



StatSoft

# Analýza způsobilosti procesu

Analýza způsobilosti je jedna z nejběžnějších analýz vyžadovaných v oblasti zpracování průmyslových dat. V tomto článku si představíme indexy způsobilosti a podrobně si vysvětlíme, jak je který index počítán a co nám říká. Cílem je dostatečně osvětlit tuto tematiku, která většinou bývá vysvětlována jen velmi povrchně.

Pokud se snažíme vylepšit výrobní proces, pak bychom měli primárně definovat a odhalovat příčiny, jež vyvolávají podstatnou a nežádoucí změnu procesu, a tím ovlivňují znaky kvality, které sledujeme. Odhalením a redukcí těchto příčin se snažíme odstranit tzv. systémovou, nenáhodnou variabilitu (viz [minulý článek](#)). Nicméně metodám, které odhalují nestabilitu procesu a pomáhají nalézt příčiny, se nyní věnovat nebudeme a v dalším budeme předpokládat, že náš proces je statisticky zvládnutý, tedy že se nám již podařilo odstranit nežádoucí nenáhodnou variabilitu.

To, co nás tentokrát zajímá a co obvykle zajímá také odběratele, jsou indexy (ukazatele) způsobilosti (process capability index PCI). Tyto indexy jsou jakousi „známkou“ jakosti, podávají nám informaci o tom, zda a do jaké míry se daří dodržovat předepsané regulační (specifikační) meze a definovanou úroveň měřené veličiny, jež je sledován (např. v regulačním diagramu). Tyto indexy spadají do tzv. analýzy způsobilosti, kterou se v tomto článku budeme zabývat. Poznamenejme, že nám půjde jen o nejčastější situaci a to, když měřenou veličinou je spojitý znak.

## Analýza způsobilosti

Analýzou způsobilosti procesu zjišťujeme nejen, zda regulovaná veličina nepřekračuje specifikace, ale také jaká je rezerva vůči specifikacím, resp. do jaké míry náš proces odpovídá technickým specifikacím nebo obchodním cílům. Pokud např. vyrábíme pístní kroužky, chtěli bychom znát, kolik z nich má rozměr spadající dovnitř specifikačních mezí, jak dobře se nám daří specifikace dodržovat.

Data, která vstupují do analýz způsobilosti jsou měření daného procesu na několika vzorcích, měříme tedy například 25 vzorků každý po 5ti pozorováních. V následujícím mějte prosím na paměti, že toto je typ dat pro naši úlohu.

## Indexy způsobilosti

Indexy způsobilosti jsou založeny na porovnávání přirozeného kolísání skutečného procesu vůči technologickému předpisu. Někdy jsou považovány za „známkou“ jakosti, kterou odběratel požaduje.

Než se podíváme na samotné indexy, přijmeme obecně akceptovanou úmluvu, podle níž se meze pro technologický předpis stanovují ve vzdálenosti  $\pm 3\sigma$  od nominální hodnoty ( $\sigma$  je sm. odchylka procesu). Toto koresponduje s mezemi klasických regulačních diagramů, pomocí kterých monitorujeme proces a snažíme se ho přivést do statisticky zvládnutého stavu. Rozdíl těchto mezí označuje rozsah procesu.

Skutečný proces se pak popisuje odhadem střední hodnoty a odhadem směrodatné odchylky, z kterých se pak jednotlivé indexy počítají. Jaké přesně jsou tyto odhady, si ukážeme za chvíli.

Podívejme se nyní na ty nejdůležitější indexy, které se běžně používají k popisu způsobilosti procesu, a po té si ukážeme aplikaci na příkladu výroby pístních kroužků.

## Index způsobilosti $C_p$

Index  $C_p$  je nejzákladnější a nejpřímější indikátor způsobilosti procesu. Jednoduše porovnáme skutečnou variabilitu procesu (naměřených dat) vůči variabilitě povolené specifikacemi. Skutečný rozsah (jmenovatel) je vyjádřen šestinásobkem odhadnuté směrodatné odchylky - šestinásobek jednoduše proto, že specifikační meze byly také stanoveny jako šestkrát teoretická variabilita (viz předpoklad na specifikační meze výše). Čítenel ve vzorci je tedy jinak řečeno 6 krát teoretická variabilita.

$C_p$  je tedy definován jako poměr specifikačního rozsahu a reálný rozsahu procesu. Tento index se dá vyjádřit jako:

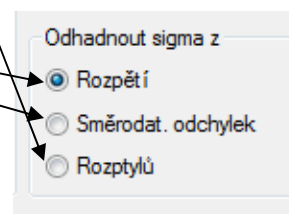
$$C_p = \frac{USL - LSL}{6s}$$

LSL a USL je dolní a horní specifikační mez,  $s$  je odhad směrodatné odchylky vypočítaný z naměřených hodnot.

Na tomto místě nám dovolte udělat malou nebo raději větší, ale velmi přínosnou odbočku.

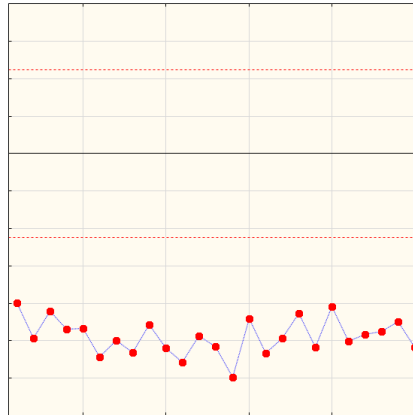
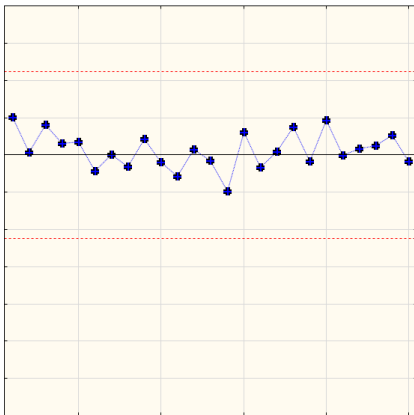
*Odbočka:* Nejzajímavější a vlastně také asi to, co v praxi vyvolává nejvíce otázek je to, jak se vlastně počítá odhad  $s$ . Vzpomeňme si, jaká máme data. V jednom vzorku máme pozorování naměřená za téměř homogenních podmínek. A hlavní pro analýzu způsobilosti je ocenit schopnost mít nízkou variabilitu dat za homogenních podmínek (opravdu pouze přirozené kolísání procesu bez dalších vlivů). Tudíž variabilita, o kterou nám jde nejvíce, je tedy variabilita uvnitř vzorků. Z každého vzorku tedy spočítáme odhad vnitroskupinové variability (z historických důvodů se nejčastěji používá početně jednoduchý způsob napočítání rozpětí – od maxima ve vzorku odečteme minimum ve vzorku) a poté z nich uděláme průměr. V softwaru je nicméně možné nastavit i odhad rozptylu z vnitroskupinových směrodatných odchylek nebo rozptylů.

Takže shrnuto, odhad  $s$  můžete dostat podle nastavení jako odmocninu z průměru rozptylů nebo jako průměr odhadnutých rozpětí vydělený konstantou  $d_2(n)$  nebo jako průměr směrodatných odchylek vzorků vydělený konstantou  $c_4(n)$ . Konstanty  $d_2(n)$  a  $c_4(n)$  jsou zde proto, aby odhady dávaly odhad stejné veličiny, tedy směrodatné odchylky. V softwaru (v nápovědě) tyto konstanty najdete v tabulce: *Unbiasing Constants  $c_4, c_5, d_2, d_3, d_4$* .



Volně řečeno, tento index vyjadřuje, jak se nám liší skutečný a povolený rozptyl dat. Hodnoty menší než 1 značí nízkou způsobilost, hodnoty nad 1 značí, že proces se chová lépe než předepsaný. V praxi se používá žebříček: nezpůsobilost (menší než 1), střední způsobilost (mezi 1 a 1,3) a vysoká způsobilost (větší než 1,3).

Je potřeba si uvědomit, že předchozí index nezachycuje neshodu střední hodnoty sledovaného procesu s předepsanou (nominální) hodnotou. Na obrázcích níže s vývojem procesu v čase můžete vidět, dvě situace se stejným indexem  $C_p$ , ale jeden z nich je mimo specifikační meze a jistě by žádného odběratele nenadchl.



Aby byla zachycena do hodnocení i poloha procesu, byl zaveden další index, který se běžně značí  $C_{pk}$ .

## Index Cpk

Úpravou indexu  $C_p$  můžeme získat index  $C_{pk}$ , který penalizuje i za to, že proces není vycentrován.

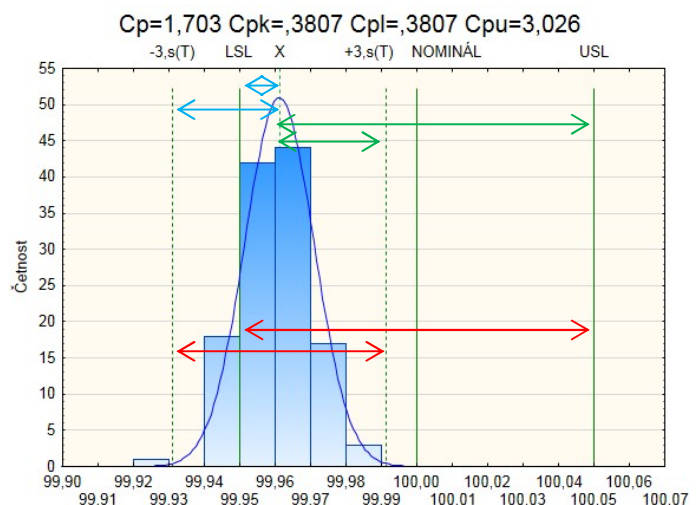
$$C_{pk} = \frac{\min(USL - \bar{x}; \bar{x} - LSL)}{3s}$$

Uvedený vzorec se skládá ze dvou částí: způsobilost dodržet horní specifikační mez a způsobilost dodržet dolní specifikační mez. Za výslednou hodnotu indexu  $C_{pk}$  považujeme menší z obou dílčích výsledků (tedy horší scénář). Ve vzorci je trojnásobek odhadu směrodatné odchytky, neboť se díváme vždy na vzdálenost meze od středu, což je při vycentrování polovina rozsahu. Tento index nás informuje o způsobilosti procesu při aktuálním seřízení výrobního zařízení.

Pokud je proces vycentrován, pak je  $C_{pk}$  shodné s  $C_p$ . Pokud se ovšem proces vzdálí od nominální hodnoty, pak je vždy  $C_{pk}$  menší než  $C_p$ .

Ilustrujme si situaci na příkladu. Na obrázku máme histogram hodnot, specifikační meze  $USL$ ,  $LSL$  i střed intervalu, tedy  $NOMINÁL$ , specifikace jsou tedy značeny plnou čarou. Dále zde máme průměr reálně naměřených hodnot, (značen jako  $X$ ), další čárkované čáry s označením  $\pm 3s(T)$  jsou poté meze vypočítané z dat.

Šipkami jsou názorně ukázány rozsahy, s kterými se počítá ve vzorcích. Vidíme tedy, že  $C_p$  je poměr červených vzdáleností,  $C_{pk}$  je minimum z poměru modrých ( $C_{pl}$ ) a zelených ( $C_{pu}$ ) vzdáleností. Přesná čísla pro tento příklad jsou v nadpisu obrázku.



## Potenciální způsobilost C<sub>pm</sub>

Jedna z posledních modifikací indexu způsobilosti je zaměřena na odstranění vlivu (náhodného) nevycentrování. Počítá se směrodatnou odchylkou ( $s_2$ ), nicméně ta je vypočítána jinak než jsme byli doposud zvyklí – místo průměru se ve vzorci přímo vyskytuje předepsaná hodnota  $T$ . Změnou je také to, že se počítá ze všech naměřených dat (celkový počet měření je  $n$ ).

$$s_2 = \sqrt{\frac{\sum_{i=1}^n (x_i - T)^2}{n - 1}}$$

Pokud použijeme tento alternativní odhad směrodatné odchylky k výpočtu indexu  $C_p$ , dostaneme nový index, který budeme označovat jako  $C_{pm}$ .

$$C_{pm} = \frac{UCL - LCL}{6 \cdot s_2}.$$

Výhodou tohoto indexu je, že oproti ostatním se rozptyl vztahuje ke stanovené referenční hladině a ne k hladině odhadnuté z dat. Pokud se podíváme na chování, vidíme, že s hodnotami blíže  $T$  roste index, stejně tak roste, pokud je rozptýlenost okolo  $T$  menší.

$C_{pm}$  vychází vždy menší než  $C_p$ , a proto je přísnější.

## Další pojmy

V souvislosti s indexy způsobilosti můžeme narazit i na další charakteristiky, které se občas požadují, následuje jejich výpis:

### Poměr způsobilosti (C<sub>r</sub>)

Tento index je vlastně ekvivalentní indexu  $C_p$ ; je roven jeho převrácené hodnotě.

$$C_r = \frac{1}{C_p}$$

Index se dá tedy slovně popsat jako podíl specifikacího rozsahu, který je využíván vlastním procesem.

### Dolní/horní potenciální způsobilost (C<sub>pl</sub>, C<sub>pu</sub>)

Hlavní nevýhodou indexu  $C_p$  (nebo  $C_r$ ) je to, že poskytuje nesprávné informace, pokud proces není vycentrován kolem nominální hodnoty (TS). „Excentricitu“ procesu je možné vyjádřit pomocí průměru procesu. Nejdříve je možné vypočítat dolní a horní potenciální způsobilost procesu:

$$C_{pl} = \frac{\bar{x} - LSL}{3s} \quad \text{a} \quad C_{pu} = \frac{USL - \bar{x}}{3s}.$$

Je zřejmé, že pokud obě hodnoty nejsou shodné, proces není vycentrován.

### Korekce (k)

Index  $C_p$  je možné upravit tak, aby reflektoval fakt, že proces není vycentrován. Můžeme vypočítat hodnotu korekce  $k$ :

$$k = \left| \frac{T - \bar{x}}{\frac{1}{2}(USL - LSL)} \right|$$

Tato korekce vyjadřuje excentricitu (nominální hodnota mínus průměr) vztaženou ke specifikačnímu rozsahu. Index  $C_p$  se pak dá ekvivalentně vyjádřit jako:

$$C_{pk} = (1 - k)C_p.$$

Poznámka: Kompletní výčet vzorců použitých v programu *STATISTICA* týkající se této problematiky najdete v nápovědě pod heslem *Z Benchmark Potential*

## Výkonnost procesu vs. způsobilost procesu

V praxi se často setkáváme se situací, kdy v procesu existuje významná variabilita mezi skupinami vzorků. V takových případech ukazatele způsobilosti dost často poskytují neobjektivní informaci o způsobilosti procesu za toto dlouhé období. Proto byly zavedeny také ukazatele výkonnosti procesu, které se počítají stejně jako ukazatele způsobilosti, jen odhad směrodatné odchylky  $s$  vychází z variability za delší časové období.  $s$  je tedy klasická směrodatná odchylka všech pozorování za měřené období a ignoruje se fakt, že data pocházejí z více vzorků.

Pokud tedy použijeme k výpočtu indexů celkovou variabilitu, říká se výsledným indexům *indexy výkonnosti* (protože popisují vlastní výkon procesu) a značí se analogicky jako indexy způsobilosti, jen s písmenem P, tedy ( $P_p$ ,  $P_{pk}$ , ...).

## Příklad

Nyní už k příkladu, neboť jak je známo, nejlépe se vše osvětlí na příkladu. Použitý datový soubor *Pistons.sta* je k dispozici v každé *STATISTICE Soubor* -> *Otevřít příklady* -> *Datasets*

Představme si, že vyrábíme pístní kroužky. Specifikace říkájí, že jejich průměr by měl být  $74 \pm 0,05$  mm. Z procesu jsme odebrali 25 vzorků po 5 pozorováních. Úkolem je zjistit, jak způsobilý je náš proces.

### Předpoklady

Zatím jsme to implicitně nezmínili, ale jako téměř každá obecná parametrická metoda i představené indexy způsobilosti jsou založeny na předpokladu normality. Pro data, která nevykazují normální rozdělení, existují taktéž indexy, nicméně se počítají trochu jinak a jsou nad rámec tohoto základního článku o indexech způsobilosti.

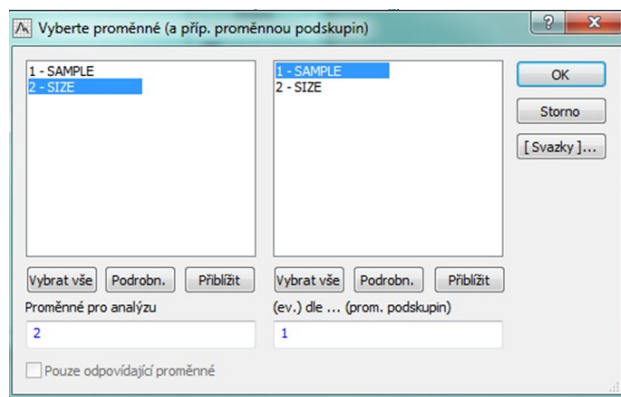
Pokud bychom provedli test normality pro naše data, zjistili bychom, že nezamítáme hypotézu o normalitě dat. Použití indexů způsobilosti, jak jsme si je představili, je tedy v pořádku.

### Indexy způsobilosti procesu

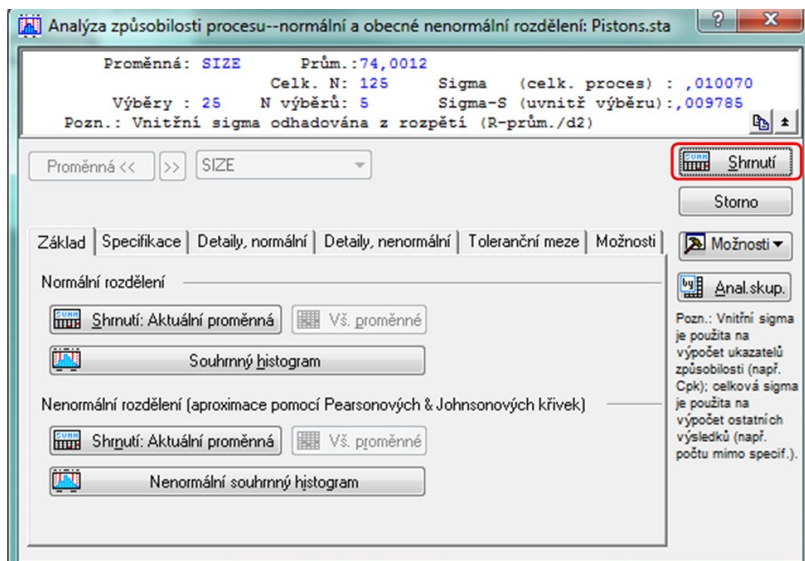
Indexy způsobilosti najdeme jako jednu z analýz v menu *Analýza procesu* v průmyslových statistikách. Zde první dvě položky dělají to, co potřebujeme. Zvolíme první, neboť máme data ve tvaru zdrojových hodnot pro každé měření. Klikněte na tlačítko *Proměnné* a zvolte proměnnou *SIZE* v poli *Proměnné pro analýzu* a *SAMPLE* jako grupovací proměnnou (proměnná *SAMPLE* totiž určuje, ke kterému vzorku pozorování patří).

Klikněte na *OK*. Všechna ostatní nastavení nechte na jejich implicitních hodnotách.

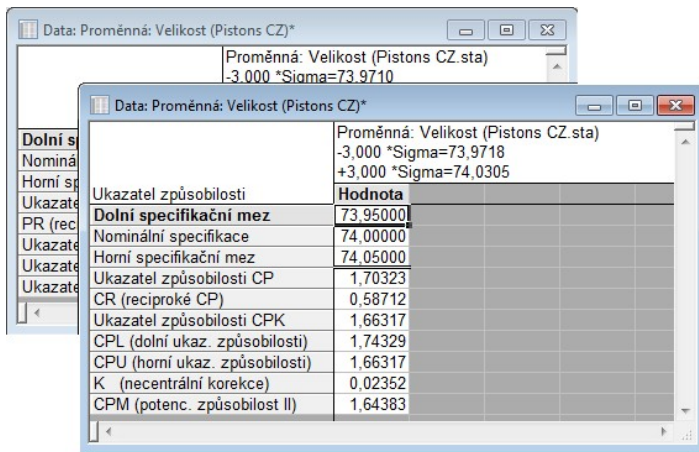
Objeví se dialogové okno *Zadání/úpravy specifikačních mezí*. Abychom mohli indexy způsobilosti počítat, je nutné nejprve zadat hodnoty technických specifikací. Do pole *Nominální* vložte 74 a do pole *Delta* hodnotu 0,05 (případně lze vložit *Nominální* a hodnoty *LSL* a *USL*).



Klikněte na tlačítko *OK* a stiskněte *Shrnutí*.



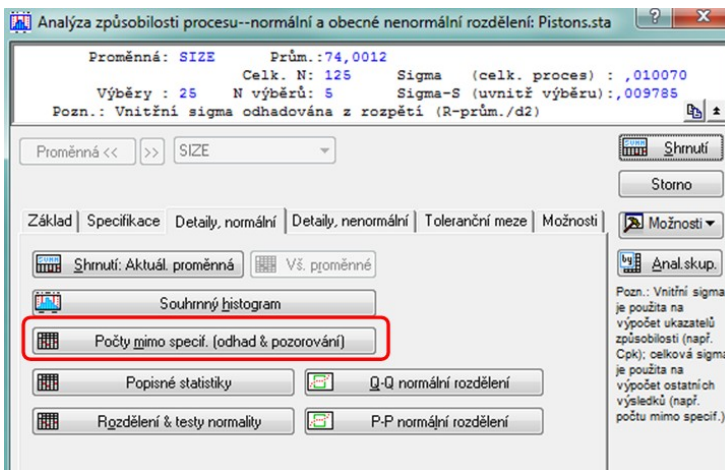
*STATISTICA* vytvoří dvě tabulky (jednu pro indexy způsobilosti a druhou pro indexy výkonnosti):



Tato tabulka obsahuje specifikace a různé indexy způsobilosti. Protože všechny indexy způsobilosti jsou vyšší než 1, proces se zdá být způsobilý při splňování našich požadavků. Jinými slovy, výrobní linka produkuje méně než 1% neshodných výrobků (to plyne z předpokladu normality, za tohoto předpokladu je ve specifikačním intervalu přibližně 99% hodnot, nicméně specifikační interval je širší než 6krát směrodatná odchylka v našich datech a tedy se do něj reálně vejde ještě více dat).

Ve skutečnosti proces využívá méně než 60% specifikačního rozsahu, jak je vidět z hodnoty indexu  $C_p$  (0,58).

Z výstupní tabulky je patrné, že proces je přibližně vycentrován kolem hodnoty 74, protože hodnota indexu  $C_{pk}$  je stále poměrně vysoká (1,66317) a necentrální korekce  $k$  (0,023) je velice blízká nule.



Je možné také provést analýzu očekávaného a pozorovaného počtu naměřených hodnot ležících mimo specifikační meze. Na záložce **Detaily, normální** klikněte na tlačítko **Počty mimo specifikace**.

Proměnná: SIZE, rozdělení: Normál. (Pistons)				
Specifikace: Dolní=73,9500 Nominál=74,00				
Prům.74,001, Sm.od.:0,1007				
	Pozorov.	Procenta Pozorov.	Očekáv.	Procenta Očekáv.
Nad USL :	0	0,00	0,000078	0,000062
Pod LSL :	0	0,00	0,000023	0,000019
Celk.	0	0,00	0,000101	0,000081

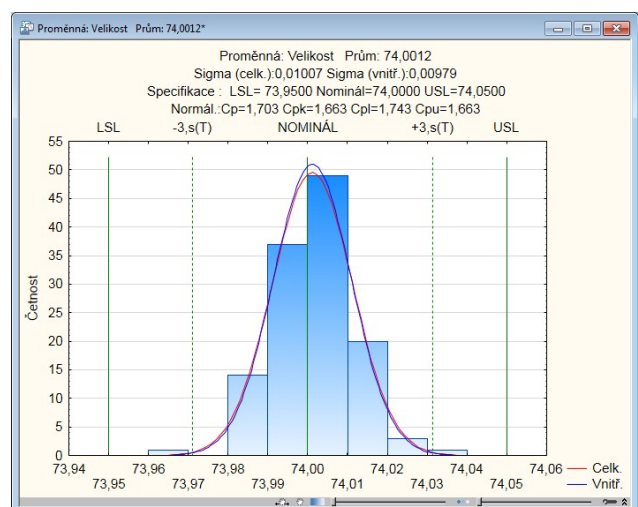
Tato tabulka obsahuje údaje o počtu a procentu pozorovaných jednotek ve vzorku, které padnou mimo specifikační meze (LSL, USL). Současně je

uveden také počet a procento jednotek, u kterých se očekává, že padnou mimo specifikační meze, podle normálního rozdělení. Pro naše data ze specifikací nevypadlo ani jedno pozorování. Podle normálního rozdělení s daným průměrem procesu a směrodatnou odchylkou bychom mohli očekávat, že pouze 0,00081% nebude odpovídat našim specifikacím.

Souhrn analýzy způsobilosti je také možné nechat zobrazit graficky v podobě histogramu. Klikněte na tlačítko **Souhrnný histogram**.

(Poznámka: Tento histogram byl vytvořen pro deset kategorií. Na záložce **Možnosti** zrušte označení políčka **Pěkné intervaly** a zvolte počet kategorií.)

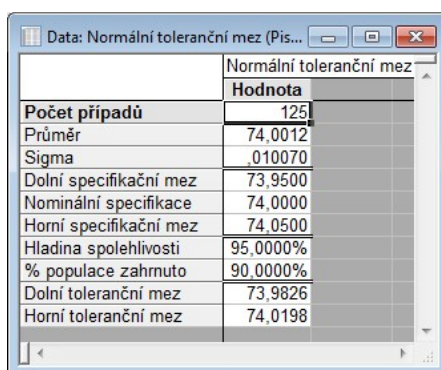
Uvedený histogram zobrazuje rozložení naměřených hodnot společně s proloženou křivkou normálního rozdělení (pro očekávané hodnoty). Současně jsou v histogramu vyznačeny specifikační meze (LSL a USL) a interval  $\pm 3s$ . Titulek grafu obsahuje hlavní indexy způsobilosti procesu.



## Závěrem

Doufáme, že jsme tímto článkem alespoň částečně osvětlili problematiku analýzy způsobilosti.

Nakonec bychom ještě doplnili obrázek o metodách analýzy způsobilosti, a zmínili metodu, která byla využívána dříve. Než byly do praxe uvedeny indexy způsobilosti, používaly se k ohodnocení charakteru výrobního procesu toleranční intervaly. Při dané velikosti vzorku ( $n$ ), průměru procesu, směrodatné odchylce procesu ( $\sigma$ ), hladině spolehlivosti a podílu populace, který má spadat do tolerančního intervalu, je možné vypočítat toleranční meze. Pro toleranční meze našeho příkladu se stačí přepnout na záložku **Toleranční meze**, určit výše zmíněné parametry (v našem případě změníme např. údaj v poli **procentu zahrnutých** na 90%) a kliknout na tlačítko **Toleranční meze, normální rozdělení**.



	Normální toleranční mez
	Hodnota
Počet případů	125
Průměr	74,0012
Sigma	,010070
Dolní specifikační mez	73,9500
Nominální specifikace	74,0000
Horní specifikační mez	74,0500
Hladina spolehlivosti	95,0000%
% populace zahrnuto	90,0000%
Dolní toleranční mez	73,9826
Horní toleranční mez	74,0198

Kdybychom měli stručně shrnout údaje v tabulce, mohli bychom říci asi toto: jsme si na 95% jistí, že 90% populace bude ležet mezi hodnotami 73,9826 a 74,0198. Tyto hodnoty jsou popsány jako **Dolní** a **Horní toleranční mez**.