



**Národní informační středisko  
pro podporu kvality**

# **Výkonnost procesů v případě nenormálně rozděleného znaku kvality**

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.

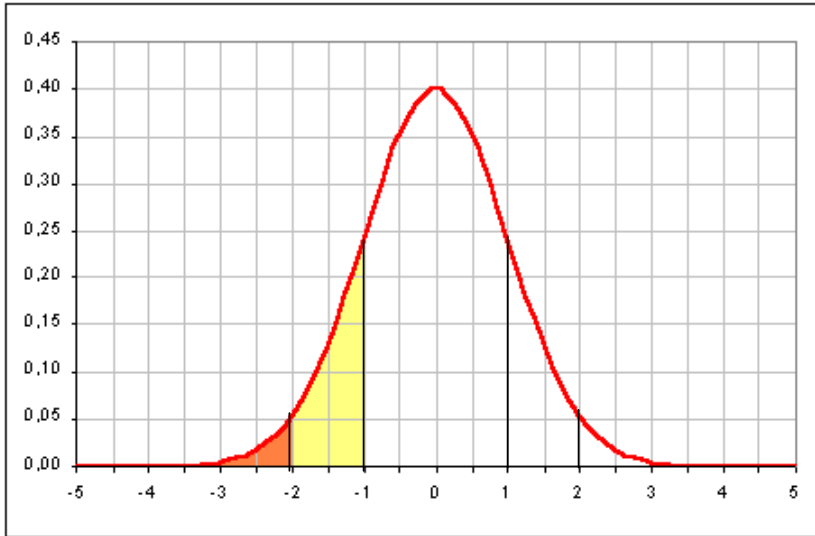
Bohužel ne vždy jsou výsledky získané statistickými metodami interpretovány správným způsobem a není tomu ani jinak u ukazatelů způsobilosti a výkonnosti. Celá řada softwarů na základě naměřených dat vrátí pouze odhady těchto ukazatelů, aniž by se mnohdy ověřily předpoklady pro jejich správné použití 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 teorii stanovené předpoklady pro použití statistických postupů v řízení kvality a varují před jejich vědomým nerespektováním. 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  $2 \cdot 15,87 \%$ , t.j. **31,74 %**.  $P_p = 0,33$

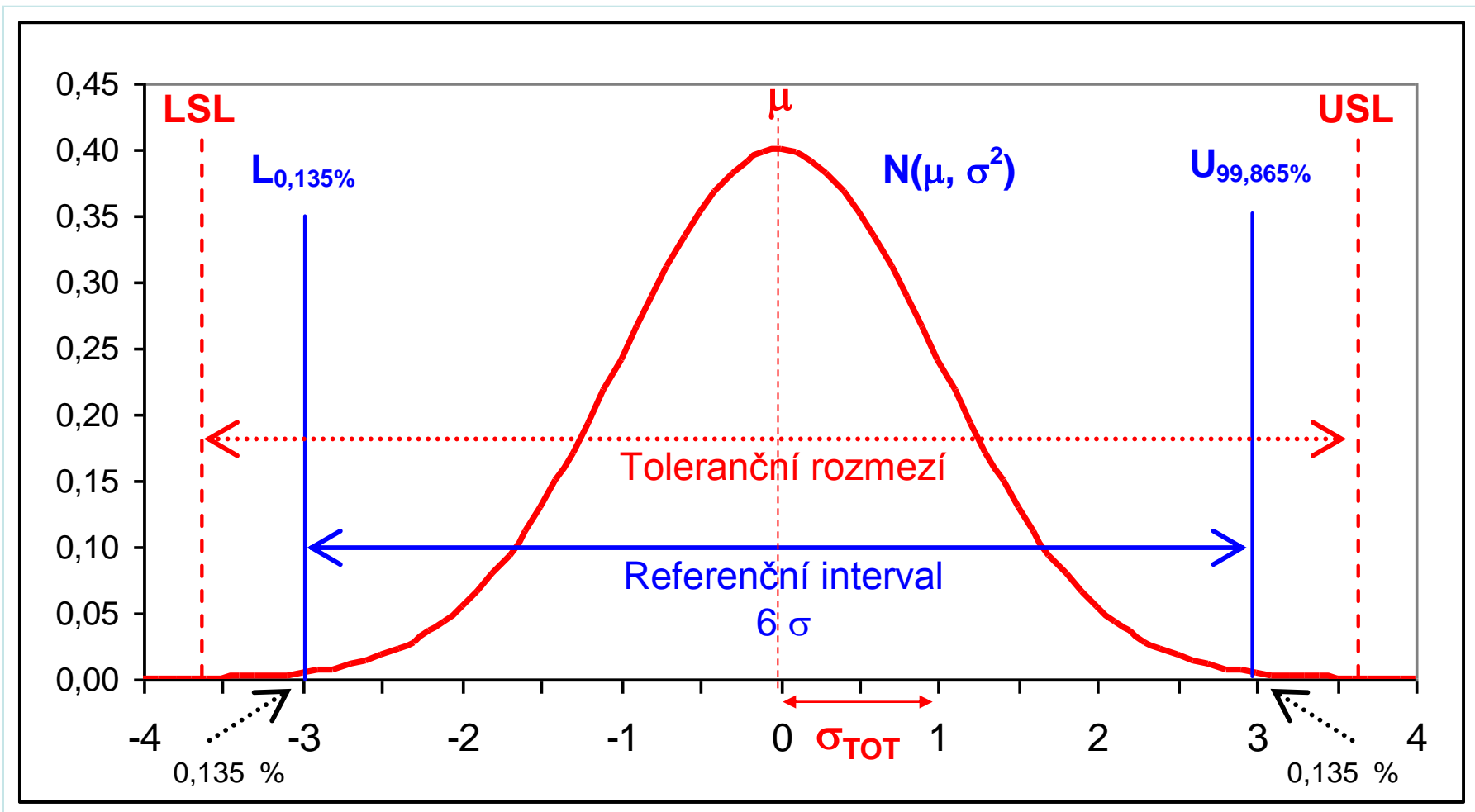
V intervalu  $\langle \mu - 2\sigma, \mu + 2\sigma \rangle$  leží **95,44 %** všech pozorování, mimo tento interval leží  $2 \cdot 2,28 \%$ , t.j. **4,56 %**.  $P_p = 0,67$

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

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

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

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

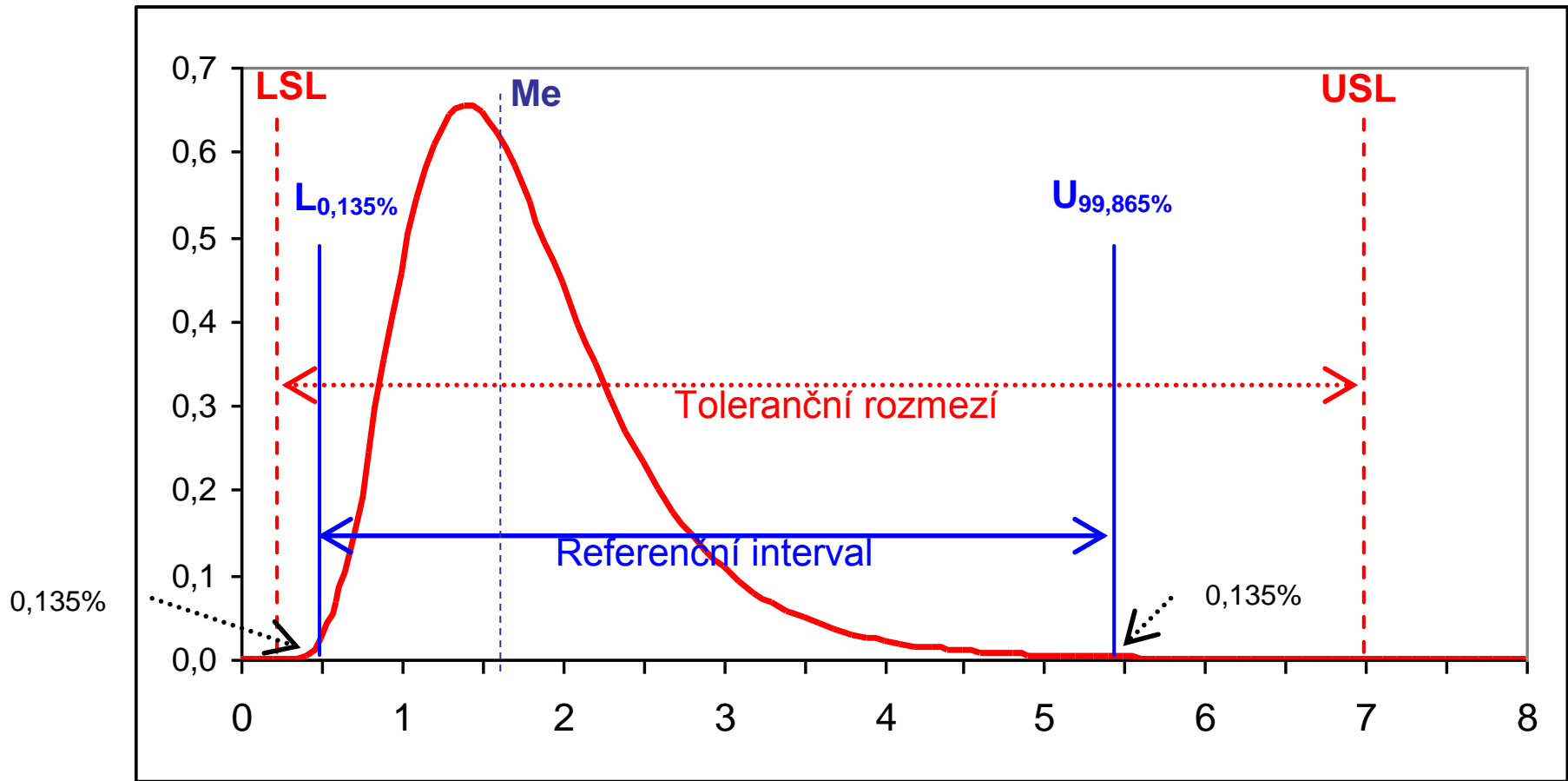


Ukazatele výkonnosti v případě normálně rozděleného znaku kvality

$$P_p = \frac{USL - LSL}{6\sigma_{TOT}}$$

$$P_{pkU} = \frac{USL - \mu}{3\sigma_{TOT}}$$

$$P_{pkL} = \frac{\mu - LSL}{3\sigma_{TOT}}$$



Ukazatele výkonnosti v případě nenormálně rozděleného znaku kvality

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



## Znamé 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.

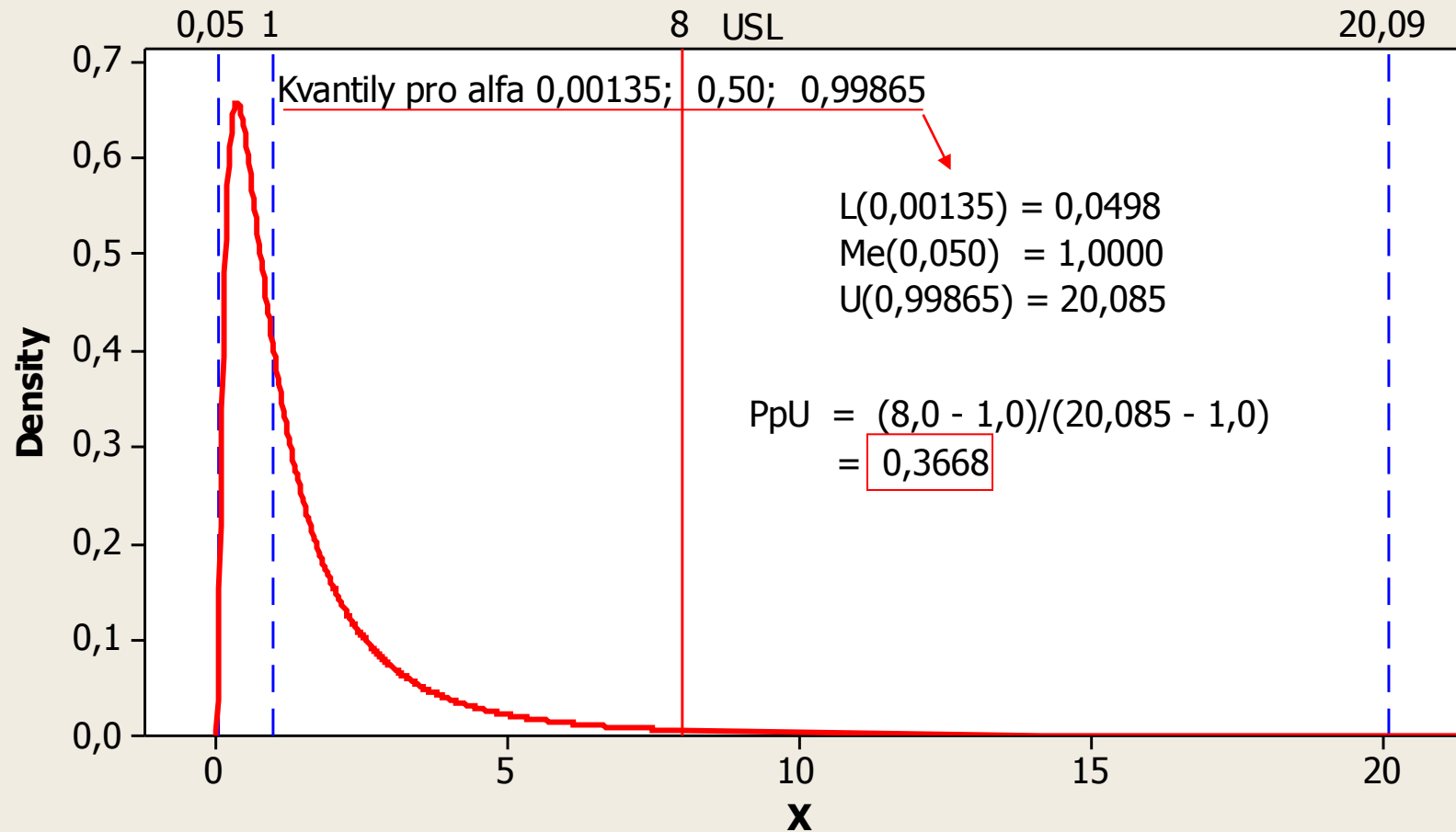
**Příklad:** Log-normální rozdělení sledovaného znaku kvality s předepsanou horní mezní hodnotou  $USL = 8,0$ .

(„Sem – únor 2010.MPJ“)

Model

## Distribution Plot

Lognormal; Loc=0; Scale=1; Thresh=0

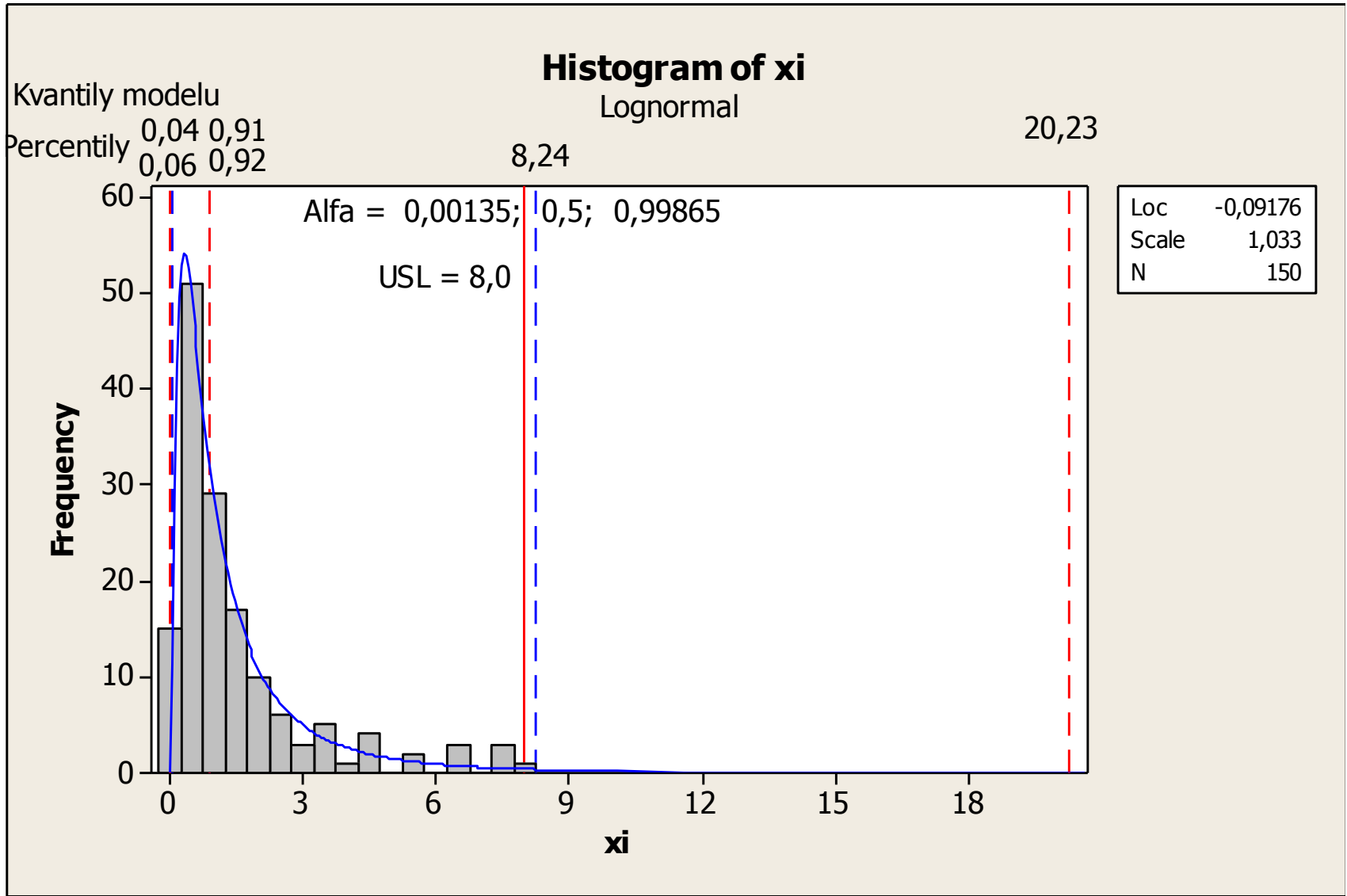


## Metoda identifikace rozdělení

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 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.

Vychází se z napozorovaných dat zapsaných např. ve sloupci xi.



## Poznámka

### Kvantily $\Leftrightarrow$ Percentily

Pro zadané  $\alpha$  odpovídá  **$\alpha$ -kvantil**  $x_\alpha$  (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  $\alpha$  odpovídá  **$\alpha$ -percentil**  $X_\alpha$  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  $\alpha$  v souvislosti s kvantily jako podíly a hodnoty  $\alpha$  v souvislosti s percentily jako procenta. Tato zásada se v praxi často nedodrží.

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

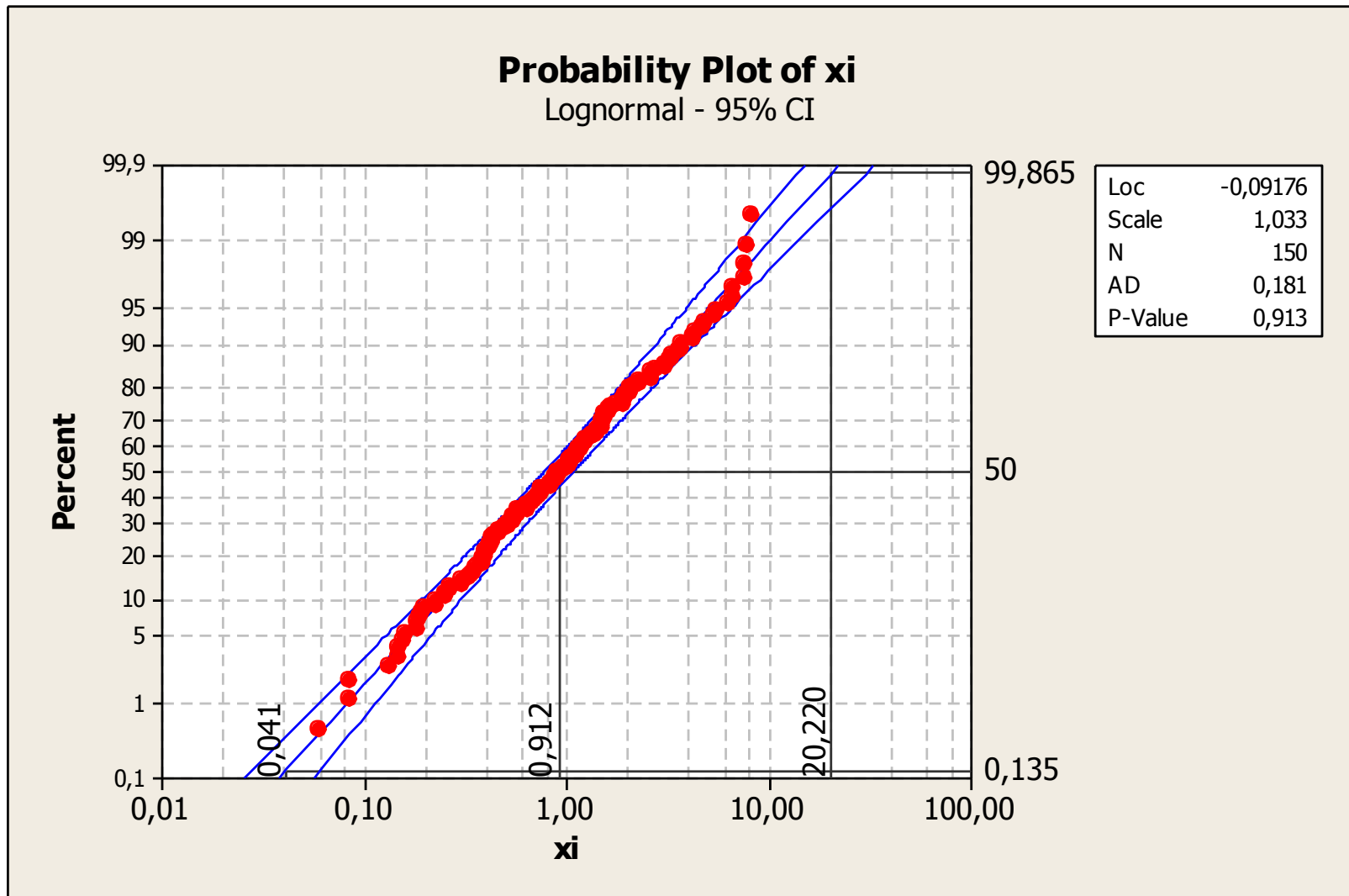
Table of Percentiles

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

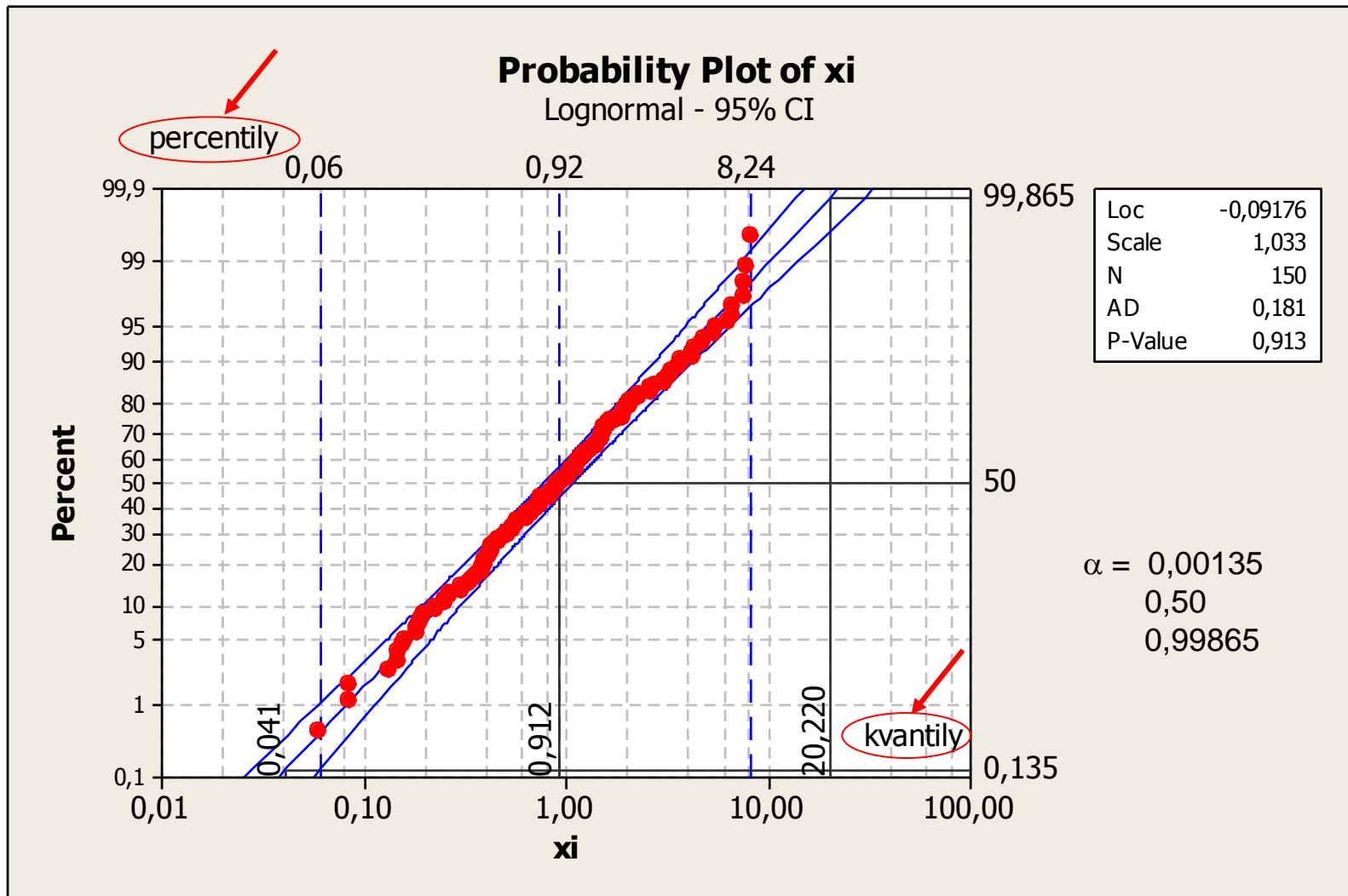
Do pravděpodobnostního papíru, kde stupnice na ose  $x$  je lineární a stupnice na ose  $y$  je pravděpodobnostní, odpovídající určitému rozdělení pravděpodobnosti, se zakreslí  $N$  naměřených hodnot, uspořádaných podle velikosti  $(x(i))$  s odpovídajícími hodnotami empirické distribuční funkce  $((i) / N)$ . 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).

## Metoda Pearsonových křivek

Jako alternativní metoda může být použita metoda normalizovaných Pearsonových křivek pomocí které se odhadne model rozdělení sledovaného znaku kvality a vypočítají odhady kvantilů

$$\hat{L}(0,00135); \hat{M}_e(0,50); \hat{U}(99,865).$$

V případě použití této metody je třeba počítat vedle výběrového průměru a výběrové směrodatné odchylky i výběrovou šikmost a výběrovou špičatost. Tato metoda, která poskytuje jen přibližné výsledky, není preferována pro svoji pracnost.

*Využití této metody je demonstrováno v Dodatku 5 k publikaci [1].*

## 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

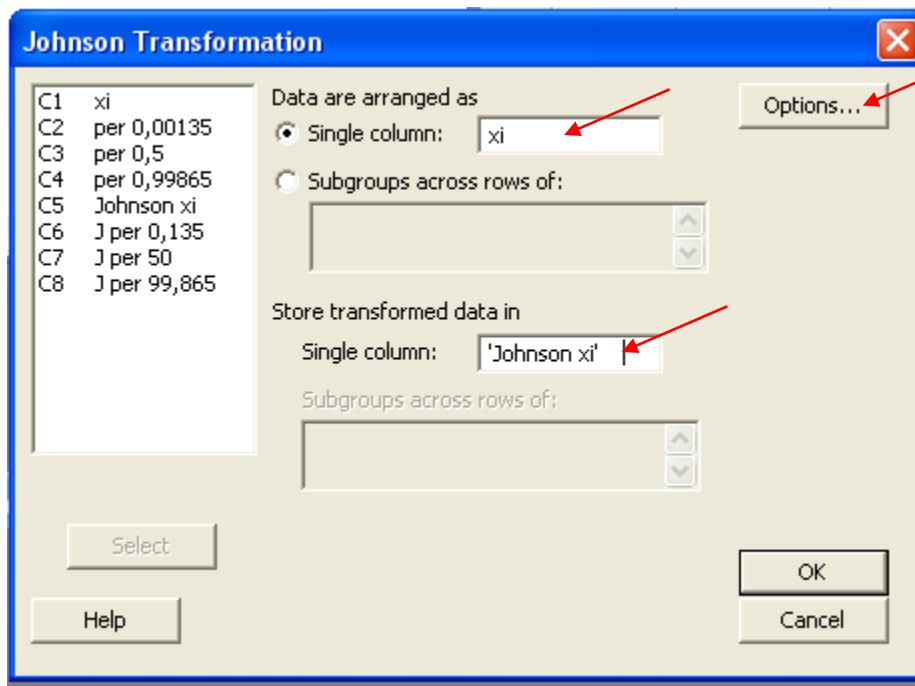
$$y = a + b \cdot \operatorname{Asinh}((x - c) / d)$$

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

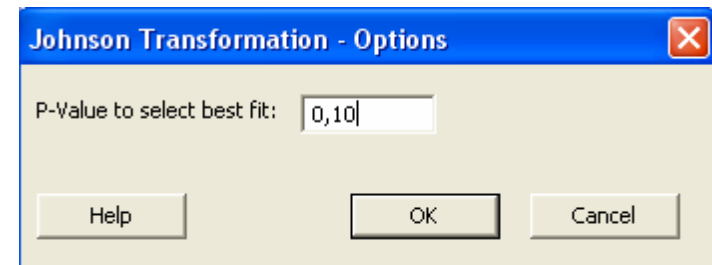
Software postupuje tak, že uvažuje všechny možné funkce Johnsonova systému, odhadne jejich parametry, transformuje data, vypočítá Anderson-Darlingovu testovou statistiku a jí odpovídající p-hodnotu a vybere tu funkci, které odpovídá největší p-hodnota.

Např. software Minitab postupuje následovně:

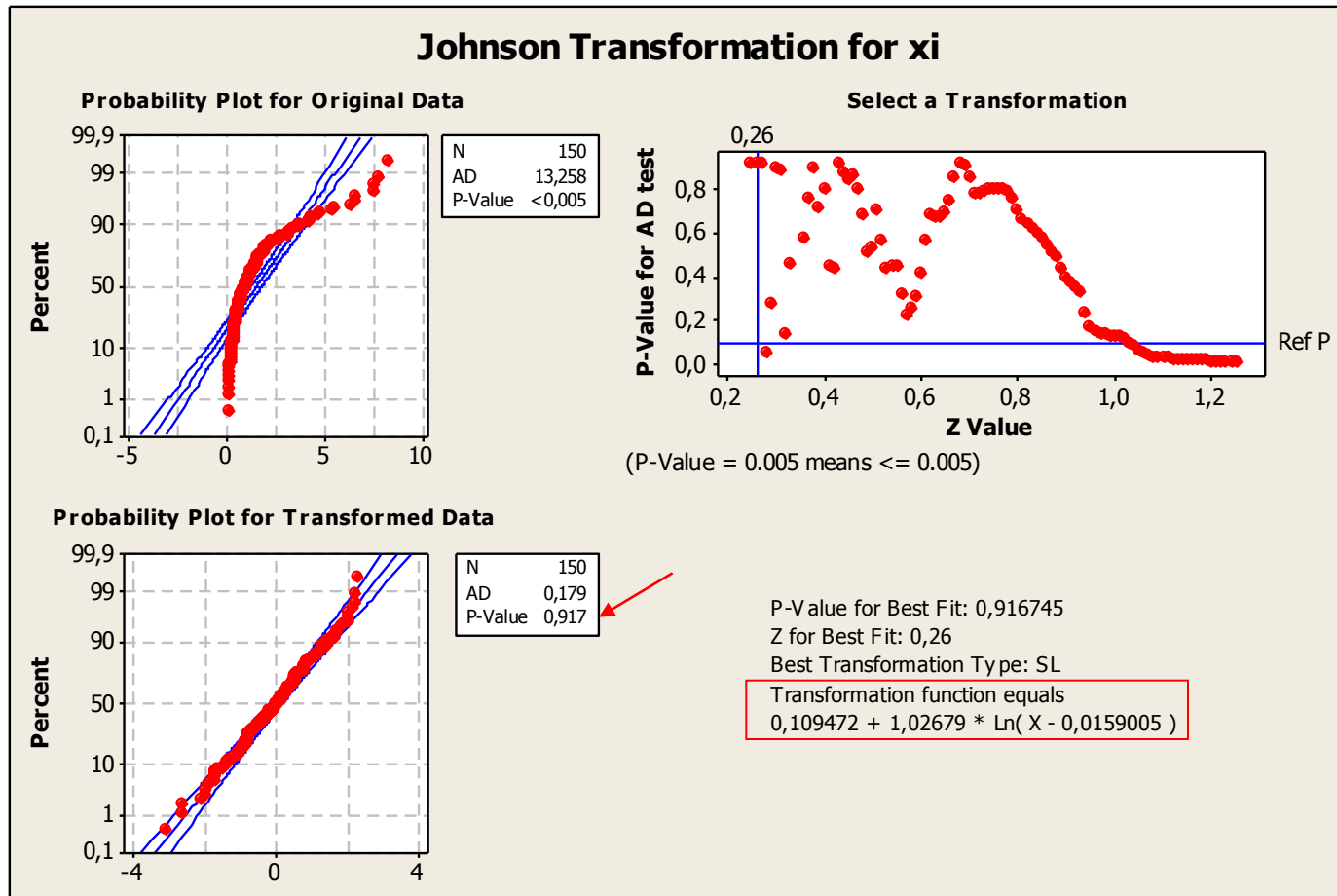
Stat > Quality Tools > Johnson transformation > a vyplníme dialogové okno.



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



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



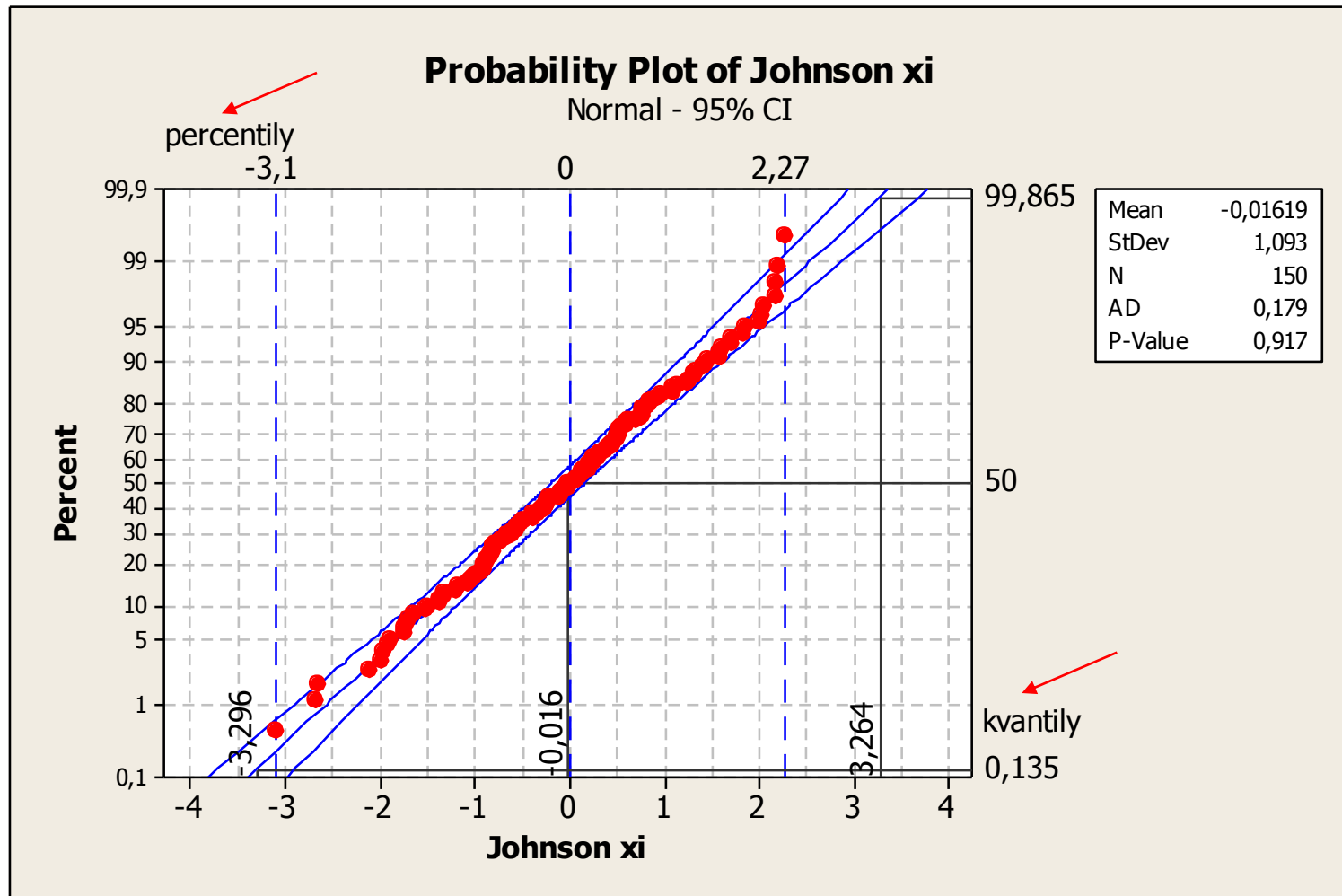
Z grafu je patrné, že p-hodnota Anderson-Darlingova testu dobré shody s normálním rozdělením je vysoká (0,917) a že tedy není důvod pochybovat, že transformovaná data jsou rozdělena normálně. Dále je zde transformační rovnice typu SL

$$0,109472 + 1,02679 \cdot \ln((x - 0,0159005)).$$

Z transformovaných dat ve sloupci C5 (Johnson xi) můžeme vypočítat odhady percentilů pro procenta 0,135; 50; 99,865 pomocí kalkulačky 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
<b>Johnson xi</b>	<b>J per 0,135</b>	<b>J per 50</b>	<b>J per 99,865</b>
1,09225	-3,10047	0,0006066	2,27349
0,53542			
0,20609			
-0,08605			



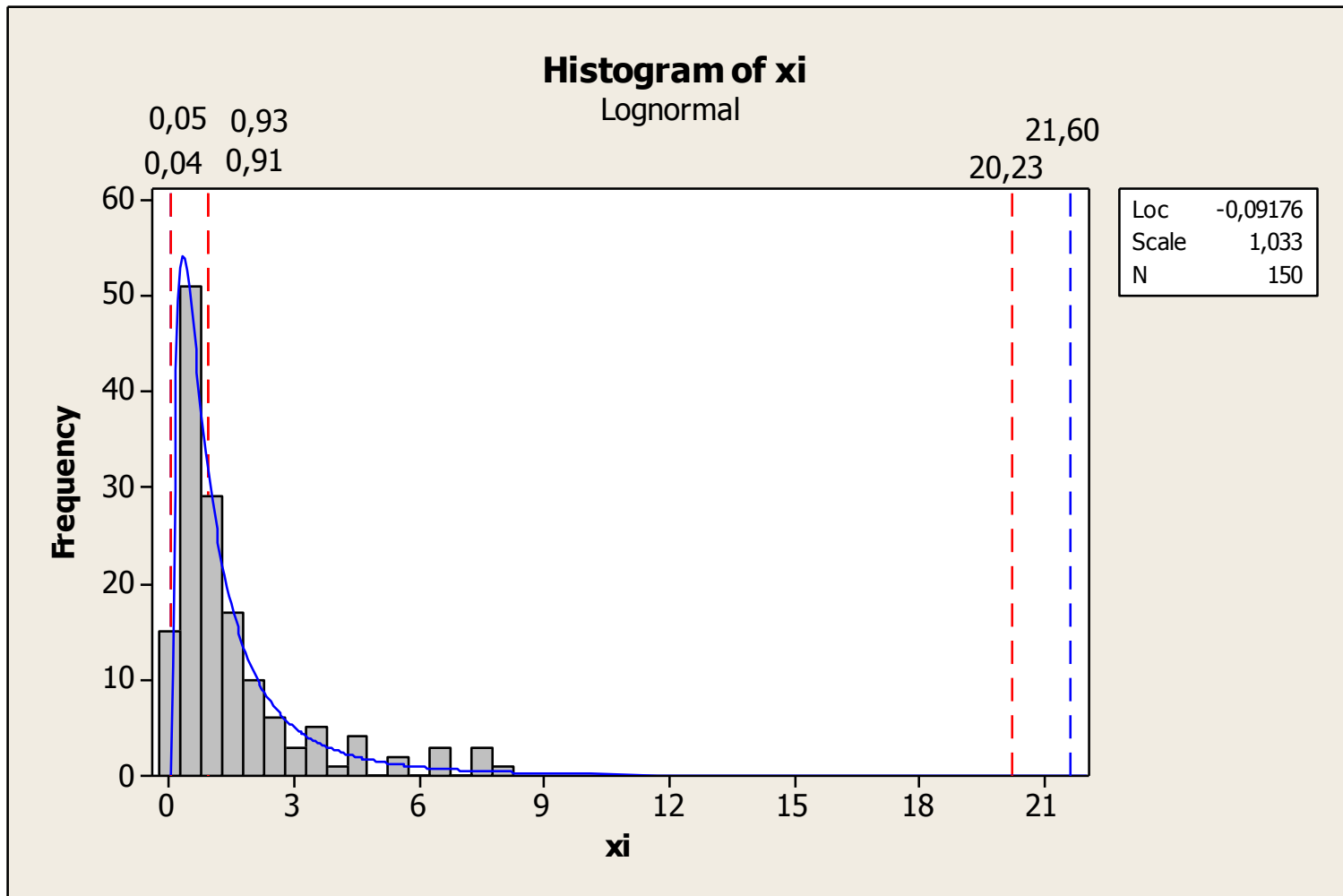
Porovnání percentilů a kvantilů transformovaných dat vypočítaných po vložení „Add - Percentil Lines“ na pravděpodobnostním grafu.

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 (sloupec C1 – xi) 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“:

The screenshot shows an Excel spreadsheet with the following data:

Typ SL	
$a+b \cdot \ln(x+c) = y$	
a =	0,109472
b =	1,02679
c =	-0,015901

The Solver tool is configured as follows:

Hledání řešení	
funkce y	x
-3,296039	0,052174
Nastavená	Měněná
obsahuje vzorec	

The Solver dialog box is open, showing the following settings:

Hledání řešení	
Nastavená buňka:	\$F\$8
Cílová hodnota:	5
Měněná buňka:	\$G\$8

Two tables show the transformation of y and x values:

Transformace	
y	x
0,016029	0,928913

Transformace	
y	x
3,263909	21,603840

Annotations in the image point to the Solver dialog box and the transformation tables:

- Buňka obsahující vzorec transformace (points to the Solver dialog box)
- Transformovaná hodnota, pro kterou hledáme zpětné řešení (points to the Solver dialog box)
- Buňka ve které bude zapsán výsledek (points to the Solver dialog box)
- Transformovaná hodnota (points to the y value in the first transformation table)
- Hodnota původní proměnné (points to the x value in the first transformation table)

## Metoda Box-Coxovy transformace

Box-Coxova transformace odhaduje hodnotu  $\lambda$  , která minimalizuje směrodatnou odchylku normalizované transformované proměnné.

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

Hodnota  $\lambda$

$$\lambda = 2$$

$$\lambda = 0,5$$

$$\lambda = 0$$

$$\lambda = -0,5$$

$$\lambda = -1$$

Transformace

$$y = x^2$$

$$y = \sqrt{x}$$

$$y = \ln x$$

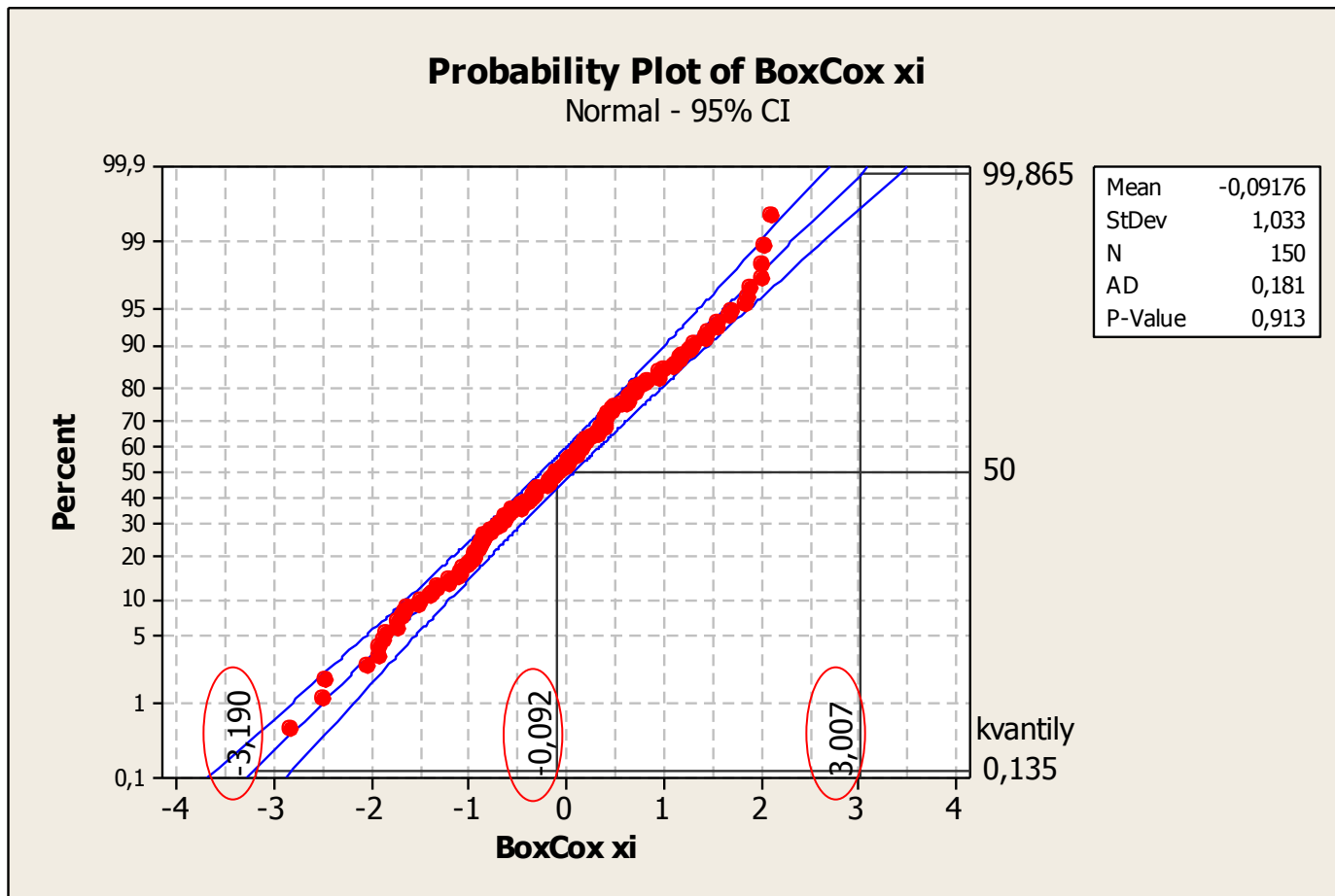
$$y = 1 / (\sqrt{x})$$

$$y = 1 / x$$

Např. software Minitab postupuje následovně:

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

Stat > Control Charts > Box-Cox Transformation > a v dialogovém okně zadáme data, sloupec C1 – „xi“, velikost případných podskupin, v tomto případě  $n = 1$ . Pod „Options“ zvolíme sloupec C2 – „BoxCox xi“, 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“.

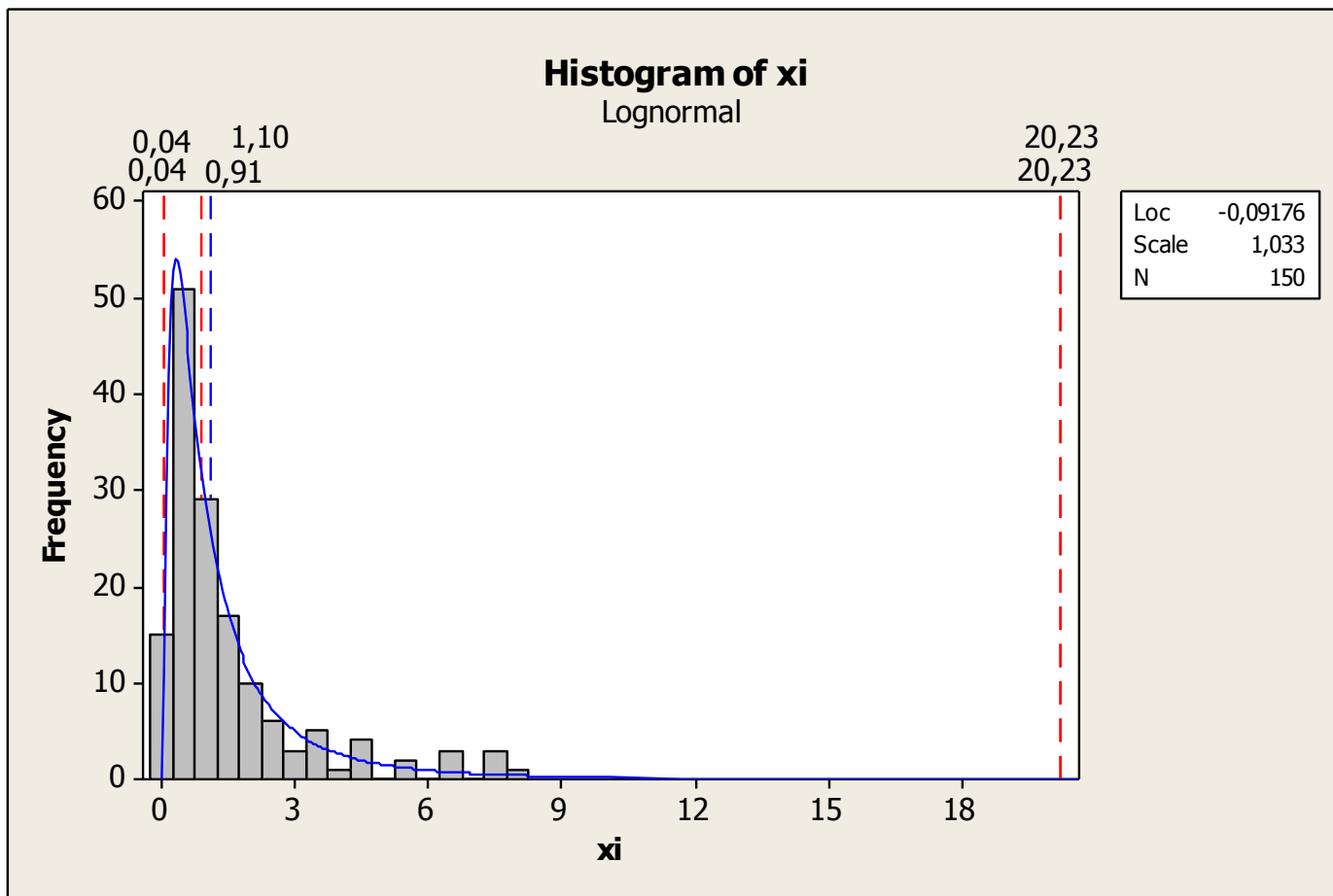
$y = \ln x$					
$\lambda =$	0	Hledání řešení			
		funkce y	X		
		-3,19081	0,04114		
		Nastavená	Měněná		
		obsahuje vzorec			
				-3,190	

		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 BoxCox 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

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

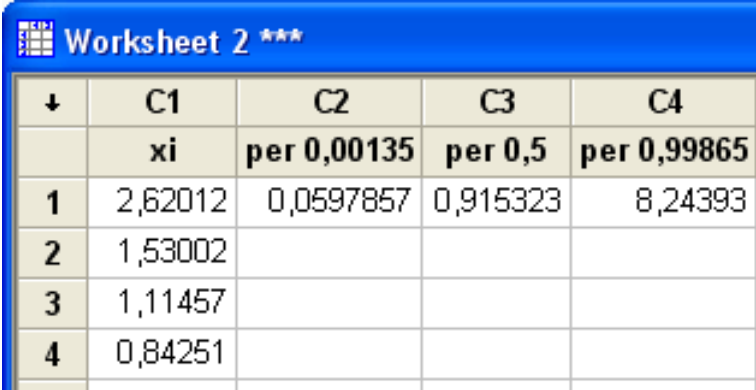
odpovídající procentům, se kterými pracují ukazatele výkonnosti.

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.

V projektu „SEM – ÚNOR 2010.MPJ“ jsou v listě „Worksheet 2“ ve sloupci C1 – Data a ve sloupcích C2 až C4 vypočtené percentily:



	C1	C2	C3	C4
	<b>xi</b>	<b>per 0,00135</b>	<b>per 0,5</b>	<b>per 0,99865</b>
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

### Process Data

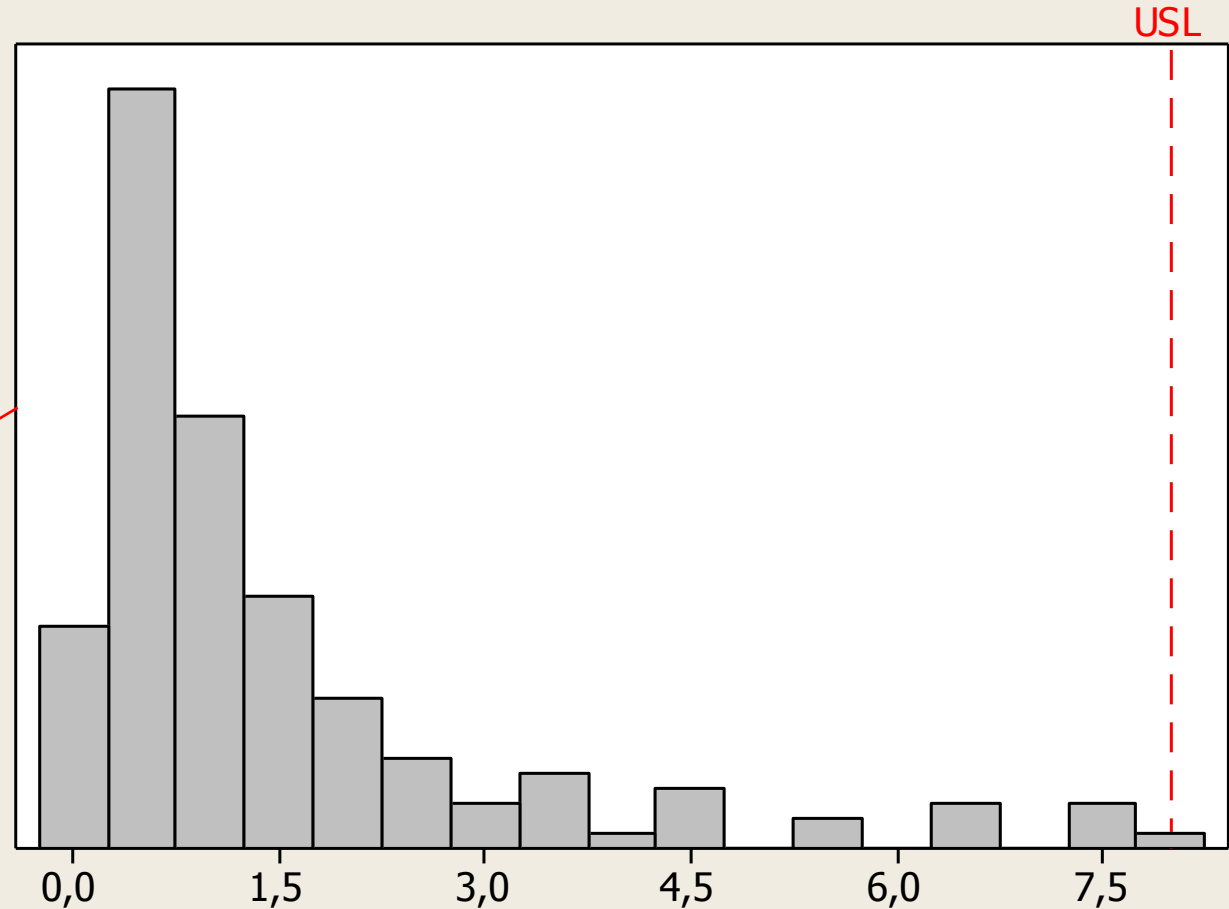
LSL	*
USL	8,00000
Sample Median	0,91532
Sample Mean	1,51631
Sample StDev	1,69720
Sample N	150

### Capability Indices

C <sub>np</sub>	*
C <sub>npl</sub>	*
C <sub>npu</sub>	0,966715
C <sub>npk</sub>	0,966715

### Observed Performance

PPM < LSL	*
PPM > USL	6666,67
PPM Total	6666,67



**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.