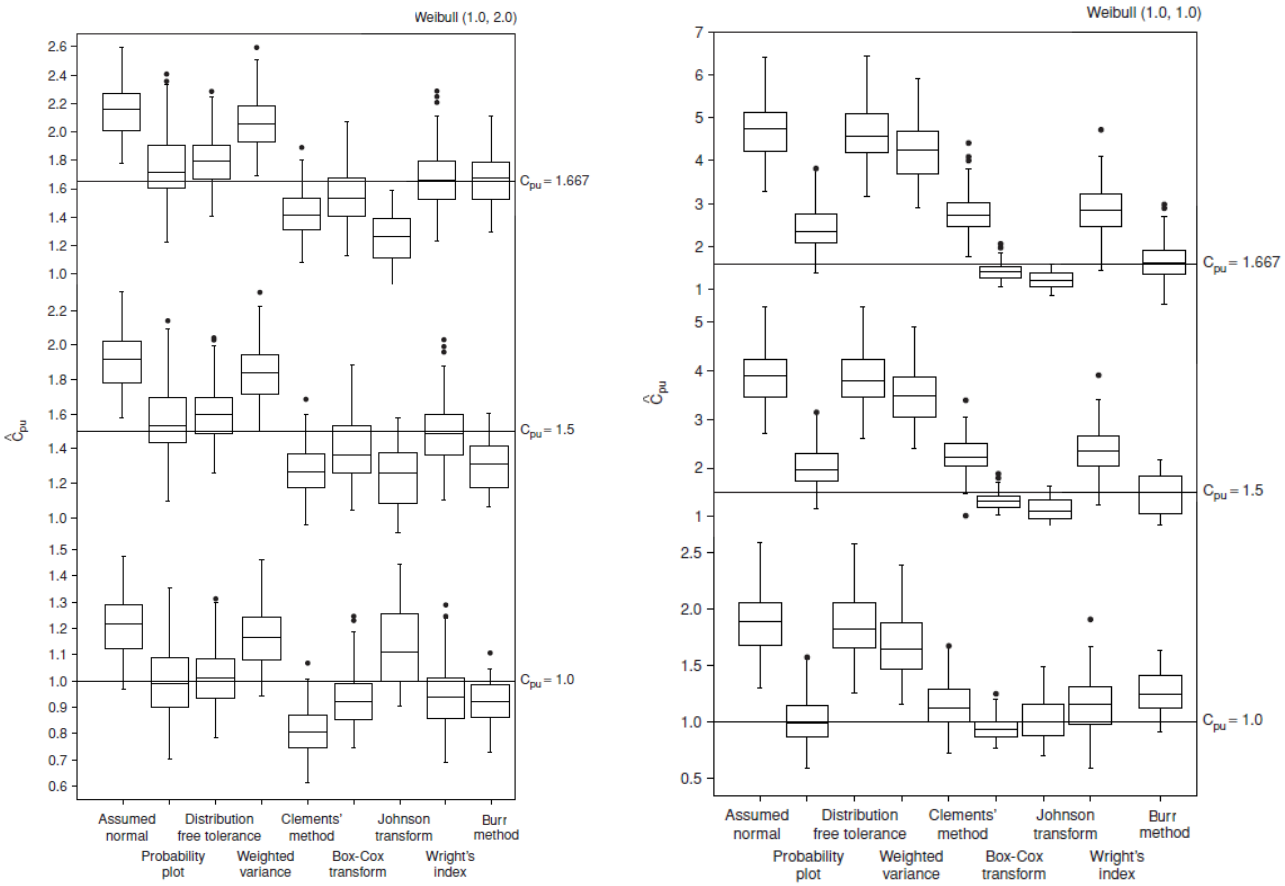
Každý běh byl proveden 100krát pro získání průměrné hodnoty $\overline{C}_{pu}$ ze 100 hodnot $\widehat{C}_{pu}$. K vyšetření, která metoda je nejlepší pro nakládání s nenormalitou, budeme porovnávat pomocí krabicových grafů, ve kterých bude zobrazena hodnota $\widehat{C}_{pu}$ pro všech devět metod v každé cílové hodnotě $C_{pu}$ a pro obě rozdělení. Tyto krabicové grafy jsou schopny graficky zobrazit několik důležitých statistických charakteristik týkajících se simulovaných hodnot $C_{pu}$, jako je průměr a rozptyl (v boxplotu je většinou poloha charakterizována mediánem, variabilita pak mezikvartilovým rozpětím), odlehlé a extrémní hodnoty. Pro metody, které by mohly zachytit cílovou hodnotu $C_{pu}$, bude v krabicových grafech průměrná hodnota protínat horizontální čáru na úrovni cílové hodnoty. [3]



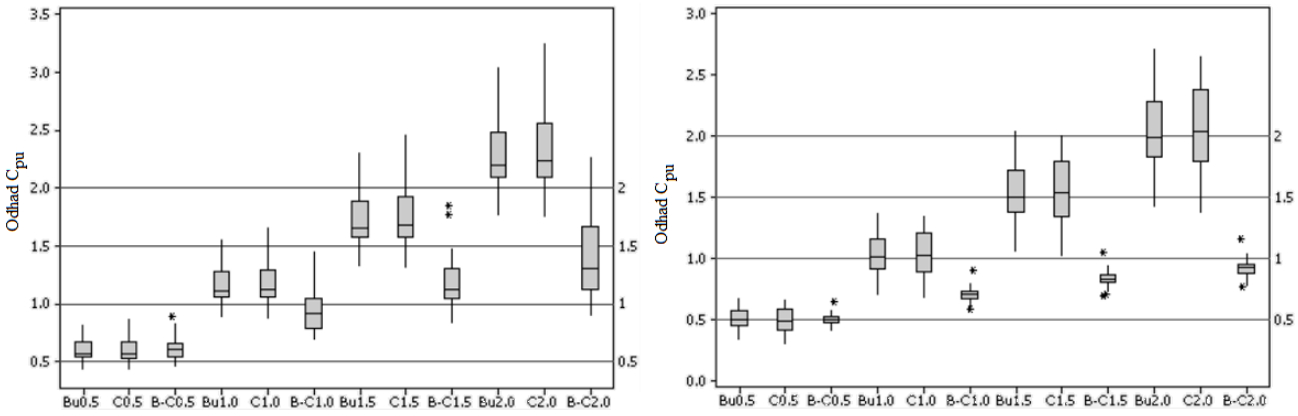

Obrázek 6: Hustota pravděpodobnosti lognormálního a Weibullova rozdělení s parametry z tabulky 1. Zdroj: inspirace z [3].




Obrázek 7: Krabicové grafy indexu pro všechny metody (u každé je 100 pozorování) pro lognormální rozdělení ($\mu = 0,\ \sigma^2 = 0{,}5,\ 0{,}3$) s cílovou hodnotou $C_{pu} = 1{,}0,\ 1{,}5,\ 1{,}667$.




Obrázek 8: Krabicové grafy indexu $\widehat{C}_{pu}$ pro všechny metody (u každé je 100 pozorování) pro Weibullovo rozdělení ($\sigma = 0{,}0,\ 1{,}0,\ \eta = 1{,}0,\ 2{,}0$) s cílovou hodnotou $C_{pu} = 1{,}0,\ 1{,}5,\ 1{,}667$.




Obrázek 9: Krabicové diagramy odhadů $C_{pu}$ hodnot s cílovou $C_{pu}$ hodnotou 0,5, 1,0, 1,5 a 2,0 pro Weibullovo rozdělení (vlevo) a krabicové diagramy odhadů $C_{pu}$ hodnot s cílovou $C_{pu}$ hodnotou 0,5, 1,0, 1,5 a 2,0 pro lognormální rozdělení (vpravo).

Tabulka 2: Šikmost a špičatost rozdělení vyšetřovaných hodnot

**Lognormální rozdělení**

| Střední hodnota | Rozptyl | Šikmost $\sqrt{\beta_1}$ | Špičatost ($\beta_2$) |
|---|---|---|---|
| 0,0 | 0,5 | 2,9388 | 21,5073 |
| 0,0 | 0,3 | 1,9814 | 10,7057 |
| 0,0 | 0,1 | 1,0070 | 4,8558 |
| 1,0 | 0,5 | 2,9388 | 21,5073 |

**Weibullovo rozdělení**

| Parametr tvaru | Parametr měřítka | Šikmost $\sqrt{\beta_1}$ | Špičatost ($\beta_2$) |
|---|---|---|---|
| 1,0 | 1,0 | 2,0000 | 9,0000 |
| 2,0 | 1,0 | 0,6311 | 3,2451 |
| 4,0 | 1,0 | −0,0872 | 2,7478 |
| 0,5 | 1,0 | 2,0000 | 9,0000 |

## Závěr a diskuze k dosaženým výsledkům

Devět použitých metod je uvedených v kapitole „Simulace metodou Monte Carlo“, je možné obecně rozdělit do dvou kategorií. Odhad $C_{pu}$ založený na předpokladu normality, pravděpodobnostní graf, tolerance nezávislá na rozdělení, metoda váženého rozptylu a Wrightův index jsou klasifikovány jako netransformační metody. Druhou kategorii tvoří transformační metody, do které spadá Clementsova metoda, Burrova metoda, Box-Coxova a Johnsonova transformace.

$C_{pu}$ jsme použili z toho důvodu, protože jsme výpočty indexů způsobilosti procesu testovali na Lognormálním a Weibullovu rozdělení, u kterých se při fitování v případě výrobních procesů předpokládá přirozená spodní mez, tedy nula. Proto odhad $C_{pu}$ bude stejný jako $C_{pk}$.

V části porovnávání výsledků simulace jsou dvě měřítka výkonnosti, konkrétně přesnost a správnost s ohledem na velikost vzorku. Pro přesnost se podíváme na rozdíl mezi simulovanou průměrnou hodnotou $\widehat{C}_{pu}$, $\overline{C}_{pu}$ a cílovou hodnotou $C_{pu}$. Pro správnost se podíváme na rozptyl nebo rozpětí simulovaných hodnot $\widehat{C}_{pu}$. Čím menší je rozptyl nebo rozpětí simulovaných hodnot $\widehat{C}_{pu}$, tím je lepší výkonnost dané metody. Z předchozích výsledků můžeme vypozorovat, že transformační metody jsou lepší než netransforma-

ční metody pro obě dvě zvolená nenormální rozdělení. Existují pouze dvě výjimky z tohoto tvrzení.

Za prvé, Clementsova metoda nepracuje stejně dobře jako další tři transformační metody. V případě Weibullova rozdělení, výkonnost Clementsovy metody je nižší než metoda pravděpodobnostního grafu.

Za druhé, z krabicových diagramů můžeme vypozorovat, že metoda pravděpodobnostního grafu je jediná netransformační metoda, která dokáže překonat transformační metody, ale pouze v případě Weibullova rozdělení s $\sigma = 1{,}0$ a $\eta = 2{,}0$. Pro lognormální rozdělení transformační metody důsledně poskytují $\overline{C}_{pu}$ hodnoty, které jsou blíže k cílové hodnotě $C_{pu}$, zatímco všechny ostatní metody tyto hodnoty výrazněji překračují. Praktický důsledek je ten, že ačkoliv netransformační metody jsou z hlediska jednoduchosti výpočtu atraktivnější, ukázaly se jako nedostatečné v zachycení způsobilosti procesu s výjimkou případů, kdy základní rozdělení je přibližně normální.

Navíc, rozdíly v simulovaných $\widehat{C}_{pu}$ hodnotách jsou převážně větší než u transformačních metod. Použití jiných netransformačních metod je jednoznačně nedostatečné pro rozdělení, která se výrazněji odchylují od normálního. To znamená, že pokud jedna z těchto metod selhává, je třeba pracovat s transformovanými empirickými daty, které se tím pádem budou více podobat normálnímu rozdělení a následně po výpočtu charakteristik tyto zpětně retransformovat. Výkonnost transformačních metod je poměrně konzistentní z hlediska správnosti a přesnosti. Je však třeba poznamenat, že transformační metody jsou velmi citlivé na velikost vzorku, jelikož rozdíly mezi odhady hodnot $\overline{C}_{pu}$ při $n = 50$ a 150 jsou 33 %, 11 % a 26 % pro Clementsovu, Box-Coxovu, a Johnsonovu transformační metodu. Tyto rozdíly jsou vyšší ve srovnání s méně než 10% rozdílem u netransformačních metod. Vzhledem k výše uvedenému, transformační metody vykazují menší variabilitu v $\widehat{C}_{pu}$, stejně jako velký rozdíl mezi hodnotami $\widehat{C}_{pu}$, vypočtenými z malých a velkých výběrů, tato skutečnost nemůže být připsána pouze přirozené statistické variabilitě. To naznačuje skutečnost, že transformační metody nejsou vhodné pro malé rozsahy výběrů. Některé možné důvody jsou následující.

1. Transformační metody obvykle zahrnují změnu již od měřítka rozsahu rozdělení do tvaru rozdělení. Tato skutečnost může vyvolat nežádoucí posun procesu, který narušuje $\widehat{C}_{pu}$, zvláště pak pro malé rozsahy vzorků.
2. Úspěch transformační metody pro účely analýzy způsobilosti závisí na její schopnosti zachytit chování v krajních částech – „chvostech“ – rozdělení (přes $USL$ percentil).

To znamená, že použití transformační metody při studiu procesní způsobilosti je vhodné, když bychom měli $n \geq 100$. Ale přesto nadřazenost Box-

Coxovy transformace je zřetelně vidět z krabicových diagramů, zejména pro lognormální rozdělení. Výše zmíněné je třeba očekávat, jelikož Box-Coxova transformační metoda poskytuje přesný převod na hodnoty přirozených logaritmů např. v případě lognormálního rozdělení platí přesně.

Srovnáme-li Box-Coxovu transformaci a Johnsonovu metodu, Box-Coxova metoda je přesnější. Box-Coxova metoda vyžaduje maximalizaci logaritmicko-věrohodnostní funkce k získání optimální hodnoty $\lambda$ a Johnsonova metoda vyžaduje odhady prvních čtyř momentů. Použití metody pravděpodobnostního grafu se doporučuje pro procesy, které vykazují pouze mírný odklon od normality. Navíc tato metoda detekuje různé anomálie v procesu, jako je bimodalita či odlehlé hodnoty. Další poznatek plynoucí ze simulace je ten, že Burrova metoda poskytuje srovnatelné výsledky jako Box-Coxova transformace. Podíváme se nyní, jak je tato metoda srovnatelná s Box-Coxovou transformací při menších výběrech. V závěru této části bylo nasimulováno 30 výběrů s $n = 50$ pro stanovení, která z transformačních metod je lepší pro menší výběry, které se v průmyslové praxi častěji vyskytují.

Simulace ukazuje, že Burrova metoda jako jediná z transformačních metod obecně umožňuje lepší odhad indexů způsobilosti procesu pro nenormální data při menších výběrech. Výsledky simulace Monte Carlo potvrdily, že výpočet klasických indexů způsobilosti procesu při nenormálně rozložených datech podávají zkreslené informace o procesu, jelikož tyto indexy způsobilosti procesu vycházejí z předpokladu, že procesní data mají normální rozdělení.

*Poděkování*: Příspěvek byl zpracován za podpory projektu financovaného z Operačního programu Vzdělávání pro konkurenceschopnost pod číslem jednacím OP VK CZ.1.07/2.3.00/20.0147 „Rozvoj lidských zdrojů v oblasti výzkumu měření a řízení výkonnosti podniků, klastrů a regionů“, který je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

## Literatura

[1] Pearn W. L., Kotz S.: Application of Clements' method for calculating second and third generation process capability indices for nonnormal Pearsonian populations. *Quality Engineering*, 7(1): 139–145, 1994. doi: 10.1080/08982119408918772

[2] Rivera L. A. R., Hubele N. F., Lawrence F. D.: $C_{pk}$ index estimation using data transformation. *Computers and Industrial Engineering,* 29(1): 55–58, 1995. doi: 10.1016/0360-8352(95)00045-3

[3] Tang L. C., Than S. E., Ang B. W.: A graphical approach to obtaining confidence limits of $C_{pk}$. *Quality and Reliability Engineering International,* 13(6): 337–346, 1997. doi: 10.1002/(SICI)1099-1638(199711/12)13:6<337::AID-QRE103>3.0.CO;2-Z

[4] Fabian F., Horálek V., Křepela J., Michálek J., Chmelík V., Chodounský J., Král J.: *Statistické metody řízení jakosti.* Praha, ČSJ, 2007. ISBN 978-80-02-01897-1.

[5] Wright P. A.: A process capability index sensitive to skewness. *Journal of Statistica Computation and Simulation,* 52: 195–203, 1995. doi: 10.1080/00949659508811673

[6] Yang K., El-Haik Basem S.: *Design For Six Sigma.* McGraw-Hill Education – Europe (United Kingdom), 2003. 624 s. ISBN 97871-41208-7.

[7] Choi I. S., Bai D. S.: Process capability indices for skewed populations. *Proceedings of the 20th International Conference on Computer and Industrial Engineering,* 1211–1214, 1996.

[8] Ishikawa K.: *What Is Total Quality Control? The Japanese Way.* Překlad Lu D. J. Englewood Cliffs, N. J.: Prentice-Hall, 1985.

[9] Clements J. A.: Process capability calculations for non-normal distributions. *Quality Progress,* 22(9): 95–100, 1989.

[10] Abouheif E.: A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research,* 1: 895–909. `http://biology.mcgill.ca/faculty/abouheif/articles/Abouheif%201999.pdf`

[11] Burr I. W.: Cumulative frequency functions. *Annals of Mathematical Statistics*, 13(2): 215–232, 1942. doi: 10.1214/aoms/1177731607

[12] Burr I. W.: Parameters for a general system of distribution to match a grid of $\alpha_3$ a $\alpha_4$. *Communications in Statistics*, 2(1): 1–21, 1973. doi: 10.1080/03610927308827052

[13] Meloun M., Militký J.: *Statistická analýza experimentálních dat.* 2. vyd. Praha: Academia, nakladatelství Akademie věd České republiky, 2004. 953 s. ISBN 80-200-1254-0.

[14] Cchan L. K., Cheng S. W., Spiring F. A.: A graphical technique for process capability. *ASQC Quality Congress Transactions,* Dallas, 268–275, 1992.

[15] Venkataraman P.: *Applied Optimization with Matlab Programming.* 2. vyd. Nakladatelství John Wiley & Sons, Inc., 2009. 526 s. ISBN 97870-08488-5.

[16] Kupka K.: *Statistické řízení jakosti.* 1. vydání. Pardubice: TriloByte, 2001. ISBN 80-238-1818-X.

[17] Meloun M., Militký J.: *Kompendium statistického zpracování dat.* 2. vyd. Praha: Academia, nakladatelství Akademie věd České republiky, 2006. 982 s. ISBN 80-200-1396-2.

[18] Hill I. D., Hill R., Holder R. L.: Fitting Johnson Curves by Moments: Algorithm AS 99. *Applied Statistics,* 25(2): 180–189, 1976. `http://www.jstor.org/stable/2346692`

[19] Johnson N. L., Kotz S., Pearn W. L.: Flexible process capability indices. *Pakistan Journal of Statistics*, 10: 23–31, 1992.

[20] Kane V. E.: Process capability indices. *Journal of Quality Technology,* 18: 41–52, 1986. `http://www.stat.ncsu.edu/information/library/mimeo.archive/isms_1992_2072.pdf`

[21] Dlouhý M., Fábry J., Kuncová M., Hladík T.: *Simulace podnikových procesů.* 1. vydání. Nakladatelství Computer Press, a. s., Brno, 2007. 201 s. ISBN 978-80-251-1649-4.

## Obsah

### Vědecké a odborné statě

*Zdeněk Půlpán, Jiří Kulička*
Jednoduchý fuzzy regresní model ..... 1

*Martin Kovářík*
Výpočet indexů způsobilosti procesu při nenormálně rozložených datech ..... 14

**Informační bulletin České statistické společnosti** vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214.

The Information Bulletin of the Czech Statistical Society is published quarterly.
The contributions in bulletin are published in English, Czech and Slovak languages.



**DOI: 10.5300/IB, http://dx.doi.org/10.5300/IB**
**ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)**

Toto číslo bylo vytištěno s laskavou podporou Českého statistického úřadu.