

STATISTICKÉ ŘÍZENÍ PROCESŮ SE SW PODPOROU

RNDr. Jiří Michálek, CSc.

Centrum pro kvalitu a spolehlivost CQR
při Ústavu teorie informace a automatizace AVČR
e-mail: michalek@utia.cas.cz

Ing. Jan Král

ISQ PRAHA s.r.o., Pechlátova 19, 150 00 Praha 5,
tel./fax: 251 553 339; e-mail: kral.jan@isq.cz

Klíčová slova: SPC, regulační diagramy; metody statistické regulace; statisticky zvládnutý proces; rozšířené (modifikované) regulační meze.

Metodická schémata pro aplikaci SPC

Cílem těchto schémat je poskytnout lidem z praxe ověřené a teoreticky podložené postupy, kterých by se měli držet při používání statistických nástrojů pro řízení a sledování znaků jakosti na výrobku. Jde totiž o to, aby se grafické a numerické výsledky statistické analýzy skutečně vztahovaly k reálné situaci, která panuje ve výrobním procesu. *Schéma 1* ve formě vývojového diagramu ukazuje základní vztah mezi daty získanými při sledování výrobního procesu a možnostmi použití nástrojů pro SPC. Z procesu lze na výrobku získávat data dvojí povahy. Jednak jsou to data diskrétní povahy, která vyjadřují buď počet neshodných výrobků či počet neshod. Hovoříme taktéž o atributivních datech. Druhým typem dat jsou data spojitého charakteru, které obvykle nesou více informace o stavu procesu nežli data atributivní. Při analýze diskrétních dat se obvykle problémy nevyskytují, zde stačí použít binomické či Poissonovo rozdělení pro jejich analýzu, větší problémy nastávají u dat spojitě povahy, neboť tam je nutno najít vhodný statistický model neboli rozdělení pravděpodobnosti obvykle ve formě hustoty, který nám pomůže s daty správně pracovat. Je tedy zapotřebí ověřit splnění základních požadavků pro správnou aplikaci nástrojů SPC, aby statistická analýza dávala relevantní výsledky, protože zcela formální postup bez ověření předpokladů, které matematická statistika vyžaduje, může přinést zcela nereálnou představu o stavu výrobního procesu.

Prvním krokem je samozřejmě sběr dat (viz *Schéma 2*). Bez nich se nelze nic dozvědět o stavu výrobního procesu. tento krok je velice důležitý, protože od kvality dat se vše ostatní odvíjí. Před vlastním odběrem dat z procesu je nutno rozhodnout, která data budeme získávat, jak je budeme měřit či zjišťovat, kdo bude sběr provádět, jak často, kam se data budou zapisovat či ukládat, jaké množství dat bude zapotřebí, aby data přinesla žádanou informaci, jaký problém chceme pomocí dat řešit či k čemu budeme data potřebovat. S tím úzce souvisí samozřejmě i analýza systému měření, především reprodukovatelnost a opakovatelnost, přesnost měření na dostatečný počet desetinných míst a též zaškolení pracovníků hlavně při využívání výpočetní techniky spolu s nějakým softwarem. Podle charakteru dat hovoříme o SPC měřením či SPC srovnáváním. Platí doporučení, pokud lze informaci získat měřením, je to vždy výhodnější především z pohledu počtu získávaných dat nežli při SPC srovnáváním, i když při srovnáváním jsou data získávána obvykle snazším způsobem.

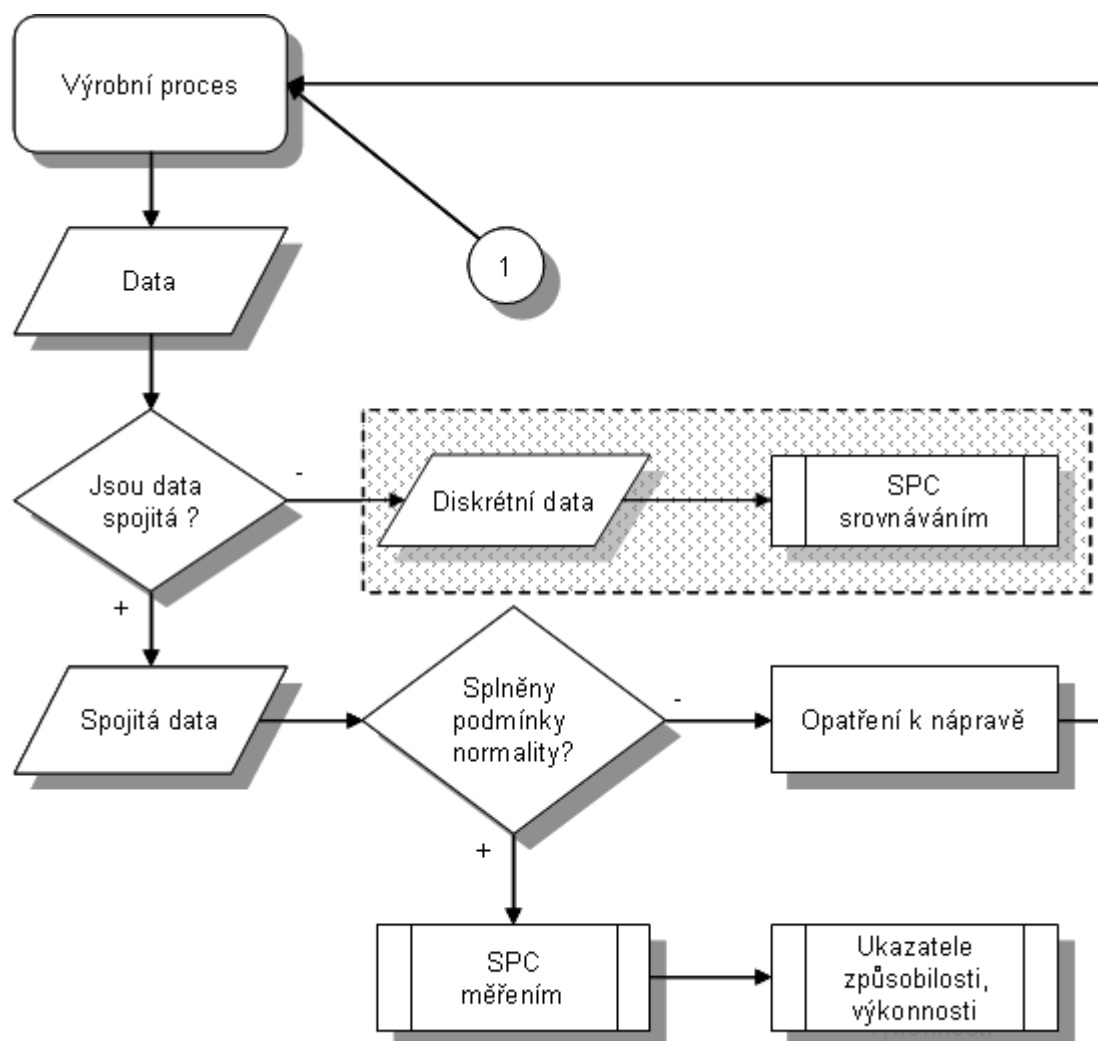


Schéma 1:

Druhým krokem je základní statistická analýza dat spolu s ověřením možnosti aplikace Shewhartových regulačních diagramů. V tomto místě se postupy odlišují podstatně podle toho, o jaký typ dat se jedná:

- U dat diskrétního charakteru není obvykle nutné provádět ověřování předpokladů, pro volbu vhodného regulačního diagramu je pouze nutné si uvědomit, zdali sledujeme počet neshodných výrobků či počet neshod.
- U dat spojitého charakteru je situace komplikovanější, neboť musíme najít vhodný statistický model pro popis dat. Nejčastěji se jedná o model normálního rozdělení, ale není to pravidlem.

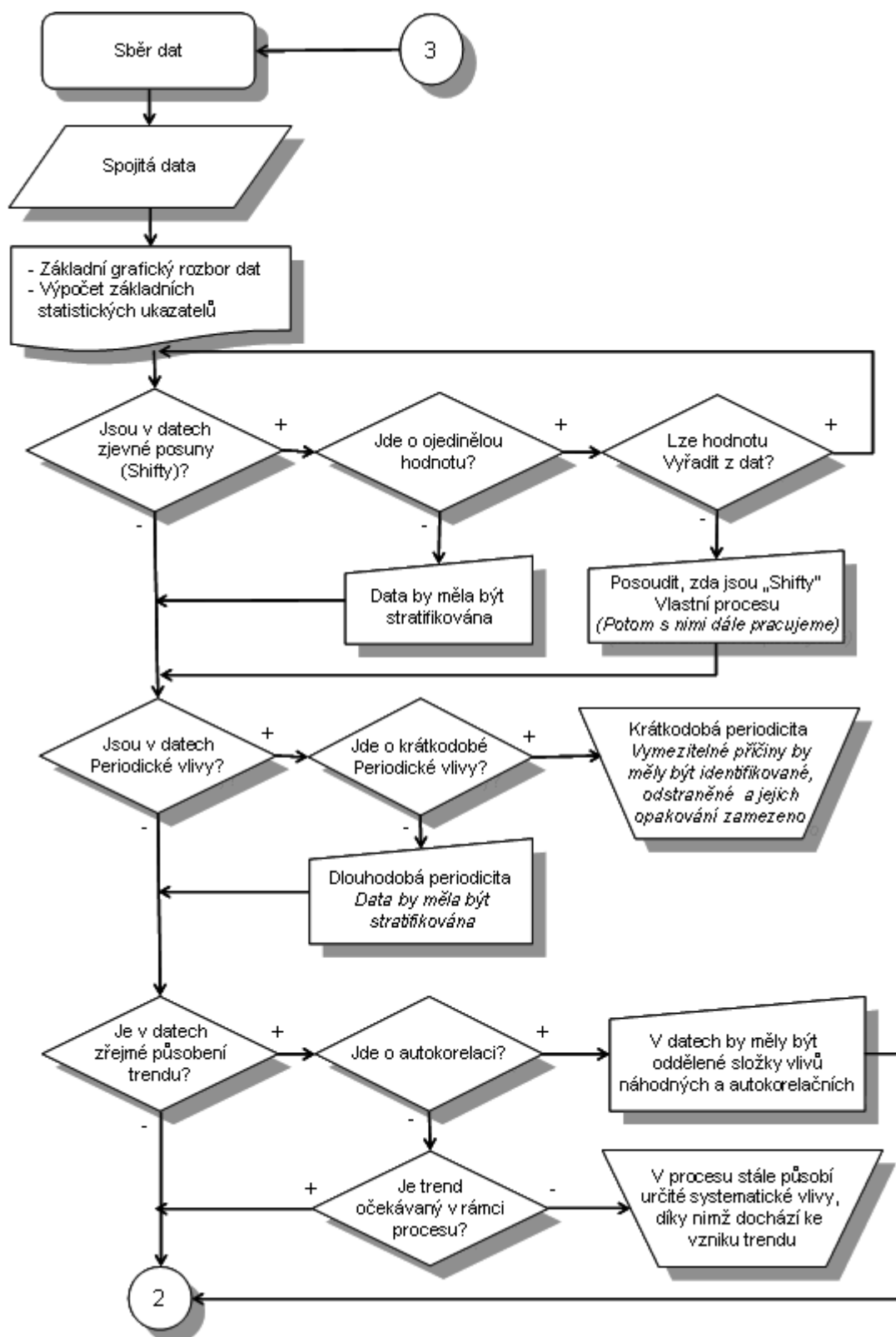


Schéma 2/1:

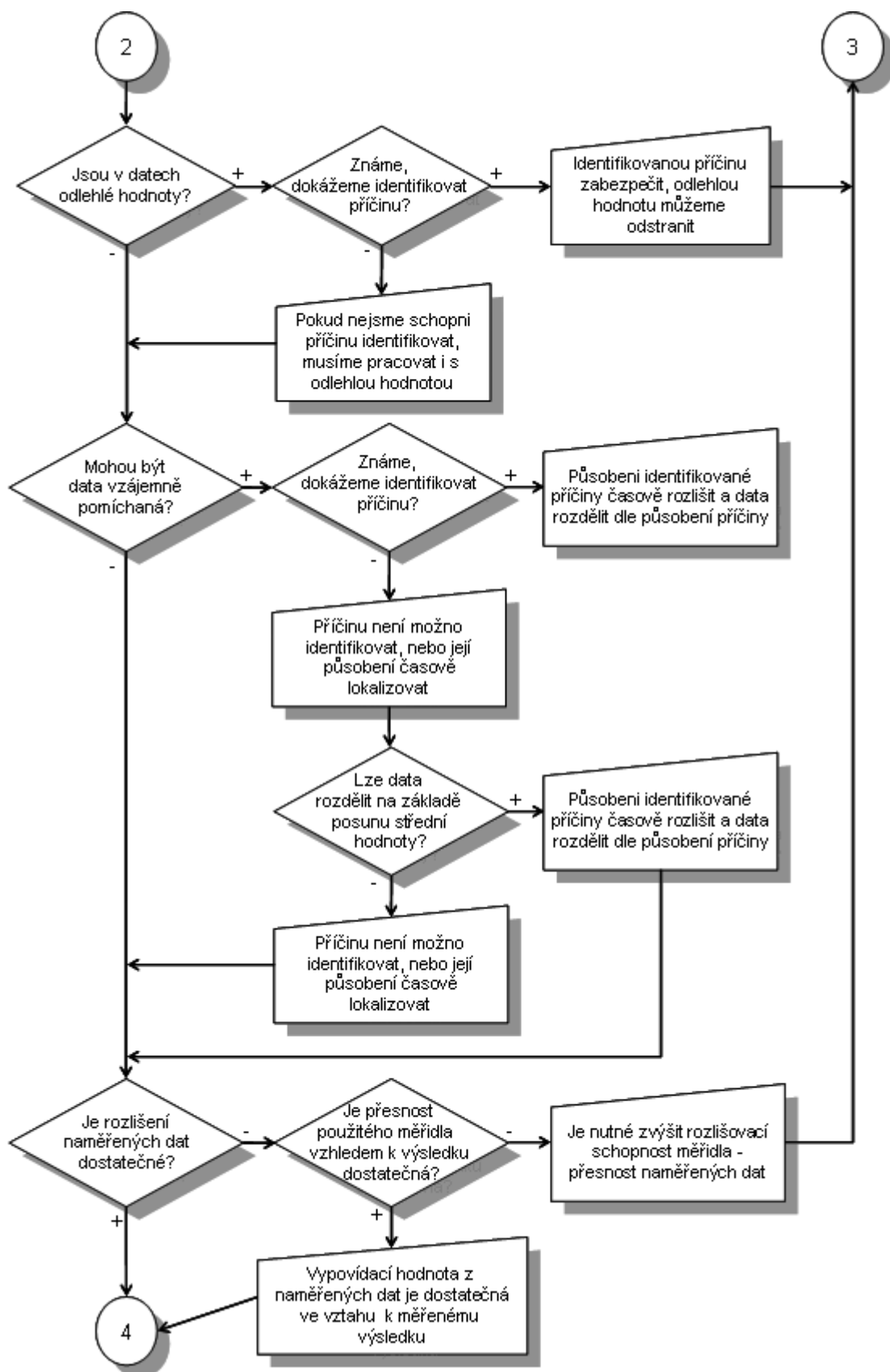


Schéma 2/2:

Začíná se grafickým rozbohem dat s nástroji jako je histogram, bodový či krabicový diagram a průběhový diagram. Podle tvaru histogramu usuzujeme na vhodný model rozdělení pravděpodobnosti, jehož volbu ověříme pomocí testu dobré shody. Je nutno mít na paměti, že konstrukce klasických Shewhartových regulačních diagramů je založena na předpokladu normálně rozdělených dat. Tento předpoklad je nutno ověřit především při aplikaci regulačního diagramu I-MR pro individuální data, i když i v případě aplikace regulačních diagramů \bar{x} -R či \bar{x} -s se nemusejí aritmetické průměry z jednotlivých logických podskupin chovat jako normálně rozdělená data. Toto nebezpečí je akutní hlavně u podskupin s malým počtem dat a kde vychází data vykazují silně asymetrické chování (histogram je silně vychýlený na jednu stranu). V těchto případech je nutno konstrukci regulačních diagramů založit na příslušných kvantilech vhodného rozdělení pravděpodobnosti, které popisuje chování individuálních dat u diagramu I-MR či aritmetických průměrů u \bar{x} -R či \bar{x} -s diagramů.

Průběhový diagram slouží k identifikaci zjevných nenáhodných seskupení dat, případně odhalení kandidátů na odlehlá pozorování. Za nenáhodná seskupení se považují posuny, periodické chování apod. Každé takové seskupení je vyvoláno nějakou reálnou příčinou, která vyvolala změnu podmínek, za nichž byla data získána. Rovněž tak u dat podezřelých z odlehlosti zřejmě došlo ke změně podmínek, za nichž byla data získána. Tyto změny by měly být identifikovány a mělo by být zajištěno, aby se pokud možno neopakovaly. Proto je vhodné či přímo nutné současně se sběrem dat vést poznámky, ve kterých bude zaznamenána jakákoliv změna či zásah týkající se výrobního procesu. Tyto poznámky pak slouží ke snazší identifikaci příčin vyvolávajících změny v procesu. Když se nepodaří odhalit příčinu v chování procesu, jedná se o nestabilitu v jeho průběhu a u takového procesu aplikace klasických regulačních diagramů Shewhartova typu obvykle selhává a je nutno použít modifikované typy diagramů, například s rozšířenými mezemi. Dalším jevem, který se může v praxi vyskytovat je závislost dat, která opět není u Shewhartových diagramů předpokládána. Je tedy vhodné u dat, která vykazují podezření na setrvačnost se přesvědčit, zdali nejsou mezi sebou jako časová řada korelována. Zvláště v případě kladné autokorelace by formální použití klasických regulačních diagramů mohlo vést ke zvýšenému počtu falešných poplachů.

Shewhartovy regulační diagramy

Základním předpokladem pro aplikaci klasických Shewhartových diagramů je stabilita procesu, neboli jeho statistická zvládnutelnost, a normalita výchozích dat. Postup, jak tyto předpoklady ověřit, je popsán ve *Schématu 3*. Stabilita procesu znamená, že parametr polohy i úroveň variability u sledovaného znaku jakosti lze považovat za konstantní v čase. Je nutné si uvědomit na základě zkušeností z praxe, že takto se chová pouze malé procento reálných výrobních procesů. Budeme hovořit o zvládnutelnosti „v užším slova smyslu“. Nutným předpokladem pro dosažení takového stavu procesu je přítomnost pouze tzv. náhodných příčin, které nelze nikdy úplně z procesu eliminovat, které neumíme odhalit a každá působí pouze nepatrnou měrou na chování procesu. Jejich přítomnost vyvolává jistou úroveň tzv. inherentní variability, která je nejhodněji odhadována pomocí odhadů variability uvnitř logických podskupin. Pokud nelze některé vymezitelné (speciální) příčiny, které vyvolávají změny v chování znaku jakosti z procesu odstranit, je nutno vlivy těchto změn zakomponovat do tvorby a používání regulačních diagramů. Toto je v praxi vcelku běžný stav. Jestliže se ale podaří dosáhnout takového stavu procesu, kdy jsou tyto změny pod kontrolou a zvládnuty, potom budeme hovořit o procesu stabilním „v širším slova smyslu“. Nejčastěji se jedná o změny v chování parametru polohy, tzn. trendy či posuny vůči požadované cílové hodnotě.

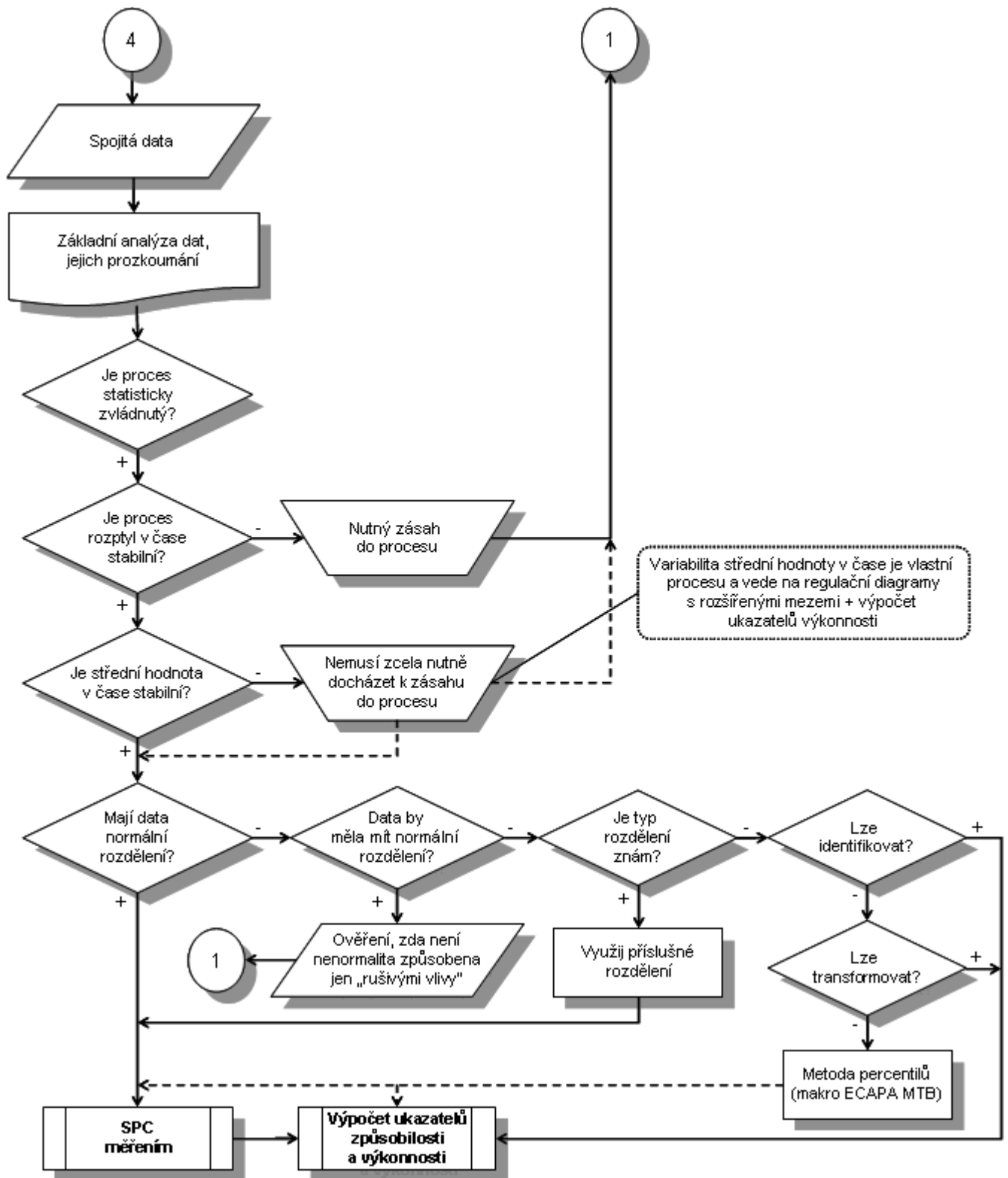


Schéma 3:

V takovém případě do chování sledovaného znaku jakosti vstupuje další forma variability, a to variabilita mezi logickými podskupinami vyvolaná právě změnami v procesu. Tato variabilita není pak zachycena klasickými regulačními diagramy, jejichž konstrukce stojí pouze na úrovni inherentní variability, a tím formální použití Shewhartových diagramů není možné, protože by vedlo k vysokému výskytu falešných poplachů, a je nutno se obrátit na regulační diagramy s rozšířenými mezemi.

Stabilitu procesu posuzujeme jednak vůči parametru polohy, a jednak vůči úrovni variability u sledovaného znaku jakosti. Nejdříve posoudíme úroveň variability. Zde je nutné mít data sbírána ve formě podskupin, každá podskupina musí obsahovat alespoň dvě pozorování. Pokud jsou data pozorována individuálně, tj. podskupina je jednoprvková, je nutno si řadu pozorování rozdělit uměle do podskupin. Nulová hypotéza zní, že v podskupinách je úroveň variability stejná, proti alternativní hypotéze, že v alespoň jedné podskupině je úroveň variability odlišná. Lze použít Bartlettův či Leveneův test, první předpokládá normalitu dat, druhý test se hodí na jakákoliv spojitá data. Pokud je nulová hypotéza zamítnuta, znamená to, že se v procesu projevuje nějaká vymezená příčina, která s úrovní variability v čase mění.

Pomocí analýzy rozptylu (ANOVA) lze za předpokladu nezamítnutí nulové hypotézy o úrovni variability rozhodnout o tom, zdali všechny logické podskupiny mají stejnou střední hodnotu, tj. nulová hypotéza, proti alternativě, že tomu tak není. Pokud obě nulové hypotézy nejsou zamítnuty a znak jakosti lze popsat normálním rozdělením, lze se obrátit k použití klasických regulačních diagramů popsaných např. v ČSN ISO normě 8258. Pokud některá nulová hypotéza je zamítnuta, je nutno hledat příčinu, která tuto nestabilitu způsobuje a pokusit se ji odstranit z procesu. Pokud příčinu odhalíme, porozumíme, jak proces ovlivňuje, ale je trvalou součástí procesu, nelze obvykle klasické regulační diagramy použít a je nutno je modifikovat.

Samozřejmě problémy nastávají, když data nelze popsat normálním rozdělením. Jak postupovat v takovém případě. První možnost je, že sledovaný znak jakosti se nedá popsat normálním rozdělením z nějakých často fyzikálních důvodů a tento rys se projevuje u tohoto znaku vždy. Zde je nutno hledat jiný typ rozdělení, který se hodí na popis chování dat. Často s jedná o logaritmicko-normální rozdělení, Weibullovo rozdělení, překlopené normální rozdělení, rozdělení maximálních či minimálních hodnot apod. Zde se bez vhodného softwaru neobejdeme. Další možnost je, že data jsou sice původně normálně rozdělena, ale pocházejí z různých zdrojů, jedná se vlastně o směs normálních rozdělení, která lze někdy od sebe oddělit (stratifikovat) podle nějakého příznaku (např. různé stroje, různí operátoři, různé dávky vstupního materiálu). Další možnost spočívá ve větším počtu dat nebo v přesnějším měření (na více desetinných míst). I tímto způsobem lze v datech normalitu objevit. Pokud data stále odolávají, přichází další možnost založená na vhodné transformaci dat na nová data, která již budou normálně rozdělena. Tento postup se dá uplatnit i při statistické regulaci, kdy se proces řídí a sleduje přes transformovaná data, na něž lze použít klasické regulační diagramy. Nejčastěji přichází v úvahu Box-Coxova transformace či třída Johnsonových transformací. Opět se neobejdeme bez vhodného statistického softwaru. Může se ale stát, že vhodná transformace nefunguje, pak de facto poslední možností jsou odhady požadovaných kvantilů získané numerickou cestou. Tento přístup má ale velké úskalí v tom, že pro spolehlivé odhady kvantilů potřebujeme poměrně velký počet dat (jedná se o stovky), což nemusí být snadné vždy opatřit.

Ukazatele způsobilosti

Základní myšlenkou definice ukazatelů způsobilosti a výkonnosti, je poměr mezi **tolerančním rozmezím**, tj. rozpětím mezi horní a dolní mezní hodnotou (USL – LSL) a **referenčním intervalem**, tj. intervalem pokrývajícím 99,73 % sledovaného znaku jakosti v procesu ($U_{99,865\%} - L_{0,135\%}$), vymežujícím přirozenou variabilitou jakostního znaku ve výrobním procesu. V případě normálně rozděleného znaku jakosti odpovídá tento interval šestinásobku směrodatné odchylky s znaku jakosti.

Horní mezní hodnotu značíme USL, dolní mezní hodnotu značíme LSL; $U_{99,865\%}$ je horní percentil, pod kterým leží 99,865% a $L_{0,135\%}$ je dolní percentil, pod kterým leží 0,135% všech hodnot sledovaného znaku jakosti. (Hovoří se o kvantilech, pokud se místo procent uvažují podíly.) Je-li znak jakosti rozdělen normálně se střední hodnotou μ a směrodatnou odchylkou σ , potom interval $U_{99,865\%} - L_{0,135\%} = 6\sigma$.

Jestliže se proces nachází ve stabilním stavu „v užším slova smyslu“, lze vyhodnocovat jeho způsobilost pomocí ukazatelů C_p a C_{pk} . Z dat vypočteme jejich odhady, správně bychom měli rozlišovat mezi použitými odhady směrodatné odchylky inherentní variability, a tyto odhady pomocí metod matematické statistiky konfrontovat s požadovanými hodnotami ukazatelů způsobilosti, jak si přeje zákazník či konstruktér. Pouze porovnání těchto bodových odhadů s požadovanými nic neřeší, je nutno předepsané hodnoty ukazatelů porovnávat s konfidenčními mezemi. Odhady ukazatelů výkonnosti P_p a P_{pk} by se v těchto případech neměly prakticky lišit od odhadů ukazatelů způsobilosti. Je nutno opět zdůraznit, že data použitá k odhadům ukazatelů musí vyhovovat normalitě, aby bylo možno použít vzorce:

$$C_p = (USL - LSL) / 6\sigma \quad C_{pk} = \min(USL - \mu, \mu - LSL) / 3\sigma$$

Běžně se proces považuje za způsobilý, pokud ukazatel C_p je alespoň 1.33, tj. pokud toleranční rozmezí odpovídá nejméně osmi směrodatným odchylkám $(USL - LSL) = 8\sigma$. Pokud je $C_p < 1.0$, považuje se proces za nezpůsobilý, v případě $C_p = 1.0$ za přibližně způsobilý.

Vzhledem k tomu, že ukazatel způsobilosti C_p nezohledňuje nastavení procesu (parametr polohy μ), budou jeho hodnoty stejné i v případě procesů, které nejsou centrovány.

Když je proces zvládnut v „širším slova smyslu“, pak ukazatele C_p a C_{pk} je nutno nahradit ukazateli P_p a P_{pk} , neboť do hry vstupuje místo inherentní variability tzv. celková či totální variabilita, která v sobě obsahuje jak variabilitu inherentní, tak i variabilitu mezi podskupinami vyvolanou nestabilitou v parametru polohy. Opět je nutno se přesvědčit, zdali lze data považovat celkově za normálně rozdělená, aby byly použity správné vzorce pro odhady ukazatelů výkonnosti.

Jestliže data nelze vysvětlit normálním rozdělením a i když ve proces v ustáleném stavu „v užším slova smyslu“, ukazatelé způsobilosti C_p a C_{pk} ztrácejí smysl, protože v odpovídajících vzorcích pro nenormální data nevystupuje směrodatná odchylka inherentní variability, ale kvantilové rozpětí a medián. Pak je nutno buď hledat jiný model rozdělení pravděpodobnosti, či data transformovat na data normálně rozdělená či numerickými metodami získat odhady požadovaných kvantilů pro odhad délky statistického pokrývného intervalu, která vystupuje místo 6σ u normálně rozdělených dat. Zde se opět neobejdeme bez vhodného statistického softwaru.

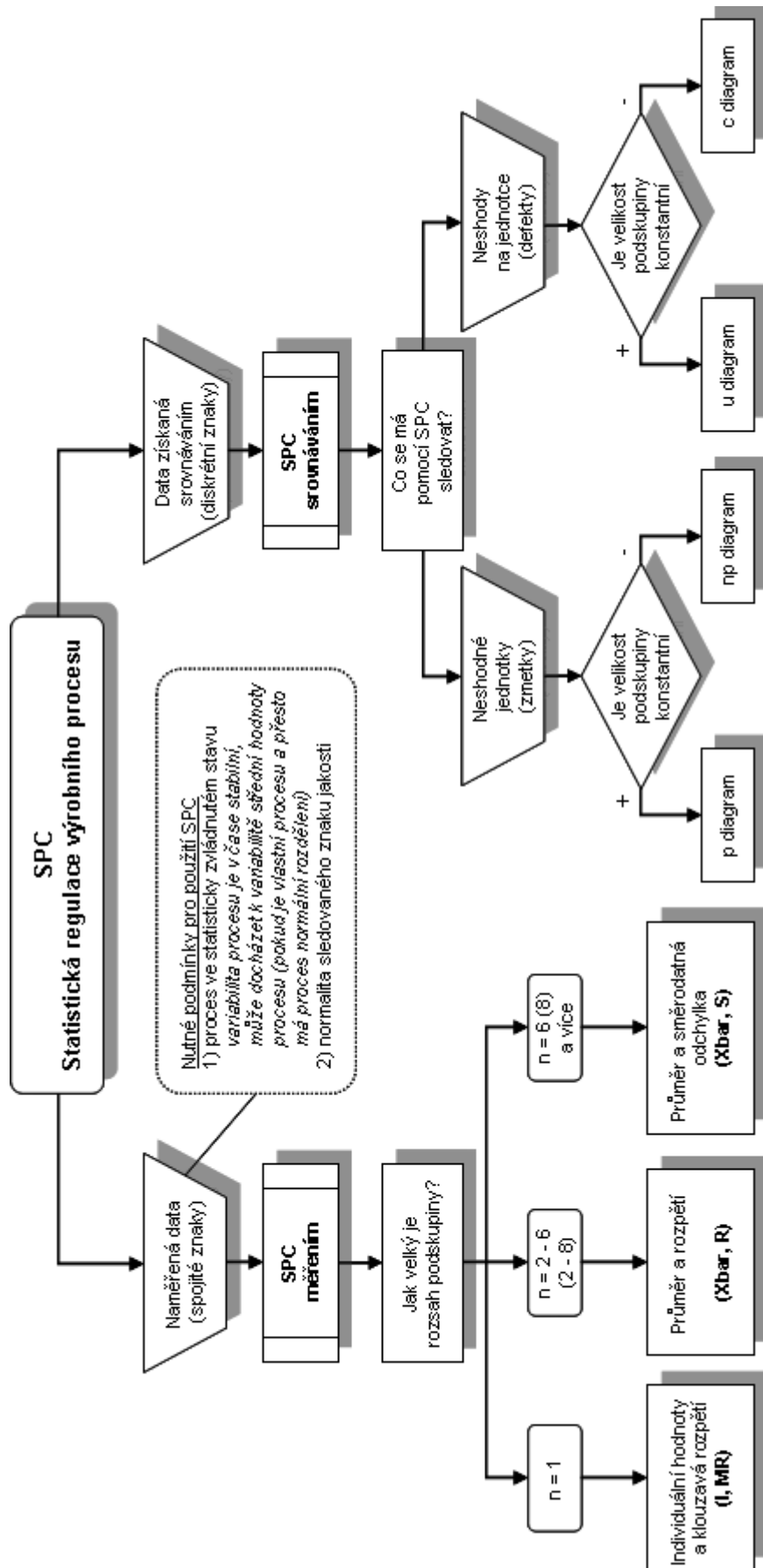


Schéma 4/1:

Proces je statisticky zvládnutý

Pokud v procesu dochází k variabilně střední hodnoty, která je neodstranitelná díky charakteru procesu, lze počítat pouze výkonnost

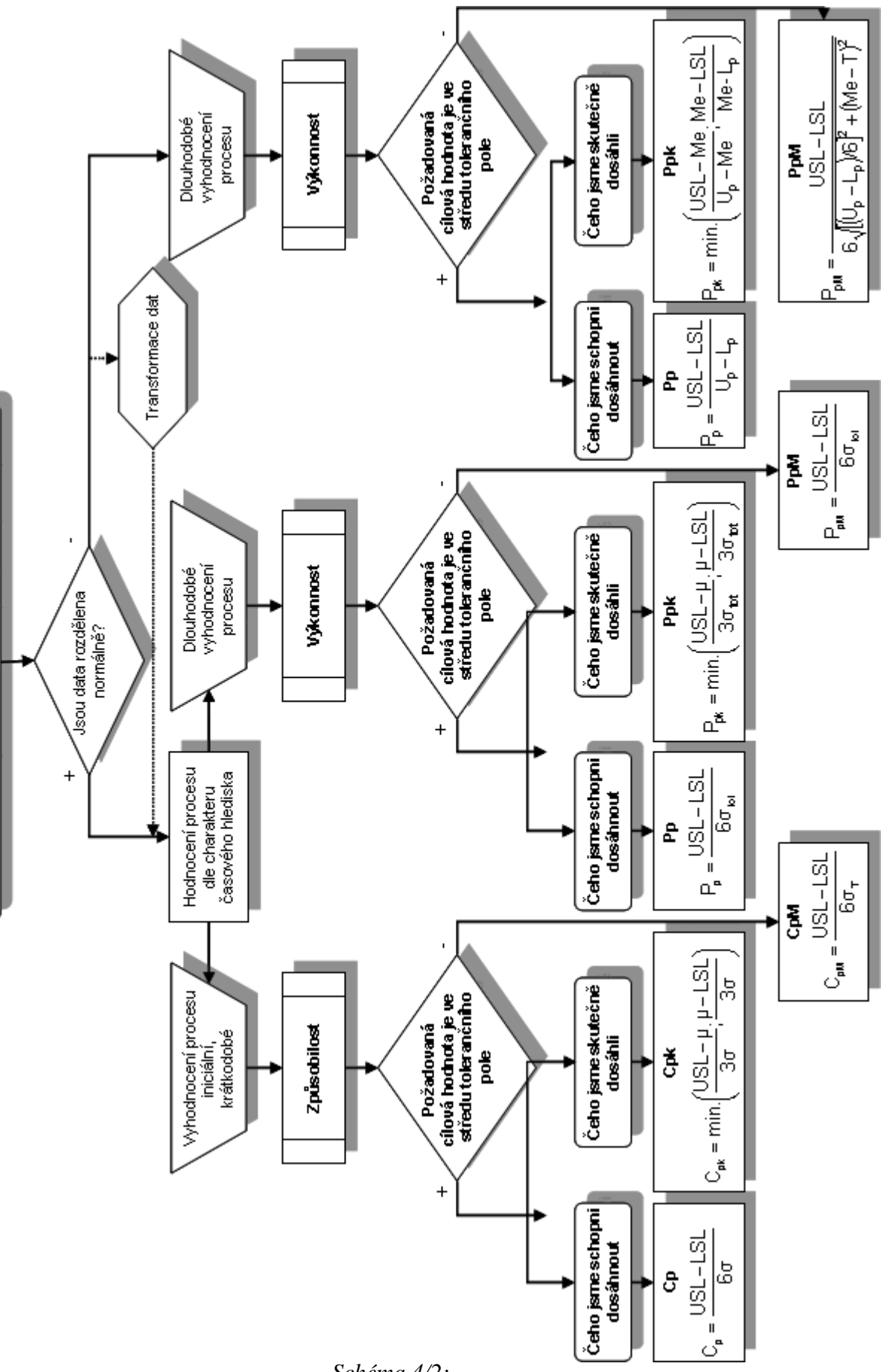


Schéma 4/2:

Ukazatele výkonnosti

Výkonnost procesu je definována na základě celkové (totální) variability procesu, charakterizované směrodatnou odchylkou σ_{TOT} , tj. variability vyvolané náhodnými příčinami a případnými neodstranitelnými zvláštními příčinami za předpokladu, že proces je statisticky zvládnut v „širším slova smyslu“, takže se v čase mění známým a odůvodněným, ale neodstranitelným způsobem střední hodnota procesu. Jedná se např. o trend vyvolaný opotřebením nástroje, o nemožnost udržet proces přesně centrováný vlivem vstupního materiálu apod.

Postupu při volbě vhodného regulačního diagramu

Schéma 4 je věnováno postupu při volbě vhodného regulačního diagramu podle typu znaku jakosti, zdali se jedná o spojitá či atributivní data. Dále rozhodujícím faktorem je i velikost podskupiny, která má být u spojitých dat konstantní, u atributivních dat se může velikost podskupiny měnit. Nevýhodou klasických regulačních diagramů je jejich poměrně dlouhá doba odezvy na změnu v chování znaku jakosti. Pro zkrácení průměrné doby odezvy na změnu neboli zmenšení chyby 2. druhu, pokud se na regulační diagram díváme jako na sekvenční test, je vhodné zvláště u procesů citlivých na změny použít modernější typy regulačních diagramů jako jsou diagramy EWMA či CUSUM.

Příklad vyhodnocení procesních dat

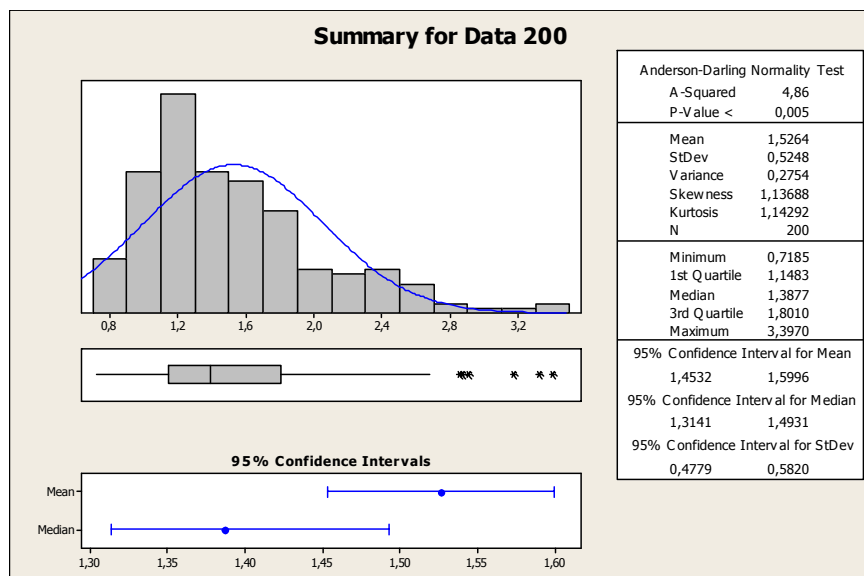
Příspěvek je zakončen příkladem zaměřeným na data, která nejsou normálně rozdělena a je nutno vyhodnotit výkonnost sledovaného procesu. Zpracování dat je provedeno pomocí softwaru Minitab 15.

Uvažujeme případ, kdy znak jakosti v procesu není rozdělen normálně. Ze znalosti procesu a z dřívějších měření je patrné, že data jsou rozdělena asymetricky. Dolní mezní hodnota byla stanovena $LSL = 0,5$ a horní mezní hodnota $USL = 5$. Požaduje se ověřit typ rozdělení znaku jakosti a odhadnout jeho parametry; vyhodnotit výkonnost procesu a navrhnout regulační diagram pro individuální hodnoty.

Z procesu bylo odebráno během několika pracovních dnů v přibližně stejných intervalech celkem 200 jednotek (počet podskupin $k = 200$ rozsahu $n = 1$). Byl sledován znak jakosti, o kterém se předpokládá, že v čase nedochází ke změně střední hodnoty ani variability (proces je statisticky zvládnut).

Úkolem je ověřit, typ rozdělení znaku jakosti, odhadnout jeho parametry a ověřit předpoklad, že proces je statisticky zvládnut. Dále vyhodnotit výkonnost procesu a vypočítat na základě napozorovaných dat parametry regulačního diagramu pro individuální hodnoty x_i a klouzavá rozpětí dvou sousedních hodnot MR_2 .

- 1) První informace o napozorovaných datech – základní výběrové charakteristiky, histogram s proloženou křivkou hustoty hypotetického normálního rozdělení pravděpodobnosti a 95%-ní konfidenční intervaly pro střední hodnotu μ a směrodatnou odchylku σ - je získána pomocí funkce „Graphical Summary“ (Stat > Basic Statistic > Graphical Summary).

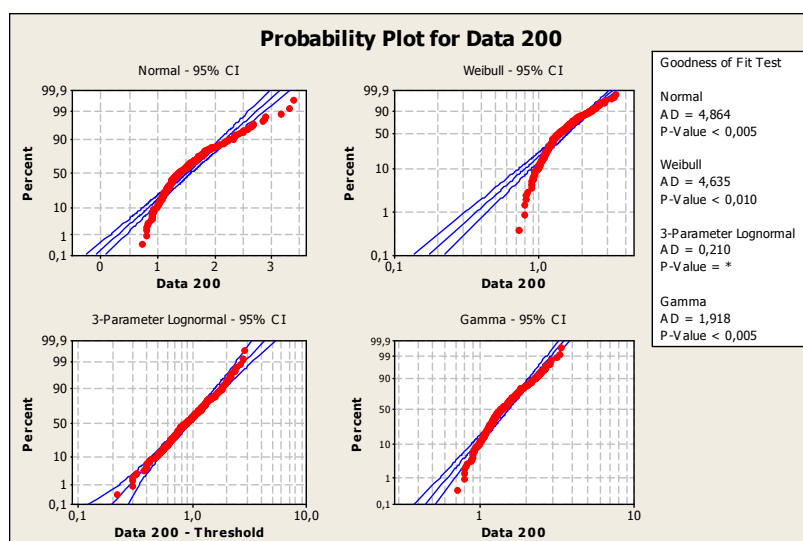


Z histogramu s proloženou hustotou pravděpodobnosti normálního rozdělení je patrné, že studovaný znak jakosti není normálně rozdělen. To potvrzuje i p-hodnota Anderson-Darlingova testu normality ($p\text{-Value} < 0,005$) která je hluboko pod obvykle používanou hladinou významnosti $\alpha = 0,05$. Asymetričnost rozdělení je patrná i z box-plot diagramu.

2) Minitab umožňuje

- a) identifikovat vhodný typ rozdělení pravděpodobnosti;
- b) transformovat původní nenormálně rozdělená data pomocí Box-Coxovy transformace;
- c) transformovat původní nenormálně rozdělená data pomocí Johnsonovy transformace.

V Minitabu (Stat > Quality Tools > Individual Distribution Identification) zvolíme typy rozdělení, která chceme pro napozorovaná data ve sloupci C1 – „Data 200“ odzkoušet (Např. normální, Weibullovo, logaritmickeo-normální, gama).



Z pravděpodobnostních grafů je patrné, že nejlépe napozorovaná data vystihuje model 3-parametrického log-normálního rozdělení. V okně Session jsou zobrazeny v tabulce číselné výsledky – parametry modelu

Distribution ID Plot for Data 200

Descriptive Statistics

N	N*	Mean	StDev	Median	Minimum	Maximum	Skewness	Kurtosis
200	0	1,52643	0,524802	1,38771	0,718456	3,39703	1,13688	1,14292

Goodness of Fit Test

Distribution	AD	P	LRT	P
Normal	4,864	<0,005		
Weibull	4,635	<0,010		
3-Parameter Lognormal	0,210	*	0,004	
Gamma	1,918	<0,005		

ML Estimates of Distribution Parameters

Distribution	Location	Shape	Scale	Threshold
Normal*	1,52643		0,52480	
Weibull		3,00762	1,70765	
3-Parameter Lognormal	-0,09951		0,50043	0,50224
Gamma		9,61373	0,15878	

* Scale: Adjusted ML estimate

a tabulka zadaných percentilů:

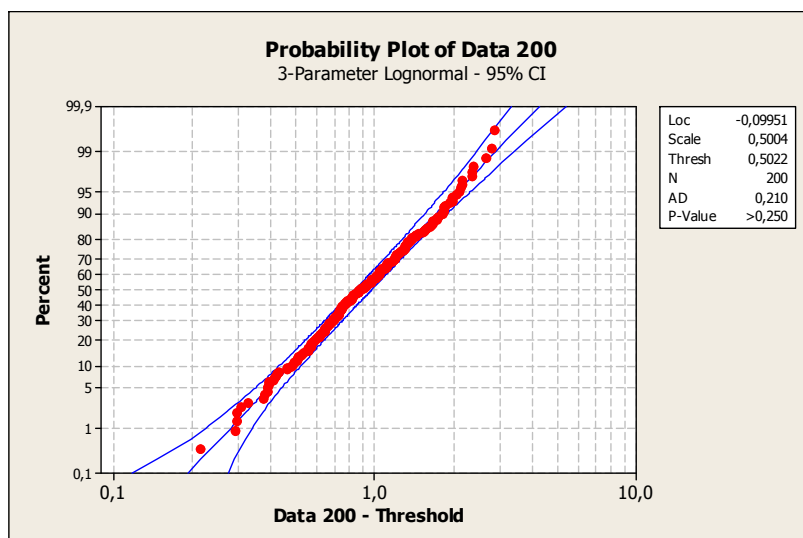
Table of Percentiles

Distribution	Percent	Percentiles	Standard		
			Error	95,0% CI	
Normal	0,135	-0,047966	0,0871617	-0,2	0,1
Weibull	0,135	0,189830	0,0231104	0,1	0,2
3-Parameter Lognormal	0,135	0,703978	0,0385791	0,6	0,8
Gamma	0,135	0,457287	0,0352708	0,4	0,5

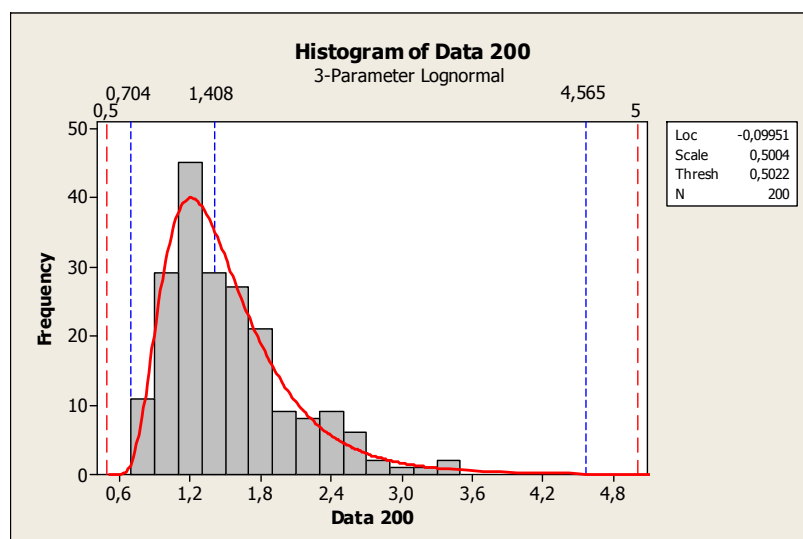
Distribution	Percent	Percentiles	Standard		
			Error	95,0% CI	
Normal	50	1,52643	0,0371091	1,5	1,6
Weibull	50	1,51174	0,0418490	1,4	1,6
3-Parameter Lognormal	50	1,40752	0,0344290	1,3	1,5
Gamma	50	1,47384	0,0340017	1,4	1,5

Distribution	Percent	Percentiles	Standard		
			Error	95,0% CI	
Normal	99,865	3,10082	0,087162	2,9	3,3
Weibull	99,865	3,19931	0,105816	3,0	3,4
3-Parameter Lognormal	99,865	4,56459	0,475397	3,7	5,6
Gamma	99,865	3,43026	0,138706	3,2	3,7

Z pravděpodobnostního grafu plyne, že je dobrá shoda empirických dat s identifikovaným modelem 3-parametrického log-normálního rozdělení s parametry „Loc“ = -0,09951; „Scale“ = 0,5004; „Thresh“ = 0,5022. Naměřené hodnoty leží v 95%-ním konfidenčním intervalu a p-hodnota Anderson-Darlingova testu je větší než 0,250.



Do histogramu z napozorovaných hodnot můžeme zakreslit hustotu pravděpodobnosti identifikovaného modelu, červeně mezní hodnoty LSL = 0,5; USL = 5,0 a modře vypočítané percentily pro 0,135 %; 50 %; 99,865 %.



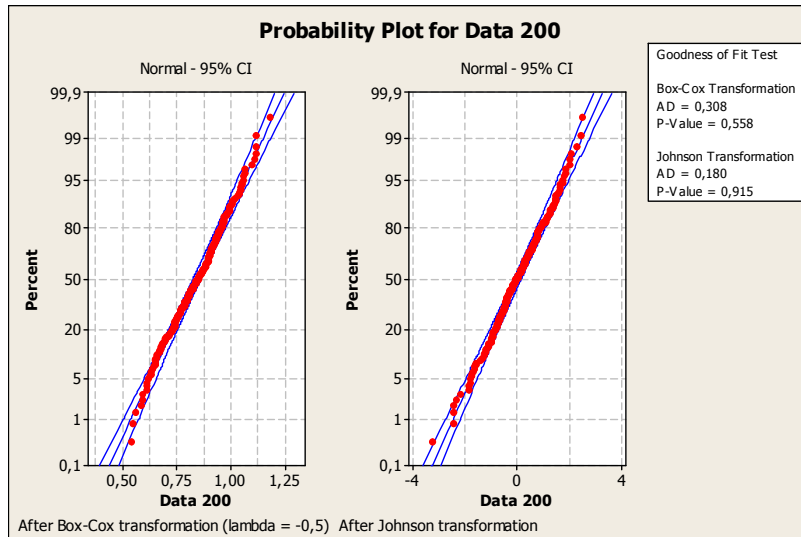
V případech, kdy se nepodaří identifikovat rozdělení studovaného znaku jakosti nebo si navrženým modelem nejsme jisti (nízká p-hodnota), můžeme použít obě výše zmíněné transformace Box-Coxovu a Johnsonovu. V tomto případě necháme transformované hodnoty zapsat např. do sloupců C2 a C3.

V Minitabu (Stat > Quality Tools > Individual Distribution Identification) zvolíme „Box-Cox transformation“ a „Johnson transformation pro napozorovaná data ve sloupci C1 – „Data 200“.

Výstupem jsou pravděpodobnostní grafy transformovaných dat, které ukazují na velmi dobrou shodu transformovaných dat s modelem normálního rozdělení, p-hodnoty jsou relativně vysoké 0,558 a 0,915.

Nejlepší výsledek se jeví při Johnsonově transformaci, kde p-hodnota je velmi vysoká 0,915.

Zdá se vhodné využít takto transformovaných dat pro další výpočty.



V okně Session jsou zobrazeny v tabulce číselné výsledky obou transformací a tabulka zadaných percentilů:

Descriptive Statistics

N	N*	Mean	StDev	Median	Minimum	Maximum	Skewness	Kurtosis
200	0	1,52643	0,524802	1,38771	0,718456	3,39703	1,13688	1,14292

Box-Cox transformation: Lambda = -0,5

Johnson transformation function:
 $4,16567 + 1,70066 * \ln((X - 0,588679) / (10,9047 - X))$

Goodness of Fit Test

Distribution	AD	P
Box-Cox Transformation	0,308	0,558
Johnson Transformation	0,180	0,915

ML Estimates of Distribution Parameters

Distribution	Location	Shape	Scale	Threshold
Box-Cox Transformation*	0,84154		0,13109	
Johnson Transformation*	-0,00619		1,05419	

* Scale: Adjusted ML estimate

Table of Percentiles

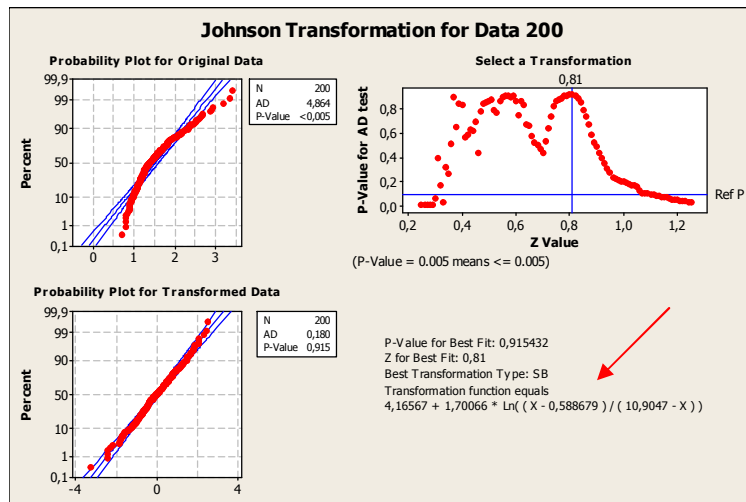
Distribution	Percent	Percentiles	Standard Error	95,0% CI
Box-Cox Transformation	0,135	0,44828	0,021772	0,4 0,5
Johnson Transformation	0,135	-3,16874	0,175086	-3,5 -2,8

Distribution	Percent	Percentiles	Standard Error	95,0% CI
Box-Cox Transformation	50	0,841540	0,0092693	0,8 0,9
Johnson Transformation	50	-0,006188	0,0745427	-0,2 0,1

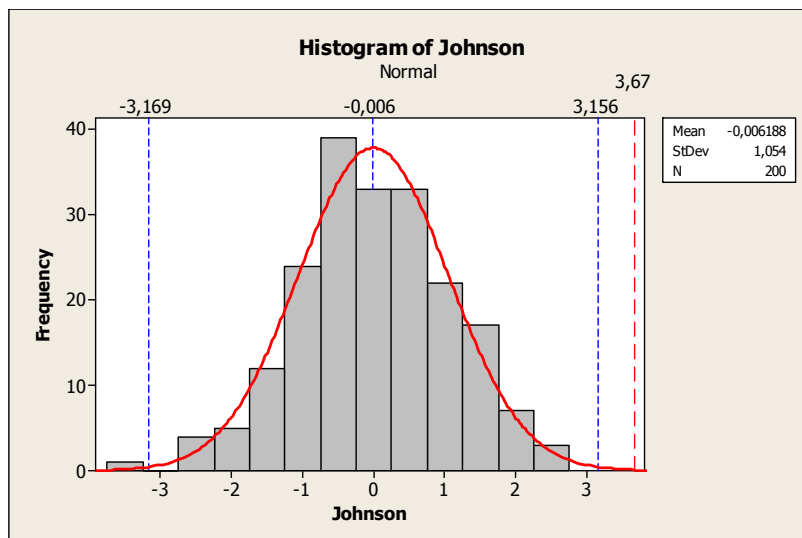
Distribution	Percent	Percentiles	Standard Error	95,0% CI
Box-Cox Transformation	99,865	1,23480	0,021772	1,2 1,3
Johnson Transformation	99,865	3,15637	0,175086	2,8 3,5

Napozorovaná data můžeme v programu Minitab transformovat následujícím postupem: (Stat > Quality Tools > Johnson Transformation) > ...).

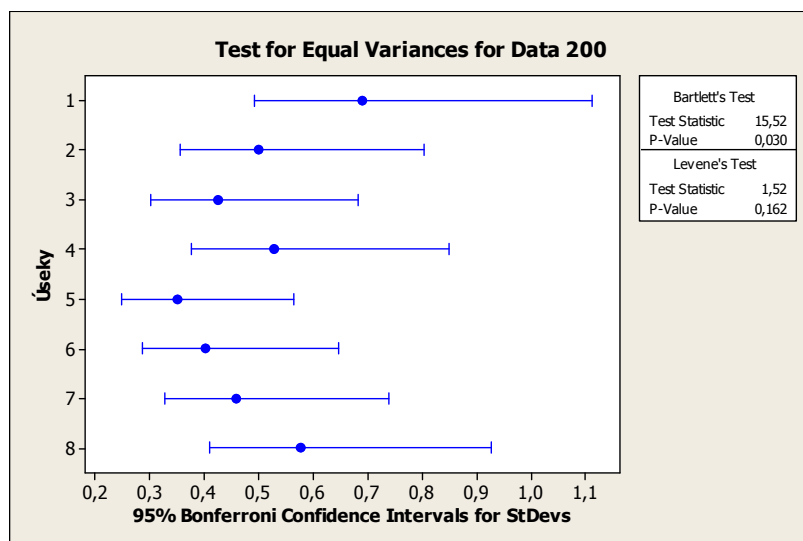
Výstupem je tabulka obsahující pravděpodobnostní grafy původních a transformovaných dat a transformační rovnice:



Histogram transformovaných dat můžeme porovnat s odpovídajícím normálním rozdělením. V grafu jsou čárkovane modře zakresleny zvolené percentily odpovídající 0135 %; 50 %; 99,865 %. Červeně čárkovane je zakreslena transformovaná horní mezní hodnota. Dolní mezní hodnota je již mimo rámeček transformace (ve výrazu se vyskytne záporný argument v přirozeném logaritmu). Transformace mezních hodnot se provede pomocí kalkulátoru (Calc > Calculator).



3) Abychom mohli ověřit, že v čase nedochází ke změně variability (rozptylu), rozdělíme pozorování do několika (v našem případě do osmi) za sebou jdoucích úseků. Úseky jsou uvedeny ve sloupci C4. Na základě Leveneova testu rovnosti rozptylů, vhodného v případě spojitého rozdělení (Stat > ANOVA > Test for Equal Variances), není důvod pochybovat, že podskupiny pocházejí ze základních souborů se stejnými rozptyly; p-hodnota (0,162) tohoto testu je nad obvykle volenou hladinou významnosti $\alpha = 0,05$. Je možno akceptovat předpoklad, že v čase se nemění variabilita sledovaného znaku jakosti.

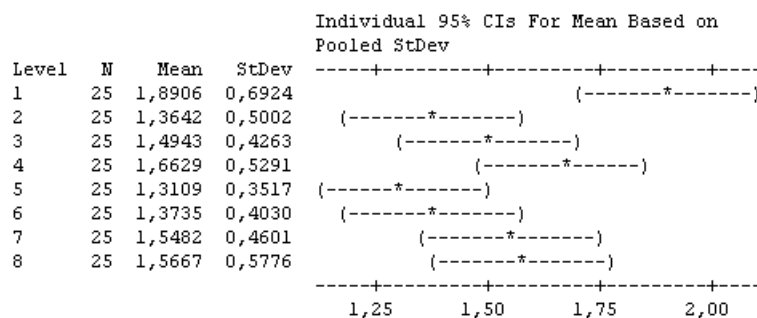


4) Na základě výsledku analýzy rozptylu (Stat > ANOVA > One Way), kde p-hodnota (0,001) je menší než obvykle volená hladina významnosti, máme důvod zamítnout předpoklad, že data v podskupinách pocházejí ze základních souborů se stejnými středními hodnotami. Tento výsledek je však třeba brát pouze orientačně, protože ANOVA předpokládá normální rozdělení analyzovaných dat. Jako vhodnější se nabízí použití neparametrického testu o rovnosti mediánů, např. test [Kruskal-Wallisův](#) (viz Poznámka 2 na konci tohoto příkladu).

One-way ANOVA: Data 200 versus Úseky

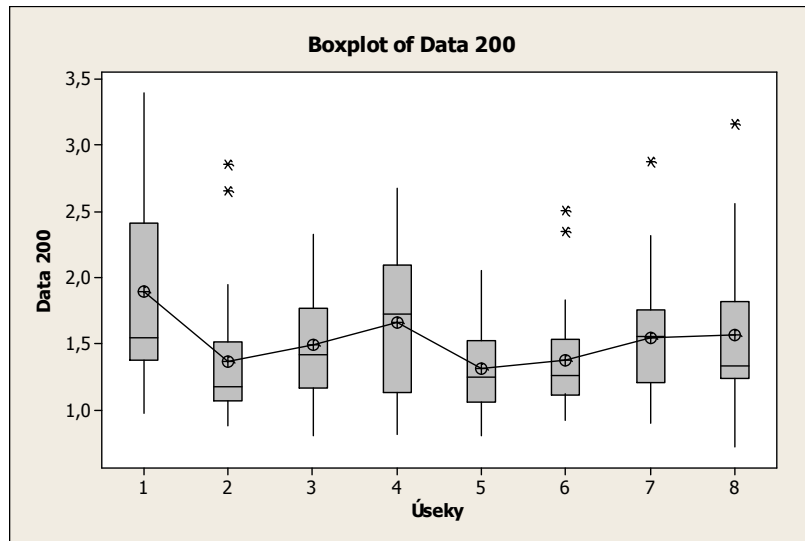
Source	DF	SS	MS	F	P
Úseky	7	6,263	0,895	3,54	0,001
Error	192	48,545	0,253		
Total	199	54,808			

S = 0,5028 R-Sq = 11,43% R-Sq(adj) = 8,20%



Pooled StDev = 0,5028

Tento výsledek dokumentuje i příslušný diagram Boxplot.

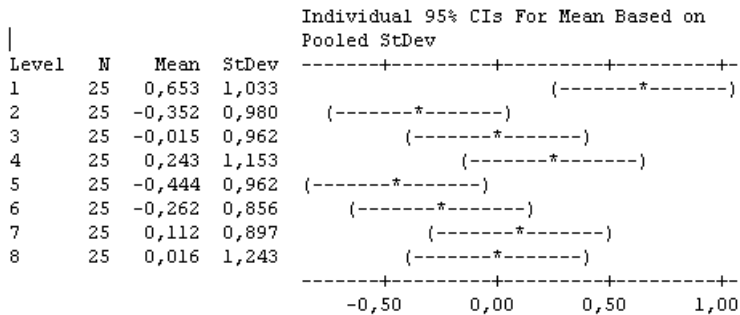


Srovnatelný výsledek dostaneme i v případě analýzy transformovaných dat:

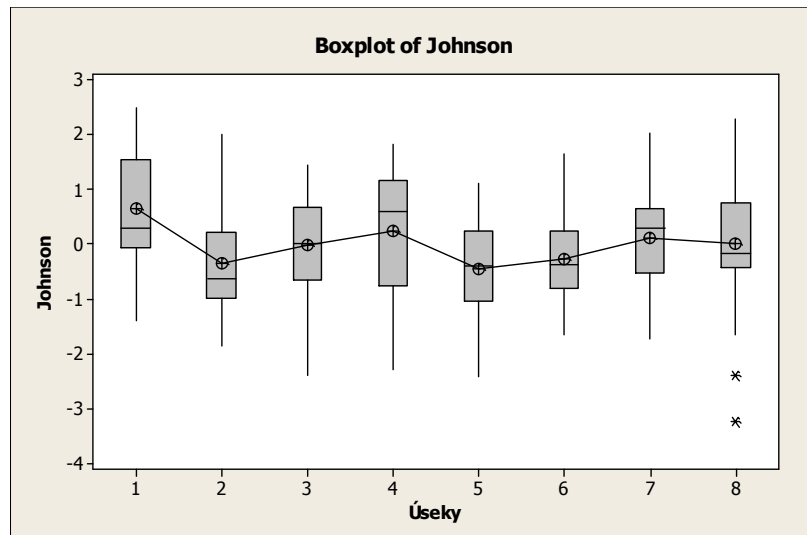
One-way ANOVA: Johnson versus Úseky

Source	DF	SS	MS	F	P
Úseky	7	22,20	3,17	3,06	0,004
Error	192	198,96	1,04		
Total	199	221,15			

S = 1,018 R-Sq = 10,04% R-Sq(adj) = 6,76%



Pooled StDev = 1,018



- 5) Výkonnost procesu ve vztahu k mezním hodnotám $USL = 5,0$ a $LSL = 0,5$ je možno vyhodnotit jedním z následujících postupů.
- a) Vycházíme z identifikovaného rozdělení pravděpodobnosti, v našem případě se jedná o 3-parametrické log-normální rozdělení, pro které jsme dostatečně přesně stanovili percentily (kvantily) pro 0,135 %; 50 %; 99,865 %, které označíme $L_p = 0,704$; $Me = 1,408$; $U_p = 4,565$. Potom odhady ukazatelů výkonnosti počítáme ze vztahů

$$P_p = \frac{USL - LSL}{U_p - L_p} ;$$

$$P_{pk} = \min \{ P_{pU}, P_{pL} \} = \min \left\{ \frac{USL - Me}{U_p - Me}, \frac{Me - LSL}{Me - L_p} \right\}.$$

V našem příkladě se jedná o odhady:

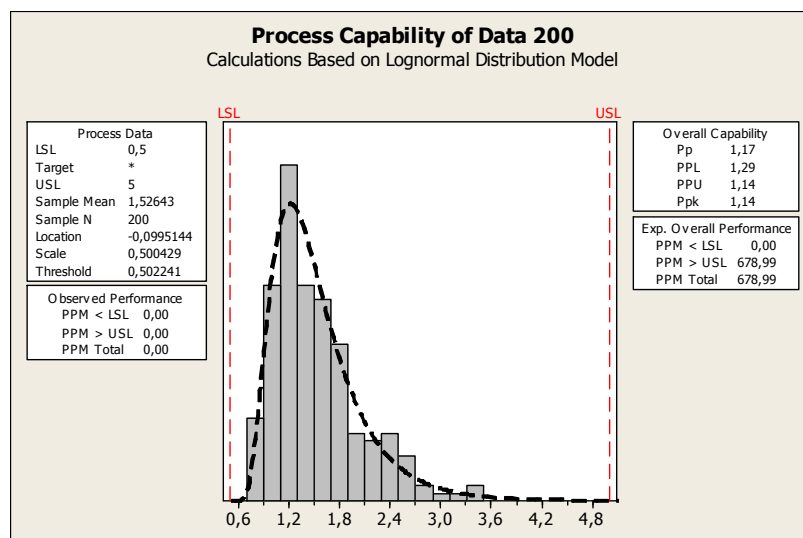
$$\hat{P}_p = \frac{USL - LSL}{U_p - L_p} = \frac{5,0 - 0,5}{4,565 - 0,704} = \frac{4,5}{3,861} = 1,166,$$

$$\hat{P}_{pU} = \frac{USL - Me}{U_p - Me} = \frac{5,0 - 1,408}{4,565 - 1,408} = \frac{3,592}{3,157} = 1,138,$$

$$\hat{P}_{pL} = \frac{Me - LSL}{Me - L_p} = \frac{1,408 - 0,5}{1,408 - 0,704} = \frac{0,908}{0,704} = 1,290,$$

a tedy $\hat{P}_{pk} = 1,138$.

- b) Minitab počítá ukazatele výkonnosti v případě nenormálně rozdělených dat postupem (Stat > Quality Tools > Capability Analysis > Nonnormal). Je-li identifikován typ rozdělení pravděpodobnosti (tj. 3-parametrické log-normální), Minitab po jeho zadání vyhodnotí příslušné ukazatele výkonnosti a očekávaný počet ppm mimo aktuální mezní hodnoty. V tomto příkladě vychází odhady $P_p = 1,17$; PPL (P_{pL}) = 1,29; PPU (P_{pU}) = 1,14 a $P_{pk} = 1,14$ (stejně jako výše) a očekávaný počet mimo USL je 679 ppm; mimo LSL je 0 ppm; celkem 679 ppm.

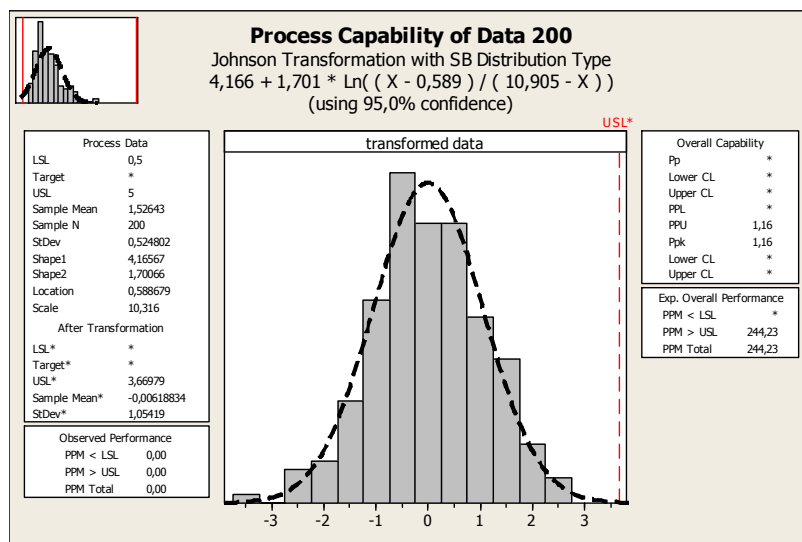


c) Minitab počítá ukazatele výkonnosti v případě nenormálně rozdělených dat rovněž pomocí Johnsonovy transformace postupem (Stat > Quality Tools > Capability Analysis > Nonnormal). Není-li identifikován typ rozdělení pravděpodobnosti (což je častý případ), Minitab nabízí volit metodu na základě Johnsonovy transformace. V tomto případě je umožněno počítat i konfidenční intervaly aktuálních ukazatelů výkonnosti. Ty není možno počítat v předešlém případě při zadání konkrétního modelu.

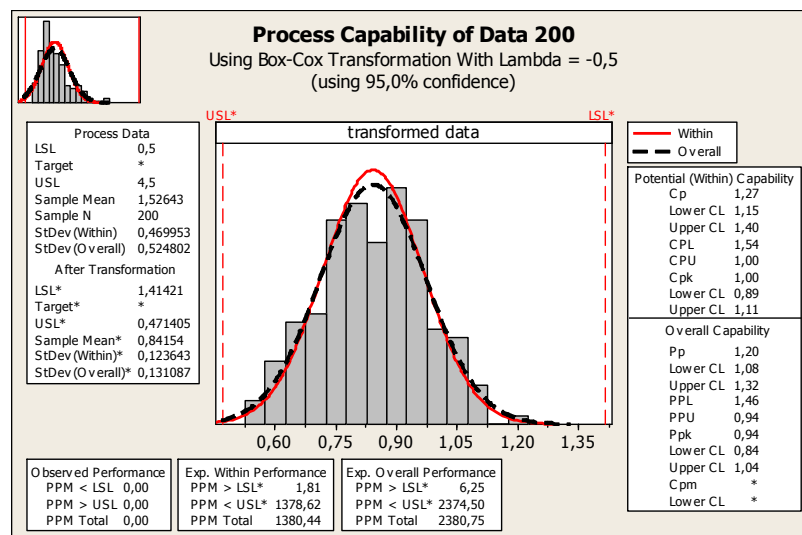
Tento postup někdy může selhat, jako v tomto případě, kdy dolní mezní hodnota LSL = 0,5 je mimo definiční obor transformace. Potom program nevyhodnotí ukazatele P_p a PPL (P_{pL}).

Transformovaná data, která jsou uložena ve sloupci C3 – „Johnson“, je možno použít k různým výpočtům, například k výpočtu kvantilů, regulačních mezí pro individuální hodnoty a pod. Do původních hodnot se zpět můžeme dostat pomocí „zpětné transformace“ (viz soubor „Zpětná transformace.xls“), kam je třeba vložit koeficienty transformační rovnice

$$4,16567 + 1,70066 * \text{Ln}((x - 0,588679) / (10,9047 - x)) .$$

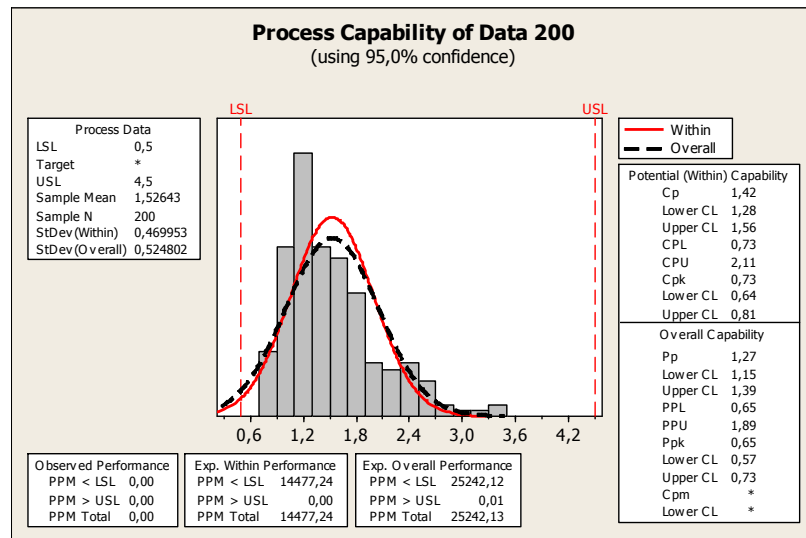


d) Minitab umožňuje počítat ukazatele výkonnosti rovněž na základě Box-Coxovy transformace (Stat > Quality Tools > Capability Analysis > Normal). To přichází v úvahu např. není-li identifikován typ rozdělení pravděpodobnosti a nepodařilo-li se najít Johnsonovu transformaci. V dialogovém okně použijeme tlačítko „Box-Cox“ a zvolíme „Use optimal lambda“.



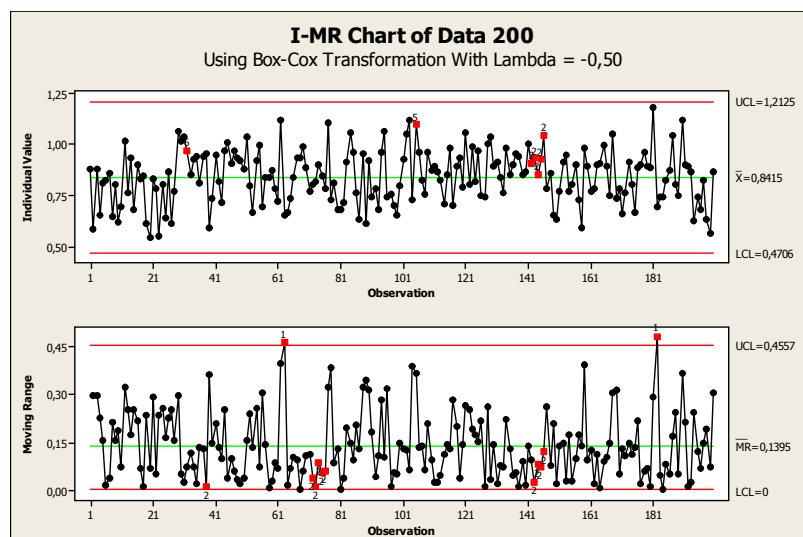
e) Počítat ukazatele výkonnosti pro znak jakosti, který není normálně rozdělen jako kdyby normálně rozdělen byl, je v praxi poměrně časté a vede to k nemalým chybám. Postup (Stat > Quality Tools > Capability Analysis > Normal) bez volby Box-Coxovy transformace vede k následujícímu výsledku:

hodnoty odhadů ukazatelů výkonnosti P a zejména ukazatelů C se v tomto případě výrazně liší od reality odpovídajícím hodnotám.



6) Vzhledem k tomu, že sledovaný znak jakosti není normálně rozdělen, nabízí software možnost navrhnout regulační diagram s využitím Box-Coxovy transformace.

Stat > Control Charts > Variables Charts for Individuals > I-MR Chart a v nabídce I-MR Options > Box-Cox zvolit Optimal lamda.



Parametry tohoto regulačního diagramu pro individuální hodnoty jsou

$$CL = 0,8415; UCL = 1,2125; LCL = 0,4706$$

a pro výběrová rozpětí jsou

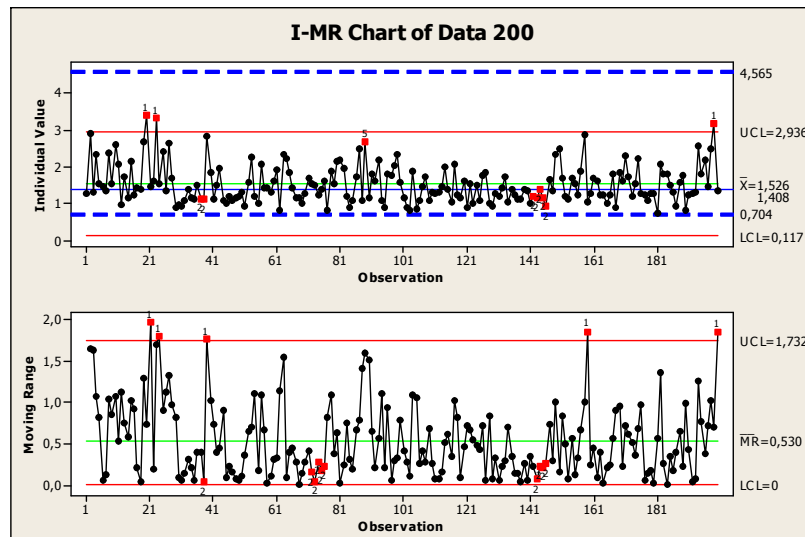
$$CL = 0,1395; UCL = 0,4557; LCL = 0.$$

Tento přístup má ovšem ten nedostatek, že do regulačního diagramu je třeba zakreslovat transformované a nikoliv přímo naměřené hodnoty.

Další možnost je využít již dříve vypočítané kvantily identifikovaného rozdělení (3-parametrické log-normální rozdělení“)

$$K_{0,135} = L_p = 0,704, K_{50} = Me = 1,408; \text{ a } K_{99,865} = U_p = 4,565$$

a použit je jako regulační meze – „zásahové“ (zakresleny čárkovaně modře do diagramu netransformovaných hodnot (meze zakreslené plně červeně jsou běžné Shewhartovy regulační meze vypočítané programem za předpokladu normálního rozdělení znaku jakosti).



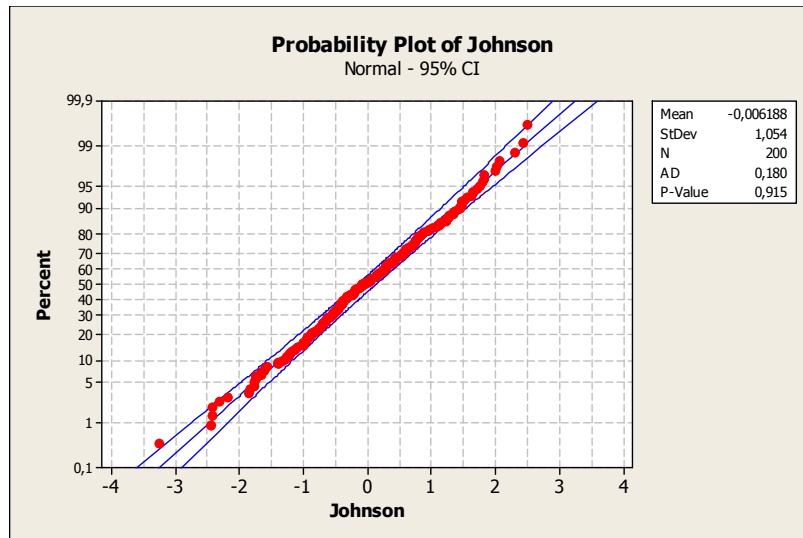
V tomto případě červeně zvýrazněné výběrové body obvykle signalizující přítomnost nenáhodné příčiny variability ztrácejí smysl, neboť jejich poloha je konfrontována se Shewhartovými mezemi.

7) Uvedeme postup, který se v tomto případě jeví jako optimální, i když pracnější, respektující i tu skutečnost, že proces není striktně statisticky zvládnut a zahrnuje i variabilitu ve střední hodnotě (viz výsledek ANOVA).

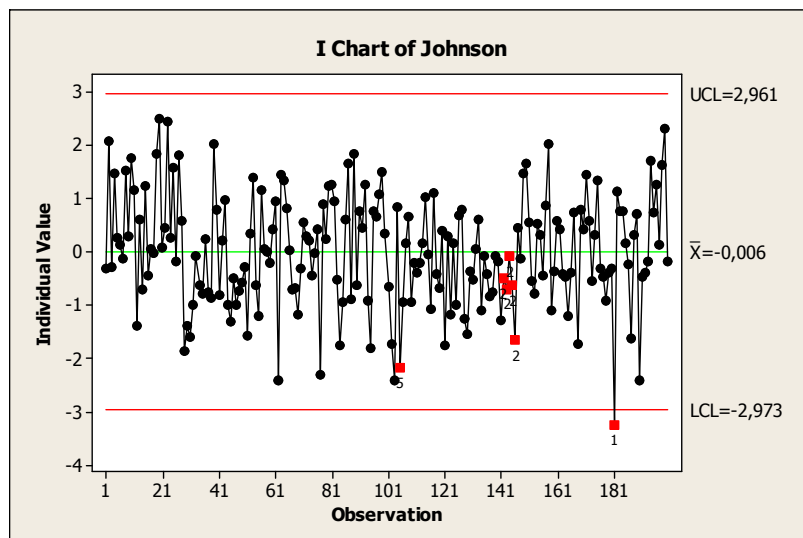
Ve sloupci C3 Worksheetu „Data“ jsou původní data po Johnsonově transformaci, která jsou normálně rozdělena. Softwarem vybraná transformační rovnice je:

$$4,16567 + 1,70066 * \text{Ln}((x - 0,588679) / (10,9047 - x)).$$

Pravděpodobnostní graf a Anderson-Darlingův test dobré shody (p-hodnota = 0,915) potvrzuje, že nemáme důvod pochybovat o normálním rozdělení dat ve sloupci C3:

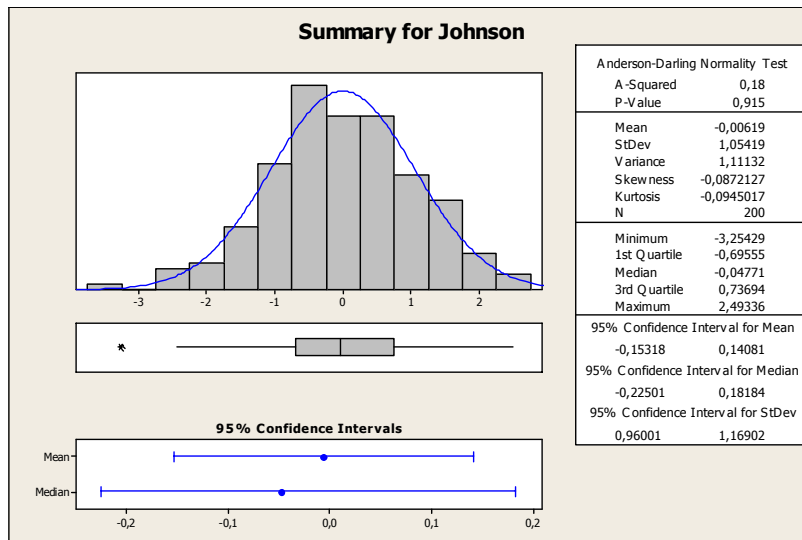


Můžeme tedy pro tato data vytvořit běžný Shewartův graf pro individuální hodnoty.



Vzhledem k tomu, že proces není statisticky zvládnut v užším slova smyslu, je třeba použít „rozšířených“ regulačních mezí, v tomto případě s použitím celkové směrodatné odchylky, na místo směrodatné odchylky odvozené z krátkodobé variability, tj. klouzavého rozpětí dvou sousedních hodnot.

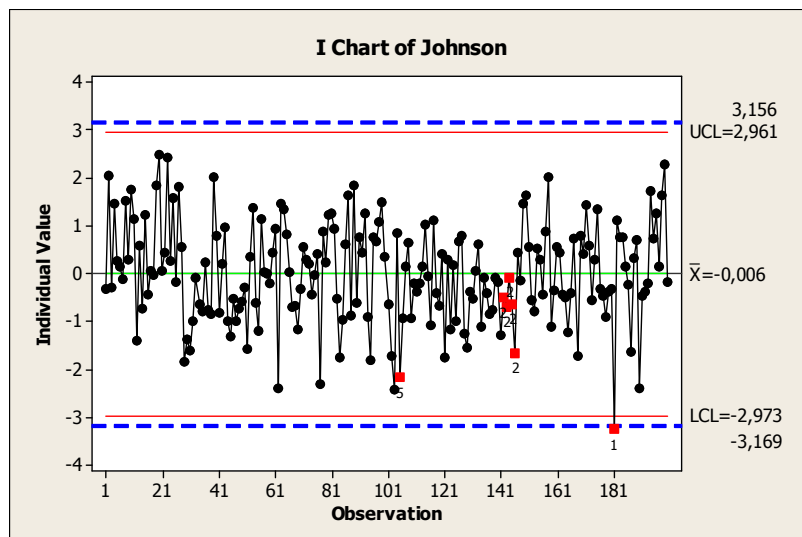
Informace o transformovaných datech – základní výběrové charakteristiky, histogram s proloženou křivkou hustoty hypotetického normálního rozdělení pravděpodobnosti a 95% konfidenční intervaly pro střední hodnotu μ a směrodatnou odchylku σ - získáme běžně pomocí funkce „Graphical Summary“ (Stat > Basic Statistic > Graphical Summary).



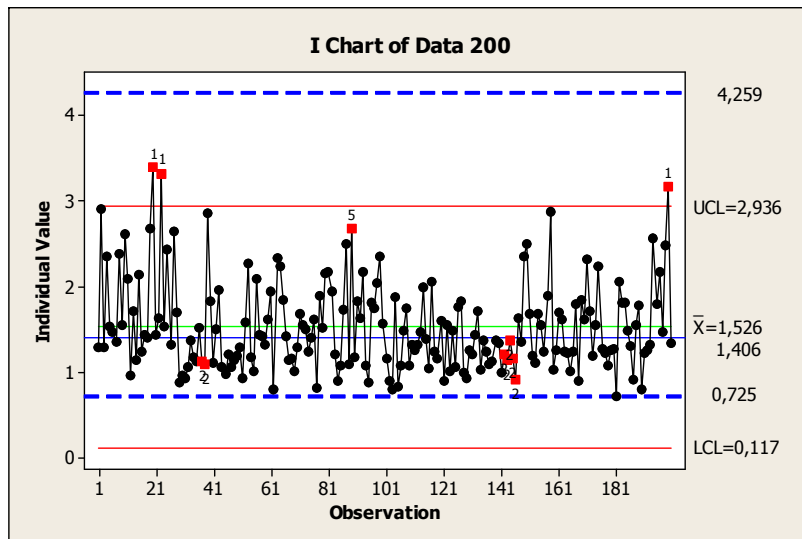
Vidíme, že celkový (totálny) průměr $\bar{x}_{TOT} = -0,00619$ a $s_{TOT} = 1,05419$. Potom regulační meze v diagramu individuálních hodnot lze rozšířit:

$$LCL = -3,1688 ; CL = -0,00619 ; UCL = 3,1564 .$$

Tyto meze zakreslíme čárkovaně modře do regulačního diagramu transformovaných hodnot (uvedeného výše).

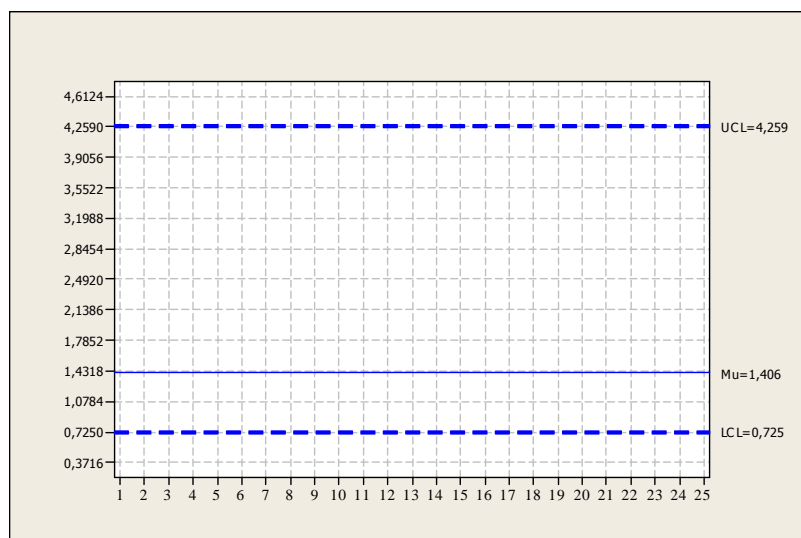


Aby se v praxi každá naměřená hodnota nemusela transformovat, provedeme „zpětnou transformaci“ vypočtených rozšířených regulačních mezí. Oba postupy dávají téměř stejné regulační meze.



Pro zakreslení prázdného formuláře výsledného regulačního diagramu pro další období můžeme použít makra „**BLANKCH.MAC**“ :

Zvolíme Editor > Enable Commands a v okně „Session“ za příkaz MTB> zapíšeme **%BLANKCH** a potvrdíme ENTER. Výstupem v okně „Session“ jsou instrukce a dotazy (zadání). Po jejich vyplnění je výstupem požadovaný regulační diagram se zakreslenými již zadanými regulačními mezemi.



Poznámka 1

Instalace makra „BLANKCH.MAC“:

V adresáři „BLANKCH“ je zapsáno makro BLANKCH.MAC ve formátu „Poznámkový blok“. Tento soubor je třeba uložit pro verzi Minitab 15 na adresu:

Program Files > Minitab 15 > English > Macros .

Pro verzi Minitab 14 na adresu:

Program Files > Minitab 14 > Macros.

Knihovna dalších uživatelských maker pro Minitab je dostupná na adrese

<http://www.minitab.com/support/macros>.

Makra mohou být prováděna buď z okna „Session“ tak, že se vybere:

„Editor“ > „Enable Commands“ > kde se zobrazí „MTB“ >
zapiše se příslušné „Macro“ > „Enter“ ;

nebo se vybere

Edit > Command Line Editor >
zapiše se příslušné „Macro“ > „Submit Commands“ .

U jednotlivých maker jsou uvedeny specifické pokyny pro jejich provedení.

Poznámka 2

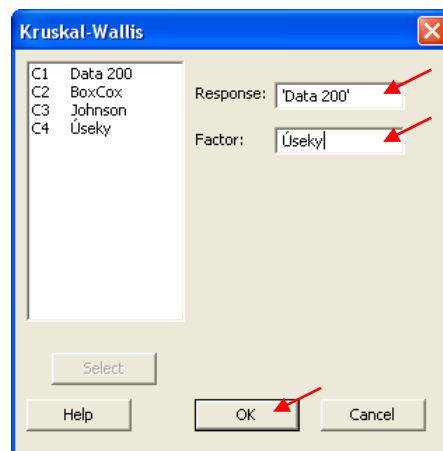
Kruskal-Wallisův test

je neparametrickou alternativou jednocestné ANOVA. Nevyžaduje, aby data byla rozdělena normálně, stačí předpokládat spojité rozdělení sledovaného znaku jakosti. Pro analýzu využívá pořadí dat namísto jejich aktuálních hodnot. Lze použít k ověření, zda dva nebo více nezávislých výběrů pochází ze základních souborů se stejným mediánem.

V softwaru Minitab postupujeme následovně:

Stat > Nonparametrics > Kruskal Wallis >

vyplníme dialogové okno a potvrdíme OK.

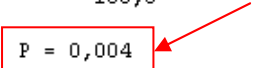


Výsledek testu je zapsán v okně Session:

Kruskal-Wallis Test on Data 200

Úseky	N	Median	Ave Rank	Z
1	25	1,546	133,6	3,05
2	25	1,167	76,1	-2,26
3	25	1,419	101,8	0,12
4	25	1,724	117,6	1,58
5	25	1,251	77,9	-2,09
6	25	1,257	84,2	-1,50
7	25	1,554	107,0	0,60
8	25	1,333	105,8	0,49
Overall	200		100,5	

H = 21,12 DF = 7 P = 0,004



Výsledná p-hodnota $< 0,05$ signalizuje důvod zamítnout hypotézu, že náhodné výběry (podskupiny) pocházejí ze základních souborů se stejným mediánem.

Použitá literatura:

- [1] Michálek J. aj. *Statistické metody řízení jakosti*, Česká společnost pro jakost. Praha 2007
- [2] Montgomery D.C.: *Introduction to Statistical Quality Control*. John Wiley. N.Y. 2000 (4.vydání)
- [3] Daimler Chrysler Corporation, Ford Motor Company, General Motors Corporation: *Statistická regulace procesů (SPC)*, Česká společnost pro jakost, Praha 2005
- [4] Michálek J.: *Vyhodnocování způsobilosti a výkonnosti výrobního procesu*, Centrum pro jakost a spolehlivost ve výrobě (CQR), Praha 2009