Výkonnost procesů v případě nenormálně rozděleného znaku kvality

Jiří Michálek

Hodnocení způsobilosti a výkonnosti výrobních procesů je prováděno především u dodavatelů do automobilového průmyslu, kde řada metod aplikované matematické statistiky již zdomácněla a je využívána v současné době pomocí celé řady softwarů domácích i cizích.

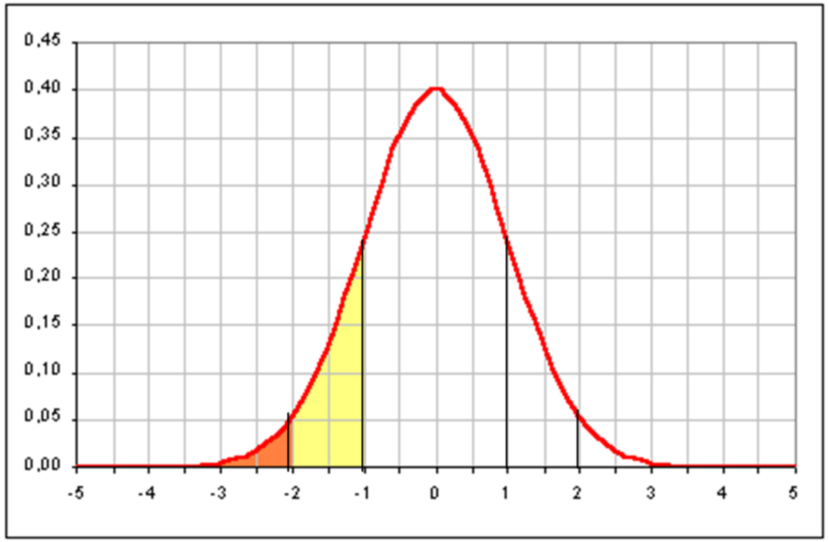
Bohužel ne vždy jsou výsledky získané statistickými metodami interpretovány správným způsobem a není tomu ani jinak u ukazatelů způsobilosti a výkonnosti. Celá řada softwarů na základě naměřených dat vrátí pouze odhady těchto ukazatelů, aniž by se mnohdy ověřily předpoklady pro jejich správné použití a rovněž závěry postavené pouze na hodnotách odhadů mohou znamenat nebezpečí pro odběratele ve větším počtu neshodných výrobků nežli je požadováno a pro dodavatele to může znamenat naopak přísnější požadavky na přesnost výrobního procesu.

Východiskem pro tuto prezentaci jsou zkušenosti a poznatky autorů ze spolupráce s výrobními podniky v rámci Centra pro jakost a spolehlivost výroby, Konzultačního střediska statistických metod NIS-PK a kurzů věnovaných statistickým metodám pořádaných Českou společností pro jakost.

Pozornost je věnována zejména těm otázkám, na které se v literatuře hledá odpověď jen těžko, jako postupy v případě nenormálně rozdělených náhodných veličin, měřených znaků kvality, což je v praxi velmi častá situace.

V řadě případů v praxi neexistuje teoreticky čisté řešení, často pro nesplnění některých předpokladů. Praktik ale musí vždy nějaké řešení zvolit, musí v konkrétním případě rozhodnout, dát na položenou otázku co nejlepší odpověď. Musí volit optimální, prakticky možné řešení ze všech, která se mu intuitivně nabízejí.

Autoři se snaží navrhovat řešení pro případy, kdy nejsou striktně splněny teorií stanovené předpoklady pro použití statistických postupů v řízení kvality a varují před jejich vědomým nerespektování. Jedná se nejčastěji o aplikaci statistických metod při nedodržení předpokladu normálního rozdělení sledovaného znaku kvality, případně nedodržení předpokladu statisticky zvládnutého procesu při implementaci Shewhartových regulačních diagramů a vyhodnocování způsobilosti a výkonnosti procesů.



V intervalu $\langle \mu - \sigma, \mu + \sigma \rangle$ leží **68,26 %** všech pozorování, mimo tento interval leží 2•15,87 %, t.j. 31,74 %. **Pp = 0,33**

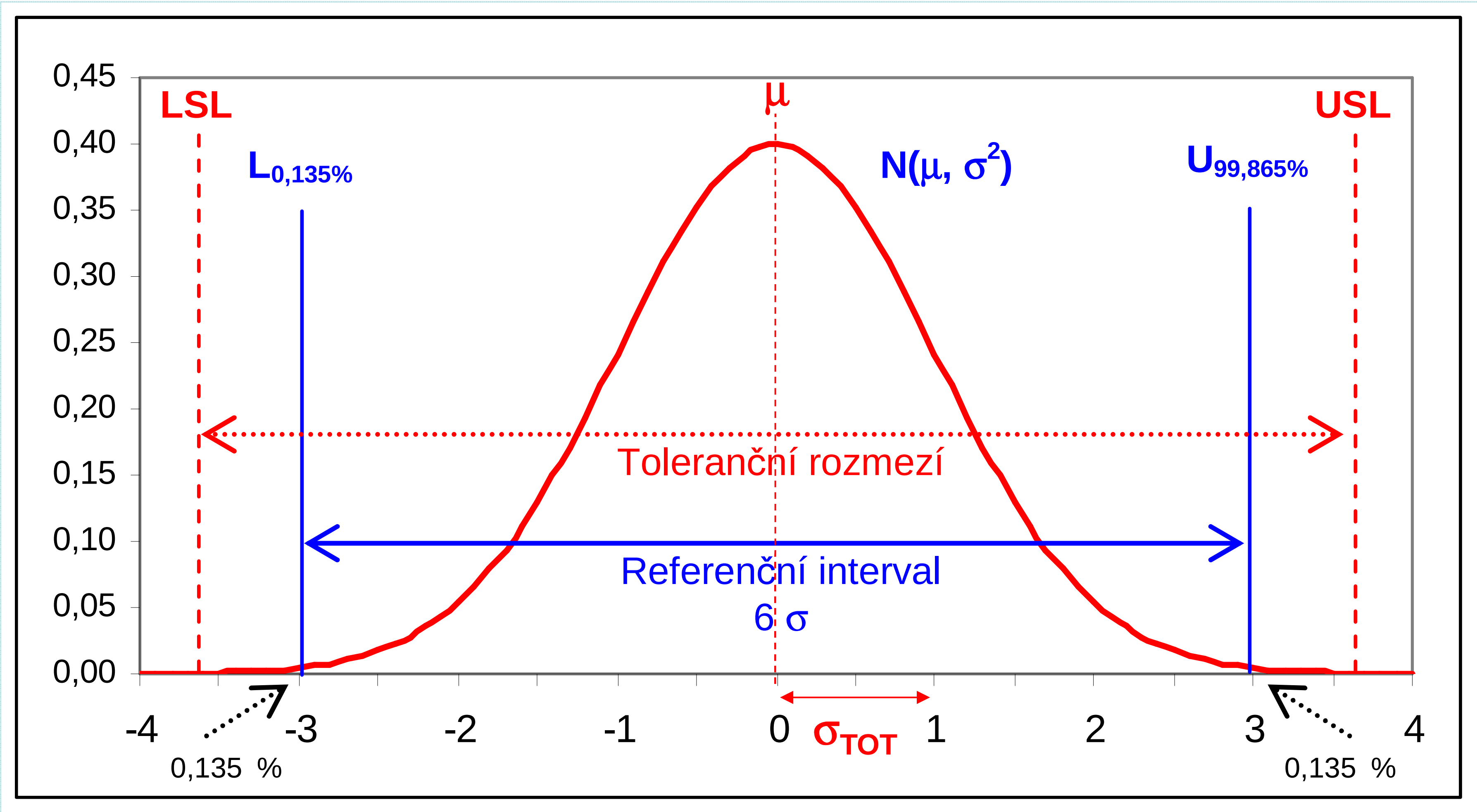
V intervalu $\langle \mu - 2\sigma, \mu + 2\sigma \rangle$ leží **95,44 %** všech pozorování, mimo tento interval leží 2•2,28 %, t.j. 4,56 %. **Pp = 0,67**

V intervalu $\langle \mu - 3\sigma, \mu + 3\sigma \rangle$ leží **99,73 %** všech pozorování, mimo tento interval leží 2•0,135 %, t.j. 0,27 % (2 700 ppm). **Pp = 1,0**

V intervalu $\langle \mu - 4\sigma, \mu + 4\sigma \rangle$ leží **99,994 %** všech pozorování, mimo tento interval leží 2•0,003 %, t.j. 0,006 % (60 ppm). **Pp = 1,33**

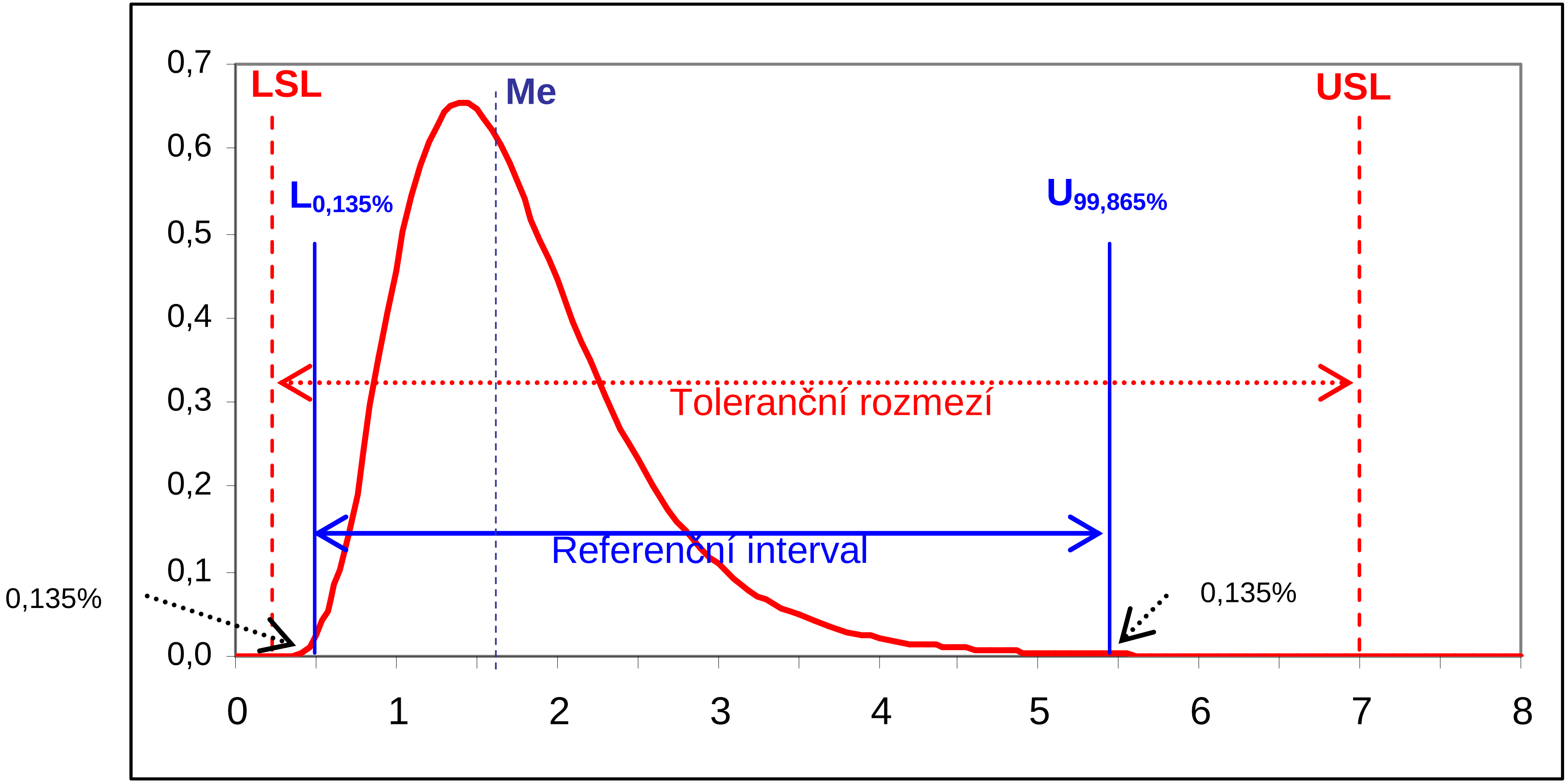
V intervalu $\langle \mu - 5\sigma, \mu + 5\sigma \rangle$ leží **99,99994** % všech pozorování, mimo tento interval leží 2•0,00003 %, t.j. 0,00006% (0,6 ppm). **Pp = 1,67**

V intervalu $\langle \mu - 6\sigma, \mu + 6\sigma \rangle$ leží **99,999999999** % všech pozorování, mimo tento interval leží 2•0,000000001%, t.j. 0,000000002% (0,002 ppm) **Pp = 2,0**



Ukazatele výkonnosti v případě normálně rozděleného znaku kvality

$$P_p = \frac{USL - LSL}{6\,\sigma_{TOT}} \qquad P_{pkU} = \frac{USL - \mu}{3\,\sigma_{TOT}}, \qquad P_{pkL} = \frac{\mu - LSL}{3\,\sigma_{TOT}};$$



Ukazatele výkonnosti v případě nenormálně rozděleného znaku kvality

$$P_p = \frac{USL - LSL}{U_{0,99865} - L_{0,00135}} \qquad P_{pU} = \frac{USL - Me}{U_{0,99865} - Me}, \qquad P_{pL} = \frac{Me - LSL}{Me - L_{0,00135}}$$

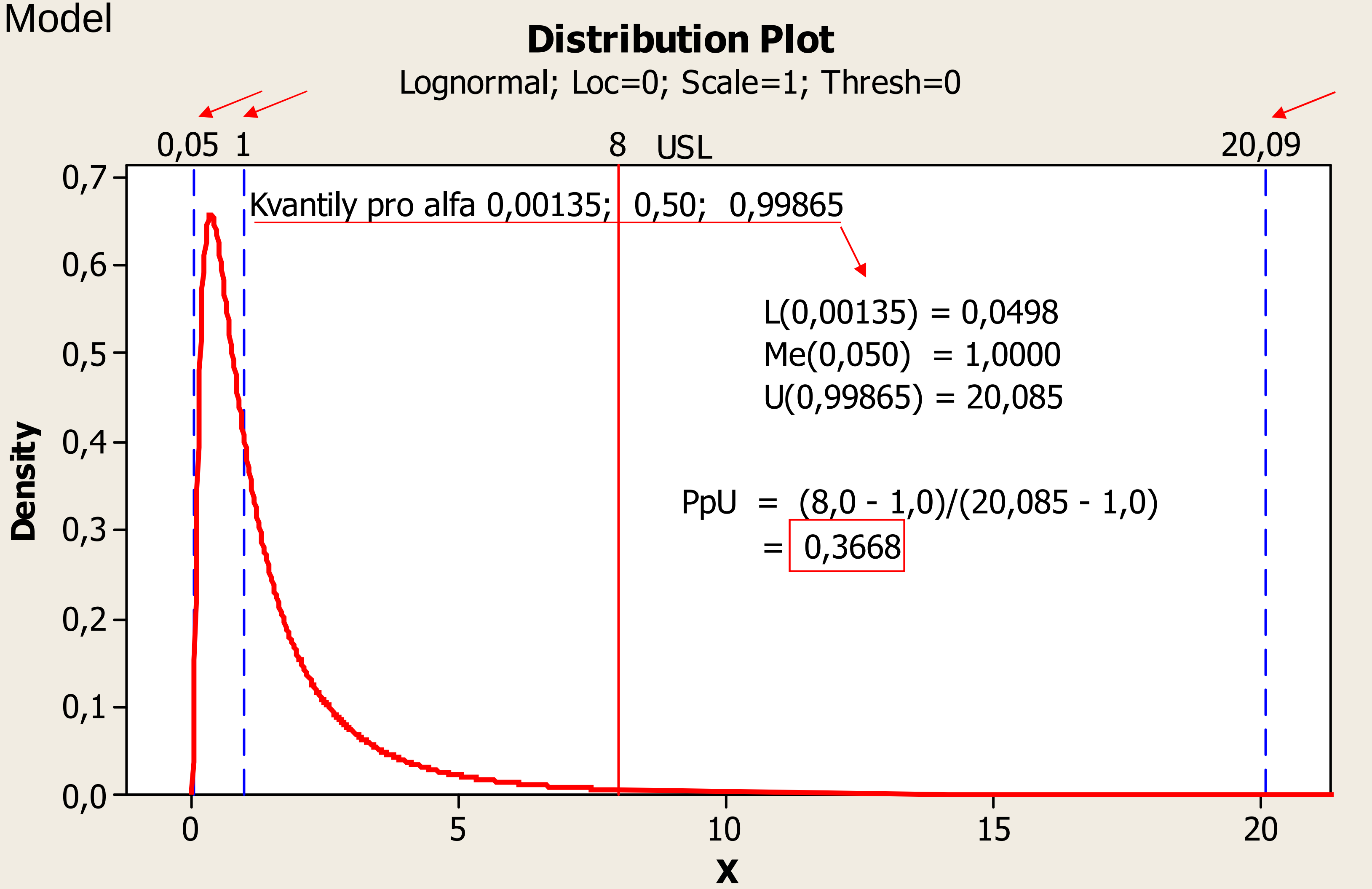
Známé nenormální rozdělení

Někdy je tvar rozdělení sledovaného znaku jakosti 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.

Příklad: Log-normální rozdělení sledovaného znaku kvality s předepsanou horní mezní hodnotou USL = 8,0.

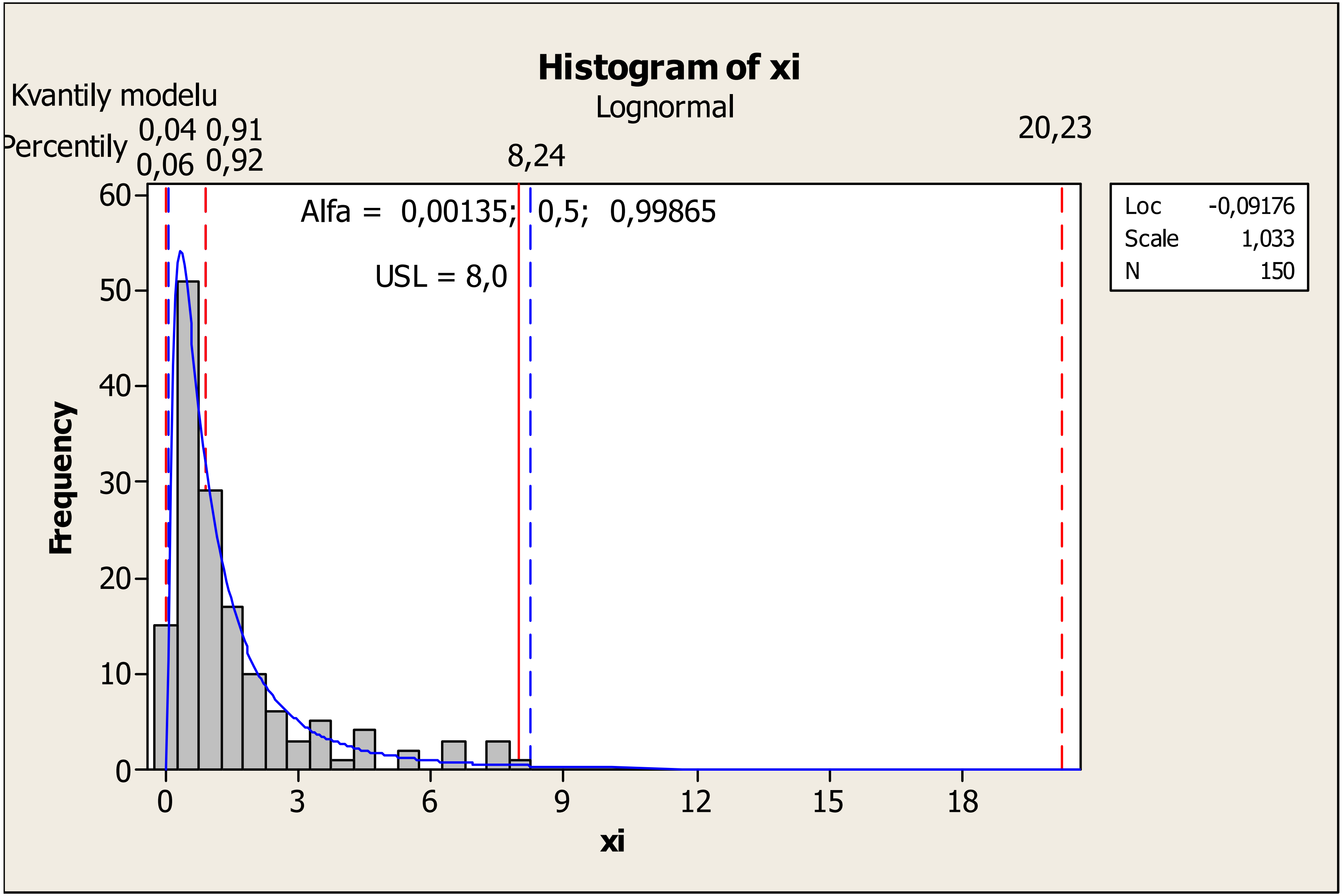
Model


Distribution Plot
Lognormal; Loc=0; Scale=1; Thresh=0
0,05
1
8
USL
20,09
Kvantily pro alfa 0,00135; 0,50; 0,99865
L(0,00135) = 0,0498
Me(0,050) = 1,0000
U(0,99865) = 20,085
PpU = (8,0 - 1,0)/(20,085 - 1,0)
= 0,3668
Density
0,0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0
5
10
15
20
X

Metoda identifikace rozdělení

Některé softwary umožňují identifikaci vhodného modelu rozdělení na základě napozorovaných dat sledovaného znaku kvality. Existují rozdělení, se kterými se často setkáváme při vyšetřování způsobilosti. Nejdříve je třeba identifikovat rodinu rozdělení, následně pak parametry rozdělení, které nejlépe vystihuje napozorovaná data. Následně je třeba stanovit a vypočítat potřebné kvantily rozdělení. Identifikované rozdělení je třeba podrobit testu dobré shody, ověřit, zda opravdu dobře vystihuje rozdělení experimentálních dat.

Např. s podporou softwaru Minitab je možno uvažovat buď všech 14 nabízených nejběžnějších rozdělení nebo specifikovat jedno až čtyři z těchto rozdělení a Box-Coxovu resp. Johnsonovu transformaci napozorovaných nenormálně rozdělených dat na data normálně rozdělená.


Histogram of xi
Lognormal
Kvantily modelu
Percentily
0,04 0,91
0,06 0,92
8,24
20,23
Alfa = 0,00135; 0,5; 0,99865
USL = 8,0
Loc -0,09176
Scale 1,033
N 150
Frequency
xi

Poznámka

Kvantily $\Leftrightarrow$ Percentily

Pro zadané α odpovídá **α-kvantil** x_α (odvozený z hypotetické kumulativní distribuční funkce $F(x_\alpha) = \alpha$.
V uvažovaném příkladu je pro $\alpha = 0{,}99865$ hodnota $x_\alpha = 20{,}220$.

Pro zadané α odpovídá **α-percentil** X_α odvozený z empirické kumulativní distribuční funkce (z napozorovaných dat) $F(X_\alpha) = \alpha$.
V uvažovaném příkladu je pro $\alpha = 99{,}865\%$ hodnota $X_\alpha = 8{,}244$.

Vhodné je zadávat hodnoty α v souvislosti s kvantily jako podíly a hodnoty α v souvislosti s percentily jako procenta. Tato zásada se v praxi často nedodržuje.

- Vybrané kvantily pro vybranou distribuční funkci log-normálního rozdělení odpovídající zadaným podílům v nabídce „Results“.

Table of Percentiles

Distribution	Percent	Percentiles	Standard Error	95,0% CI	
Lognormal	0,135	0,0411641	0,0081576	0,0	0,1

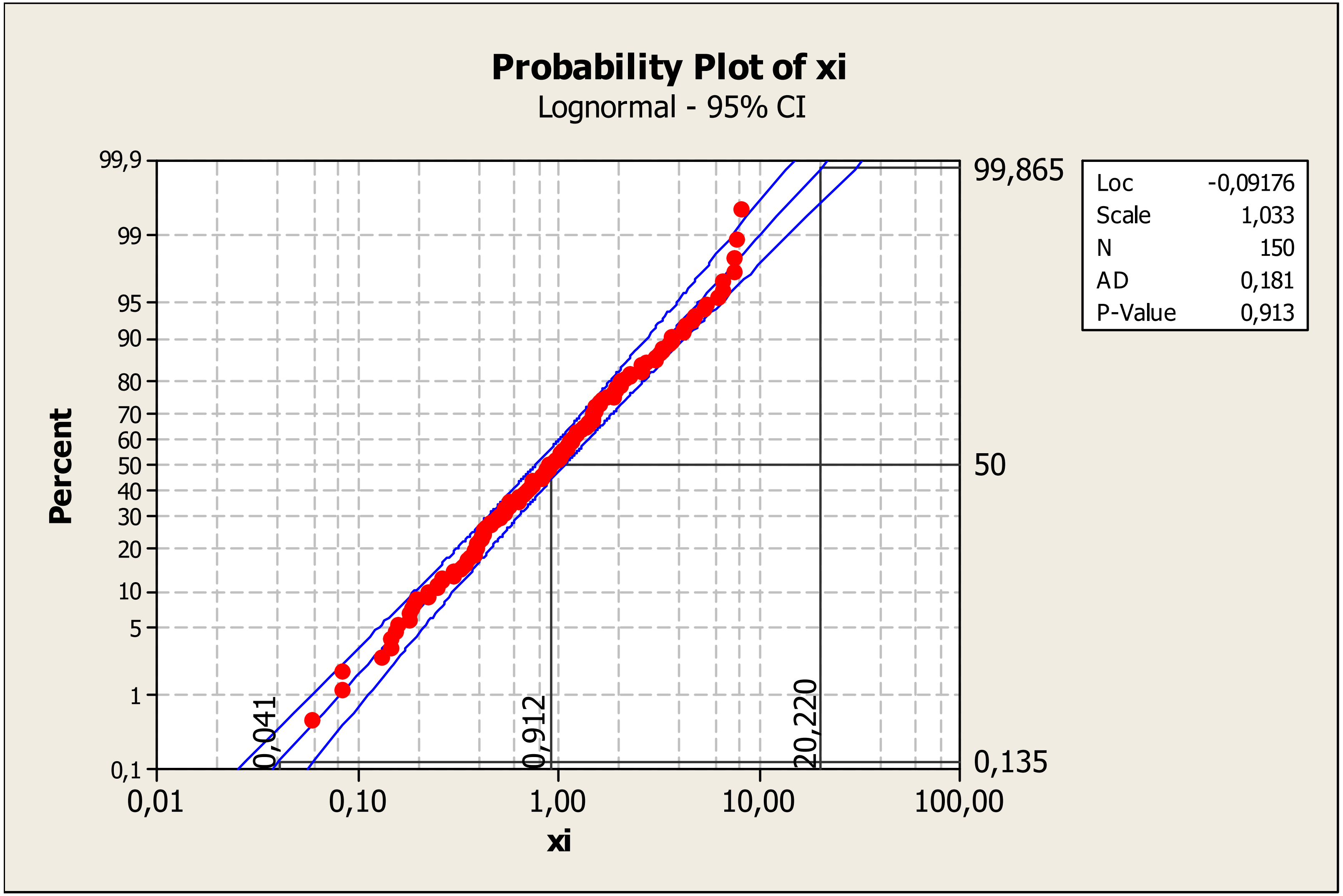
Distribution	Percent	Percentiles	Standard Error	95,0% CI	
Lognormal	50	0,912324	0,0769356	0,8	1,1

Distribution	Percent	Percentiles	Standard Error	95,0% CI	
Lognormal	99,865	20,2199	4,00706	13,7	29,8

Ve výstupní tabulce by bylo lépe psát kvantily místo percentilů a místo procent uvádět podíly. Tabelované hodnoty vycházejí z odhadnutého modelu (distribuční funkce).

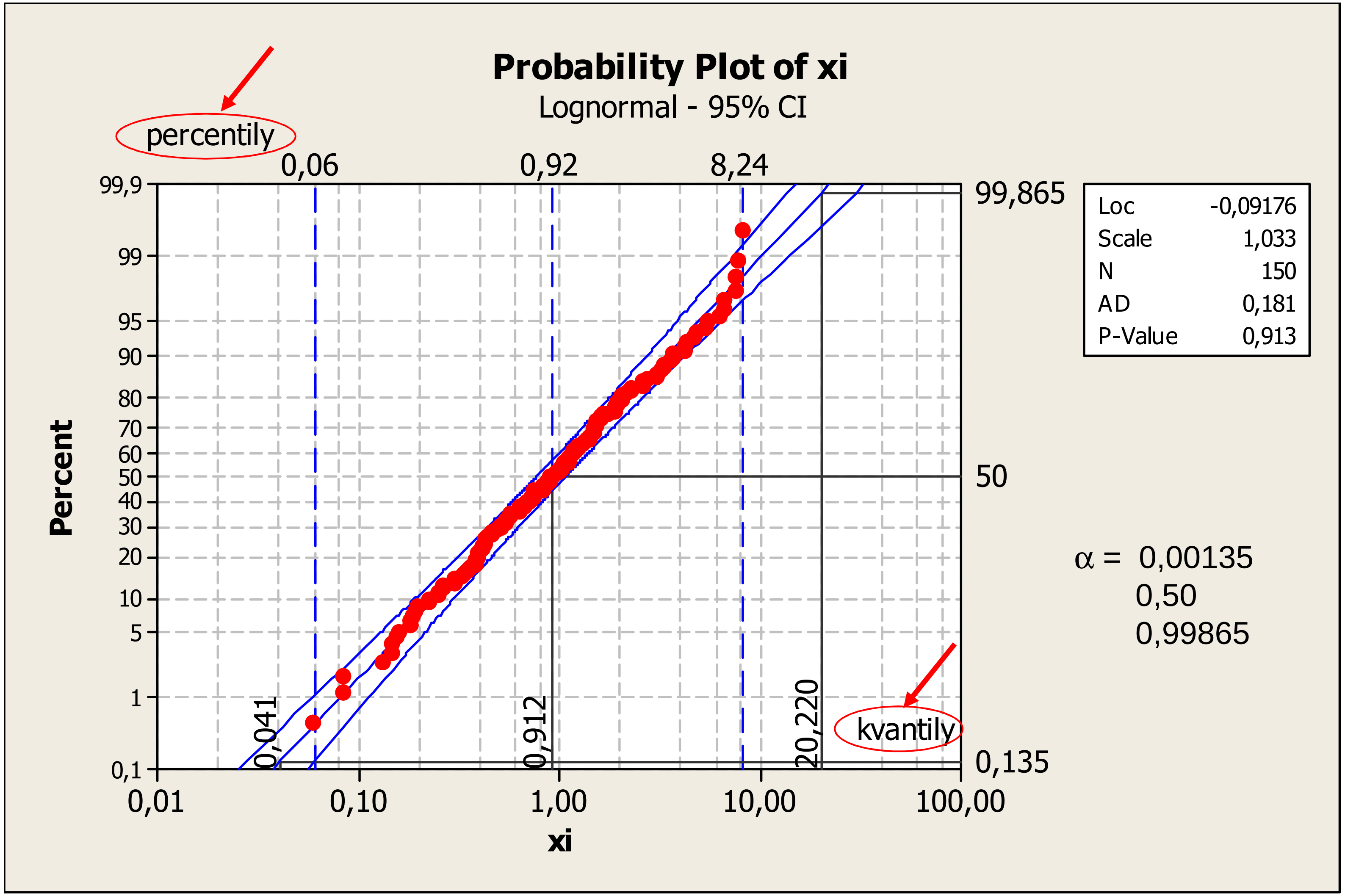
Metoda pravděpodobnostního papíru

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 nejlepším 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.



Odhad kvantilů: L(0,00135) = 0,041; Me(0,5) = 0,912; U(0,99865) = 20,220.

Výsledné hodnoty vycházejí z odhadnutého modelu (distribuční funkce).



Porovnání percentilů (odvozených z napozorovaných dat) a kvantilů, (odvozených z odhadnutého modelu).

Teoretické podklady pro transformaci dat na data normálně rozdělená 1

Mějme tedy na výrobku dán nějaký měřitelný znak kvality s jeho specifikacemi USL a LSL (samozřejmě lze uvažovat i jednostranné případy).
Symbolicky znak označme jako X.

Dále máme k dispozici vhodnou transformaci T(.), která musí mít určité vlastnosti. Jedná se o to, aby byla definovaná na otevřeném (omezeném či nikoliv) intervalu s kladnou derivací v každém bodě.

Tím je zajištěno, že transformace T(.) je ostře rostoucí. Kdyby transformace měla derivace záporné,bude se uvažovat transformace opačná.

Nechť $Y = T(X)$ je transformovaný nový znak kvality mající pouze pomocný charakter.
Boxova-Coxova i Johnsonova transformace požadavky na monotonii splňují.
Transformace T(.) je vybrána tak, že znak Y má obecně rozdělení $N(\mu,\sigma^2)$.
U Johnsonovy transformace je dokonce $\mu = 0$ a $\sigma = 1$.

Mohla by být uvažována jakákoliv jiná transformace s požadovanými vlastnostmi, která veličinu X transformuje do nějakého známého rozdělení, nemusí být normální.

Teoretické podklady pro transformaci dat na data normálně rozdělená 2

Boxova-Coxova transformace je dána vztahem $T(x) = (x^{\lambda} – 1)/\lambda$, kde λ je z nějakého intervalu (-a, a), např. v Minitabu je a = 5.
Pro $\lambda = 0$ je $T(x) = \ln x$. (v Minitabu je uvažována zjednodušená verze
$T(x) = x^{\lambda}$).
Derivace $T'(x) = x^{\lambda-1}$ je kladná v oblasti $(0,\infty)$, kde je transformace definována.
Zkráceně B-C transformace vyžaduje pouze kladná data, tzn. pokud bychom chtěli transformovat i záporná data, je nutno je vhodně posunout na kladnou polopřímku.
Jestliže znak Y bude mít normální rozdělení s parametry μ a σ, pak původní znak X bude mít hustotu rozdělení pravděpodobnosti ve tvaru

$$f(x)= \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(\frac{x^{\lambda}-1}{\lambda}-\mu)^2}{2\sigma^2})x^{\lambda-1}$$

Teoretické podklady pro transformaci dat na data normálně rozdělená 3

Neznámé parametry μ a σ jsou obvykle nahrazeny aritmetickým průměrem a výběrovou směrodatnou odchylkou počítanými z hodnot znaku Y.
Tento tvar hustoty poslouží k tomu, abychom odhadli požadované kvantily $L_{0,00135}$, Me a $U_{0,99865}$ nutné pro výpočet odhadů ukazatelů výkonnosti.

Zcela analogicky vypadá situace s Johnsonovou transformací, kde je ale nutné uvažovat tři třídy transformací SB, SU a SL. Snadno lze zjistit, že všechny tři případy splňují požadavky na monotonnost a jejich derivace jsou:

SL $\quad T'(x) = b/(x + c)$

SB $\quad T'(x) = b(d + c)/((d – x)(x + c))$

SU $\quad T'(x) = b((d^2 + (x-c)^2)^{-0,5}$.

Teoretické podklady pro transformaci dat na data normálně rozdělená 4

Protože Johnsonova transformace převádí původní data na data rozdělená N(0,1), pak označíme-li φ(.) hustotu tohoto rozdělení,pak hustota rozdělení původních dat má tvar

$$f(x) = \varphi(T(x))T'(x),$$

podle typu vybrané transformace.

Samozřejmě, že parametry transformace jsou odhadnuty z dat a volbu optimální možnosti provede software automaticky a vrátí i vzorec pro vybranou transformaci.

Průběh hustoty rozdělení lze pak graficky znázornit a použít pro výpočet požadovaných kvantilů pro odhady ukazatelů výkonnosti.

Teoretické podklady pro transformaci dat na data normálně rozdělená 5

Samozřejmě nás zajímá pravděpodobnost možného výskytu hodnot sledovaného znaku kvality nad USL a pod LSL, tedy

$$\mathrm{Pst}\{ X < LSL\}, \text{ resp. } \mathrm{Pst}\{ X > USL\}.$$

Díky monotonii transformace T máme ihned, že

$$\mathrm{Pst}\{ X < LSL\} = \mathrm{Pst}\{ Y < T(LSL)\},$$

a zcela analogicky

$$\mathrm{Pst}\{ X > USL\} = \mathrm{Pst}\{ Y > T(USL)\}.$$

To ale znamená, že např. pravděpodobnost výskytu hodnoty znaku X nad USL vyjádřená třeba v ppm jednotkách je totožná s pravděpodobností výskytu transformované hodnoty Y nad transformovanou horní specifikací T(USL) opět vyjádřenou v ppm. Protože znak Y je normálně rozdělen, není problém tento počet v ppm spočítat. Je nutno si ale uvědomit, že transformace T obsahuje parametry, které musíme odhadnout z dat, a tudíž se do výpočtu promítá náhoda.

Teoretické podklady pro transformaci dat na data normálně rozdělená 6

Obdobně lze spočítat, že

$$Pst\{ a < X < b\} = G(T(b)) - G(T(a)),$$

kde G(.) je distribuční funkce pro znak Y.

Pro výpočet ukazatelů Pp a Ppk potřebujeme znát příslušné kvantily $L_{0,00135}$, Me a $U_{0,99865}$. Je-li $0 < \alpha < 1$, pak

$$\alpha = Pst\{ X < x_\alpha\} = Pst\{ Y < T(x_\alpha)\}.$$

To ale znamená, že $T(x_\alpha)$ musí být odpovídající α-kvantil normálního rozdělení znaku Y. Tedy $T(x_\alpha) = G^{-1}(\alpha)$ a díky tomuto vztahu lze vyjádřit α-kvantil x_α následovně

$$x_\alpha = T^{-1}(G^{-1}(\alpha)).$$

Teoretické podklady pro transformaci dat na data normálně rozdělená 7

Tento vztah může být využit pro vyjádření ukazatele Pp v řeči původného znaku X:

$$Pp = (USL – LSL)/(T^{-1}(G^{-1}(0{,}99865) - T^{-1}(G^{-1}(0{,}00135)).$$

V případě B-C transformace jsou parametry μ a σ normálního rozdělení nahrazeny jejich odhady aritmetický průměr a výběrová směrodatná odchylka, v případě Johnsonovy transformace lze přímo položit $G^{-1}(0{,}99865) = 3$ a $G^{-1}(0{,}00135) = -3$.

Pro výpočet ukazatele Ppk je nutný ještě odhad pro medián, který lze získat buď z uspořádání dle velikosti původních dat nebo jako $G^{-1}(0{,}50)$.

Z výše uvedeného plyne, že transformace T je pouze pomocný nástroj a že o reálném stavu procesu nevypovídají hodnoty znaku Y, ale je nutno vše zpět přepočítat do řeči původního znaku X.

Metoda Johnsonovy transformace

Pokud hodnoty sledovaného znaku kvality nejsou normálně rozdělené, je možno použít Johnsonovu transformaci tak, že nová transformovaná data jsou potom rozdělena normálně N(0,1). Johnsonova transformace vybere jednu ze tří typů rovnic v závislosti na tom, zda náhodná veličina je „ohraničená“ – typ SB; je „lognormální“ – typ SL; je „neohraničená“ – SU. Jedná se o rovnice:

Typ SB

$$y = a + b \cdot \ln((x + c) / (d - x))$$

pro $b > 0;\ -c < x < d$.

Typ SL

$$y = a + b \cdot \ln(x+c)$$

pro $b > 0;\ -c < x$.

Typ SU

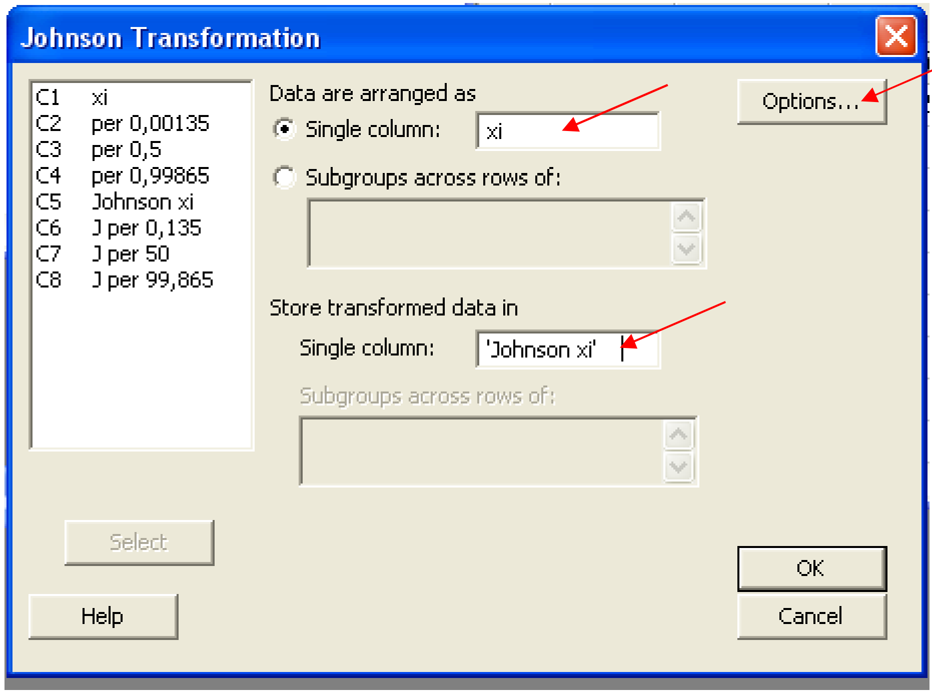
$$y = a + b \cdot \operatorname{Asinh}((x - c) / d)$$

pro $b > 0;\ d > 0$.

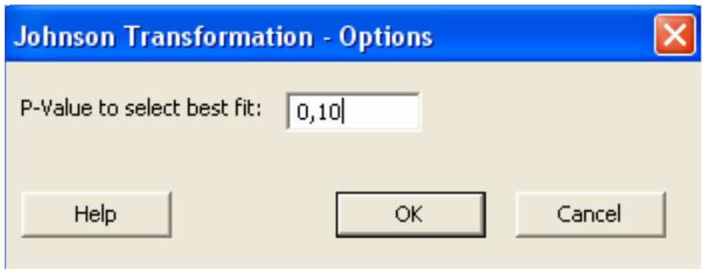
Software postupuje tak, že uvažuje všechny možné funkce Johnsonova systému, odhadne jejich parametry, transformuje data, vypočítá Anderson-Darlingovu testovou statistiku a jí odpovídající p-hodnotu a vybere tu funkci, které odpovídá největší p-hodnota.

Např. software Minitab postupuje následovně:

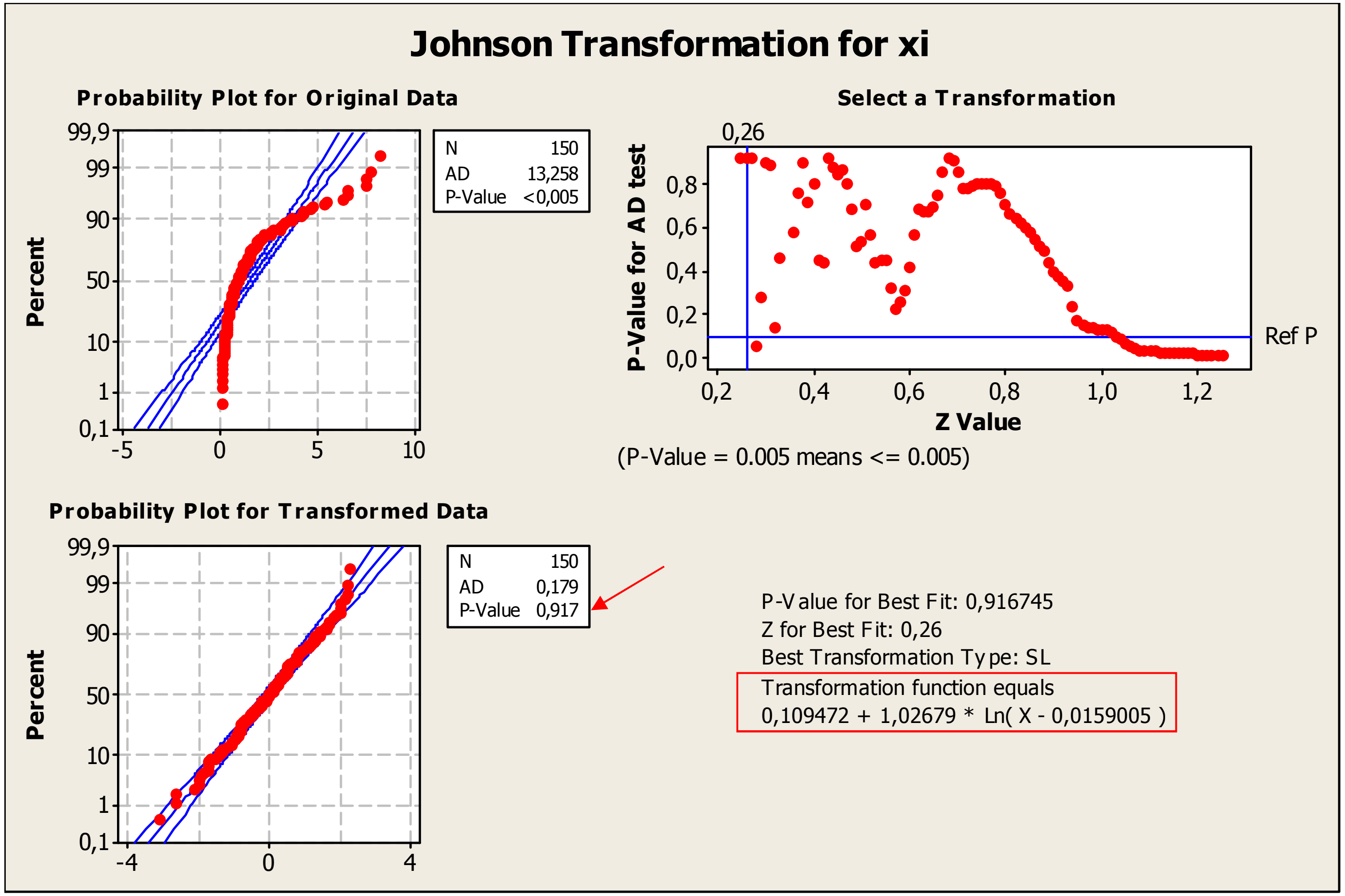
Stat > Quality Tools > Johnson transformation > a vyplníme dialogové okno.



V nabídce „Options“ můžeme změnit přednastavenou minimální hodnotu hladiny významnosti $\alpha = 0,10$.



Transformovaná data se zapíší do pracovního listu a zobrazí se graf s výsledkem transformace



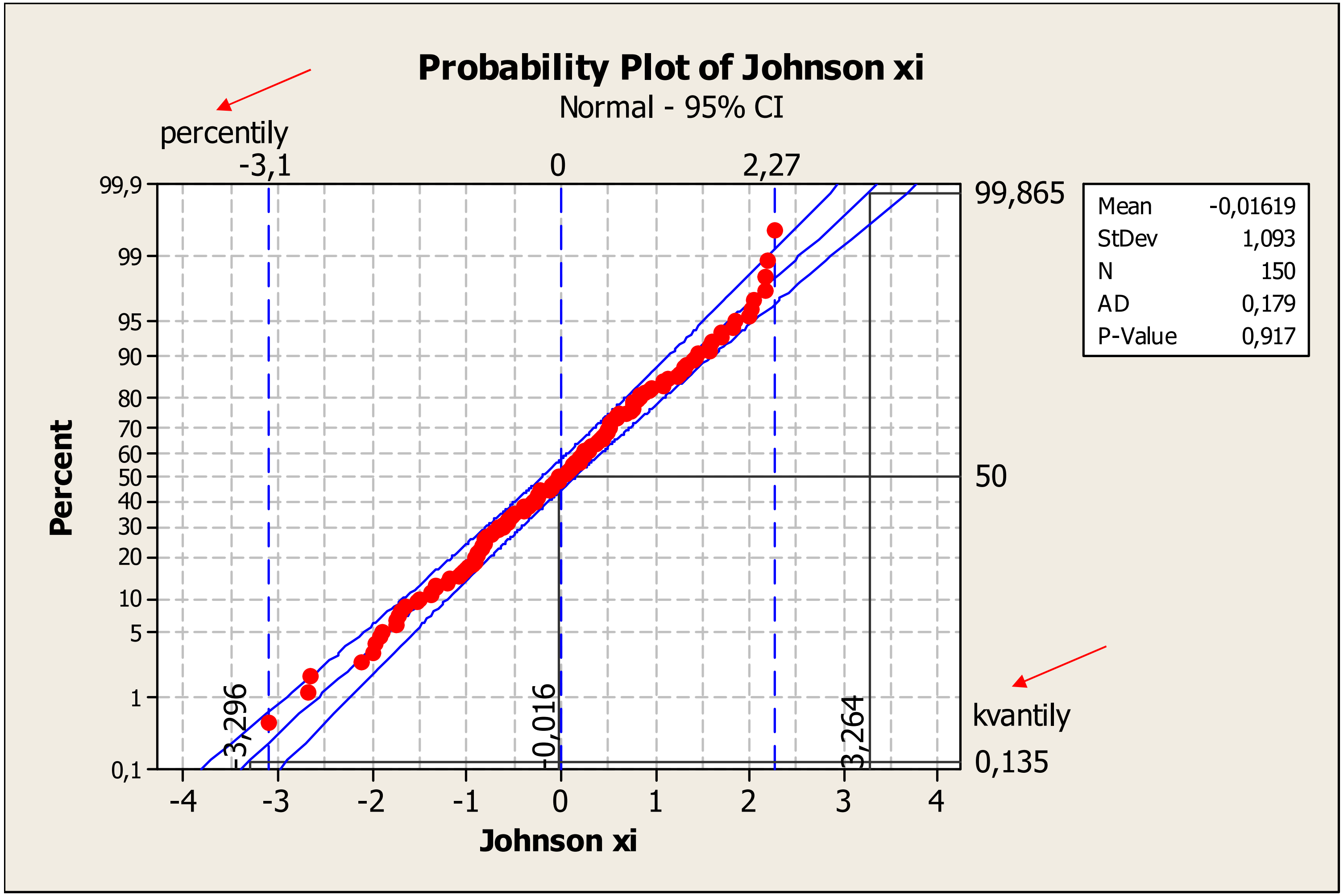
Z grafu je patrno, že p-hodnota Anderson-Darlingova testu dobré shody s normálním rozdělením je vysoká (0,917) a že tedy není důvod pochybovat, že transformovaná data jsou rozdělena normálně. Dále je zde transformační rovnice typu SL

$$0{,}109472 + 1{,}02679 \cdot \mathrm{Ln}((x - 0{,}0159005).$$

Z transformovaných dat ve sloupci C5 (Johnson xi) můžeme vypočítat odhady percentilů pro procenta 0,135; 50; 99,865 pomocí kalkulátoru a uložit je např. do sloupců C6; C7; C8 postupem:

Calc > Calculator > vyplníme dialogové okno, kde do počítaného výrazu (Expression) dosadíme funkci „PERCENTILE“ a za její argumenty „number“ dosadíme sloupec C5 (Johnson) a za „probability“ dosadíme 0,00135. Výsledek bude uložen v první buňce sloupce C6. Tak postupně vypočítáme i další odhady percentilů do prvních buněk sloupců C7 a C8.

C5	C6	C7	C8
Johnson xi	J per 0,135	J per 50	J per 99,865
1,09225	-3,10047	0,0006066	2,27349
0,53542			
0,20609			
-0,08605			



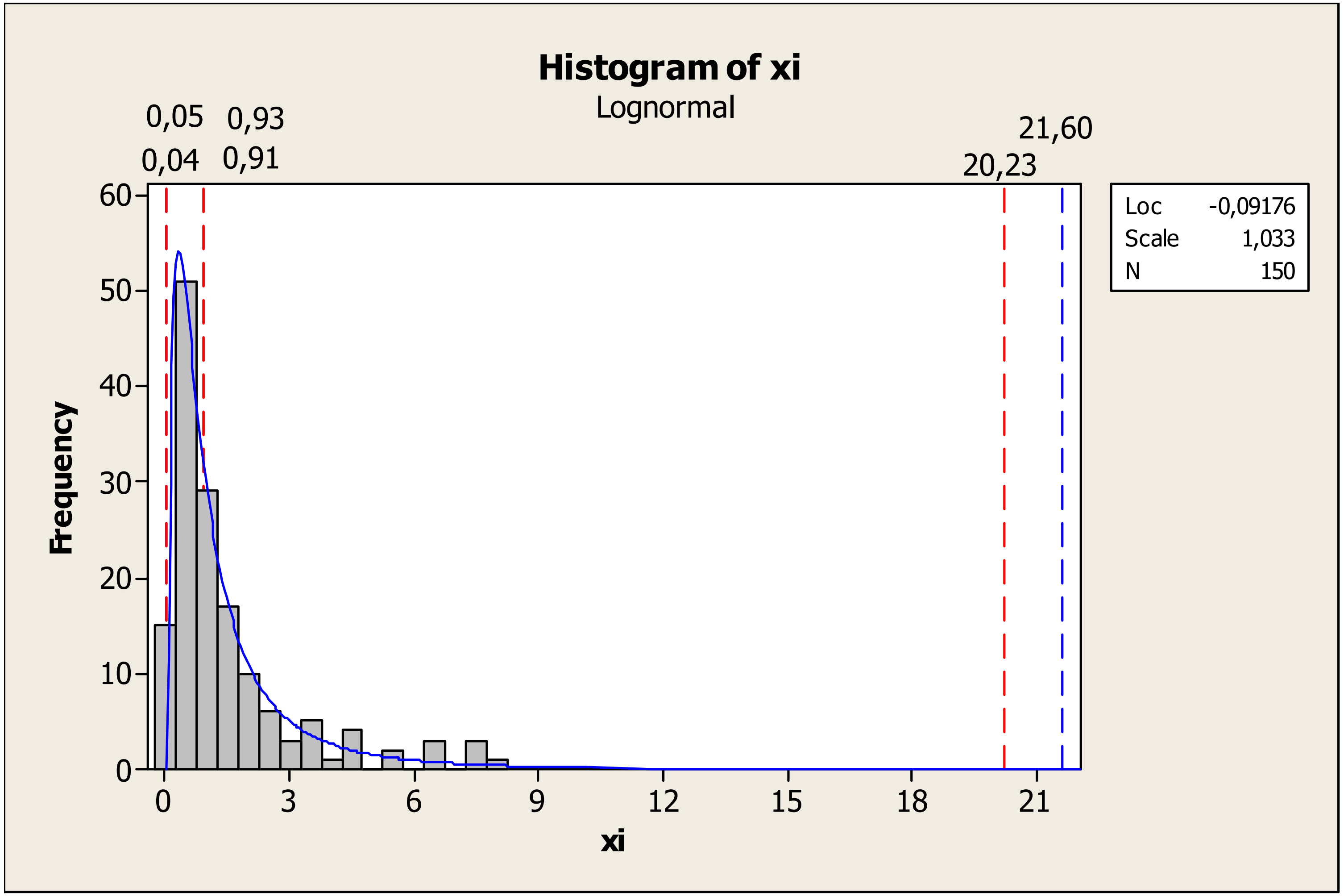
Porovnání percentilů a kvantilů transformovaných dat vypočítaných po vložení „Add - Percentil Lines“ na pravděpodobnostním grafu.

Abychom mohli odpovídající odhady kvantilů zakreslit do původních hodnot, musíme provést zpětnou transformaci a vypočítat odhady kvantilů v původní proměnné. To je třeba učinit v Excelu pomocí nástroje „Hledání řešení“.

Na následujícím obrázku je histogram původních dat se zakreslenými odhady kvantilů:

červeně - kvantily vypočítané z odhadnutého modelu lognormálního rozdělení;

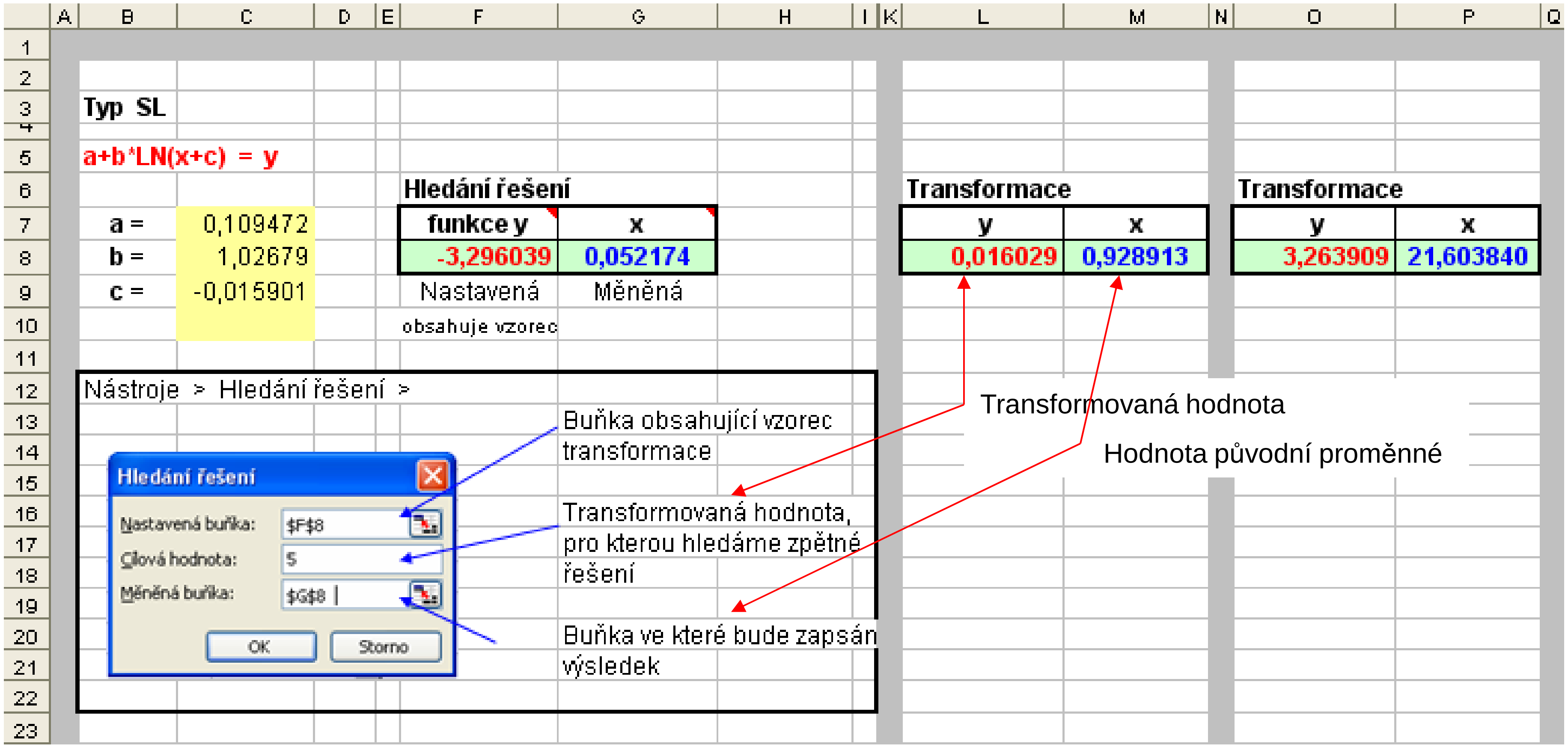
modře - kvantily vypočítané pomocí zpětné Johnsonovy transformace.



Kvantily vypočítané z modelu lognormálního rozdělení (červené 0,04; 0,91; 20,23) a na základě zpětné Johnsonovy transformace (modré 0,05; 0,93; 21,60).

Zpětná Johnsonova transformace

Zpětnou transformaci je možno provést v MS Excelu pomocí nástroje „Hledání řešení“:



Metoda Box-Coxovy transformace

Box-Coxova transformace odhaduje hodnotu λ , která minimalizuje směrodatnou odchylku normalizované transformované proměnné.

Software prozkoumá mnoho transformací, z nichž uvedeme jen ty, které jsou pro zaokrouhlené hodnoty λ. y je transformovaná hodnota proměnné (naměřeného znaku kvality) x.

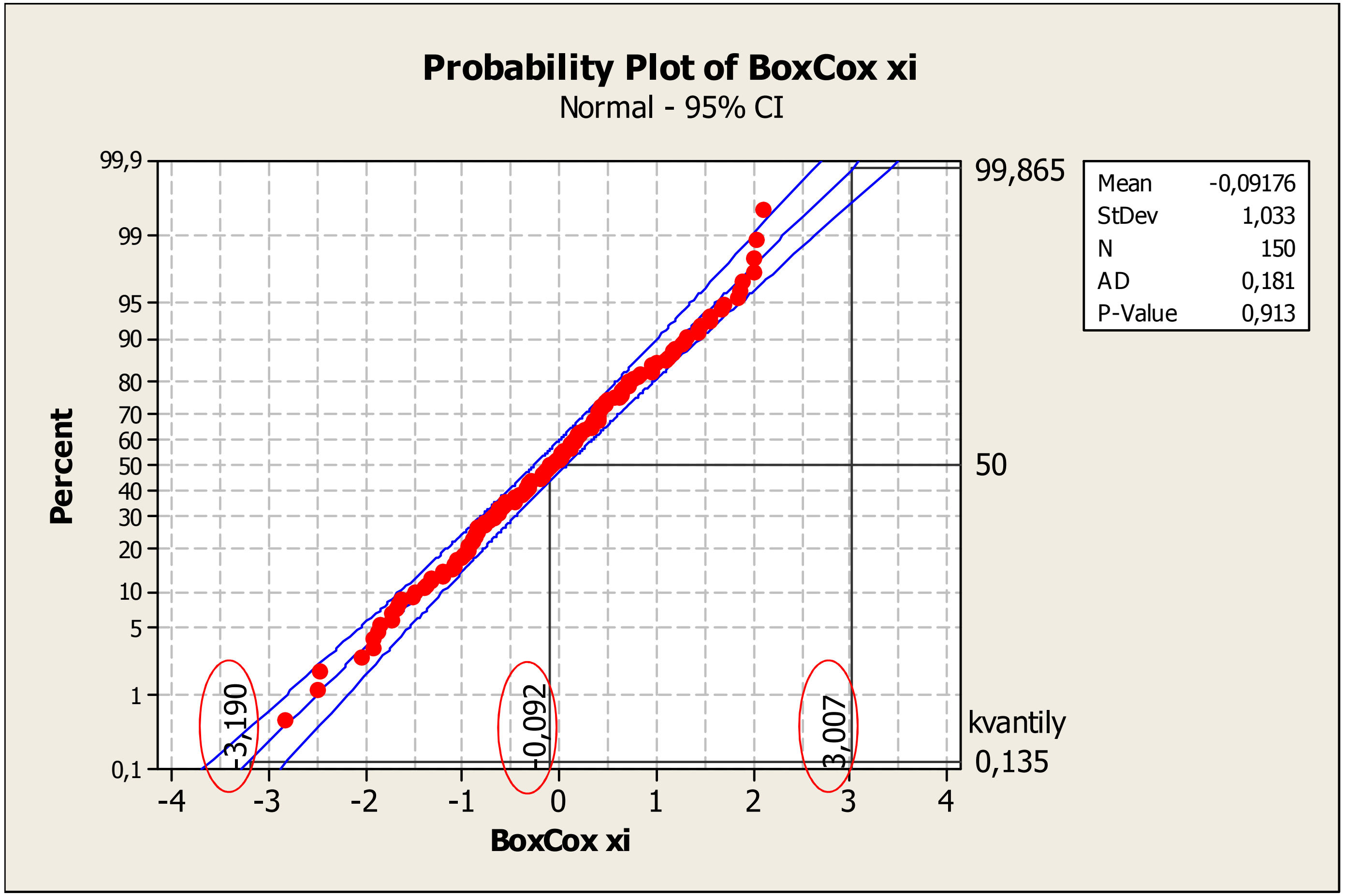
Hodnota λ	Transformace
$\lambda = 2$	$y = x^2$
$\lambda = 0,5$	$y = \sqrt{x}$
$\lambda = 0$	$y = \ln x$
$\lambda = -0,5$	$y = 1 / (\sqrt{x})$
$\lambda = -1$	$y = 1 / x$

Např. software Minitab postupuje následovně:

nenormálně rozdělená data jsou ve sloupci C1 („Data“). Transformovaná data zapíšeme do sloupce C2 (Box-Cox). Zvolíme postupně

Stat > Control Charts > Box-Cox Transformation >

v dialogovém okně zadáme data, velikost případných podskupin, v tomto případě n = 1. Pod „Options“ zvolíme jiný sloupec , kam zapíšeme transformovaná data a potvrdíme výpočet s optimální hodnotou lambda.



Odhad kvantilů transformovaných dat vypočítaných po vložení příslušných procent do „Add - Percentil Lines“ na pravděpodobnostním grafu.

Zpětná Box-Cox transformace

Abychom mohli odpovídající odhady kvantilů zakreslit do původních hodnot, musíme provést zpětnou transformaci a vypočítat odhady kvantilů v původní proměnné. To je třeba učinit v Excelu pomocí nástroje „Hledání řešení“.

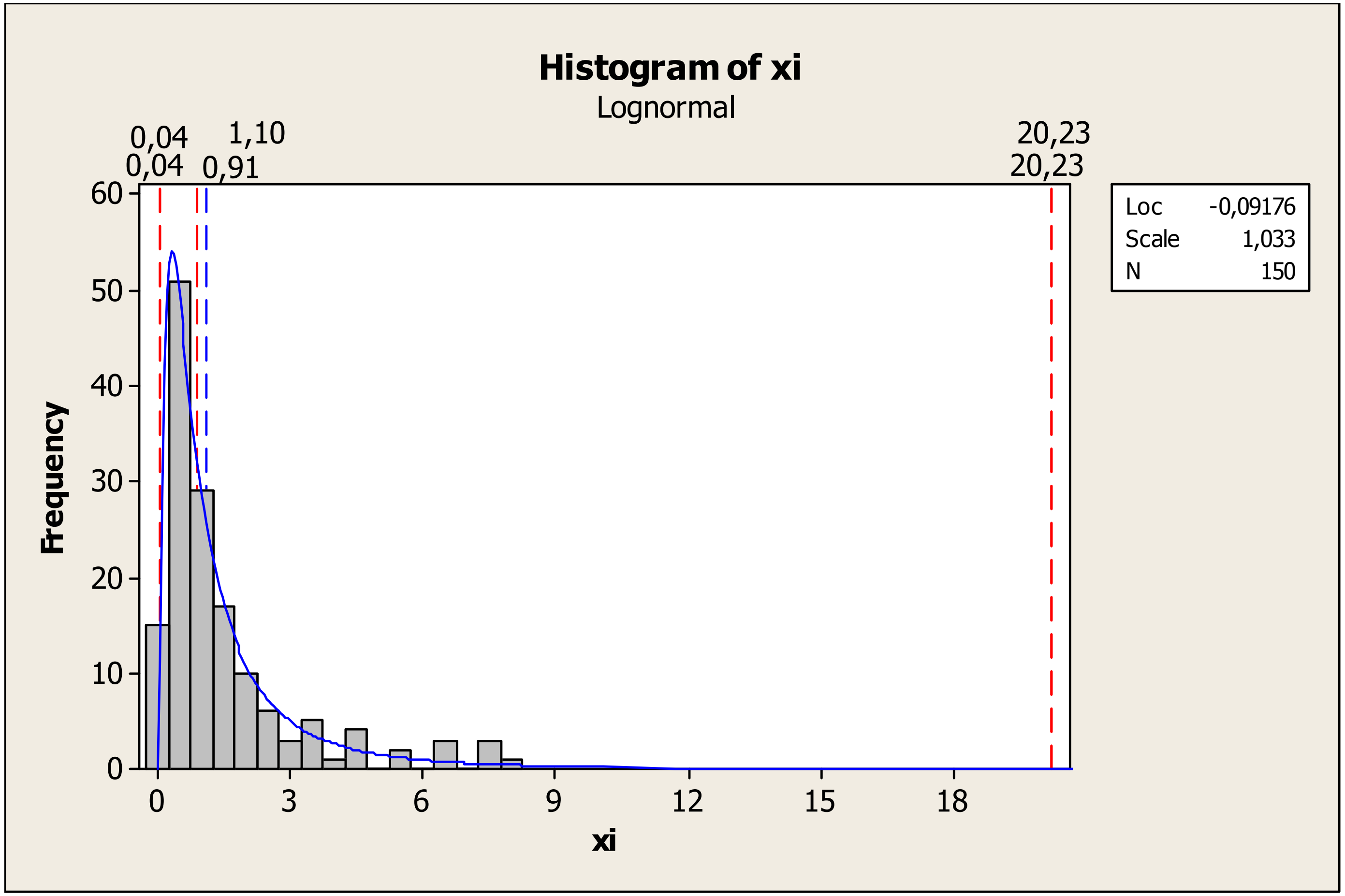
y = ln x					
λ =	0		Hledání řešení		
			funkce y	X	
			-3,19081	0,04114	
			Nastavená	Měněná	
			obsahuje vzorec		
					-3,190

funkce y	X
0,09202	1,09638
	0,092

funkce y	X
3,00699	20,22645
	3,007

Kvantily transformovaných hodnot

Kvantily původních hodnot



Kvantily vypočítané z modelu lognormálního rozdělení (červené 0,04; 0,91; 20,23) a na základě zpětné Box-Cox transformace (modré 0,04; 1,10; 20,23).

Metoda empirických percentilů

V případech, kdy neznáme rozdělení pravděpodobnosti sledovaného znaku kvality nebo jeho parametry, nebo se nedaří vhodné rozdělení najít, potom se nabízí přibližná metoda využití empirických percentilů vypočítaných z napozorovaných dat. Jedná se o percentily

$$\tilde{L}_{0{,}135\,\%} \qquad \widetilde{Me}_{50\,\%} \qquad \tilde{U}_{99{,}865\,\%}$$

odpovídající procentům, se kterými pracují ukazatele výkonnosti.

Aby tato metoda byla alespoň orientačně použitelná, je třeba, aby rozsah napozorovaných dat byl dostatečný, nejméně N = 100. Problém je v tom, že odhady percentilů, které odpovídají malému či velkému procentu vykazují značnou variabilitu, neboť jsou obvykle odvislé od malého počtu nejmenších či největších hodnot ve výběru. Tím jsou i odhady ukazatelů výkonnosti zatíženy značnou variabilitou.

Výpočet percentilů z dat můžeme provést pomocí kalkulátoru na základě funkce „Percentile (number; probability)“, kam za „number“ dosadíme sloupec s daty (např. C1) a za „probability“ odpovídající hodnoty 0,00135; 0,50; 0,99865.

U jednostranných specifikací se pravděpodobnosti volí 0,0027 nebo 0,9973 podle zadané specifikační meze.

Worksheet 2 ***

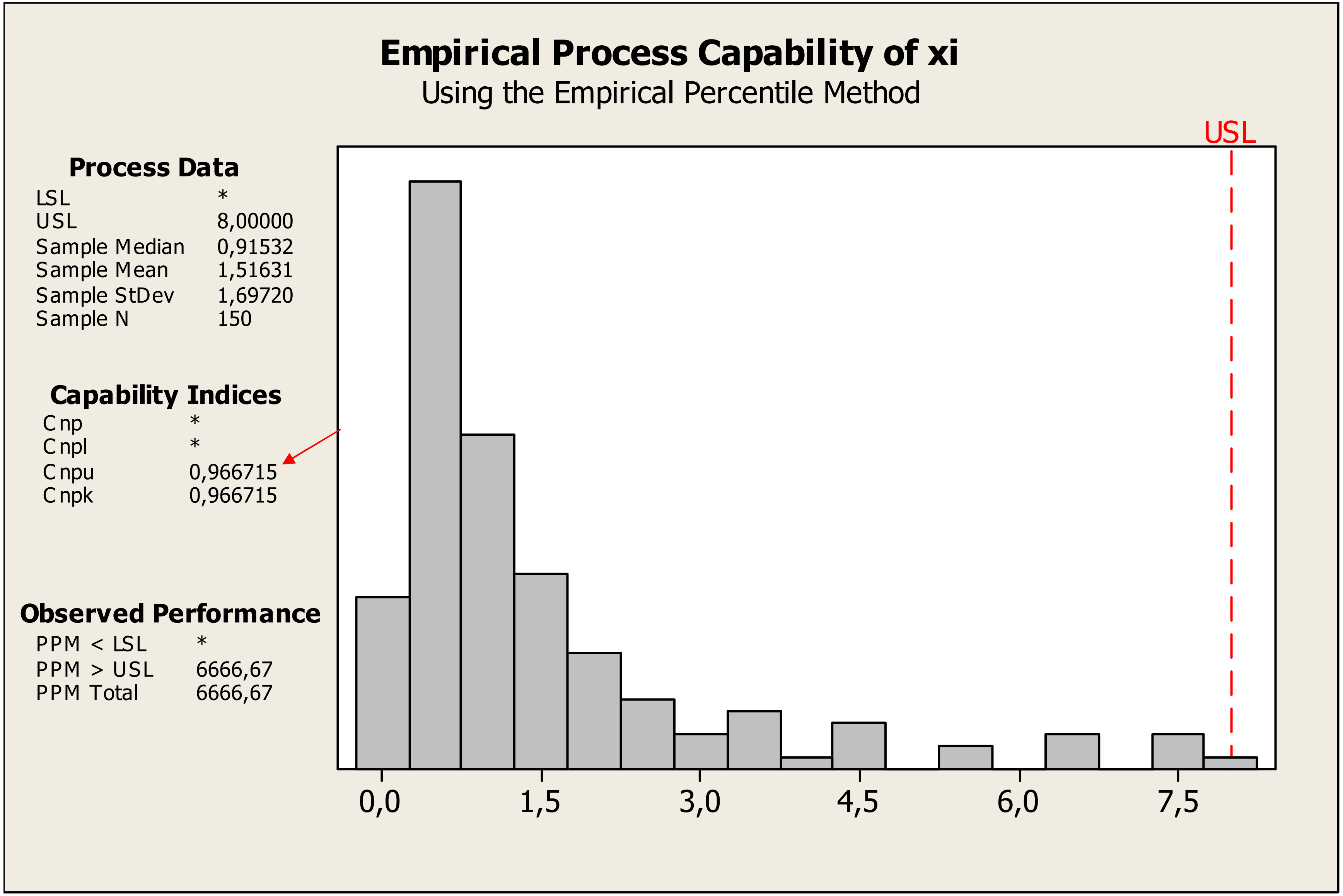
↓	C1	C2	C3	C4
	xi	per 0,00135	per 0,5	per 0,99865
1	2,62012	0,0597857	0,915323	8,24393
2	1,53002			
3	1,11457			
4	0,84251			

Potom můžeme vypočítat odhad ukazatele výkonnosti P_{pU} na základě percentilů z dat ve sloupci C1 při USL = 8 :

$$\hat{P}_{pkU} = \frac{USL - \tilde{M}e_{0,50}}{\tilde{U}_{0,99865} - \tilde{M}e_{0,50}} = \frac{8 - 0,915323}{8,24393 - 0,915323} = 0,96672$$

Stejného výsledku, ale rychleji, dosáhneme pomocí makra „ECAPA.MAC“
následujícím postupem:
Edit > Command Line Editor > zapíšeme
%ECAPA C1;
USL 8.
a stiskneme „Submit Commands“.

Výsledkem je graf s vypočítanými odhady ukazatelů výkonnosti, které odpovídají těm, které jsme vypočetli.


Empirical Process Capability of xi
Using the Empirical Percentile Method
USL
Process Data
LSL *
USL 8,00000
Sample Median 0,91532
Sample Mean 1,51631
Sample StDev 1,69720
Sample N 150
Capability Indices
Cnp *
Cnpl *
Cnpu 0,966715
Cnpk 0,966715
Observed Performance
PPM < LSL *
PPM > USL 6666,67
PPM Total 6666,67
0,0
1,5
3,0
4,5
6,0
7,5

Hodnocení výkonnosti procesu pomocí statistických tolerančních intervalů

Statistický toleranční interval je takový interval, který s předem zadanou pravděpodobností (např. 95%) alespoň pokrývá určitou předem zvolenou část základního souboru (např. 99%-ní část). Pokud bychom znali dokonale rozdělení náhodné veličiny, pak tvrzení o pokrytí části základního souboru je 100%-ní. Např. u normálního rozdělení tvrzení, že interval $(\mu-3\sigma, \mu+3\sigma)$ pokrývá 99,73% hodnot základního souboru, je při známých hodnotách parametrů pravdivé. Když parametry neznáme a odhadujeme z dat, je nutné z pravdivosti tvrzení slevit, např. na 95%.

Minitab 17 nabízí dvojí řešení: a) pro normální rozdělení
b) pro nenormální rozdělení pomocí neparametrické metody

Příklad konstrukce statistického tolerančního intervalu – Minitab 17

Tolerance Interval: C1

Method

Tolerance interval type	One-sided (Upper)
Confidence level	95%
Percent of population in interval	99,73%

Statistics

Variable	N	Mean	StDev
C1	60	0,910	1,347

95% Upper Tolerance Bound, Using Normal Method

Variable	Upper Bound
C1	5,406

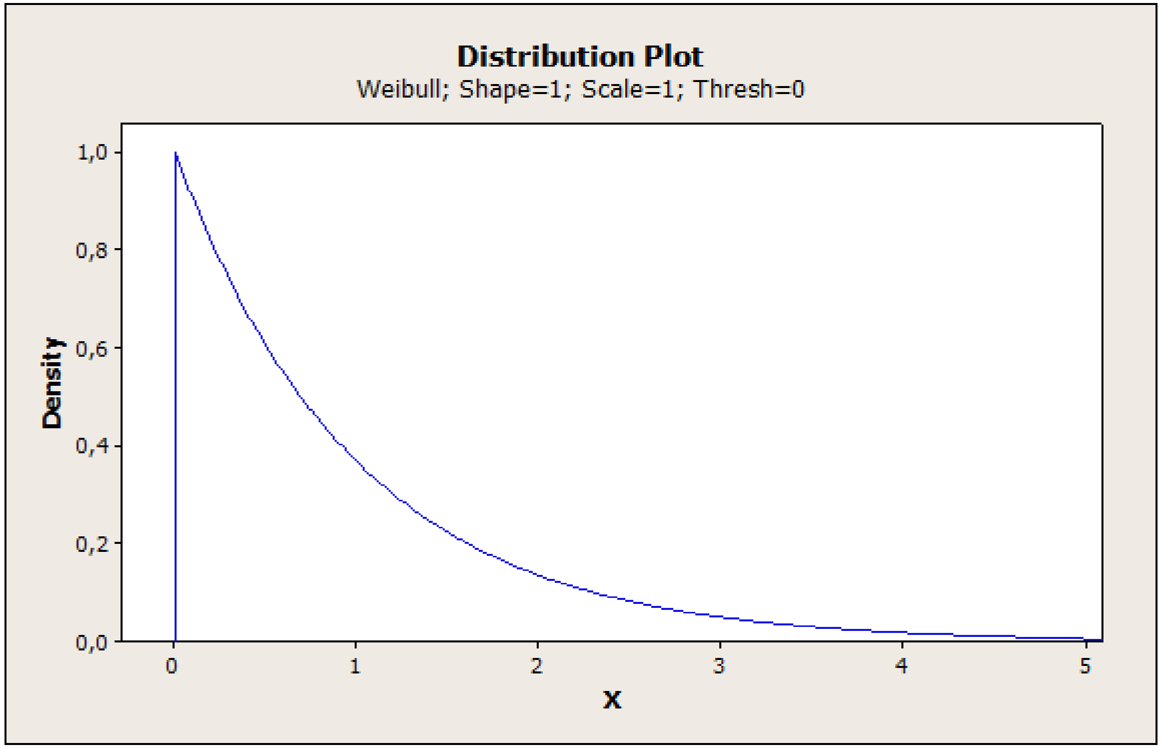
95% Upper Tolerance Bound, Using Nonparametric Method

Variable	Upper Bound	Achieved Confidence
C1	7,928	15,0

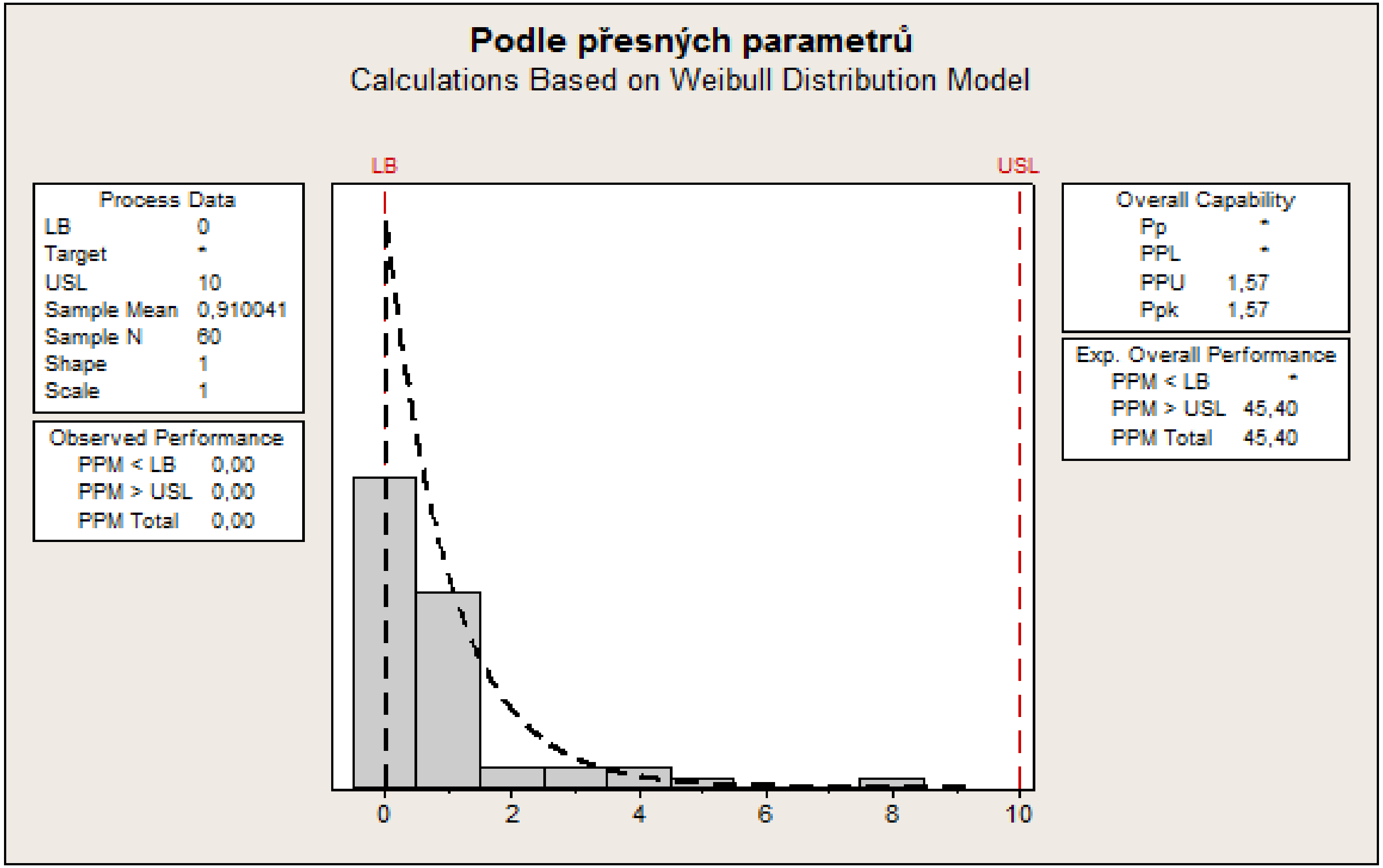
Pozn. Pro nenormálně rozdělená data se vyžaduje minimálně 100 hodnot

Ilustrativní příklad - Minitab 17

Data byla vygenerována z rozdělení Weibull (1,1)
Celkem bylo získáno 60 hodnot – jedná se o minimální počet pro vyšetřování výkonnosti procesu



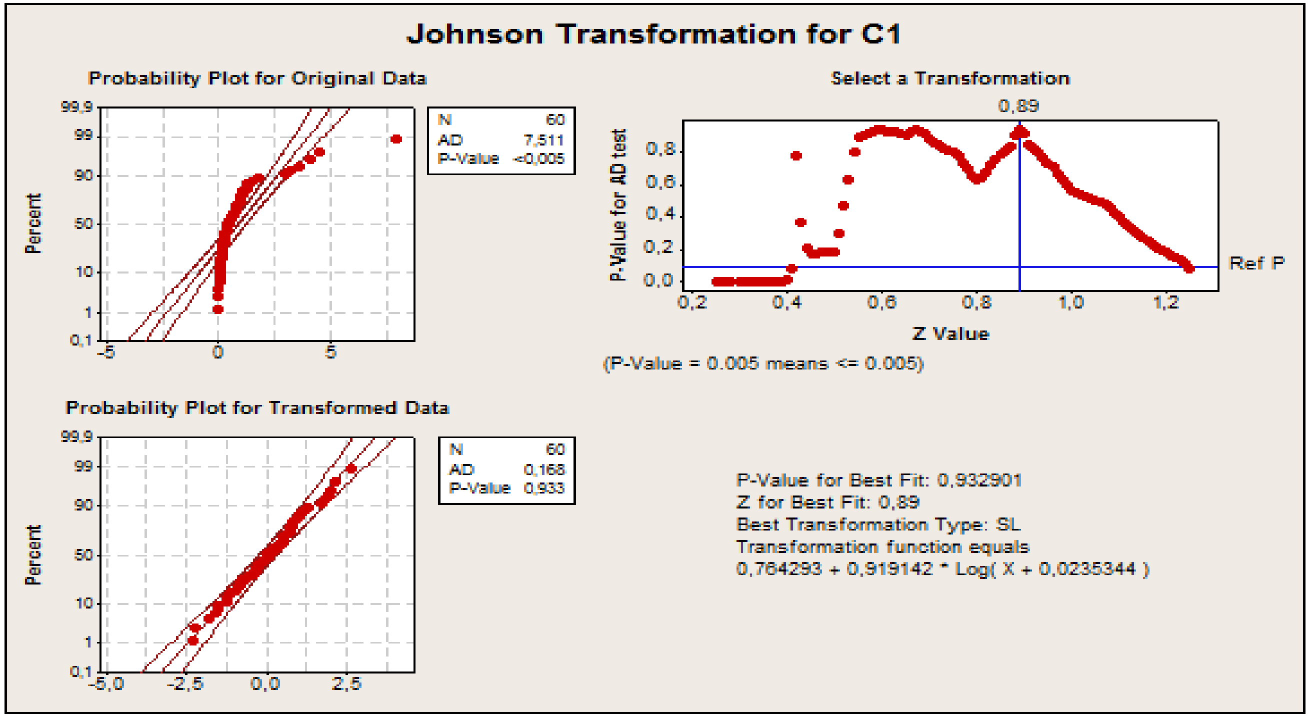
Hodnota USL byla stanovena na 10 – PpkU = 1,57



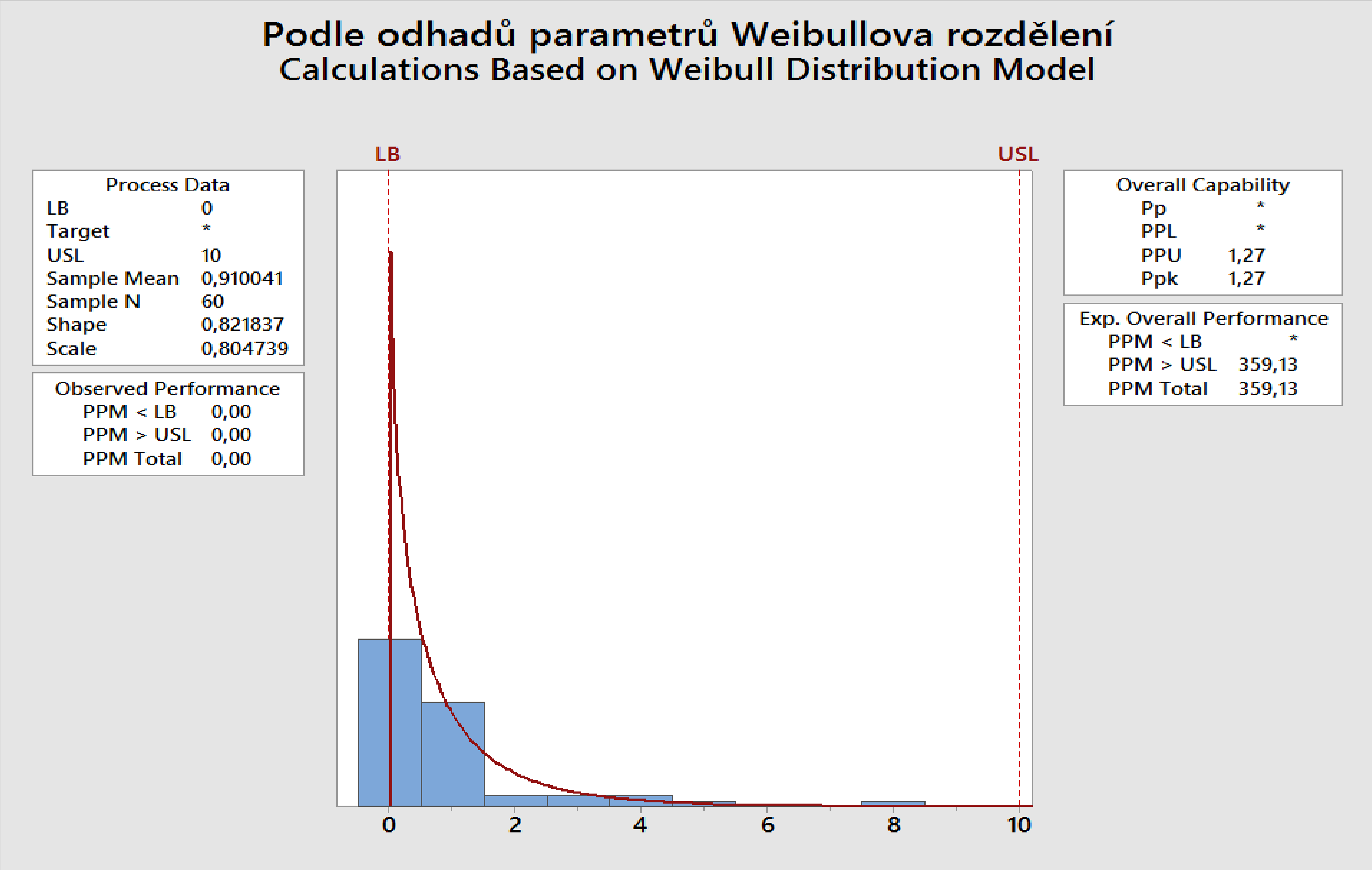
Volba vhodného modelu pomocí Minitab 17

Byly doporučeny 3 modely, a to logaritmicko-normální rozdělení, Weibulovo rozdělení a loglogistické rozdělení

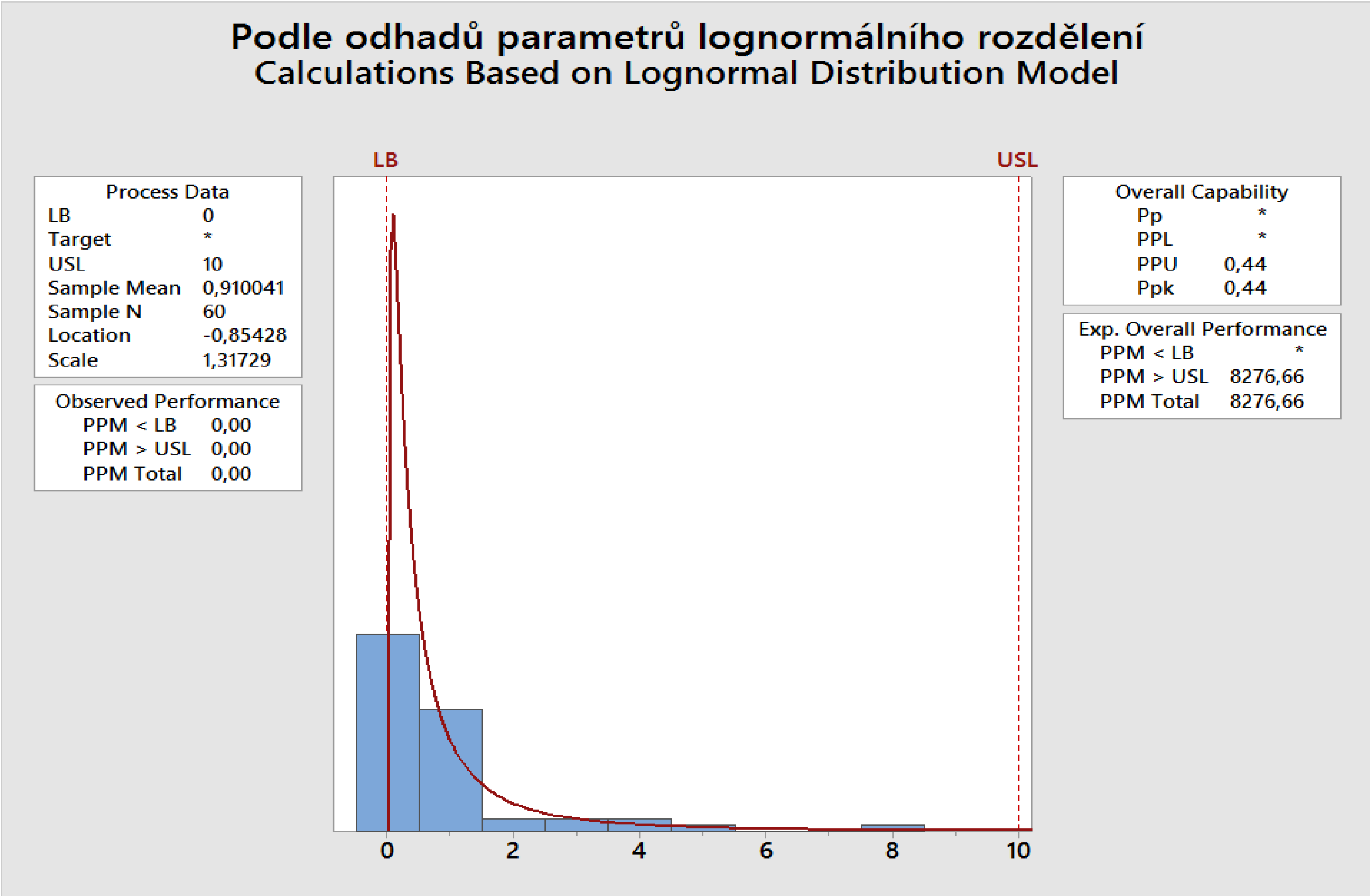
Rovněž byly doporučeny jak Boxova-Coxova transformace, tak i Johnsonova transformace typu SL



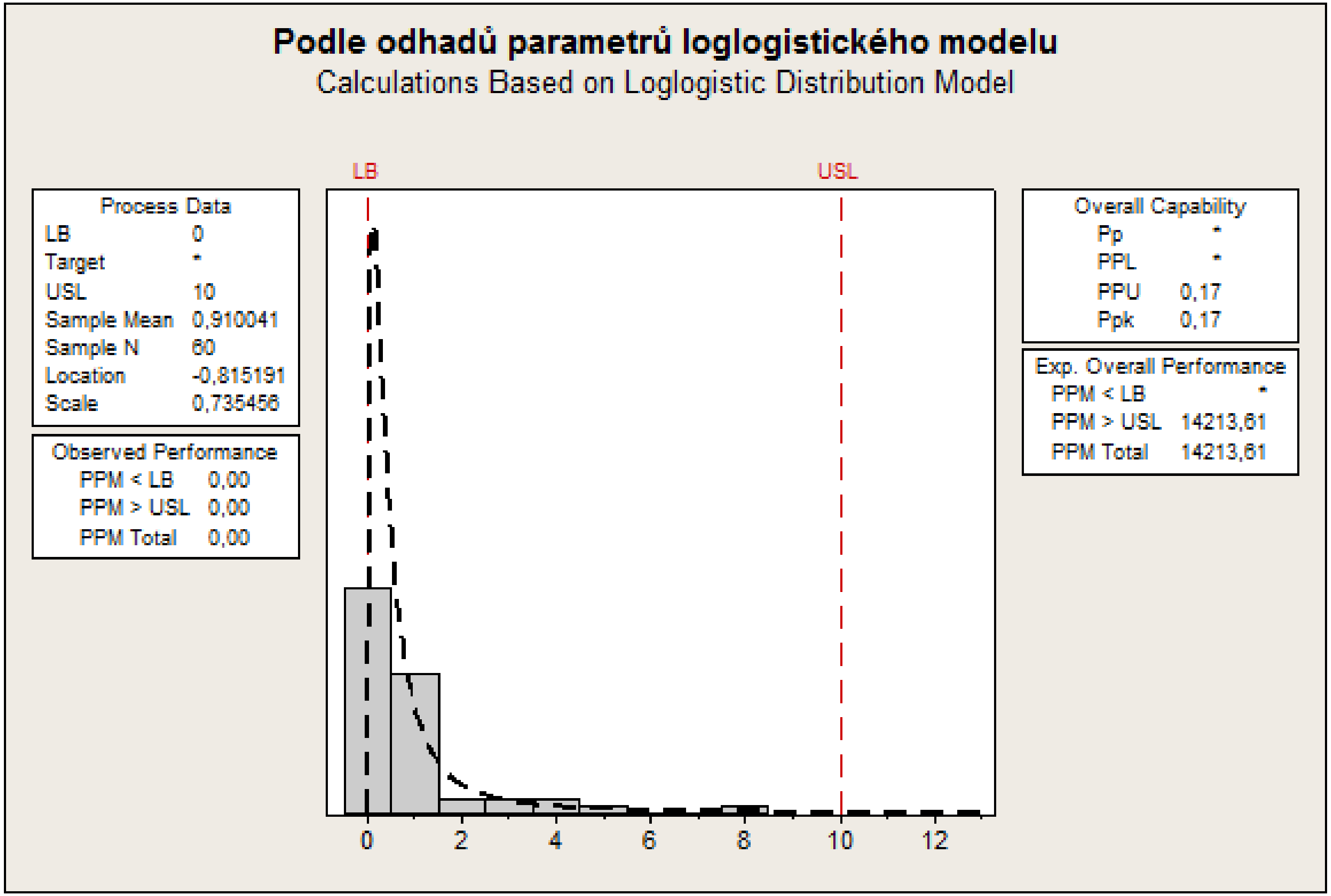
Volba vhodného modelu pomocí Minitab 17



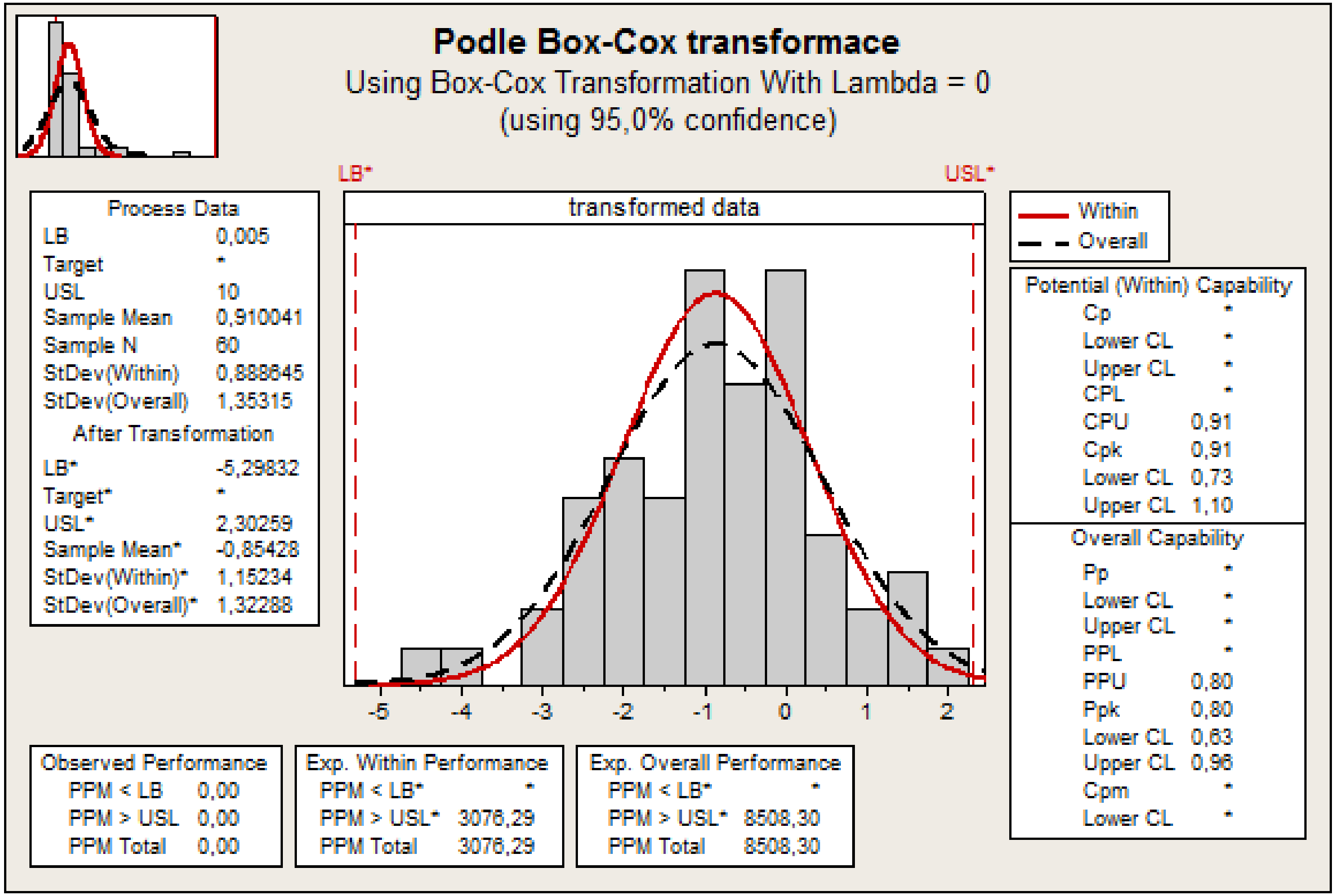
Volba vhodného modelu – Minitab 17



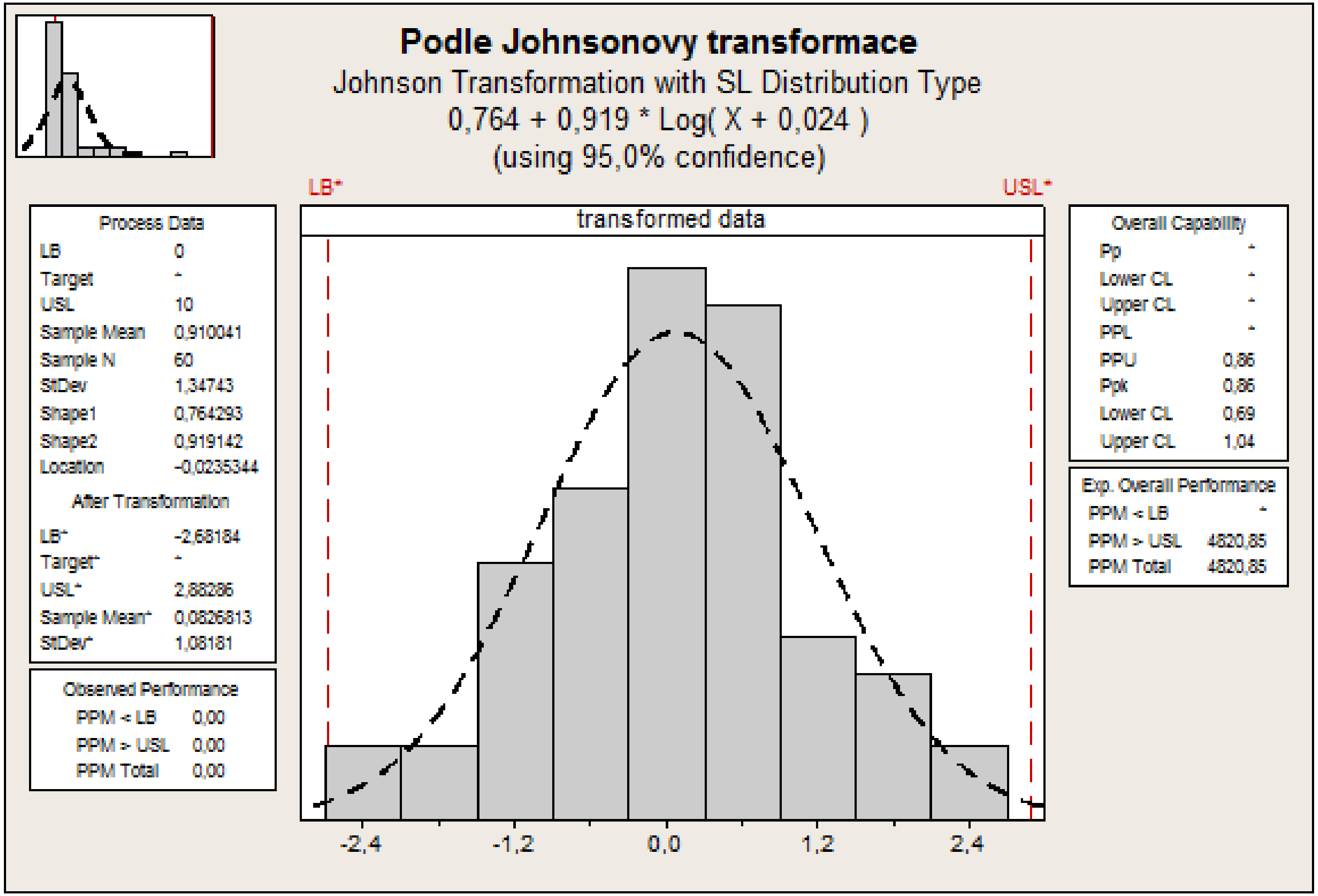
Volba vhodného modelu – Minitab 17



Volba vhodného modelu – Minitab 17



Volba vhodného modelu – Minitab 17



Odhad ukazatele Ppk pomocí tolerančního intervalu a funkce Percentile – Minitab 17

Podle vzorce

PpkU = (USL – Med)/(Upper Bound – Med)

V našem případě u dat C1 je tedy odhad

PpkU = (10- 0,445012)/(7,928- 0,445012)

PpkU = 1,2769

Tentýž odhad získáme pomocí funkce Percentile(*number*, *probability*) v nástroji Calculator v Minitab 17, když za *number* zadáme naměřené hodnoty a za *probability* 0,9973.
Je zajímavé, že tyto oba dva postupy, i když jsou z pohledu mat. statistiky jednodušší, dávají rozhodně lepší odhady nežli výše uvedené sofistikovanější metody.

Jak správně použít Boxovu-Coxovu a Johnsonovy transformace?

Z příkladu je vidět, že použité transformace silně podhodnotily výkonnost procesu. Je to způsobeno tím, že Minitab 17 (ale i předchozí verze) špatně transformace využívá, protože informace získané transformacemi nejsou převedeny do řeči původních dat pomocí inverzních transformací. Obecně platí, že ukazatele výkonnosti nejsou vůči použitým transformacím invariantní. Z tohoto důvodu byl v Excelu vytvořen nástroj Zpětné transformace, které pomohou odhadnout požadované kvantily u původních dat správným způsobem.

Nejdříve potřebujeme zjistit u transformovaných dat odhady požadovaných kvantilů. U našeho jednostranného případu se jedná o odhad 99,73%-ního kvantilu.

Optimální Boxova-Coxova transformace má parametr $\lambda = 0$

tedy ($T(x) = \ln x$)

Optimální Johnsonova transformace má tvar

$$T(x) = 0{,}7643 + 0{,}9191 \cdot \log(x + 0{,}0235)$$

Výpočet pomocí inverzní Johnsonovy transformace

Typ SL

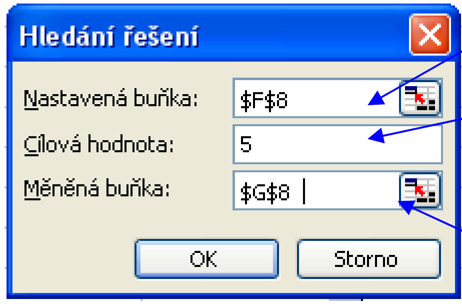
$a+b*LN(x+c) = y$

a =	0,7643
b =	0,9191
c =	0,05042

Hledání řešení

funkce y	**x**
2,669929	**7,901107**
Nastavená	Měněná
obsahuje vzorec	

Nástroje > Hledání řešení >



Buňka obsahující vzorec transformace

Transformovaná hodnota, pro kterou hledáme zpětné řešení

Buňka ve které bude zapsán výsledek

Odhad Ppk = (10 – 0,4450) / (7,9011- 0,4450) = 1,2815

Med = 0,4450 pro původní data

Výpočet pomocí inverzní Boxovy-Coxovy transformace

Opět pomocí nástroje Hledání řešení v Excelu zjistíme správný odhad požadovaného kvantilu pro původní data

y = ln x

λ = **0**

Hledání řešení

funkce y	**x**
2,07032	**7,92733**
Nastavená	Měněná
obsahuje vzorec	

7,9273300

Odhad ukazatele Ppk = (10 – 0,4450) / (7,9273 – 0,4450)= 1,2770
Med =0,4450 pro původní data

Postup výpočtu odhadu Ppk pomocí Minitab 17 při použití transformací

Po nalezení vhodné transformace jsou odhady ukazatelů Pp či Ppk počítány pomocí obvyklých vzorců pro normálně rozdělená transformovaná data a výsledná hodnota je prezentována jako odhad ukazatele pro původní data.
Tento postup je bohužel chybný, protože ukazatele nejsou obecně invarianty použité transformace.

Jako invariant lze použít odhad pravděpodobnosti náhodného jevu

$$P\{ X< USL\} = P\{ T(X)<T(USL)\}$$

a pak použít vzorec $PpkU = \Phi^{-1}(P\{ T(X)<T(USL)\}) / 3$.
Nutno ale dodat, že tento postup nedává tytéž výsledky jako postup založený na zpětné transformaci.

Nastává tedy otázka, který z uvedených postupů je ten správný?

Porovnání obou přístupů k definici ukazatelů výkonnosti pro nenormálně rozdělená data

Přístup A: vyžaduje nalezení odhadu kvantilového rozpětí mezi Up a Lp, kde Up je 99,865%-ní kvantil a Lp je 0,135%-ní kvantil rozdělení pravděpodobnosti sledovaného znaku kvality. Tyto odhady lze získat buď

1. Nalezením vhodného modelu rozdělení pravděpodobnosti (parametrický přístup)
2. Pomocí statistických tolerančních intervalů nebo odhadem percentilů (neparametrický přístup)
3. Nalezením vhodné transformace dat, ale pak je nutné použít zpětnou transformaci pro správný výpočet odhadu ukazatele

Je nutné si ale uvědomit, že tento přístup nedává žádnou informaci o odhadu počtu neshodných mimo specifikační meze, tj. např. z hodnoty Pp= 1,42 nelze usuzovat, že očekávaný počet mimo specifikace je lepší nežli 64 ppm

Porovnání obou přístupů k definici ukazatelů výkonnosti pro nenormálně rozdělená data

Přístup B: vyžaduje odhad pravděpodobnosti výskytu neshodných mimo specifikace. Tento odhad lze získat buď

1. Nalezením vhodného modelu rozdělení pravděpodobnosti, pomocí něhož je tato pravděpodobnost odhadnuta
2. Nalezením vhodné transformace na data normálně rozdělená a odhadem pravděpodobnosti výskytu neshodných kusů pomocí transformovaných dat a výpočtem odpovídajícího kvantilu rozdělení N(0,1)

Tento přístup umožňuje z hodnoty odhadu ukazatele výkonnosti odhadnout i počet neshodných kusů (např. v ppm). Tato možnost vychází z invariantnosti pravděpodobnosti výskytu neshodných kusů vůči použité transformaci

Závěr: Při vyšetřování výkonnosti výrobního procesu u znaků kvality, které nelze popsat normálním rozdělením,

především závisí na zkušenosti pracovníka, který hodnocení provádí

a na informacích z dřívějšího hodnocení výrobního procesu. Bohužel neexistuje jednoznačné doporučení, jak postupovat, každý případ je nutno řešit zcela individuálně.

A bohužel nelze se spolehnout ani na renomovaný software!