

Kvantitativní přístupy v regionální analýze

5RE301 – Regionální analýza
Rozšířená verze k základní prezentaci



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Literatura

- Hindls, R., Hronová, S., Seger, J. a Fischer, J. Statistika pro ekonomy. Praha: Professional Publishing, 2011, 420 s.
- Marek, L. a kol. Statistika pro ekonomy: aplikace. Praha: Professional Publishing, 2007, 485 s.
- Krkošková, Š., Ráčková, A., Zouhar, J. Základy ekonometrie v příkladech. Praha: Oeconomica, 2010, 276 s.

Tři pohledy na regionální data

- Zkoumáme více regionů ve stejném okamžiku – prostorová analýza – prostorová neboli průřezová data (cross-sectional)
- Zkoumáme jeden region během časového období – časová analýza – časová data (time series)
- Kombinace obou – analýza panelových dat (panel data)

Zkoumání závislosti ukazatelů

- Zkoumáme závislost číselných ukazatelů
- Testujeme vždy nezávislost
- Při zamítnutí testované hypotézy existuje závislost

- Regresní analýza – y, x – jednostranná
- Analýza časové řady – y, t

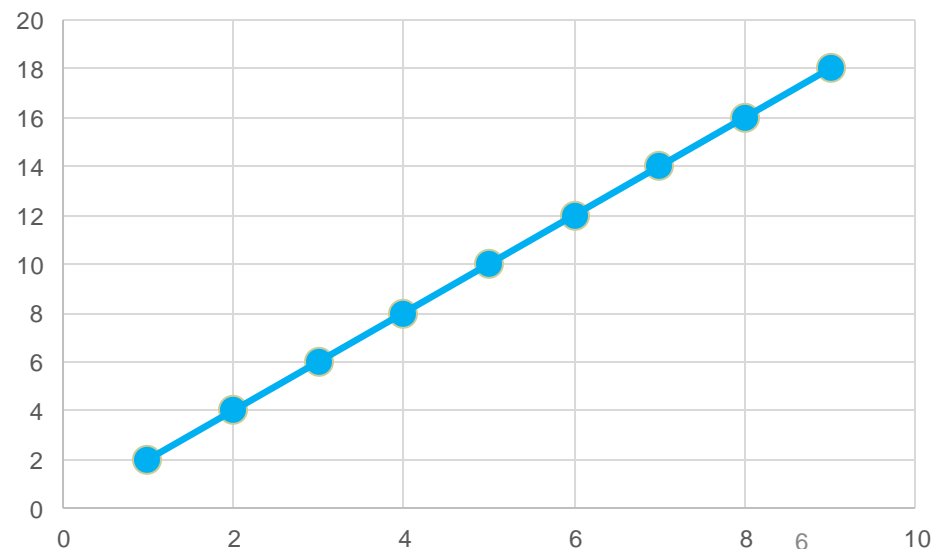
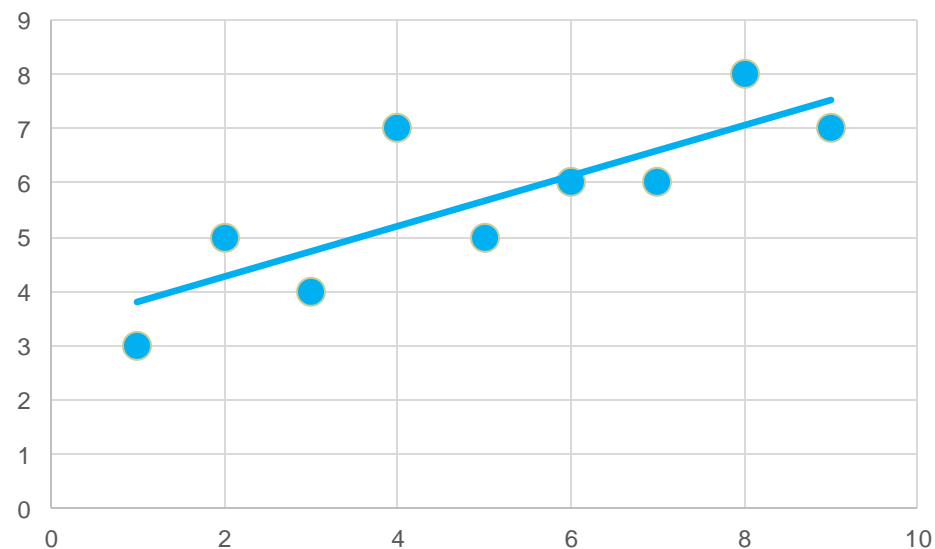
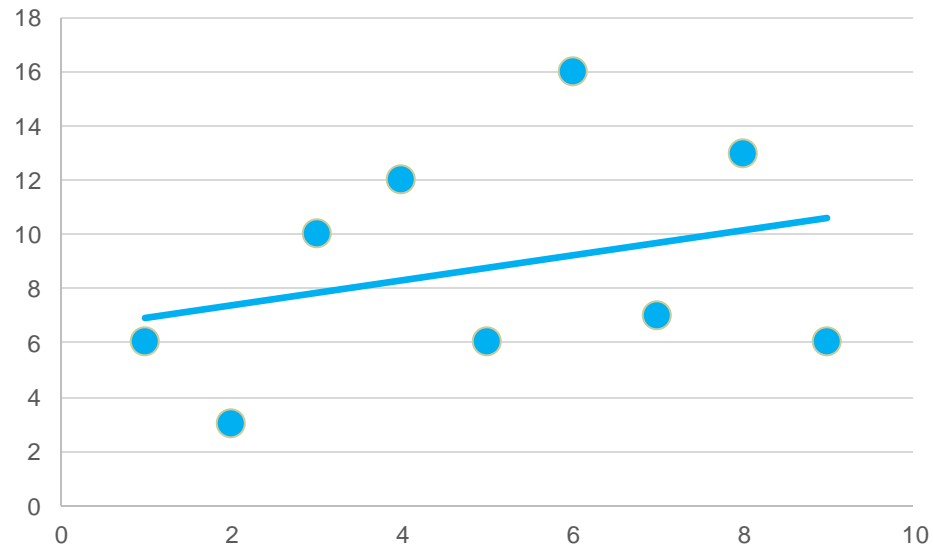
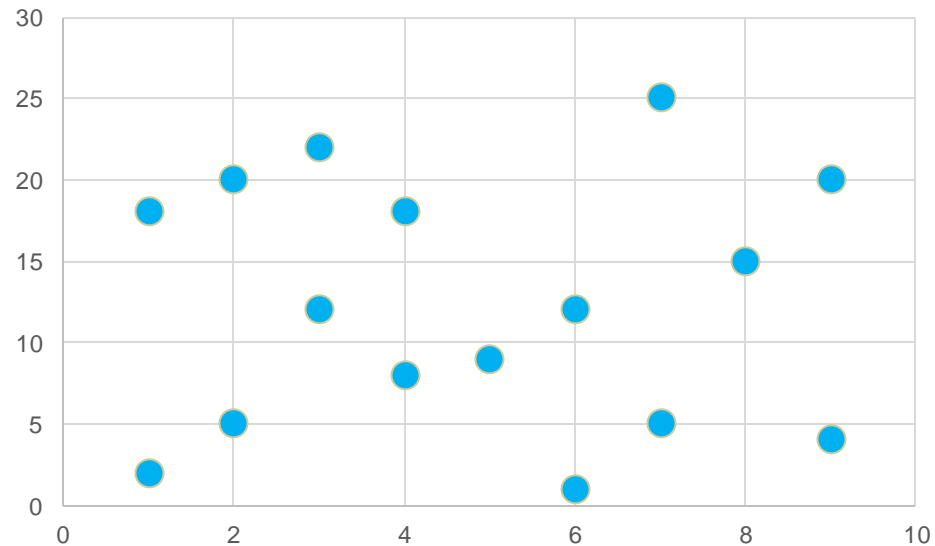
Závislost číselných proměnných

- Deterministická (funkční, pevná, pevně určená)
- Stochastická (statistická, pravděpodobnostní)

- Různě silná závislost

- Statistický pohled
- Ekonometrický pohled

Různá síla závislosti



Regresní analýza

- Podle počtu nezávislých proměnných
 - Jednoduchá (simple regression)
 - Vícenásobná (multiple regression)
- Proměnné (obvykle y a x):
 - Závisle x nezávisle proměnná
 - Vysvětlovaná x vysvětlující proměnná
 - Endogenní x exogenní proměnná

Postup v regresní analýze

- Formulovat apriorní představy o charakteru regresní funkce – hypotéza + tvar rovnice

$$y = f(x)$$

- Odhadnout parametry na základě dat
- Posoudit kvalitu – verifikace statistická a ekonomická

Přímková regrese

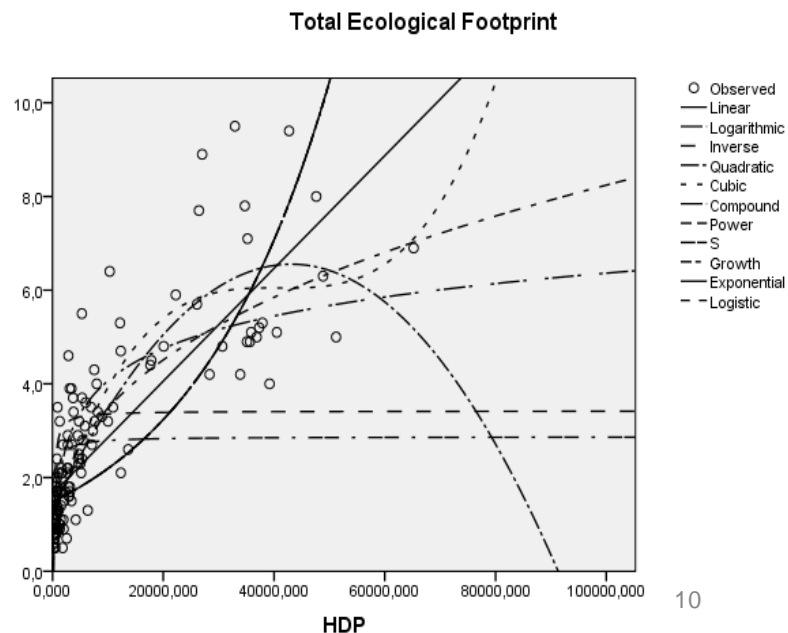
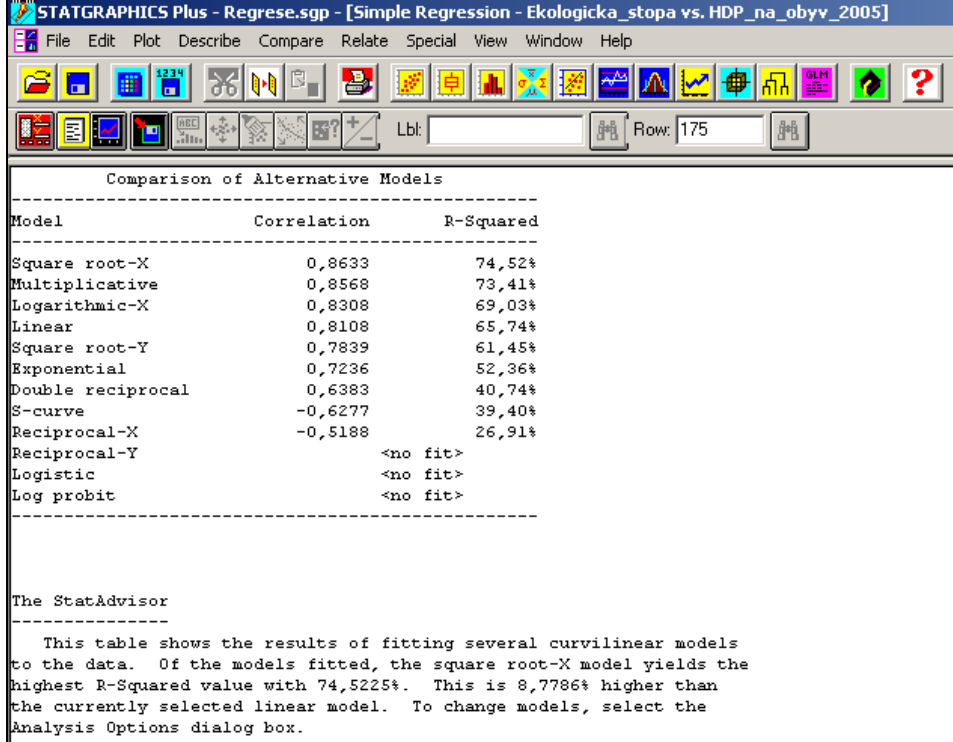
- Nejjednodušší
- Předpokládáme lineární vztah

$$y_i = ax_i + b + e_i$$

- Proměnné – endogenní (y) a exogenní (x_i)
 - Známe, y náhodné, x nenáhodné (pod kontrolou)
- Regresní parametry (a, b)
 - Odhadujeme v modelu
- Rezidua (e_i)
 - „Zbytek“, který minimalizujeme

Regresní funkce

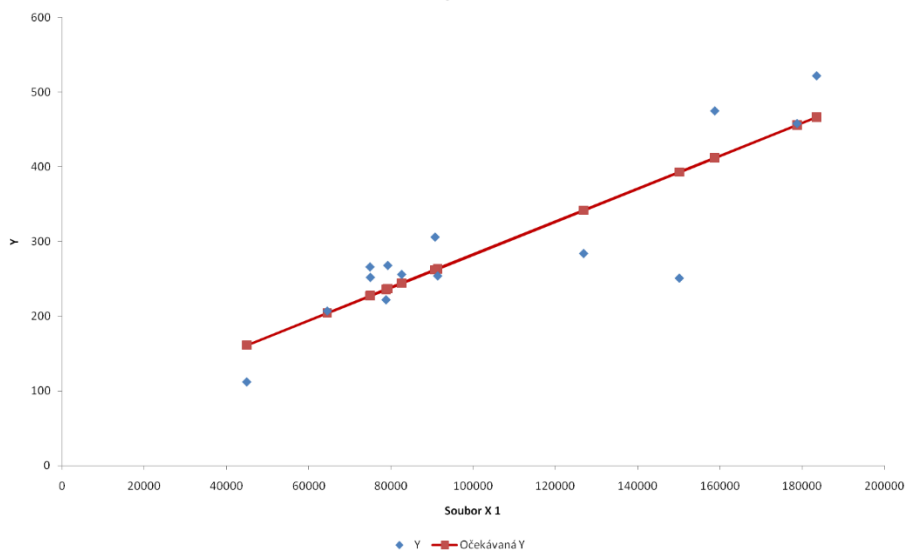
- Lineární funkce
 - Přímková
 - Kvadratická
 - Vícenásobná (rovina)
 - Hyperbolická
 - Logaritmická
- Nelineární funkce
 - Exponenciální
 - Mocninná
 - Zobecněná exponenciální
 - Logistická



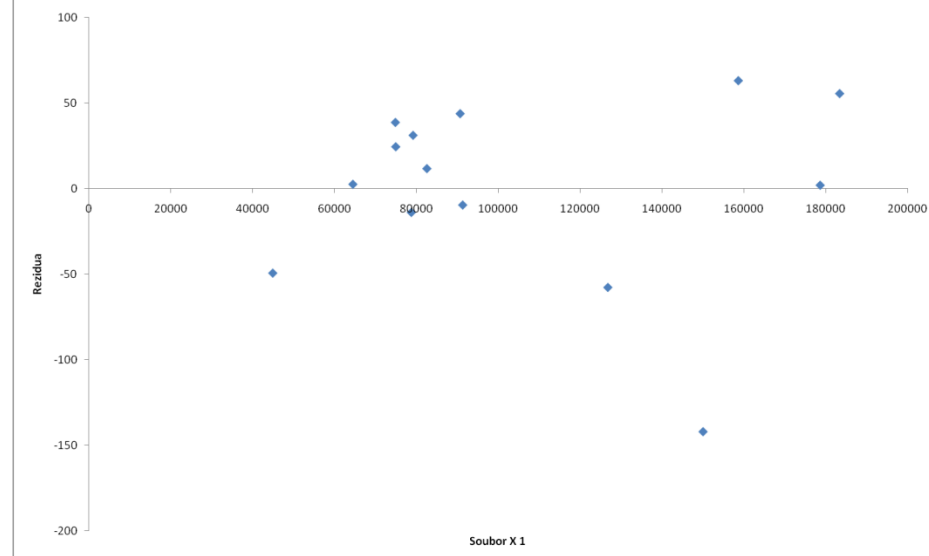
Metoda nejmenších čtverců (MNČ)

- Ordinary Least Squares (OLS)
- Odhad průběhu regresní funkce
- Cílem je minimalizace součtu čtverců reziduí

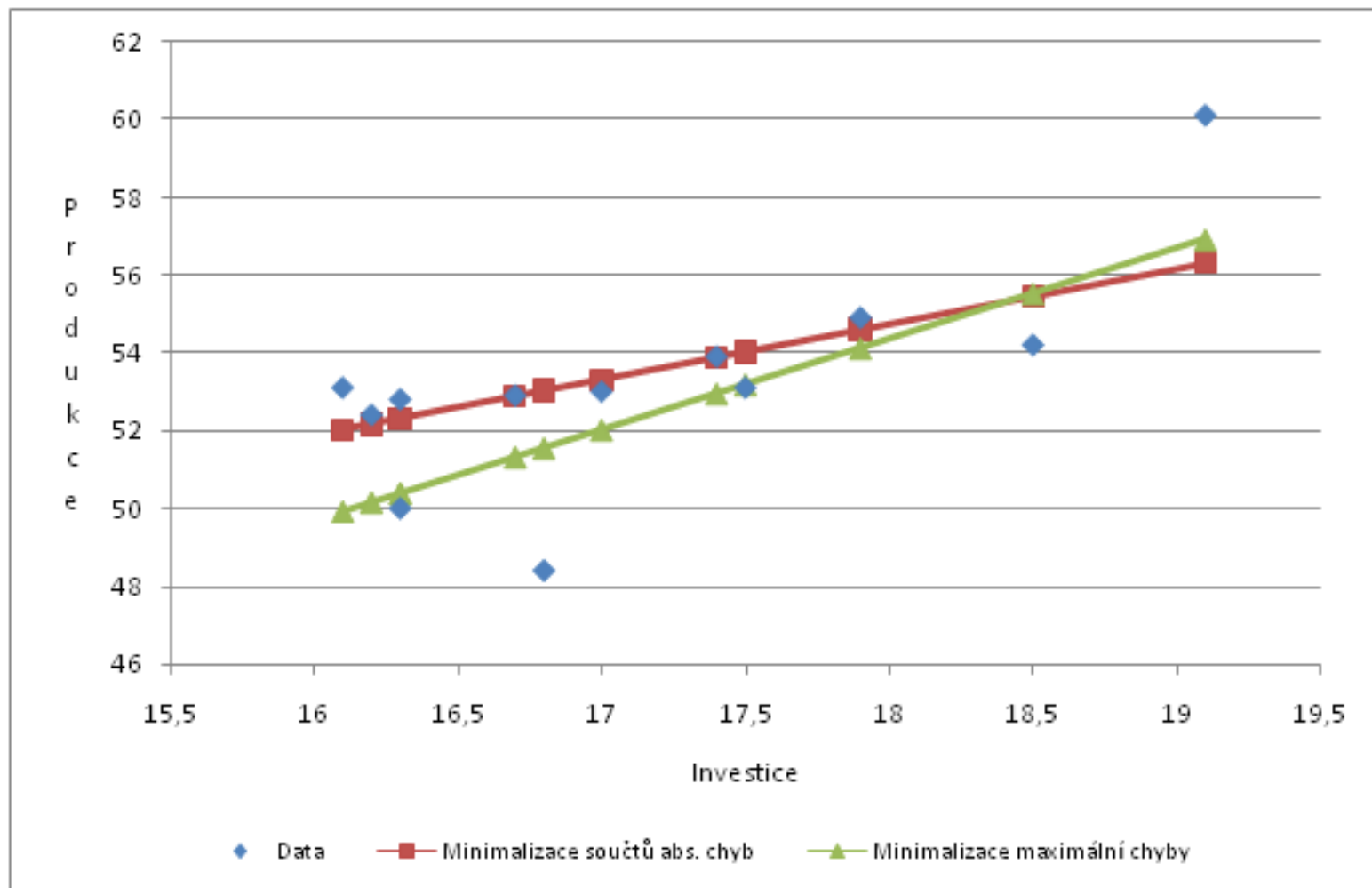
Soubor X 1 Graf porovnání hodnot



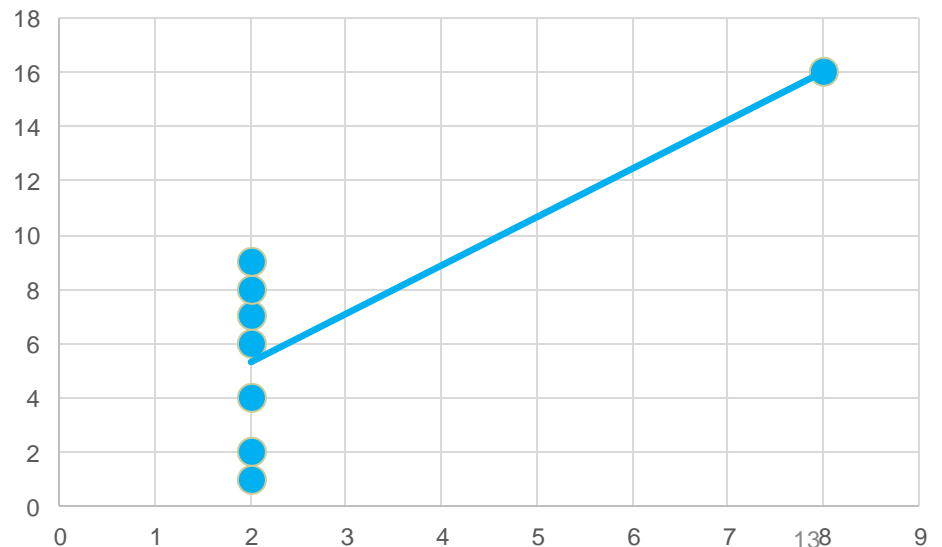
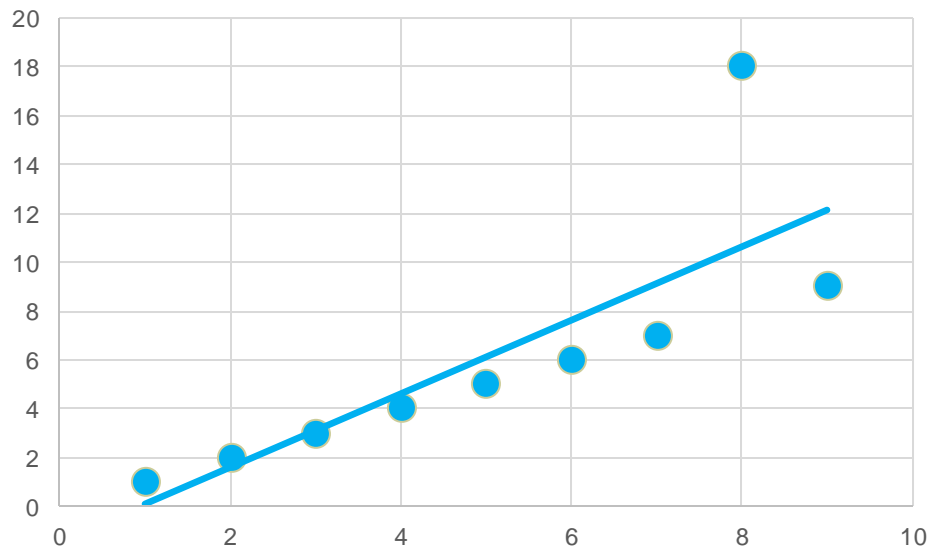
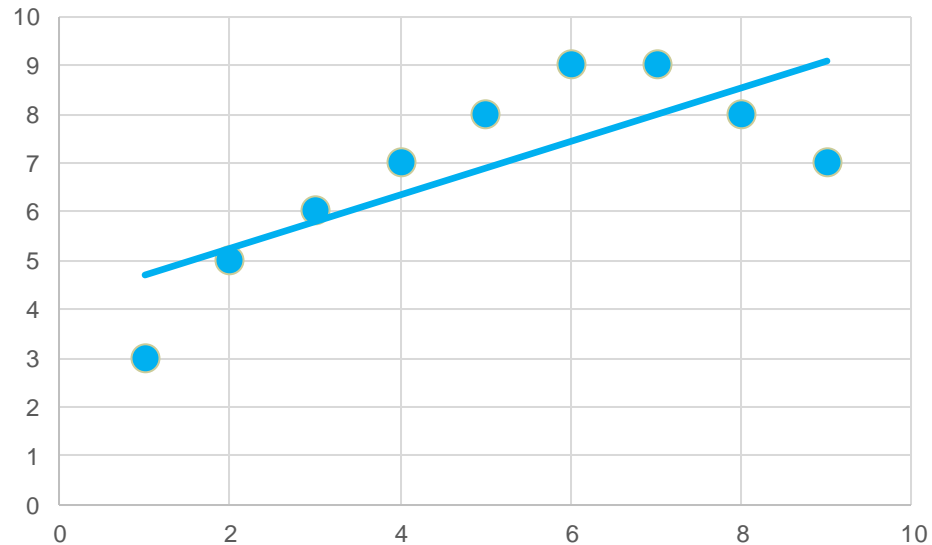
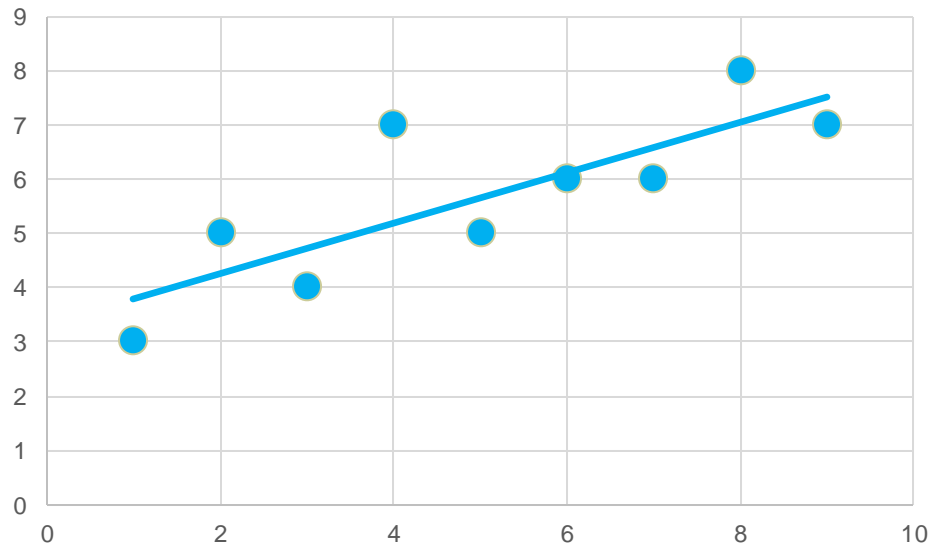
Soubor X 1 Graf s rezidui



Neexistuje ale jen MNČ



Odlehlé hodnoty, nevhodná funkce



Významnost parametrů

- Celkový F-test – model má/nemá smysl

$$F = \frac{\frac{S_T}{k-1}}{\frac{S_R}{n-k}} \quad F_{1-\alpha}(k-1; n-k)$$
$$H_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0$$

- Individuální t-testy – parametr je/není významný

$$t = \frac{b_i}{s_{b_i}} \quad t_{1-\alpha/2}(n-k)$$
$$H_0 : \beta_i = 0$$

Statistická verifikace

- Kolik procent závislosti vysvětluje model?

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

- Koeficient determinace s rostoucím počtem vysvětlujících proměnných roste
- Upravený koeficient vícenásobné determinace

$$\overline{R^2} = 1 - (1 - R^2) \frac{n - 1}{n - (k + 1)}$$

Vícenásobná regrese

- Závislost y na více x_i
- Gaussovy-Markovovy předpoklady
- Porušení předpokladů – problémy:
 - Multikolinearita
 - Autokorelace
 - Heteroskedasticita

Multikolinearita (1)

- Netýká se populace, ale konkrétního výběru
- Silná závislost mezi vysvětlujícími proměnnými
 - Lineární vztah
 - Lineární závislost
- Při opakovaných výběrech se parametry značně liší

Multikolinearita (2)

- Problémem může být nedostatečný výběr, který zkresluje odhady a má za následek velké rozptyly chyb
- Jak zjistit? => Skripta.
- Některá možná řešení:
 - Zvětšit rozsah výběru
 - Odstranit z modelu proměnnou, která multikolinearitu způsobuje

Multikolinearita (3)

- Postupná regrese (Stepwise regression)
 - Forward (dopředná) – postupné přidávání proměnných do modelu
 - Backward (zpětná) – postupná eliminace proměnných z modelu

Praktická část – výzkum a vývoj

- Patenty, VTP, výzkum a vývoj, vysoké školství

	Patentove_prihlasky	VTP	Vydaje_na_VaV	Zamestnanci_VaV	VS_celkem	Platne_patenty	Udelene_patenty
Středočeský	71	5	9782	5894	3	186	25
Jihočeský	38	4	1967	2692	2	54	9
Plzeňský	25	2	1767	2376	2	101	12
Karlovarský	3	0	98	160	1	23	4
Ústecký	13	3	808	1277	1	76	10
Liberecký	43	1	1517	1727	1	123	8
Královéhradecký	25	2	1213	2656	1	92	18
Pardubický	24	1	2002	3012	1	102	25
Vysočina	18	1	698	931	2	27	5
Jihomoravský	78	6	6047	12277	10	198	22
Olomoucký	26	2	1433	3108	3	82	2
Zlínský	48	4	1633	2362	2	59	9
Moravskoslezský	57	7	2661	5395	4	180	16

Zdroje: ČSÚ

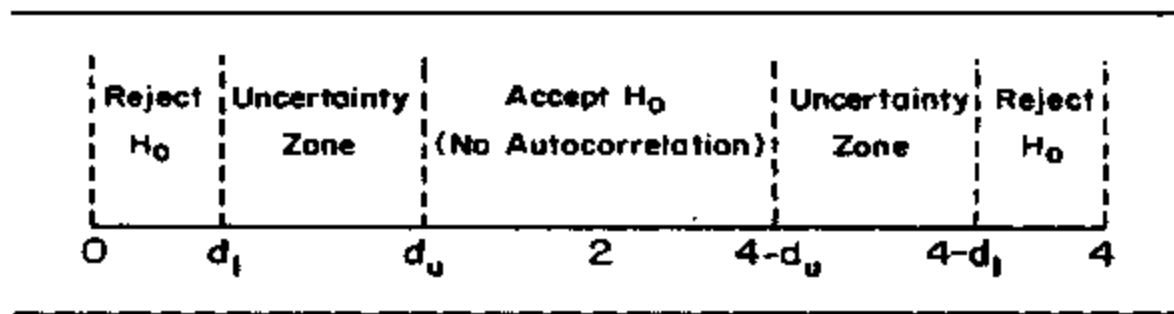
Autokorelace (1)

- Existuje závislost mezi náhodnými složkami
- Závislost mezi různými hodnotami jedné proměnné
- Prostorová autokorelace – průřezová data
- Častěji u časových řad
- Časové řady – stejné trendové chování
- Čím delší časová řada, tím lépe
- Pozitivní / negativní autokorelace

Autokorelace (2)

- Na grafu rezidua tvoří systém
- Důležitá proměnná není do modelu zahrnuta
- Durbin-Watsonova statistika
 - Nevhodná pro málo pozorování nebo mnoho x_i

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$



Zdroj:

<http://userwww.sfsu.edu/~efc/classes/biol710/timeseries/TimeSeriesAnalysis.html>

Autokorelace (3)

- Durbin-Watsonova statistika je tabelována pro různé rozsahy souborů, hladiny významnosti a počty regresorů

Durbin-Watson Statistic: 5 Per Cent Significance Points of dL and dU

n	k'=1		k'=2		k'=3		k'=4		k'=5		k'=6		k'=7		k'=8		k'=9		k'=10		
	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	
6	0.610	1.400	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0.700	1.356	0.467	1.896	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0.763	1.332	0.559	1.777	0.367	2.287	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	-----	-----	-----	-----	-----	-----	-----	-----	-----
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	-----	-----	-----	-----	-----	-----	-----
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	-----	-----	-----	-----	-----
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	-----	-----	-----
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438	-----
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304	-----

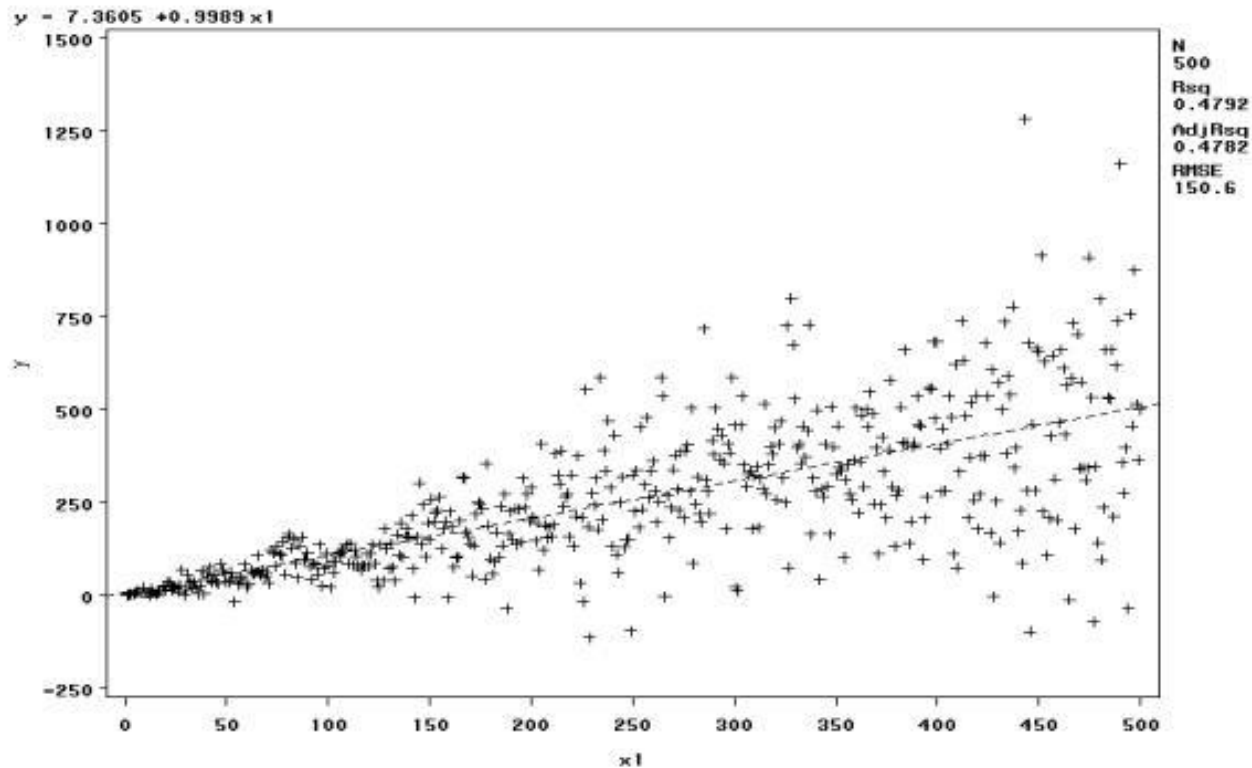
Zdroj tabulek: www.nd.edu/~wevans1/econ30331/Durbin_Watson_tables.pdf

Heteroskedasticita (1)

- Je porušena podmínka konečného a konstantního rozptylu náhodných složek
- Různý rozptyl pro skupiny pozorování
- Častěji u průřezových dat
 - Malé, střední, velké firmy
 - Lidé s nízkými / vysokými příjmy

Heteroskedasticita (2)

- Grafické testování:



Zdroj: http://economics.fundamentalfinance.com/econ_img/b_heterosked.jpg

Heteroskedasticita (3)

- Neparametrické testy
 - Zkoumáme pořadí X_i a e_i
 - Rozptyl ve dvou různých skupinách
- Parametrické testy
 - Pomocné regrese
- Více skripta Základy ekonometrie v příkladech

Analýza časových řad

- Zkoumání závislosti daného ukazatele na časové proměnné
- Průběh ukazatele v čase
- Různé klasifikace časových řad:
 - Podle typu ukazatele, který se sleduje
 - Podle délky intervalu sledování hodnot
 - Podle průběhu hodnot časové řady

Časové řady podle typu ukazatele

- **Intervalová** časová řada

- řada ukazatele, jehož hodnoty závisí na délce časového intervalu sledování
- např. objem výroby, spotřeba surovin, HDP atd.

- **Okamžiková** časová řada

- řada ukazatele, který se vztahuje k jistému časovému okamžiku
- např. počet pracovníků jisté firmy k určitému datu, peněžní zásoba

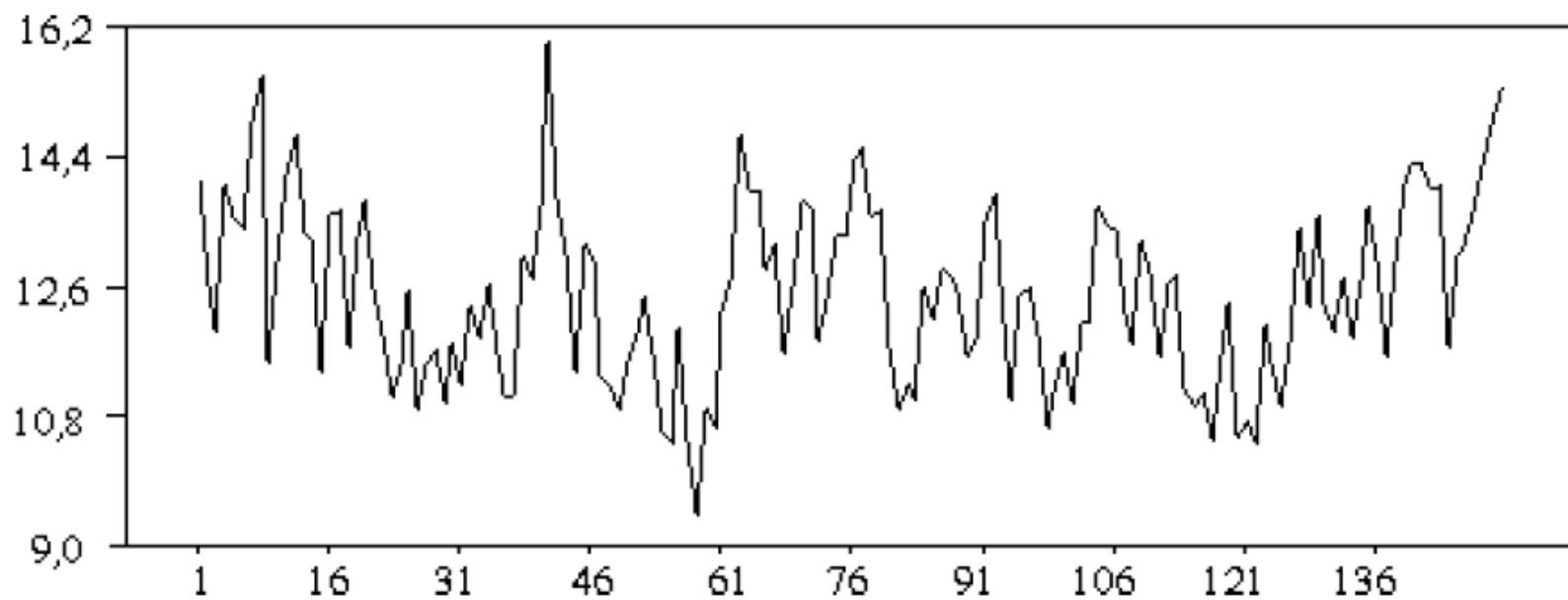
Časové řady podle délky intervalu

- **Dlouhodobé** časové řady
 - hodnoty sledované v ročních či delších časových úsecích
- **Krátkodobé** časové řady
 - hodnoty sledované v úsecích kratších než jeden rok (např. měsíční a čtvrtletní časové řady), zpravidla obsahují sezónní složku
- **Vysokofrekvenční** časové řady
 - kratší než 1 týden (hodinové časové řady, denní časové řady)

Časové řady podle průběhu hodnot

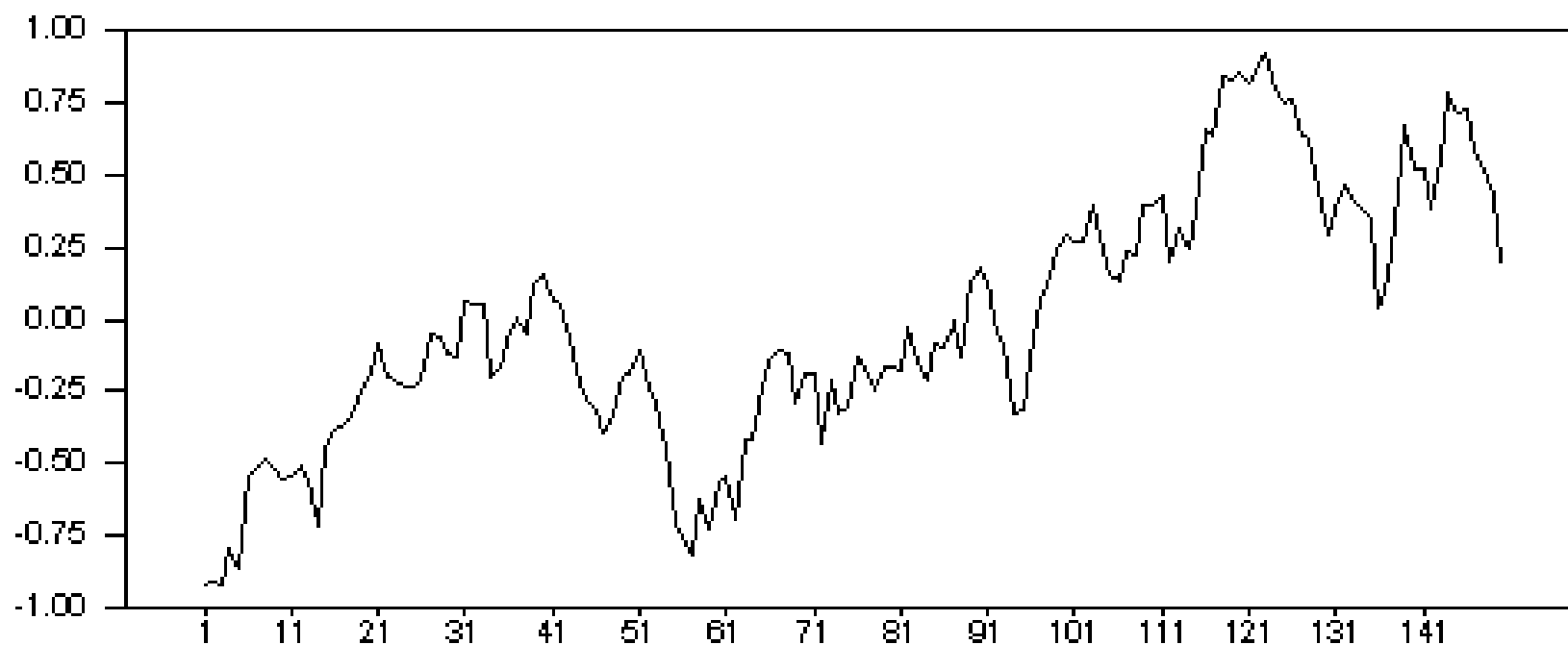
- **Stacionární časové řady**
 - časové řady s krátkou pamětí (vliv šoku z minulého období časem postupně vytrácí)
- **Nestacionární časové řady**
 - řady s dlouhou pamětí (vliv šoku z minulého období se projevuje neustále)
- Stacionarita – ustálenost v čase (invariance), vyloučení vlivu času
- V ekonomické realitě je většina řad nestacionárních

Stacionární časová řada



Zdroj: http://nb.vse.cz/~arlt/publik/A_CRTI011_98.pdf

Nestacionární časová řada



Zdroj: http://nb.vse.cz/~arlt/publik/A_CRTI011_98.pdf

Dekompozice časových řad (1)

- Čtyři složky časových řad
 - Trendová složka
 - Sezónní složka
 - Cyklická složka
 - Náhodná složka
- Nemusí být v časové řadě přítomné
- Aditivní model
- Multiplikační model

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

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

Dekompozice časových řad (2)

- Trendová složka
 - zachycuje dlouhodobé změny v chování časové řady, tedy dlouhodobý růst či dlouhodobý pokles
- Sezónní složka
 - popisuje periodické změny v časové řadě, které se odehrávají v rámci jednoho kalendářního roku a každý rok se opakují

Dekompozice časových řad (2)

- Cyklická složka
 - popisuje dlouhodobé fluktuace kolem trendu
 - zachycuje dlouhodobou fázi poklesu či růstu, která je mnohem delší než jeden rok
- Náhodná složka
 - je tvořena náhodnými výkyvy časové řady
 - pod tuto složku zařazujeme všechny vlivy, které na časovou řadu působí a které nedokážeme systematicky podchytit a popsat



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



Národohospodářská fakulta VŠE v Praze



This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.