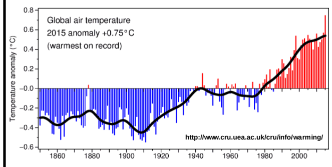
MASARYKOVA UNIVERZITA

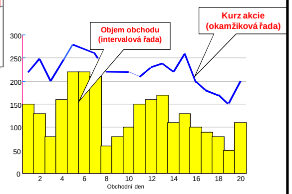
# Z2069 Statistické metody a zpracování dat II

## Analýza časových řad

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

## Příklady časových řad a jejich použití




- vývoj cen akcií
- objem obchodování na burze
- průměrný roční odtok vody z povodí
- vývoj počtu obyvatelstva určité lokality
- maximální denní srážkové úhrny na určité stanici

## Základní pojmy

**Časová řada** je chronologicky uspořádaná posloupnost hodnot určitého statistického ukazatele.

$$y_t = f(t) \qquad y_1, y_2, \cdots, y_n$$

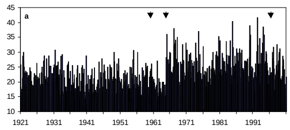
$y_t$, kde $t = 1, 2, \ldots, n$

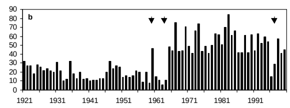
*y = ukazatel*
*t = časová proměnná*
*n = počet členů řady*

Pomocí časových řad můžeme zkoumat **dynamiku** jevů v čase.
Mají základní význam pro **analýzu příčin**, které na tyto jevy působily a ovlivňovaly jejich chování v minulosti, tak pro **předvídání** jejich budoucího vývoje.

## Problémy při sestavování časových řad

- Problém volby časových bodů pozorování
- Problémy s délkou časové řady
- Problémy s kalendářem
- Problémy s nesrovnatelností jednotlivých měření





Uvedené problémy mohou vést k narušení **homogenity** časové řady

Maximální denní nárazy větru a počty dnů s nárazy větru na stanici Praha, Karlov v období 1921-1990

## Transformace časové řady

Jedná se o úpravu původní časové řady, tak aby

1. splňovala podmínky pro následnou analýzu (např. stacionarita atd.)
2. zvýrazňovala dále analyzovanou složku

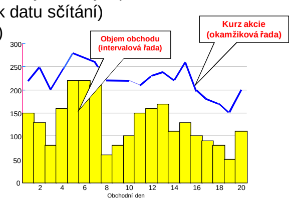
**Běžné druhy transformací:**

- přidání konstanty $\quad y = y + C$
- linearizace řady $\quad y = \ln(y)$
- odečtení průměru $\quad y = y - \bar{y}$
- standardizace $\quad y = \left(\dfrac{y - \bar{y}}{s_d}\right)$
- odečtení tzv. trendové funkce (viz. dále)

## Základní typy časových řad

Časové řady **deterministické** - neobsahují prvek náhody (sin(x)) a **stochastické** (realizace náhodného procesu)

Časové řady **absolutních** veličin (přímo zjišťovaných)
- **okamžikové** (počet obyvatel – k datu sčítání)
- **intervalové** (denní úhrn srážek)




Časové řady **odvozené**
- průměrných veličin (řada klouzavých průměrů)
- poměrných – relativních veličin (řada hektarových výnosů)

Časové řady **ekvidistantní** a **neekvidistantní**

## Okamžikové časové řady

Jsou spojité v čase, záleží u nich na rozhodném okamžiku šetření. Hodnota nezávisí na délce intervalu, za který je znak zjišťován.

Okamžikové ukazatele za několik intervalů nesčítáme. Je však pro ně typické počítání průměrů v čase. Průměr okamžikové veličiny za určité období označujeme jako tzv. **chronologický průměr.**
Nejprve spočteme průměr za časové okamžiky $t_{i-1}$ a $t_i$, pro i=2 až n.
Z těchto hodnot určíme průměr pro celou řadu:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2}y_n}{n-1}$$

Uvedený vztah platí v případě, že délka všech intervalů je konstantní. Pokud ne, je nutné jednotlivé dílčí průměry vážit délkami intervalů a vypočítat vážený chronologický průměr.

## Intervalové časové řady

Jednotlivé hodnoty se vztahují k **časovým úsekům** a přímo závisí na jejich délce.

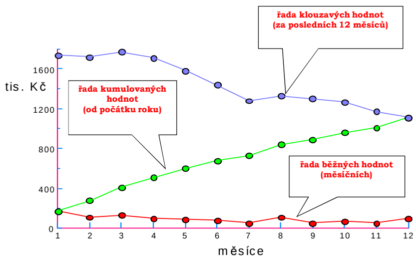
Hodnotu intervalového ukazatele zjištěnou za časový interval ($t_{i-1}$, $t_i$) označme $q_i$ a přiřazujeme-ji ke středu časového intervalu.
Časovou řadu hodnot $q_i$ označujeme intervalovou **řadou běžných hodnot**.

Za delší časové období lze intervalové ukazatele shrnovat a vytvářet **součtové (kumulativní) řady**. Součtová řada vznikne postupným sčítáním hodnot za sebou jdoucích časových intervalů. Podle průběhu součtové řady můžeme posoudit rovnoměrnost vývoje hodnot znaku.

Požadavkem sestavování intervalových časových řad je **konstantnost** délky časového intervalu. V řadě případů tento požadavek není splněn (např. počet dnů v měsíci).

Dalším typem součtových časových řad jsou **řady klouzavých úhrnů**. Jsou vhodné ke srovnání úrovně řady ve sledovaném období s úrovní řady období předešlého.

## Z - diagram




Řady běžných hodnot, řady kumulovaných hodnot a řady klouzavých úhrnů lze znázornit v tzv. Z-diagramu

## Odvozené ukazatele časové řady

Při práci s časovými řadami je typické, že často pracujeme ne přímo s původní časovou řadou, ale s nějakou její **transformací**.

**Absolutní přírůstek** (první diference) $\Delta y_t = y_t - y_{t-1}$

Jsou-li členy v řadě absolutních přírůstků prakticky konstantní, potom hodnoty řady lineárně rostou (klesají).

**Relativní přírůstek** $\delta_i = \frac{\Delta y_i}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} = \frac{y_i}{y_{i-1}} - 1$

Informuje nás o rychlosti (tempu) růstu

## Odvozené ukazatele časové řady

Koeficient růstu (**řetězový index**): vyjadřuje, o kolik procent vzrostla hodnota časové řady v okamžiku $t_i$ ve srovnání s hodnotou řady v čase $t_{i-1}$.

$$k_i = \delta_i + 1 = \frac{y_i}{y_{i-1}} \cdot 100(\%)$$

**Průměrný koeficient růstu**: pro celou řadu se vypočte jako geometrický průměr jednotlivých hodnot koeficientů růstu.

$$\bar{k} = \sqrt[n-1]{k_1 \cdot k_2 \ldots k_{n-1}} = \sqrt[n-1]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \ldots \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Uvedený výpočet je vhodný pouze v případě stálého a přibližně stejného růstu hodnot řady.

## Odvozené ukazatele časové řady

Pro účely srovnání různých časových řad se jejich hodnoty převádějí na tzv. **bazické indexy** (indexy se stálým základem):

$$k_i' = \frac{y_i}{y_z} \cdot 100(\%)$$

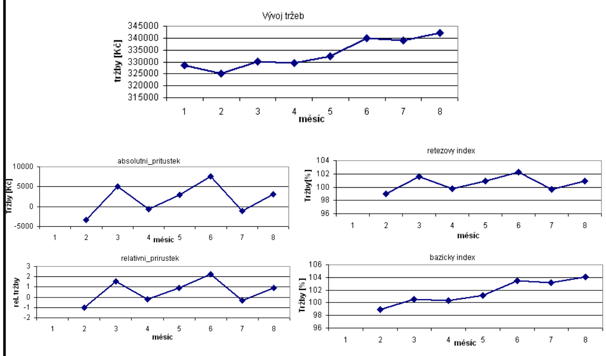
Hodnota $y_z$ je obvykle prvním nebo posledním členem časové řady (základ).

## Odvozené ukazatele časové řady

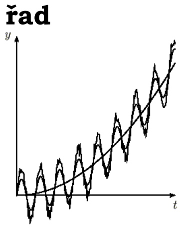
| | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 6 | měsíc | t | tržba | absolutní přírůstek | relativní přírůstek | řetezový index koeficent růstu | bazický index (z=1) |
| 7 | leden | 1 | 328541 | . | . | . | 100.0 |
| 8 | únor | 2 | 325154 | -3387 | -1.0309 | 99.0 | 99.0 |
| 9 | březen | 3 | 330244 | 5090 | 1.5654 | 101.6 | 100.5 |
| 10 | duben | 4 | 329570 | -674 | -0.2041 | 99.8 | 100.3 |
| 11 | květen | 5 | 332489 | 2919 | 0.8857 | 100.9 | 101.2 |
| 12 | červen | 6 | 340025 | 7536 | 2.2665 | 102.3 | 103.5 |
| 13 | červenec | 7 | 338962 | -1063 | -0.3126 | 99.7 | 103.2 |
| 14 | srpen | 8 | 342110 | 3148 | 0.9287 | 100.9 | 104.1 |

- Absolutní přírůstek =D8-D7
- Relativní přírůstek =E8/D7*100
- Řetězový index =D8/D7*100
- Bazický index =D8/$D$7*100

## Odvozené ukazatele časové řady




## Základy analýzy časových řad



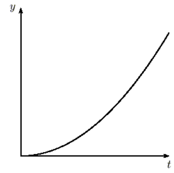
Hlavní cíle analýzy časových řad
1. odhalení zákonitostí a **příčin dosavadního vývoje**
2. **prognóza** chování časových řad

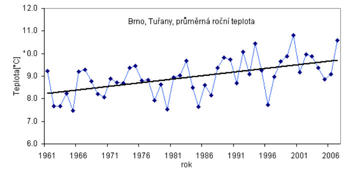
Každá řada může obsahovat čtyři základní složky:

- **trend** $(T_t)$
- **periodická (sezónní) složka** $(S_t)$
- **cyklická složka** $(C_t)$
- **náhodná složka** $(\varepsilon_t)$

První tři složky tvoří systematickou část řady.

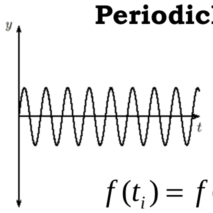
## Trendová složka časové řady



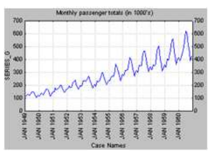



- **Trend** je obecná tendence vývoje zkoumaného jevu za dlouhé období.
- Je výsledkem dlouhodobých a stálých procesů (v měřítku posuzované délky časové řady).
- Trend může být lineární či nelineární.
- Trend může být rostoucí, klesající nebo může existovat řada bez trendu (s nulovým trendem).
- Časové řady bez trendu se označují jako stacionární.

## Periodická složka časové řady



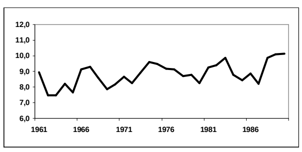
$$f(t_i) = f(t_i + T)$$

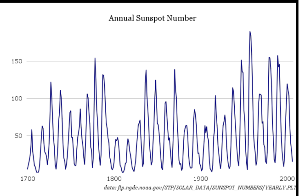



- **Periodická složka** je pravidelně se opakující odchylka od trendové složky s pevnou délkou **periody T**.
- Perioda této složky je menší než celková velikost sledovaného období.
- Typickým případem jsou **sezónní kolísání** a nebo řady denních, měsíčních, čtvrtletních ukazatelů.
- Příčiny sezónnosti jsou různé, většinou však dobře definovatelné.
- Sezónnost je typická pro časové řady ekonomických ukazatelů.

## Cyklická složka

$$f(t_i) \approx f(t_i + \overline{T})$$



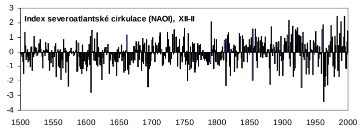



- **Cyklická složka** udává kolísání okolo trendu v důsledku dlouhodobého cyklického vývoje.
- Cyklická složka může vykazovat změny v délce a amplitudě cyklu.
- Délka cyklu je tedy většinou neznámá. (př. demografický trend, kolísání teploty vzduchu).
- Délka cyklu je tedy delší než 1 rok. V některých případech se označuje jako „střednědobý trend“.
- Bývá typickou součástí časových řad meteorologických prvků (př. problém globálního oteplování) či hydrologických jevů.

## Náhodná složka časové řady

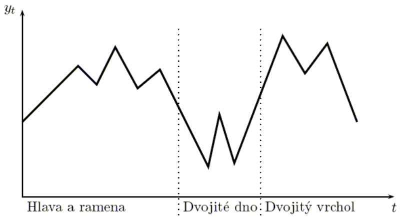

- **Náhodná (stochastická) složka** se nedá popsat žádnou funkcí času.
- "Zbývá" po vyloučení trendu, sezónní a cyklické složky.
- Jejím zdrojem jsou v **jednotlivostech** nepostižitelné jevy.
- Lze ji však popsat pravděpodobnostně.

## Grafické metody analýzy časových řad




- Prvotní analýza spočívá v grafickém znázornění průběhu řady.
- Graf slouží k prvotnímu posouzení tendence změn či k hledání opakujících se jevů („patterns“).
- I tyto jednoduché metody umožňují velmi krátkodobou předpověď.
- Graf však velmi dobře může znázorňovat nehomogenity, porovnávat dvě či více řad mezi sebou, …
- Slouží k výběru vhodné metody analýzy.

## Grafické metody analýzy časových řad

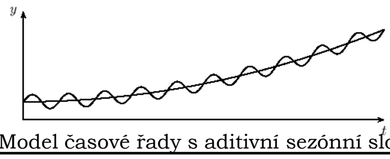



Vývoj kurzu akcií – příklad výskytu jednoduchých obrazců (patterns) v časové řadě

## Klasický (formální) model analýzy časových řad

- Klasický model je pouze popisem jednotlivých složek časové řady jako forem pohybu, ne poznáním příčin.
- Jedná se o **dekompozici** na jednotlivé složky a jejich formální popis např. tzv. **aditivním modelem**:
- Základem je popis systematické složky (trendu, cyklických a periodických kolísání).

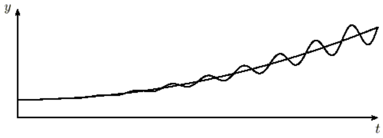
$$y_t = Y_t + \varepsilon_t = T_t + S_t + C_t + \varepsilon_t$$




Model časové řady s aditivní sezónní složkou

## Multiplikativní model

$$y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t$$




Model časové řady s multiplikativní sezónní složkou

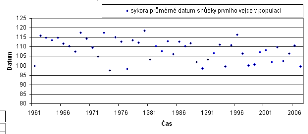
## Analýza trendu

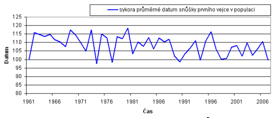
A. **Klasický přístup** založený na matematicko-statistickém modelování. Modelované parametry jsou **KONSTANTNÍ** v čase. Neadaptivní metody – např. regresní modely. Umožňují snadnou předpověď.

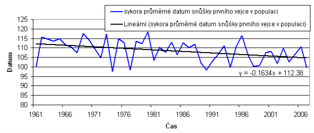
B. **Adaptivní přístup** – parametry se v čase **VYVÍJEJÍ**. Například charakter lineárního trendu se mění (mění se směrnice trendu). Za jednoduchou adaptivní metodu lze považovat i metodu klouzavých průměrů (viz. dále).

## Analýza trendu – základní metody vyrovnávání:

- analytické (popis časové řady funkcí)
- mechanické (klouzavé průměry)
- …







analytické vyrovnání trendu

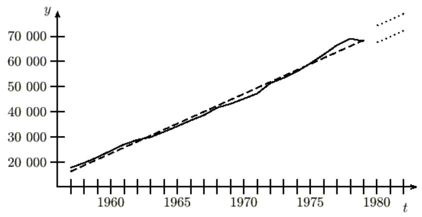
## Analytické vyrovnávání trendu matematickou křivkou

$$y_t = Tr_t + E_t$$

- Patří mezi neadaptivní metody. Vychází z předpokladu, že se trend po celou sledovanou dobu nemění a že je možné ho popsat některým typem matematické křivky.
- Identifikace trendu se redukuje na výběr správného typu matematické křivky a odhad jejích parametrů.
- Na problém analýzy trendu lze pohlížet jako na speciální případ **regresní závislosti**, kdy nezávisle proměnnou je čas.
- Časovou řadu vyrovnáváme křivkou, která nejlépe vystihuje její vývojový trend. Výpočet parametrů křivky se děje **metodou nejmenších čtverců**.

## Lineární trend

$$y_t = b_0 + b_1 t$$




Parametr $b_1$ představuje přírůstek hodnoty $y$ připadající na jednotkovou změnu časové proměnné. Řada se vyznačuje konstantními absolutními přírůstky (první diference).

## Lineární trend

Hodnoty parametrů $b_0$ a $b_1$ získáme metodou nejmenších čtverců obdobně jako v případě jednoduché lineární regrese, tedy:

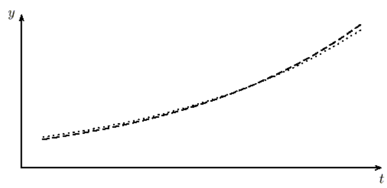
$$b_1 = \frac{\sum_{t=1}^{n} t y_t - \bar{t}\sum_{t=1}^{n} y_t}{\sum_{t=1}^{n} t^2 - n\bar{t}^2} \qquad b_0 = \bar{y} - b_1 \bar{t}$$

**Předpověď** budoucí hodnoty (**bodová** předpověď) má tvar:

$$\hat{y}_T = b_0 + b_1 T$$

## Exponenciální trend

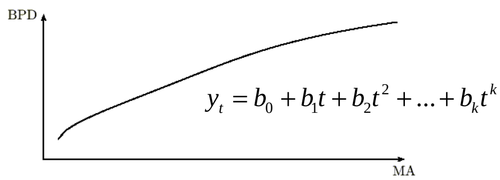
$$y_t = b_0 \cdot b_1^t$$




Parametr $b_1$ představuje průměrný přírůstek hodnot $y_t$. Ty se chovají jako členy geometrické posloupnosti. Protože se již nejedná o funkci lineární v parametrech, lze k odhadu exponenciálního trendu využít metody nejmenších čtverců pouze po její **logaritmické transformaci:**

$$\log y_t = \log b_0 + t \cdot \log b_1$$

## Polynomický trend



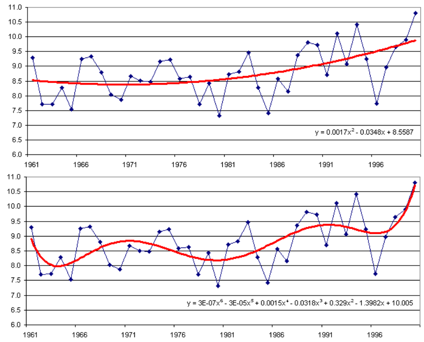

$$y_t = b_0 + b_1 t + b_2 t^2 + \ldots + b_k t^k$$

Při volbě stupně polynomu je třeba postupovat opatrně. Vyšší stupeň zajišťuje těsnější proložení empirických hodnot křivkou, vede ale k nestabilitě trendu.

Vyšší polynomy se většinou vůbec nehodí k extrapolacím.

K odhadu parametrů lze využít MNČ.

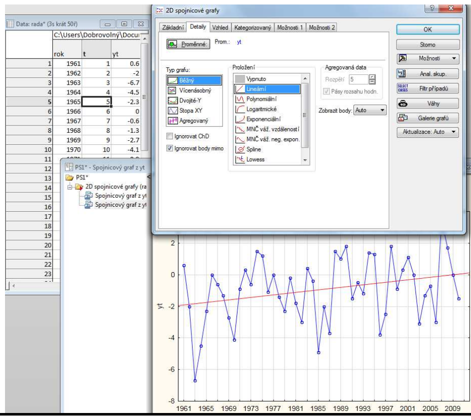
## Polynomický trend




Rozdíly v proložení časové řady polynomem 2. (nahoře) a 6. (dole) stupně

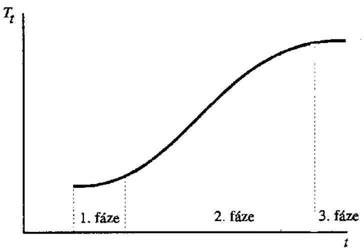
## Proložení trendu

*Grafy – 2 D spojnicové grafy*



## Logistická křivka

$$y_t = \frac{1}{k + b_0 \cdot b_1^t}$$




Křivka má tři úseky, první je charakterizován pozvolným vzestupem, druhá v okolí inflexního bodu prudkým růstem a třetí určitou vrcholovou stagnací. (patří mezi tzv. S-křivky).

## Verifikace modelu

Je zapotřebí zhodnotit statistickou **významnost** odhadnutých **parametrů** modelu i **modelu** jako celku.

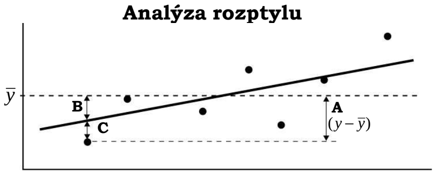
MNČ – podstatou je, že model vždy vysvětlí pouze **část** variability (proměnlivosti) pozorovaných dat.

Je nutné zjistit (testovat), zda model jako celek dává lepší vysvětlení, než je možné očekávat jako důsledek náhody a to na jisté hladině významnosti.

**Koeficient determinance $R^2$** – základní ukazatel vhodnosti použitého modelu (vzorec a interpretace viz. korelační počet)

**Analýza rozptylu**

## Analýza rozptylu




$$s_y^2 = s_{\hat{y}}^2 + s_{y-\hat{y}}^2$$

A. Rozptyl empirických hodnot (celkový) $$s_y^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$$

B. Rozptyl vyrovnaných hodnot (modelový) $$s_{\hat{y}}^2 = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2$$

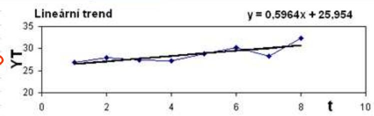
C. Rozptyl reziduální $$s_{y-\hat{y}}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_i - \hat{y}_t)^2$$

## Analýza lineárního trendu a testování vhodnosti - EXCEL

**Model vysvětluje více než 63 % proměnlivosti studované charakteristiky v čase**

VÝSLEDEK

| Regresní statistika | |
|---|---|
| Násobné R | 0,797711 |
| Hodnota spolehlivosti R | 0,636343 |
| Nastavená hodnota spolehlivosti R | 0,575733 |
| Chyba stř. hodnoty | 1,192911 |
| Pozorování | 8 |




ANOVA

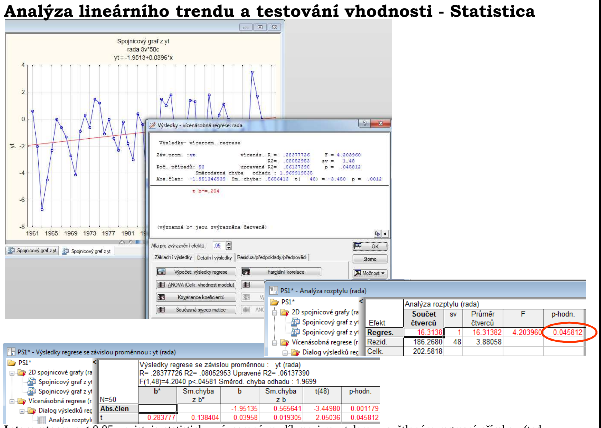
**p-hodnota**

| | Rozdíl | SS | MS | F | Významnost F |
|---|---|---|---|---|---|
| Regrese | 1 | 14,94054 | 14,94054 | 10,49905[illegible]5 | 0,017681891 |
| Rezidua | 6 | 8,538214 | 1,423036 | | |
| Celkem | 7 | 23,47875 | | | |

| | Koeficienty | Chyba stř. hod. | t stat | Hodnota P | Dolní 95% | Horní 95% | Dolní 95,0% | Horní 95,0% |
|---|---|---|---|---|---|---|---|---|
| Hranice | 25,95357 | 0,929508 | 27,92182 | 1,39605E-07 | 23,67914451 | 28,228 | 23,67914 | 28,228 |
| t | 0,596429 | 0,18407 | 3,240225 | 0,017681891 | 0,146024926 | 1,046832 | 0,146025 | 1,046832 |

**Interpretace:** $p < 0,05$ - existuje statisticky významný rozdíl mezi rozptylem vysvětleným regresní přímkou (tedy modelem trendu) a zbytkovým (reziduálním) rozptylem – zvolený model trendu je vhodný

**Analýza lineárního trendu a testování vhodnosti - Statistica**




**Interpretace:** $p < 0,05$ - existuje statisticky významný rozdíl mezi rozptylem vysvětleným regresní přímkou (tedy modelem trendu) a zbytkovým (reziduálním) rozptylem – zvolený model trendu je vhodný

## Kritéria pro volbu vhodného modelu trendové funkce I.

A. Volba vhodné trendové funkce by v prvé řadě měla vycházet z **věcné analýzy zkoumaného jevu**.

Ta nám umožní zaměřit se na určité typy (skupiny) funkcí či některé jiné předem vyloučit – jde o funkci rostoucí či klesající, má inflexní bod či je nekonečně rostoucí.

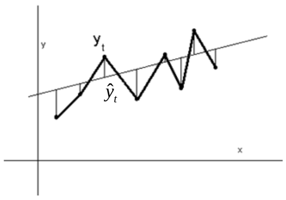
Pro použitou trendovou funkci je důležité, zda má (logistický trend) či nemá (lineární trend – růst řady není ničím omezen) asymptotu. Je to důležité pro předpovídání chování časové řady.

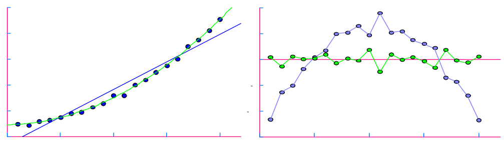
## Kritéria pro volbu vhodného modelu trendové funkce II.

**B. Analýza grafu** časové řady a **analýza reziduí**.

**$y_t$ – empirické hodnoty**

**$\hat{y}_t$ – teoretické hodnoty – vyrovnané trendovou funkcí**



## Kritéria pro volbu vhodného modelu trendové funkce III (objektivní)

Spočívají v minimalizaci předem zvoleného kritéria (jako v případě regresní analýzy). Za toto kritérium se nejčastěji bere součet čtverců odchylek empirických hodnot $y_t$ od hodnot vyrovnaných $\hat{y}_t$ (**součet čtvercových chyb**):

$$SSE = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$$

Z uvažovaných funkcí se vybírá ta s nejmenší hodnotou reziduálního součtu čtverců.

**POZOR** – jde o formální kritérium. Např. použijeme-li polynom vysokého stupně, může být reziduální součet čtverců i nulový, avšak zcela nepoužitelný.

## Objektivní kritéria pro volbu vhodného modelu trendové funkce III

Počítačové programy obvykle nabízejí následující míry vhodnosti zvolené trendové funkce:

**Střední chyba odhadu (M.E. – Mean Error)**

$$M.E. = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)}{n}$$

**Střední čtvercová chyba odhadu (M.S.E. – Mean Square Error)**

$$M.S.E. = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}$$

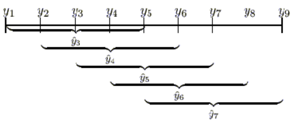
**Je to nejpoužívanější kritérium.**

## Informativní testy pro volbu vhodné trendové křivky:

| Trend | Informativní test |
|---|---|
| lineární | První diference ($y_{t+1}$ - $y_t$) jsou přibližně konstantní |
| kvadratický | Druhé diference ($y_{t+2}$ - $2y_{t+1}$ + $y_t$) jsou přibližně konstantní |
| exponenciální | Podíly sousedních hodnot ($y_{t+1}/y_t$) resp. první diference logaritmů tvaru ($\log y_{t+1}$ - $\log y_t$) jsou přibližně konstantní |
| logistický | Křivka prvních diferencí ($y_{t+1}$-$y_t$) se podobá křivce normální hustoty, podíly ($1/y_{t+2}$ - $1/y_{t+1}$)/($1/y_{t+1}$ - $1/y_t$) jsou přibližně konstantní |
| Gompertzova křivka | Podíly ($\log y_{t+2} - \log y_{t+1}$)/($\log y_{t+1} - \log y_t$) jsou přibližně konstantní |

## Mechanické vyrovnávání trendu metodami klouzavých průměrů

Používá se v případě, že se trend mění a nelze ho vyrovnat „globálně“ jednou matematickou křivkou. Metoda je vhodná pro neperiodické řady, neumožňuje extrapolaci hodnot.



Vlastní průměry se používají jako **prosté** či **vážené**.
V některých případech lze použít klouzavých mediánů.
Klouzavé průměry mohou být **necentrované** a **centrované**

## Metody klouzavých průměrů

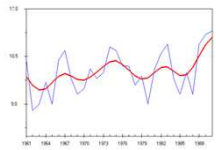
Jako klouzavé průměry obecně označujeme lineární kombinace členů původní řady.

**prosté** $\frac{1}{5}(y_{t-2}+y_{t-1}+y_t+y_{t+1}+y_{t+2})$

**vážené** $\frac{1}{8}(y_{t-2}+2y_{t-1}+2y_t+2y_{t+1}+y_{t+2})$

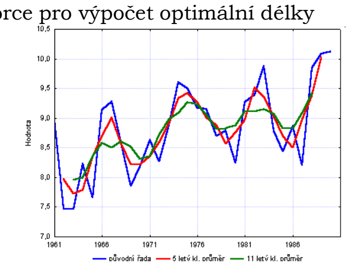
Patří mezi tzv. **adaptivní** přístupy k trendové složce časové řady.

Tzv. **polynomické** klouzavé průměry umožňují vyrovnání hodnot na počátku a konci časové řady



## Volba řádu klouzavých průměrů

- Subjektivní posouzení charakteru dat
- Délka klouzavých průměrů by měla odpovídat periodě sezónních či cyklických fluktuací
- Vzorce pro výpočet optimální délky




Obsahuje-li řada sezónní složku, je vhodné volit řád klouzavých průměrů tak, aby zahrnoval celou délku periody sezónní složky.

## Centrované klouzavé průměry

Ve většině případů se používají klouzavé průměry liché délky, u **sudé délky** je problém s přiřazením hodnot časovému okamžiku.

V ekonomických časových řadách, které často obsahují sezónní složku délky 4 (řady čtvrtletních hodnot) či 12 (řady měsíčních hodnot), se tento problém řeší tzv. **centrováním**.

Výsledné klouzavé průměry pro sudou délku klouzavé části vypočteme jako **průměry dvou sousedních klouzavých průměrů** liché délky.

## Centrované klouzavé průměry

**Příklad:** Abychom vystihli roční chod určitého ukazatele, chceme pro řadu měsíčních hodnot použít klouzavých průměrů délky 12.

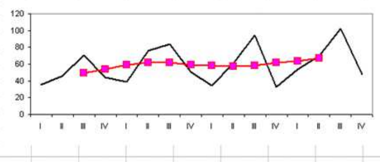
Shlazená hodnota však spadá doprostřed mezi „červen“ a „červenec“. Další shlazená hodnota pak mezi „červenec“ a „srpen“.

Tyto dva jednoduché klouzavé průměry vezmeme a zprůměrňujeme. Výsledek pak už můžeme přiřadit k „červencové“ hodnotě. Tedy vytváříme klouzavé průměry o délce 13:

$$\hat{y}_t = \frac{1}{2}\left(\frac{1}{12}(y_{t-6}+y_{t-5}+\ldots+y_{t+5})+\frac{1}{12}(y_{t-5}+y_{t-4}+\ldots+y_{t+6})\right)=$$
$$\frac{1}{24}(y_{t-6}+2y_{t-5}+2y_{t-4}+\ldots+2y_{t+5}+y_{t+6})$$

## Centrované klouzavé průměry

| | | | | |
|---|---|---|---|---|
| 1998 | I | 35 | | |
| | II | 46 | 49,0 | |
| | III | 71 | 50,0 | **49,5** |
| | IV | 44 | 57,5 | **53,8** |
| 1999 | I | 39 | 60,8 | **59,1** |
| | II | 76 | 62,3 | **61,5** |
| | III | 84 | 61,0 | **61,6** |
| | IV | 50 | 57,3 | **59,1** |
| 2000 | I | 34 | 59,8 | **58,5** |
| | II | 61 | 55,5 | **57,6** |
| | III | 94 | 60,5 | **58,0** |
| | IV | 33 | 62,8 | **61,6** |
| 2001 | I | 54 | 64,8 | **63,8** |
| | II | 70 | 68,5 | **66,6** |
| | III | 102 | | |
| | IV | 48 | | |



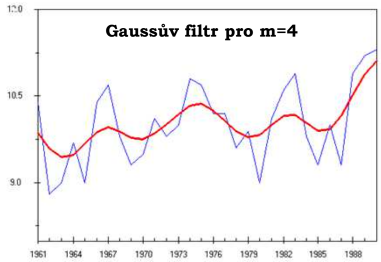
Obecně místo jednoduchých klouzavých průměrů délky 2m vytváříme centrované klouzavé průměry délky 2m+1 podle tohoto obecného vzorce:

$$\hat{y}_t = \frac{1}{4m}(y_{t-m}+2y_{t-m+1}+\ldots+2y_{t+m-1}+y_{t+m})$$

m - polovina řádu klouzavých průměrů (shlazovacího okna)

## Vážené klouzavé průměry

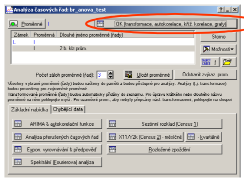
- Jednotlivé členy úseku řady přiřazeny váhy.
- Tyto váhy většinou lineárně klesají směrem od středního (vyrovnávaného) členu.
- Váhy mohou mít také např. podobu tzv. gaussova filtru.

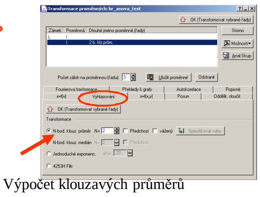



| Člen řady | váha |
|---|---|
| $y_{t-4}$ | 0,014 |
| $y_{t-3}$ | 0,048 |
| $y_{t-2}$ | 0,117 |
| $y_{t-1}$ | 0,201 |
| $y_t$ | 0,241 |
| $y_{t+1}$ | 0,201 |
| $y_{t+2}$ | 0,117 |
| $y_{t+3}$ | 0,048 |
| $y_{t+4}$ | 0,014 |

## Analýza časových řad v programu Statistica

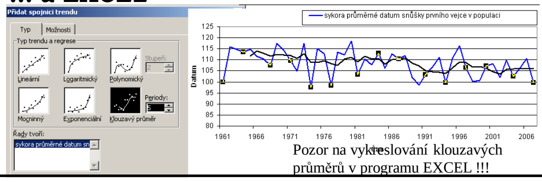
*Statistiky-Pokročilé lineární/ nelineární modely-Časové řady/Predikce*





Výpočet klouzavých průměrů

## … a EXCEL



Pozor na vykreslování klouzavých průměrů v programu EXCEL !!!

## Analýza sezónní složky časových řad (sezónní očišťování)

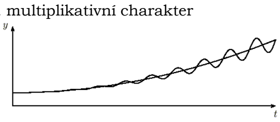
1. klasický přístup k sezónní dekompozici
2. úvod do autokorelační analýzy

**Sezónní složka $S_t$** je typická pro časové řady, jejichž interval pozorování je kratší než jeden rok (sezóna může mít délku týden, měsíc, roční období).

Objevuje se v řadách ekonomických (tržby, produkce, …), ale i v řadách meteorologických prvků (roční chod teploty vzduchu).

Řada obsahující sezónní složku se vyznačuje pravidelným opakováním hodnot kolem trendu a toto opakování může mít délku např. 7 dnů (do týdne), 12 měsíců či 4 roční období (do roku).

Sezónní složka může mít aditivní resp. multiplikativní charakter



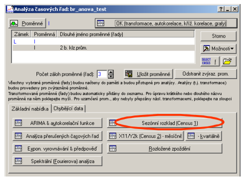
## Obecný model řady při sezónním očišťování

Trendovou a cyklickou složku považujeme za jeden celek:

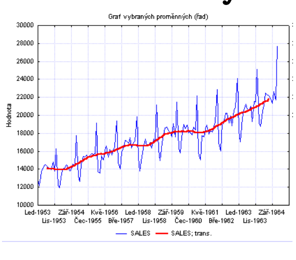
aditivní model: $$Y_t = TC_t + S_t + \varepsilon_t$$

$Y_t$ je pozorovaná hodnota časové řady v čase t.

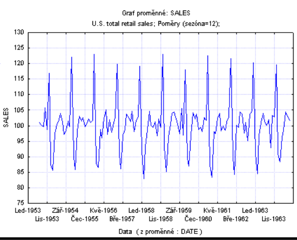
*Statistiky-Pokročilé lineární/ nelineární modely-Časové řady/Predikce*



## Jednotlivé kroky analýzy sezónní složky

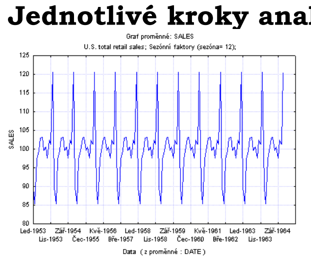


**1.** Z **originální** řady obsahující sezónní složku je vypočtena řada **klouzavých průměrů** s délkou klouzavých průměrů rovnou délce sezónní složky.

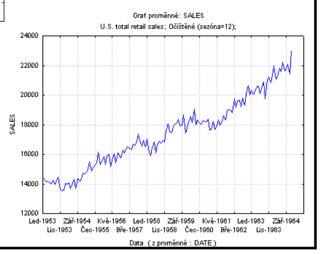


**2.** Vytvoříme novou řadu jako rozdíl **řady původní a řady shlazené**.

## Jednotlivé kroky analýzy sezónní složky

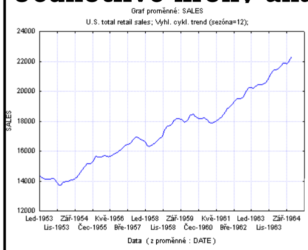


**3.** Tzv. **sezónní komponenty** jsou vypočteny jako průměr pro každý člen v rámci sezóny. Výsledné hodnoty představují průměrnou sezónní složku v časové řadě.

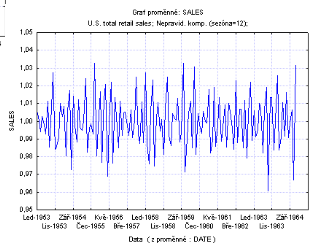


**4. Sezónně očištěná řada** (tedy řada obsahující vedle náhodné složky ještě složku $TC_t$) se potom vyjádří jako rozdíl řady originální a sezónní komponenty.

## Jednotlivé krokv analýzy sezónní složky




**5. Složka $TC_t$** se většinou aproximuje řadou shlazenou váženým klouzavým průměrem délky 5 se symetrickými vahami (1, 2, 3, 2, 1).




**6.** Obdobně lze izolovat **náhodnou složku** jako rozdíl řady sezónně očištěné a řady se zvýrazněnou složkou $TC_t$ ( viz. bod 5).

## Autokorelace časových řad

**Autokorelační analýza** - metoda, kterou lze zkoumat vzájemné vztahy mezi hodnotami jedné časové řady.

Může sloužit jako metoda k definování sezónní a cyklické složky časových řad.

Jejím základem je výpočet **autokorelačního koeficientu**, resp. **autokorelační funkce**.

## Autokorelační koeficient

**Autokorelační koeficient** $r_k$ je relativní míra proměnlivosti členů časové řady posunutých o určitou hodnotu *k*. Definuje vztah mezi členy časové řady $y_t$ a $y_{t+k}$.

Posun *k* se z angličtiny označuje jako **lag**. Je to tedy korelační koeficient vypočtený mezi jednotlivými členy časové řady, mezi kterými je *k-1* jiných pozorování tedy *lag = k* a označujeme ho jako autokorelační koeficient k-tého řádu.

Pro $k = 0$ je hodnota $r_0 = 1$ - je to vlastně hodnota korelačního koeficientu.

## Základní pojmy

**Rozptyl (variance)** – míra variability (proměnlivosti) statistického znaku *x*

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

**Kovariance** – absolutní míra vzájemné variability dvou statistických znaků *x; y*

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

**Korelace** - relativní míra vzájemné variability dvou statistických znaků *x; y*

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

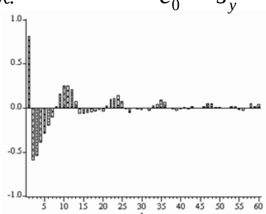
## Základní vztahy

**Autokovariance** – absolutní míra proměnlivosti členů časové řady *y* posunutých o určitou hodnotu *k*.

$$c_k = \frac{\sum_{i=1}^{n-k}(y_i - \bar{y})(y_{i+k} - \bar{y})}{n-k-1}$$

**Autokorelace** – relativní míra proměnlivosti členů časové řady *y* posunutých o určitou hodnotu *k*.

$$r_y(k) = \frac{c_k}{c_0} = \frac{c_k}{s_y^2}$$



**Autokorelační funkce** – hodnoty $r_y(k)$ pro $k=1,2,\dots M$, kde $M < N/2$, $N$ – délka řady

## Autokorelační funkce

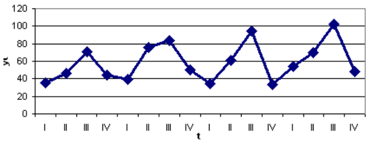
**Autokorelační funkce** (ACF) je potom závislost mezi hodnotami autokorelačního koeficientu a hodnotami posunu k.

Vyjadřuje se formou grafu – tzv. **korelogramu** (viz. obrázek). Na ose x jsou hodnoty lag (k), na ose y hodnoty autokorelačního koeficientu.

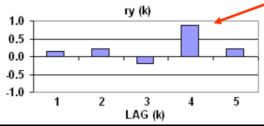
Hodnoty autokorelační funkce se pohybují v intervalu –1,1.

ACF je vhodným nástrojem k posouzení, zda časová řada obsahuje cyklickou či periodickou složku a také zda je či není řadou náhodných čísel – tedy do jaké míry je možné ji extrapolovat (předpovídat).

## Princip autokorelace časové řady




| | t1 | t2 | t3 | t4 | t5 | t6 | t7 | t8 | t9 | t10 | t11 | t12 | t13 | t14 | t15 | t16 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| t | I | II | III | IV | I | II | III | IV | I | II | III | IV | I | II | III | IV | | | | | | LAG (k) | ry (k) |
| yt | 35 | 46 | 71 | 44 | 39 | 76 | 84 | 50 | 34 | 61 | 94 | 33 | 54 | 70 | 102 | 48 | | | | | | 0 | **1.00** |
| | | 35 | 46 | 71 | 44 | 39 | 76 | 84 | 50 | 34 | 61 | 94 | 33 | 54 | 70 | 102 | 48 | | | | | 1 | **0.14** |
| | | | 35 | 46 | 71 | 44 | 39 | 76 | 84 | 50 | 34 | 61 | 94 | 33 | 54 | 70 | 102 | 48 | | | | 2 | **0.23** |
| | | | | 35 | 46 | 71 | 44 | 39 | 76 | 84 | 50 | 34 | 61 | 94 | 33 | 54 | 70 | 102 | 48 | | | 3 | **-0.19** |
| | | | | | 35 | 46 | 71 | 44 | 39 | 76 | 84 | 50 | 34 | 61 | 94 | 33 | 54 | 70 | 102 | 48 | | 4 | **0.89** |
| | | | | | | 35 | 46 | 71 | 44 | 39 | 76 | 84 | 50 | 34 | 61 | 94 | 33 | 54 | 70 | 102 | 48 | 5 | **0.21** |
| | | | | | | | ... | | | | | | | | | | | | | | | ..... | |



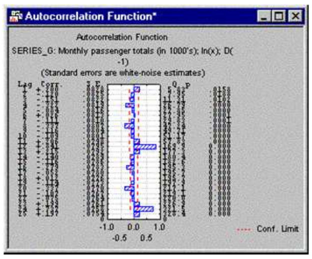

## Interpretace ACF I

**Korelogram bývá doplňován intervaly spolehlivosti, kterými lze hodnotit statistickou významnost autokorelačních koeficientů.**

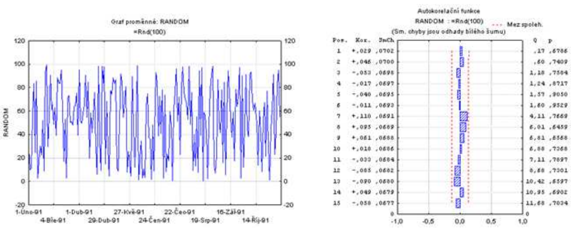
**95 % interval spolehlivosti ACF lze z dostatečnou přesností zkonstruovat ze vztahu:**

$$\pm 2 / \sqrt{N}$$

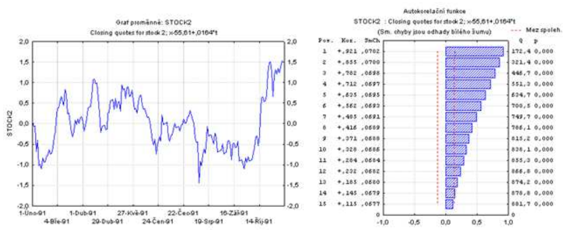
***N* – délka časové řady**




### Časová řada náhodných čísel (bílý šum) a její autokorelační funkce




### Časová řada bez periodické složky se silnou autokorelací a její autokorelační funkce




### Časová řada obsahující výraznou sezónní složku a její autokorelační funkce