



MATEMATICKÁ STATISTIKA II.

P14
2007-01-08

METODY KONSTRUKCE PŘEDPOVĚDÍ ČASOVÝCH ŘAD:

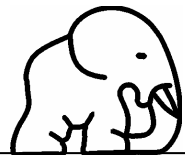
- ✓ Při analýze ČŘ se setkáváme s případy, kdy je třeba doplnit chybějící údaj v ČŘ nebo provést korekci hodnoty daného ukazatele v ČŘ.

Interpolace:

- ✓ **Interpolací** rozumíme přibližné určení chybějící hodnoty sledovaného ukazatele ČŘ za předpokladu, že známe jeho sousední hodnoty.
- ✓ Interpolaci je možné v podstatě provést dvěma způsoby a to:
 - Prostřednictvím použití dvou sousedních hodnot ČŘ:
 - Pokud předpokládáme, že absolutní roční přírůstek je poměrně stálý, můžeme dopočítat chybějící hodnotu jako jednoduchý aritmetický průměr dvou sousedních hodnot.
 - Pokud vyjdeme z předpokladu, že se údaje za sledované období měnily rovnoměrně, při poměrně stálém koeficientu růstu, je možné chybějící hodnotu určit jako součin předcházející hodnoty ČŘ a průměrného koeficientu růstu.
 - Prostřednictvím využití více či všech hodnot ČŘ, kdy pomocí metody nejmenších čtverců určíme parametry trendové funkce, ze kterých potom odhadneme chybějící údaj.

Extrapolace:

- ✓ Neodmyslitelnou součástí analýzy časových řad představuje konstrukce předpovědí budoucího vývoje zkoumaného ukazatele, tzn. **extrapolace** hodnot ČŘ.
- ✓ Pod extrapolací rozumíme určení hodnot ČŘ za interval známých hodnot ČŘ.
- ✓ Extrapolaci je možné provést jak ve směru budoucím (perspektivní extrapolace), tak ve směru minulého vývoje (retrospektivní extrapolace).
- ✓ Stejně jako interpolace, tak i extrapolace se musí uskutečňovat jen v časovém horizontu, ve kterém působí stejné zákonitosti a podmínky vývoje.
- ✓ Metod, jak tyto předpovědi stanovit, lze nalézt poměrně mnoho - od metod regresní analýzy přes strukturní analýzu a metody založené na systémovém přístupu až k metodám extrapolace časových řad. Statistická praxe nejčastěji používá metody extrapolace časových řad.
- ✓ Podstata klasických extrapolačních metod spočívá v tom, že se studuje historie prognózovaného objektu a zákonitosti jeho vývoje v minulosti a přítomnosti se přenesou do budoucnosti (vychází z principu, že budoucnost vyplývá z přítomnosti).
- ✓ Jsou konstruovány na základě předpokladu o neměnnosti nebo alespoň relativní stability existujících tendencí vývoje zkoumaného jevu. U procesů, které se v čase stabilně vyvíjí, lze tento princip s úspěchem při konstrukci předpovědí aplikovat.
- ✓ Naopak v případě, že během prognózovaného období probíhají podstatné kvalitativní změny, je použití extrapolačních modelů velmi problematické.
- ✓ Přesto však má extrapolace klasických modelů vývojových tendencí nesporný poznávací význam a řadu předností:
 - Při konstrukci se používá relativně jednoduchý matematicko-statistický aparát, který umožňuje poměrně snadnou algoritmizaci použitých metod.
 - Informační materiál, potřebný pro sestavení analyzovaných časových řad, je většinou dosažitelný (prognózovaná veličina se uvádí jako závisle proměnná, za nezávisle proměnnou se přejímá čas) - k analýze a prognóze tedy stačí informace o vývoji analyzovaného jevu v minulosti.
 - Vlastní sestavení předpovědi extrapolací je poměrně rychlé a jednoduché.
 - Nezanedbatelnou přednost představuje rovněž skutečnost, že při konstrukci extrapolačních předpovědí není nutné uskutečňovat prognózy dalších jevů, které vysvětlují extrapolovaný jev.
- ✓ Přes uvedené přednosti nejsou extrapolační metody použitelné univerzálně. Jejich slabé místo představuje již dříve zmíněný předpoklad o neměnnosti dosavadních vývojových tendencí prognózovaného jevu. K tomuto slabému místu přistupují i některé další nedostatky:
 - Extrapolační prognostické metody neposkytují systémové prognózy, každý jev se posuzuje izolovaně.
 - Kvalitu analýzy a prognózy ovlivňuje v rozhodující míře zvolený typ modelu.



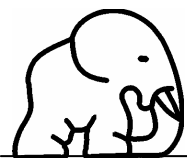
- ✓ Při hodnocení užitečnosti konstrukce extrapolčních prognóz je třeba si uvědomit, že takto získaná předpověď nepředstavuje izolovaný základ pro jakékoliv rozhodnutí, ale měla by být porovnána s předpověďmi získanými jinými prognostickými metodami. Teprve ze vzájemného porovnání prognóz získaných různými metodami vyplyne globální prognóza sledovaného jevu.
- ✓ Nejčastěji se extrapolace uskutečňuje za účelem vyjádření dalšího pravděpodobného vývoje zkoumaného ukazatele. Za předpokladu, že ve složení faktorů ovlivňujících vývoj sledovaného ukazatele a v intenzitě jejich působení nenastanou výraznější změny, je možné predikovat vývoj pro nejbližší období na základě vypočtené trendové funkce, kdy konkrétní způsob extrapolace závisí od charakteru ČR (řada má či nemá trend, periodické kolísání apod.).
- ✓ Při řešení praktických úloh se snažíme získat prognózu, která by byla únosná, přijatelná a dostatečně přesná. Jde tedy o to, sestavit předpověď, která by vedla k přijatelně nízké odchylce prognózy od skutečnosti.
- ✓ Po provedení odhadu je možné stanovit tzv. chybu „ex post“, kdy vlastně porovnáváme veličinu y_{t+i} s jejím extrapolčním odhadem P_{t+i} . $\Delta = P_{t+i} - y_{t+i}$
- ✓ Zkonstruovaná předpověď bude tím přesnější, čím bude tato chyba předpovědi menší.
- ✓ Chybu předpovědi lze rozložit na dvě složky $\Delta = \Delta_m - \Delta_p$, $\Delta_m = P_{t+i} + Y_{t+i}$.
- ✓ Toto je chyba „ex ante“, kterou lze nazvat **modelovou chybou předpovědi**.
- ✓ Jejím zdrojem je skutečnost, že v časovém okamžiku t (tj. v současnosti), kdy prognózu o i časových jednotek dopředu zkonstruujeme, nevíme, jaký vývojový mechanismus bude chování předvídané veličiny v budoucnu utvářet a do jaké míry se bude tento „nový“ pohyb lišit od pohybu, který jsme charakterizovali v období interpolace určitým teoretickým modelem.
- ✓ Y_{t+i} proto představuje předpověď teoretické veličiny, která by se realizovala za předpokladu, že ČR sledovaného ukazatele by nezměnila v období předpovědi své chování proti období interpolace.
- ✓ Druhou složkou chyby ex post je chyba $\Delta_p = Y_{t+i} - y_{t+i}$, která je **chybou vlastního prognostického modelu**, které se dopustíme tím, že použitý prognostický model se bude odlišovat od skutečné veličiny.
- ✓ Protože prakticky nelze získat bezchybnou předpověď, tj. dospět k ideálnímu stavu, kdy $\Delta_m = 0$ (prognostický model obsahuje náhodnou složku a neznámé parametry), je třeba na chybu Δ_m pohlížet jako na náhodnou veličinu, jejíž velikost lze regulovat pouze pravděpodobnostně.
- ✓ Pokud máme ČR bez trendu a bez periodického kolísání, provedeme extrapolaci přímo na základě **průměru hodnot**.
- ✓ V případě neperiodických ČR s trendem používáme pro extrapolaci vypočtenou **trendovou funkci**.
- ✓ Při tomto způsobu se předpověď stanoví na základě dosavadního vývoje, kdy se vlastně provede v časovém okamžiku t odhad neznámé veličiny y pro čas i , kde $i > 0$ je zadaný horizont předpovědi.
- ✓ U periodických ČR se postupuje obdobně, jenom je potřeba danou předpověď **upravit příslušným průměrným sezónním indexem**.
- ✓ Základem správné prognózy je vhodná volba analytické funkce, charakterizující dosavadní vývoj zkoumaného ukazatele. ČR však musí být přitom přiměřeně dlouhá, aby odrážela základní vývojovou tendenci, která se nezakládá jen na krátkodobých výkyvech a odchylkách.
- ✓ Kromě toho by měla ČR vykazovat jednoznačný trend, který je možné vyjádřit co nejjednodušší funkcí (funkcí s co nejmenším počtem parametrů).
- ✓ Trendovou funkci lze použít na prognózy krátkodobé, u dlouhodobějších prognóz je postup složitější a nevystačíme u nich jen se znalostí průběhu trendové funkce.
- ✓ Musíme si však uvědomit, že v každé ČR existuje i náhodná složka. Proto je potřeba každý prognostický odhad doplnit také charakteristikou intenzity náhodného (nepravidelného) kolísání.
- ✓ Z velikosti těchto charakteristik je možné posuzovat i velikost chyby odhadu.
- ✓ Kvalitu vyrovnání a tím i extrapolace nejčastěji posuzujeme pomocí **reziduální odchylky**.

- ✓ Reziduální odchylku lze stanovit následujícím výrazem $s_{y_t} = \sqrt{\frac{\sum_{t=1}^n (y_t - y'_t)^2}{n - p - 1}}$, případně je možné určit také

variační koeficient $v_{y_t} = \frac{s_{y_t}}{\bar{y}} \cdot 100$, kde n – délka časové řady, p – počet parametrů trendové funkce, y_t –

skutečné hodnoty ČR, y'_t – teoretické (vyrovnané) hodnoty ČR, s_{y_t} – směrodatná odchylka hodnot y_t .

- ✓ Průměrná reziduální odchylka charakterizuje průměrnou velikost náhodného kolísání hodnot ČR. Předpověď může být bodová nebo intervalová.



Bodová a intervalová předpověď:

- ✓ **Bodovou předpověď** (y_{t+i}) se nazývá odhad vyjádřený jediným číslem a získaný přímým dosazením časového údaje, pro který má být předpověď provedena, do trendové funkce. Tento odhad svou přílišnou jednoznačností však neodpovídá proměnlivé povaze sledovaných jevů a také nepřihlíží k přítomnosti a možnému vlivu náhodné složky.
- ✓ Na zohlednění náhodného kolísání a vyjádření přípustné chyby odhadu je vhodnější použít **intervalovou prognózu**.
- ✓ Intervalová prognóza spočívá v tom, že se stanoví interval spolehlivosti, ve kterém se prognózovaná hodnota s předem zadanou pravděpodobností může nacházet.
- ✓ Za předpokladu, že reziduální odchylky mají průměrnou hodnotu rovnou nule, konstantní rozptyl a jsou stochasticky lineárně nezávislé, je možné dokázat, že predikovaná hodnota zkoumané ČŘ se bude nacházet s pravděpodobností $1 - \alpha$ v intervalu $P(y'_{t+i} - t_\alpha \cdot s_{y'_{t+i}} \leq y_{t+i} \leq y'_{t+i} + t_\alpha \cdot s_{y'_{t+i}}) = 1 - \alpha$, kde y'_{t+i} je bodová předpověď na období $(t + i)$, t_α je kritická hodnota Studentova t-rozdělení pro hladinu významnosti α a $(n - 2)$ stupně volnosti, $s_{y'_{t+i}}$ je směrodatná chyba předpovídané hodnoty, která se odhaduje ze vzorce

$$s_{y'_{t+i}} = s_y \cdot \sqrt{(1 - I^2) \cdot \frac{n(n^2 - 1) + 12i^2}{(n^2 - 1) \cdot (n - 2)}}, \text{ kde } I^2 \text{ je index determinace, } s_y \text{ je směrodatná odchylka hodnot ČŘ a } i = 1,$$

2, ... je horizont předpovědi.

- ✓ Intervalové předpovědi mají jisté nevýhody spočívající v tom, že ČŘ, se kterými pracujeme, jsou často poměrně krátké, což vede v důsledku působení náhodných vlivů ke zvýšeným hodnotám směrodatné odchylky predikovaných hodnot $s_{y'_{t+i}}$ a tím zpravidla k širokým intervalům spolehlivosti, jejichž praktická informační hodnota je potom velmi malá.
- ✓ Je třeba ovšem podotknout, že při stanovení intervalové předpovědi se mohou objevit i určité problémy.
 - Intervalová předpověď vychází z předpovědi bodové. Pokud pracujeme souběžně s více modely, může vzniknout svým způsobem paradoxní situace. Je totiž dost pravděpodobné, že se bodové předpovědi jednotlivých modelů budou mnohdy značně od sebe kvantitativně odlišovat. Tím se otvírá otázka, ke které variantě bodové předpovědi vlastně vůbec intervalovou předpověď sestojit.
 - Tradičně konstruované intervalové předpovědi vycházejí ze základního předpokladu, že časovou řadu lze popsat gaussovským pravděpodobnostním modelem, kdy i prognostický model je tvořen náhodnými veličinami pocházejícími z normálního rozdělení. Ve skutečnosti však předpoklady o normalitě ve většině praktických, zejména ekonomických aplikacích, nejsou příliš reálné.
- ✓ Pro řešení tohoto vážného problému, který podmiňuje kvalitu prognóz, lze nalézt určité pozitivní východisko: je třeba uvažovat o takovém typu předpovědi, v němž se interval nestanovuje na pravděpodobnostním základě, nýbrž tak, že meze intervalové předpovědi jsou chápány jako krajní varianty různých bodových předpovědi. Tento přístup vede ke konstrukci **předpovědního rozpětí**.
- ✓ **Konstrukce předpovědního rozpětí** vychází z myšlenky, že všechny přijatelné modely časové řady lze uspořádat extrémně v tom smyslu, že část z nich bude představovat optimistické bodové předpovědi a část předpovědi pesimistické. V obecném případě nutno předpokládat, že se mezi těmito extrémy mohou objevit i předpovědi kvalitativně neutrální.
- ✓ Jestliže se hodnoty předpovězené jednotlivými „přijatými“ modely transformují tak, aby vycházely ze stejného místa na počátku předpovědi (z tzv. „referenčního bodu“ nebo „referenční hodnoty“), a u každého z těchto modelů se převedou jím prognózované hodnoty na tempa růstu, která se aplikují na referenční bod, vznikne obor hodnot, který lze označit jako **předpovědní rozpětí**. Předpovědní rozpětí je možné podle uvedeného pojetí chápat jako jistou agregaci metod vedoucích k získání různých předpovědí sledovaného ukazatele.
- ✓ Při konstrukci se v první fázi vyberou takové modely, které lze z hlediska minulého vývoje přijmout. Pro každý přijatý model se sestaví tabulka, která uvede, jaká je jím vypočítaná (teoretická) hodnota příslušející poslední skutečné hodnotě (označí se $Y_{..}$) a prognózy pro nejbližší období.
- ✓ V druhém kroku se určí jednotný společný výchozí bod všech modelů (referenční bod) a od něho se odvíjí dynamika specifická pro každý jednotlivý model. Protože každý model v hodnotě $Y_{..}$ vyjadřuje svou úroveň na konci reálné časové řady, bude referenční bod (RB) jejich výslednicí (průměrem).
- ✓ Poslední krok spočívá v myšlence vypočítat tempa růstu, která jsou indikována jednotlivými modely, a pak maximálními, resp. minimálními hodnotami těchto temp vynásobit referenční bod a tím stanovit horní, resp. dolní meze předpovědního rozpětí.



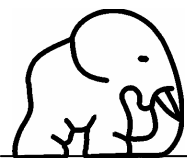
- ✓ Výsledkem předpovědního rozpětí je vždy nestochastický intervalový tvar prognózy.

Kombinované bodové předpovědi:

- ✓ Různé předpovědní modely obvykle mají tendenci poskytovat různé předpovědi pro tytéž řady. K získání co nejvhodnějších poznatků a vlastností pro předpovědi se vedle konstrukce předpovědního rozpětí nabízí také jednodušší metoda založená na myšlence agregovat informace získané různými metodami kombinováním bodových předpovědí.
- ✓ Nejsnazším způsobem kombinování předpovědí je výpočet prostého aritmetického průměru předpovědí získaných různými metodami. Při použití p metod má kombinovaná bodová předpověď provedená v období t na i období dopředu tvar $P_{t+i} = \frac{P_{(1)t+i} + P_{(2)t+i} + \dots + P_{(p)t+i}}{p}$, kde $P_{(h)t+i}$ je předpověď z metody h ($h = 1, 2, \dots, p$),
i horizont předpovědi $i = 1, 2, \dots, m$.
- ✓ Všechny předpovědi mají v této metodě přiřazeny stejné váhy $1/p$. Tento přístup tedy nebere v úvahu fakt, že některé metody jsou schopné poskytovat přesnější předpovědi než jiné metody.
- ✓ Postup kombinování předpovědí lze obecně doporučit, neboť snižuje podstatně riziko chyby zejména v případech náhlých změn, kdy různé metody poskytují někdy i výrazně odlišné hodnoty extrapolčních předpovědí.
- ✓ Pokud se kombinují předpovědi z metod, které nadhodnocují předvídané hodnoty, a z metod, které naopak mají tendenci k podhodnocování, může se výrazně snížit riziko velké chyby předpovědi.

Hodnocení přesnosti předpovědí:

- ✓ Pokud se prognózují hodnoty sledovaných ukazatelů, pak je třeba ověřit, jak byly odhady přesné. K hodnocení přesnosti odhadů se přistupuje až po uplynutí prognózovaného období, ale zároveň to umožňuje opravit zvolený model pro popis trendu a výpočet odhadů sledovaného ukazatele.
- ✓ Mezi jednoduché způsoby hodnocení přesnosti odhadů patří **absolutní chyba předpovědi** (Δ_{t+i}), což je vlastně rozdíl mezi předpovídanou (P_{t+i}) a skutečnou hodnotou (y_{t+i}) pro daný čas t a i horizont předpovědi. Jestliže $\Delta_{t+i} < 0$, jde o podceňující předpověď, jestliže naopak $\Delta_{t+i} > 0$, jde o nadceňující předpověď; teoreticky si lze představit i předpověď bezchybnou $\Delta_{t+i} = 0$.
- ✓ Jako jednu ze souhrnných měr přesnosti předpovědi je možné uvést **průměrnou chybu předpovědi**
$$\bar{D} = \frac{1}{m} \sum_{i=1}^m (P_{t+i} - y_{t+i})$$
. Zavede-li se značení $\bar{P} = \frac{1}{m} \sum_{i=1}^m P_{t+i}$ a $\bar{y} = \frac{1}{m} \sum_{i=1}^m y_{t+i}$ pro průměrnou předpověď a průměrnou skutečnost, lze zjistit, že $\bar{\Delta} = \bar{P} - \bar{y}$, tzn. že průměrná chyba předpovědi má kompenzační charakter, a tudíž se k charakterizování přesnosti předpovědi příliš nehodí.
- ✓ Tímto nedostatkem netrpí průměrná **čtvercová chyba předpovědi** $S_{\Delta}^2 = \frac{1}{m} \sum_{i=1}^m (P_{t+i} - y_{t+i})^2$. Jde o nezápornou veličinu. Své hraniční nulové hodnoty nabude v případě vesměs bezchybných předpovědí.
- ✓ Její odmocnina S_{Δ} se interpretuje jako **průměrná hodnota chyby předpovědi** (bez ohledu na znaménko). Jestliže $-S_{\Delta} < \Delta_{t+i} < +S_{\Delta}$, lze s vysokou pravděpodobností usuzovat, že t -tá následná chyba předpovědi nabývá takové hodnoty, kterou je možné považovat za náhodnou odchylku od nuly. Prakticky to znamená, že konstruovaná předpověď vedoucí k chybě, vyjádřené průměrnou hodnotou chyby předpovědi S_{Δ} , může být považována za uspokojivou, protože se od bezchybné předpovědi (tj. od předpovědi s nulovou následnou chybou) liší jen o náhodnou chybu.
- ✓ Neplatí-li naopak uvedená situace, je třeba považovat použitý způsob konstrukce předpovědi za chybný nebo přinejmenším se nelze o jeho použití s určitostí vyjádřit.
- ✓ Vedle absolutních chyb předpovědi se používá též **relativních chyb předpovědi** definovaných jako
$$\delta_{t+i} = \frac{\Delta_{t+i}}{y_{t+i}} = \frac{P_{t+i} - y_{t+i}}{y_{t+i}}$$
 pro horizont předpovědi i a počátky předpovědi $t = 1, 2, \dots, m$. Důvodem pro konstrukci těchto měr je jejich bezrozměrnost, a tudíž možnost po vynásobení stem následné chyby předpovědi vyjádřit v procentech předvídané skutečnosti.



- ✓ Jako velmi používanou mírou variability relativních chyb předpovědi lze uvést **Theilův koeficient nesouladu**

$$T^2 = \frac{\sum_{t=1}^m (P_{t+i} - y_{t+i})^2}{\sum_{t=1}^m y_{t+i}^2} .$$

Uvedená charakteristika nabývá nezáporných hodnot a dolní nulové hranice nabývá

pouze v případě, že $P_{t+i} = y_{t+i}$, tj. v případě bezchybných prognóz. Čím více se koeficient nesouladu odchyluje od nuly, tím více se soustava hodnocených předpovědí liší od ideálních bezchybných předpovědí. Odmocninu z koeficientu nesouladu (T) lze pak interpretovat jako **relativní chybu předpovědi v procentech**.

Příklad:

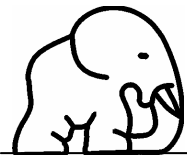
- ✓ Máme k dispozici časovou řadu agrárního vývozu v letech 1991 – 2003. Pro popis trendu byly jako nejvhodnější zvoleny funkce mocninná a kvadratická (viz. příklad u modelování trendu).
- ✓ Na základě všech použitých modelů trendu byly následně stanoveny bodové odhady na čtyři roky dopředu, tzn. na období 2004 – 2007 a pro vybrané modely i předpověď intervalová.
- ✓ Přehled bodových předpovědí je uveden v následující tabulce (bodové předpovědi byly získány dosazením za časovou proměnnou pro prognózované období do daného modelu).
- ✓ Bodové předpovědi agrárního vývozu na období 2004 – 2007 (mil. Kč):

| Model trendu | Bodová předpověď | | | |
|------------------------|------------------|----------|----------|----------|
| | 2004 | 2005 | 2006 | 2007 |
| Lineární funkce | 53 437,8 | 55 777,0 | 58 116,2 | 60 455,4 |
| Kvadratická funkce | 49 540,2 | 50 209,0 | 50 655,0 | 50 878,4 |
| Exponenciální funkce | 57 740,9 | 61 828,5 | 66 205,4 | 70 892,2 |
| Jednoduché exp. vyr. | 47 732,5 | 47 732,5 | 47 732,5 | 47 732,5 |
| Lineární exp. vyr. | 49 660,1 | 50 936,0 | 52 211,9 | 53 487,8 |
| Holtovo lin. exp. vyr. | 53 724,1 | 56 101,0 | 58 477,9 | 60 854,9 |
| Kvadratické exp. vyr. | 49 865,0 | 50 916,2 | 51 854,7 | 52 680,6 |
| ARIMA model | 50 081,8 | 52 629,3 | 54 782,7 | 57 179,3 |
| Lomená funkce | 42 296,8 | 42 440,7 | 42 566,6 | 42 677,7 |
| Logaritmická funkce | 47 735,2 | 48 549,5 | 49 311,1 | 50 026,6 |
| Mocninná funkce | 49 493,1 | 50 734,3 | 51 923,6 | 53 066,1 |

- ✓ Z tabulky je patrné, že některé odhady vedou k pomalému nárůstu objemu vývozu, některé k rychlejšímu růstu a některé ke konstantní či klesající úrovni agrárního vývozu.
- ✓ Pro stanovení budoucí úrovně agrárního vývozu bylo dále použito kombinování bodových předpovědí s těmito kombinacemi:
- Kombinace nejmenších a největších bodových předpovědí.
 - Kombinace odhadů ze dvou nejlepších trendových modelů.
 - Kombinace předpovědí z analytického vyrovnání.
 - Kombinace předpovědí z modelů exponenciálního vyrovnání.
 - Kombinace předpovědí ze všech modelů.
- ✓ Všechny kombinace byly stanoveny jako prostý aritmetický průměr z vybraných modelů.
- ✓ Kombinované předpovědi agrárního vývozu pro období 2004 – 2007:

| Model trendu | Bodová předpověď | | | |
|--------------------------------------|------------------|-----------|-----------|-----------|
| | 2004 | 2005 | 2006 | 2007 |
| Nejmenší a největší bodová předpověď | 50 018,85 | 52 134,60 | 54 386,00 | 56 784,95 |
| Dva nejlepší modely | 49 516,65 | 50 471,65 | 51 289,30 | 51 972,25 |
| Analytické vyrovnání | 50 040,67 | 51 589,80 | 53 129,70 | 54 666,10 |
| Exponenciální vyrovnání | 50 212,70 | 51 663,00 | 53 011,94 | 54 387,02 |
| Všechny modely | 50 118,86 | 51 623,09 | 53 076,15 | 54 539,23 |

- ✓ V tomto případě kombinování vede k vyrovnání jednotlivých bodových odhadů, tzn. snižuje se kolísání jednotlivých prognóz pro konkrétní rok.
- ✓ Následně byl použit postup stanovení předpovědního rozpětí.



| Model | Y ₀₃ | P ₀₄ | P ₀₅ | P ₀₆ | P ₀₇ |
|--------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Lineární funkce | 51 098,6 | 53 437,80 | 55 777,00 | 58 116,20 | 60 455,40 |
| Kvadrat. funkce | 48 648,7 | 49 540,20 | 50 209,00 | 50 655,00 | 50 878,40 |
| Exponen. funkce | 53 923,6 | 57 740,90 | 61 828,50 | 66 205,40 | 70 892,20 |
| Jednod. exp. vyr. | 45 937,9 | 47 732,50 | 47 732,50 | 47 732,50 | 47 732,50 |
| Lineární exp. vyr. | 48 384,2 | 49 660,10 | 50 936,00 | 52 211,90 | 53 487,80 |
| Holtovo exp. vyr. | 51 876,3 | 53 724,10 | 56 101,00 | 58 477,90 | 60 854,90 |
| Kvadr. exp. vyr. | 48 701,1 | 49 865,00 | 50 916,20 | 51 854,70 | 52 680,60 |
| ARIMA model | 51 572,6 | 50 081,80 | 52 629,30 | 54 782,70 | 57 179,30 |
| Lomená funkce | 42 130,8 | 42 296,80 | 42 440,70 | 42 566,60 | 42 677,70 |
| Logaritm. funkce | 46 860,6 | 47 735,20 | 48 549,50 | 49 311,10 | 50 026,60 |
| Mocninná funkce | 48 193,6 | 49 493,10 | 50 734,30 | 51 923,60 | 53 066,10 |

- ✓ Y₀₃ představuje vyrovnané hodnoty pro rok 2003, které byly získány dosazením do trendové funkce.
- ✓ Referenční bod je potom stanoven jako průměr ze všech hodnot Y₀₃, tzn. referenční bod má hodnotu 48 848.
- ✓ Následně byly pro jednotlivé bodové odhady stanoveny koeficienty růstu, tzn. k jaké relativní změně v rámci daných odhadů bude docházet.
- ✓ Dále je pak potřeba seřadit tyto koeficienty růstu podle velikosti a vybrat koeficient nejmenší a největší.

| Model | k _{04/03} | k _{05/04} | k _{06/05} | k _{07/06} |
|--------------------|--------------------|--------------------|--------------------|--------------------|
| Lineární funkce | 1,0457782 | 1,0437743 | 1,0419384 | 1,0402504 |
| Kvadrat. funkce | 1,0183253 | 1,0135001 | 1,0088829 | 1,0044102 |
| Exponen. funkce | 1,0707909 | 1,0707921 | 1,0707910 | 1,0707918 |
| Jednod. exp. vyr. | 1,0390658 | 1,0000000 | 1,0000000 | 1,0000000 |
| Lineární exp. vyr. | 1,0263702 | 1,0256927 | 1,0250491 | 1,0244370 |
| Holtovo exp. vyr. | 1,0356193 | 1,0442427 | 1,0423682 | 1,0406478 |
| Kvadr. exp. vyr. | 1,0238988 | 1,0210809 | 1,0184322 | 1,0159272 |
| ARIMA model | 0,9710932 | 1,0508668 | 1,0409164 | 1,0437474 |
| Lomená funkce | 1,0039401 | 1,0034021 | 1,0029665 | 1,0026100 |
| Logaritm. funkce | 1,0186639 | 1,0170587 | 1,0156871 | 1,0145099 |
| Mocninná funkce | 1,0269642 | 1,0250782 | 1,0234417 | 1,0220035 |

- ✓ Následně je referenční bod vynásoben nejmenším a největším koeficientem růstu, čímž dostaneme meze rozpětí pro rok 2004. Podobným způsobem postupujeme i pro další roky.
- ✓ Pro roky 2004 – 2007 tak lze získat uvedené rozpětí.

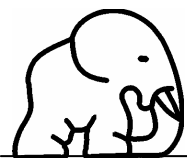
| Referenční bod = 48 848 | P ₀₄ | P ₀₅ | P ₀₆ | P ₀₇ |
|-------------------------|-----------------|-----------------|-----------------|-----------------|
| Dolní mez | 47 435,96 | 47 435,96 | 47 435,96 | 47 435,96 |
| Horní mez | 52 305,99 | 56 008,85 | 59 973,77 | 64 219,42 |

- ✓ Postup stanovení intervalového odhadu bude ukázán pouze pro lineární trend, stejné výpočty však lze aplikovat i na ostatní modely trendu.

$$s_{y'_{2004}} = 9607,33 \cdot \sqrt{(1 - 0,948222^2) \cdot \frac{13 \cdot (13^2 - 1) + 12 \cdot 1^2}{(13^2 - 1) \cdot (13 - 2)}} \quad s_{y'_{2004}} = 3326,287242$$

$$s_{y'_{2005}} = 9607,33 \cdot \sqrt{(1 - 0,948222^2) \cdot \frac{13 \cdot (13^2 - 1) + 12 \cdot 2^2}{(13^2 - 1) \cdot (13 - 2)}} \quad s_{y'_{2005}} = 3353,441058$$

$$s_{y'_{2006}} = 9607,33 \cdot \sqrt{(1 - 0,948222^2) \cdot \frac{13 \cdot (13^2 - 1) + 12 \cdot 3^2}{(13^2 - 1) \cdot (13 - 2)}} \quad s_{y'_{2006}} = 3398,215284$$



$$s_{y_{2007}} = 9607,33 \cdot \sqrt{(1 - 0,948222^2) \cdot \frac{13 \cdot (13^2 - 1) + 12 \cdot 4^2}{(13^2 - 1) \cdot (13 - 2)}}$$

$$s_{y_{2007}} = 3459,925921$$

- ✓ $t_{0,05 (13-2)} = 2,201$
- ✓ Bodové předpovědi pro jednotlivé roky:

| | |
|------|---------|
| 2004 | 53437,8 |
| 2005 | 55777,0 |
| 2006 | 58116,2 |
| 2007 | 60455,4 |
- ✓ **Rok 2004:**

$$\Delta_{2004} = 2,201 \cdot 3326,287242 = 7321,15822$$

$$P(53437,8 - 7321,15822 \leq y_{t+i} \leq 53437,8 + 7321,15822) = 0,95$$

$$P(46116,64178 \leq y_{t+i} \leq 60758,95822) = 0,95$$
- ✓ **Rok 2005:**

$$\Delta_{2005} = 2,201 \cdot 3353,441058 = 7380,923769$$

$$P(55777 - 7380,923769 \leq y_{t+i} \leq 55777 + 7380,923769) = 0,95$$

$$P(48396,07623 \leq y_{t+i} \leq 63157,92377) = 0,95$$
- ✓ **Rok 2006:**

$$\Delta_{2006} = 2,201 \cdot 3398,215284 = 7479,47184$$

$$P(58116,2 - 7479,47184 \leq y_{t+i} \leq 58116,2 + 7479,47184) = 0,95$$

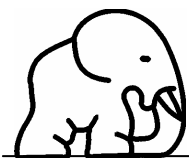
$$P(50636,72816 \leq y_{t+i} \leq 65595,67184) = 0,95$$
- ✓ **Rok 2007:**

$$\Delta_{2007} = 2,201 \cdot 3459,925921 = 7615,296951$$

$$P(60455,4 - 7615,296951 \leq y_{t+i} \leq 60455,4 + 7615,296951) = 0,95$$

$$P(52840,10305 \leq y_{t+i} \leq 68070,69695) = 0,95$$
- ✓ Vzhledem k tomu, že jsou známy hodnoty pro rok 2004 a 2005, lze vyhodnotit přesnost jednotlivých odhadů.
- ✓ Hodnota agrárního vývozu pro rok 2004 měla hodnotu 61 526 mil. Kč a pro rok 2005 pak 77 885 mil. Kč.
- ✓ K hodnocení jsou použity jak individuální chyby předpovědi, tak i chyby souhrnné.
- ✓ Z absolutních a relativních chyb je zřejmé, že ani jeden z modelů nedokázal spolehlivě určit velikost agrárního vývozu pro další dva roky.
- ✓ Stejně tak intervalová předpověď a předpovědní rozpětí nezahrnují ve svém rozmezí skutečnou hodnotu.

| Model trendu | 2004 | | | 2005 | | |
|----------------------------|------------------|----------------------------|----------------------------|------------------|----------------------------|----------------------------|
| | Bodová předpověď | Absolutní chyba předpovědi | Relativní chyba předpovědi | Bodová předpověď | Absolutní chyba předpovědi | Relativní chyba předpovědi |
| Lineární funkce | 53 437,80 | -8 088,20 | -13,146 | 55 777,00 | -22 108,00 | -28,3854 |
| Kvadratická funkce | 49 540,20 | -11 985,80 | -19,4809 | 50 209,00 | -27 676,00 | -35,5344 |
| Exponenciální funkce | 57 740,90 | -3 785,10 | -6,15203 | 61 828,50 | -16 056,50 | -20,6157 |
| Jednoduché exp. vyr. | 47 732,50 | -13 793,50 | -22,419 | 47 732,50 | -30 152,50 | -38,7141 |
| Lineární exp. vyr. | 49 660,10 | -11 865,90 | -19,286 | 50 936,00 | -26 949,00 | -34,601 |
| Holtovo lineární exp. vyr. | 53 724,10 | -7 801,90 | -12,6807 | 56 101,00 | -21 784,00 | -27,9694 |
| Kvadratické exp. vyr. | 49 865,00 | -11 661,00 | -18,953 | 50 916,20 | -26 968,80 | -34,6264 |
| ARIMA model | 50 081,80 | -11 444,20 | -18,6006 | 52 629,30 | -25 255,70 | -32,4269 |
| Lomená funkce | 42 296,80 | -19 229,20 | -31,2538 | 42 440,70 | -35 444,30 | -45,5085 |



| | | | | | | |
|--------------------------------------|-----------|------------|----------|-----------|------------|----------|
| Logaritmická funkce | 47 735,20 | -13 790,80 | -22,4146 | 48 549,50 | -29 335,50 | -37,6651 |
| Mocnná funkce | 49 493,10 | -12 032,90 | -19,5574 | 50 734,30 | -27 150,70 | -34,86 |
| Nejmenší a největší bodová předpověď | 50 018,85 | -11 507,15 | -18,7029 | 52 134,60 | -25 750,40 | -33,0621 |
| Dva nejlepší modely | 49 516,65 | -12 009,35 | -19,5191 | 50 471,65 | -27 413,35 | -35,1972 |
| Analytické vyrovnaní | 50 040,67 | -11 485,33 | -18,6674 | 51 589,80 | -26 295,20 | -33,7616 |
| Exponenciální vyrovnaní | 50 212,70 | -11 313,30 | -18,3878 | 51 663,00 | -26 222,00 | -33,6676 |

| Ukazatel | 2004 | 2005 |
|-------------------------------------|-------------|-------------|
| Průměrná chyba předpovědi | -11450,05 | -26301,49 |
| Průměrná čtvercová chyba předpovědi | 140855515 | 707811936,1 |
| Průměrná hodnota chyby předpovědi | 11868,25661 | 26604,73522 |
| Theilův koeficient nesouladu | 0,595355645 | 1,866939379 |
| Relativní chyba předpovědi | 0,771592927 | 1,366359901 |

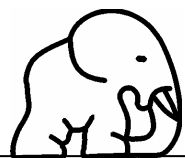
- ✓ Výsledky ukázaly, že v rámci sledovaného ukazatele došlo k výraznějšímu vlivu ostatních faktorů, především faktorů ekonomických.
- ✓ Proto jednotlivé odhady nelze hodnotit jako spolehlivé a přesné a pro další období je potřeba opravit dosavadní modely trendu a stanovit nové odhady pro nejbližší období.

POPIS NÁHODNÉ SLOŽKY:

- ✓ Náhodnou složku ε_t , kterou lze vyjádřit ve tvaru $\varepsilon_t = y_t - y_t^*$, lze chápat jako výsledek působení blíže nespecifikovaného souboru náhodných (stochastických) vlivů.
- ✓ Zdrojem této složky jsou nepodchycené nebo nepodchytitelné drobné a vzájemně nezávislé náhodné vlivy, které se v rámci ČR vykompenzují.
- ✓ Lze tedy předpokládat, že jejich střední hodnoty jsou nulové, tj. že platí $E(\varepsilon_t) = 0$, $t = 1, 2, \dots, n$.
- ✓ Vedle předpokladu o střední hodnotě náhodných poruch je nutné formulovat ještě předpoklady o jejich variabilitě, jejich vzájemné závislosti, popř. o jejich zákonu rozdělení.
- ✓ Nejméně náročnou a také nejpoužívanější je hypotéza o homoskedasticitě náhodných poruch.
- ✓ Předpokládá se, že náhodné poruchy s nulovými středními hodnotami mají konstantní rozptyl a jsou vzájemně lineárně nezávislé, tj. platí

$$D(\varepsilon_t) = \sigma^2 \quad t = 1, 2, \dots, n$$

$$E(\varepsilon_t \varepsilon_{t'}) = 0 \quad t, t' = 1, 2, \dots, n, t \neq t'$$
- ✓ Jsou-li tyto předpoklady splněny, tvoří řada ε_t tzv. **bílý šum**.
- ✓ Druhým používaným předpokladem o náhodné složce je předpoklad o **heteroskedasticitě** náhodných poruch, kde se předpokládá, že náhodné poruchy s nulovými středními hodnotami jsou vzájemně nezávislé s měnlivými rozptyly $D(\varepsilon_t) = \sigma^2 w_t^{-1}$.
- ✓ Rozptyl je v této formulaci úměrný veličině w_t^{-1} , kde veličiny w_t nazýváme vahou pozorování splňující požadavek $\sum_{t=1}^n w_t^{-1} = n$.
- ✓ Třetím používaným předpokladem, se kterým se setkáváme, je předpoklad o autoregresi náhodných poruch. Vychází z představy, že $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$, $0 < \rho < 1$, kde ρ je autokorelační koeficient sousedních náhodných poruch, který je považován za konstantní, a u_t ($t = 1, 2, \dots, n$) je posloupnost náhodných poruch s nulovými středními hodnotami a konstantními rozptyly, přičemž tyto poruchy jsou vzájemně nezávislé. Náhodná porucha v čase t se tedy skládá ze dvou složek: ze složky závislé na předchozí poruše ε_{t-1} a z náhodné složky u_t .

**Durbin – Watsonův test autokorelace:**

- ✓ Pomocí tohoto testu ověřujeme, zda jsou náhodné poruchy nezávislé.
- ✓ Proti nulové hypotéze o nezávislosti náhodných poruch (autokorelace není) stavíme alternativní hypotézu tvrdící, že náhodné poruchy jsou závislé, přičemž závislost je vyjádřena autoregresním schématem $\varepsilon_t = \varepsilon_{t-1} + u_t$.
- ✓ Jako testové kritérium se u tohoto testu používá statistika

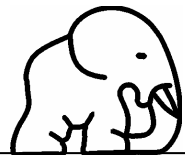
$$d = \frac{(e_2 - e_1)^2 + \dots + (e_{n-1} - e_{n-2})^2 + (e_n - e_{n-1})^2}{e_1^2 + e_2^2 + \dots + e_n^2} = \frac{\sum_{t=1}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}.$$

- ✓ Hodnoty této statistiky se pohybují v intervalu od nuly do čtyř. V případě nezávislosti náhodných poruch se statistika pohybuje okolo čísla 2, v případě přímé závislosti jsou její hodnoty kolem nuly a v případě nepřímé závislosti kolem 4.
- ✓ Pro daný počet pozorování n a počet strukturálních parametrů modelu (např. pro lineární trend je počet strukturálních parametrů, tj. počet parametrů, u kterých je časová proměnná, roven jedné) jsou tabelovány pro případ přímé závislosti a různé hladiny významnosti dvojice kritických hodnot d_L a d_U . V případě, že platí $d < d_L$, zamítáme na zvolené hladině významnosti nulovou hypotézu ve prospěch alternativní hypotézy a můžeme předpokládat, že mezi náhodnými poruchami je přímá závislost.
- ✓ Nedostatkem tohoto testu je, že v případě, že $d_L < d < d_U$, tento test „mlčí“, tj. nesvědčí ani pro nulovou, ani pro alternativní hypotézu.
- ✓ V praktických situacích obvykle speciální tabulky D-W testu po ruce nemáme, a proto vcelku spolehlivě vystačíme s přibližným vyhodnocením testového kritéria.

| DW | Výsledek |
|--------------------------|---|
| $4 - d_L < DW < 4$ | H_0 se zamítá – autokorelace |
| $4 - d_U < DW < 4 - d_L$ | Neumíme rozhodnout, je třeba zvýšit n |
| $2 < DW < 4 - d_U$ | Přijímá se H_0 – autokorelace není |
| $d_U < DW < 2$ | Přijímá se H_0 – autokorelace není |
| $d_L < DW < d_U$ | Neumíme rozhodnout, je třeba zvýšit n |
| $0 < DW < d_L$ | H_0 se zamítá – autokorelace |

Adaptivní přístupy k modelu časové řady:

- ✓ Další kvalitativně jiný přístup k popisu trendu představují adaptivní metody. Dosud uváděné postupy vycházely z předpokladu, že v průběhu celé popisované doby se **parametry modelu nemění** (hovoří se o modelech s konstantními parametry).
- ✓ Pokud se pomocí těchto modelů konstruuje i předpovědi, tak jde o předpovědi vycházející ze situace, že i v budoucnu nedojde ke změnám systému.
- ✓ Jinak řečeno, vychází se ze situace, že informativní hodnota (aktuálnost) údajů pocházejících z počátku i z konce zkoumaného období je stejná. Bude-li se tedy takto extrapolovat např. trend, bude předpověď jenom kopií minulosti.
- ✓ Při předvídání budoucího průběhu ekonomických procesů se však stává předpoklad neměnnosti analytického tvaru modelu a jeho parametrů velmi omezující a někdy dokonce neudržitelný.
- ✓ Velmi často nastane situace, kdy během analyzovaného období se hodnoty strukturálních parametrů v čase mění. Tyto změny pak vedou k desaktualizaci modelu s konstantními parametry.
- ✓ Znamená to, že jeho strukturální parametry už neodrážejí skutečné kvantitativní relace mezi endogenní proměnnou a časem a jeho použití k prognózám v takové situaci může vést k závažným systematickým chybám.
- ✓ Uvedené důvody byly podnětem pro konstrukci adaptivních modelů (někdy se hovoří o modelech s měnlivými parametry).
- ✓ Klasickým modelům s konstantními parametry jsou blízké tím, že neobjasňují kauzální mechanismus vývoje analyzované proměnné, ale pouze popisují její průběh v čase.
- ✓ Od klasických modelů s konstantními parametry se však zásadně liší tím, že nepředpokládají stabilitu analytického tvaru ani strukturálních parametrů v čase, a dokonce ani spojitost trendové funkce.
- ✓ V zásadě jediný předpoklad, nutný pro konkrétní užívání adaptivních metod v procesu předvídání, představuje časová stacionarita rozdělení chyb prognózy.



- ✓ Modely tohoto typu rychle reagují na strukturální změny, k nimž dochází v čase, a jsou velmi vhodné při prognózování průběhu časových řad, které se vyznačují nepravidelnostmi a zlomy v trendu.
- ✓ Adaptivní modely vychází z předpokladu, že pro konstrukci prognózy budoucího vývoje mají cenu nejnovější pozorování časové řady.
- ✓ Proto těmto nejnovějším pozorováním se přiřazují největší váhy, a dřívější pozorování se buď úplně vyřazují ze zkoumání nebo se jim přiřazují menší váhy ve srovnání s později pozorovanými hodnotami.
- ✓ Adaptivní modely tedy berou v úvahu „stárnutí“ informací.
- ✓ Statistická teorie zná více těchto postupů. Mezi nejznámější, které přináší v praktických aplikacích dobré výsledky, patří **metody exponenciálního vyrovnávání**.

Exponenciální vyrovnávání:

- ✓ Předpokládejme, že v časovém okamžiku n , který představuje pozorování v přítomném čase, máme k dispozici řadu empirických hodnot y_{n-k} ($k = 0, 1, \dots, n-1$), kde jednotlivá k interpretujeme jako „stáří“ (věk) pozorování. Vycházíme opět z aditivního modelu časové řady, tj. platí $y_{n-k} = T_{n-k} + \varepsilon_{n-k}$.
- ✓ Hodnotu trendové složky T_{n-k} lze přitom popsat funkcí $T_{n-k} = a_0 - a_1 k + a_2 k^2 + \dots + (-1)^k a_k k^k$, kde k je časová proměnná, kterou lze chápat jako „věk“ pozorování z hlediska časového okamžiku n .
- ✓ Odhady parametrů této trendové funkce lze získat na základě metody nejmenších čtverců ve formulaci

$$\sum_{k=0}^{n-1} (y_{n-k} - T_{n-k})^2 = \min.$$

- ✓ Při tomto způsobu vyrovnávání se každému empirickému pozorování při vyrovnávání přisuzuje stejná váha, tj. předpokládá se, že pozorování blízká časovému bodu n (tj. současnosti) jsou pro odhad parametrů a_k a tím i pro možnou konstrukci prognózy budoucího vývoje analyzovaného ukazatele stejně důležitá jako pozorování pro poměrně vysoké hodnoty k , tj. pro pozorování starší. Přitom lze důvodně předpokládat, že empirická pozorování „čerstvější“ (bližší časovému okamžiku n) budou více ovlivňovat budoucí vývoj analyzované řady než pozorování starší.

- ✓ Měla by se tedy těmto „čerstvějším“ pozorováním při odhadu parametrů a_k přiřazovat větší váha než pozorováním starším. Za této situace je nutné předchozí podmínku formulovat ve tvaru

$$\sum_{k=0}^{n-1} (y_{n-k} - T_{n-k})^2 w_k = \min, \text{ kde } w_k \text{ představují váhy, které jsou nepřímo úměrné „stáří“ pozorování, tj. se}$$

vzrůstajícím věkem pozorování se váha snižuje.

- ✓ Předpokládá se přitom, že váha w_k je exponenciální funkcí typu $w_k = \alpha^k$, $0 < \alpha < 1$, $k = 0, 1, \dots, n-1$, kde veličina α se nazývá vyrovnávací konstanta. Váhy w_k jsou tedy exponenciální funkcí věku pozorování.
- ✓ V praxi se lze setkat se třemi různými způsoby exponenciálního vyrovnávání:

- S Brownovým exponenciálním vyrovnáváním.
- S Holtovým lineárním exponenciálním vyrovnáváním.
- S Wintersovým sezónním vyrovnáváním.

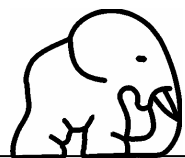
- ✓ U Brownova exponenciálního vyrovnávání se ještě rozlišuje:

- Vyrovnávání jednoduché, kdy trend je možno považovat v krátkých úsecích řady za konstantní.
- Dvojitě (lineární), kdy se trend v časové řadě modeluje po částech přímkou.
- Trojitě (kvadratické), kdy je trend v časové řadě popisován po částech parabolou.

- ✓ Pro získání vyrovnaných hodnot se u **Brownova vyrovnávání** pracuje s vyrovnávací konstantou α z intervalu $(0; 1)$. U **Holtova vyrovnávání** se odhadují dvě vyrovnávací konstanty α a β opět z intervalu $(0; 1)$. Konstanta α se používá k vyrovnávání úrovně časové řady, konstanta β k vlastnímu vyrovnání trendu. Zatímco tyto dvě metody modelují trend v časové řadě, **Wintersovo vyrovnávání** pokrývá vedle trendu rovněž sezónní složku. Používá se tedy pro sezónní časové řady. Při jeho konstrukci se vychází z multiplikativního modelu časové řady, kde se trendová složka popisuje lineární trendovou funkcí a sezónní složku se kvantifikuje pomocí modelu proporcionální sezónnosti. Pro tento způsob vyrovnávání se odhadují tři vyrovnávací konstanty α , β , γ z intervalu $(0; 1)$.

Korelace časových řad:

- ✓ Pokud se sleduje současně několik časových řad, často vznikne otázka, zda mezi těmito časovými řadami neexistují takové souvislosti, které by dovolovaly vysvětlit změny v jedné časové řadě změnami v druhé časové řadě, popř. v několika dalších časových řadách.
- ✓ Ze statistického hlediska se potom jedná o **problematiku korelace časových řad**.



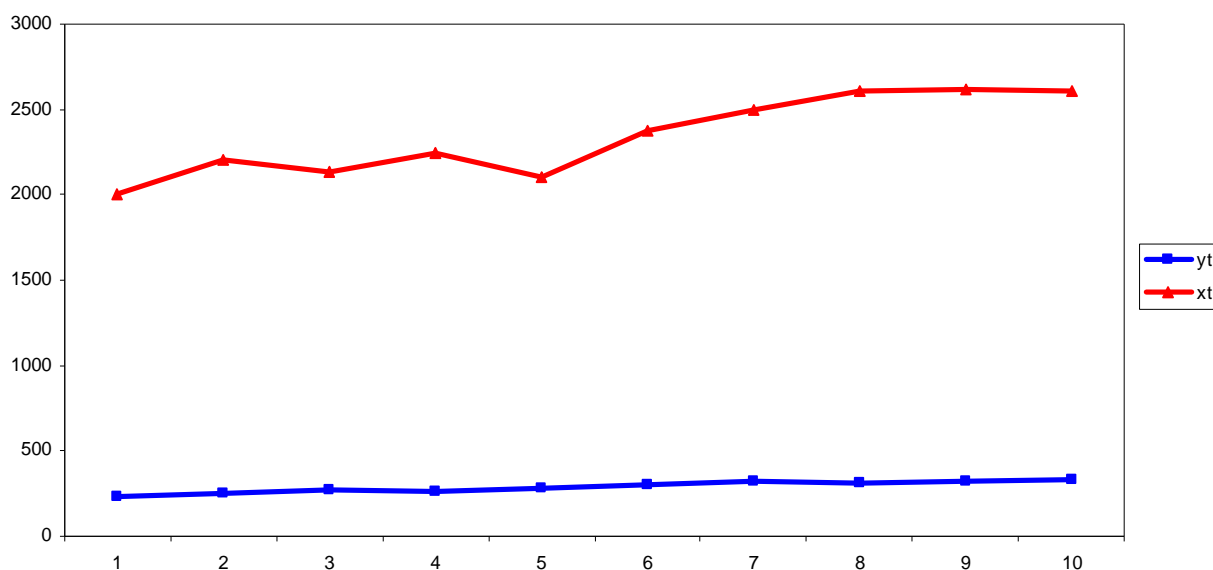
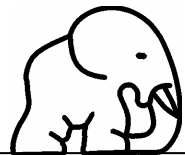
- ✓ Při zkoumání korelace časových řad musíme mít na zřeteli, že statistický soubor časové řady je daný hodnotami postupně na sebe navazujících období. Bez zjevné věcné souvislosti může mít velmi podobný vývoj v daném čase i jiná časová řada. V takovýchto případech vypočtené poměrně vysoké míry korelace nevyjadřují souvislost mezi jevy, ale jenom souběžnost průběhu dvou či více ČŘ.
- ✓ Při zkoumání vztahů mezi časovými řadami se zpravidla vychází z předpokladu, že je lze popsat určitým aditivním modelem, tj. že každou časovou řadu je možné vyjádřit jako součet pravidelné a nepravidelné složky.
- ✓ Pokud se zkoumá určitý příčinný vztah mezi řadami, tak nestačí hodnotit pouze celkovou vývojovou tendenci nebo sezónní kolísání, protože tyto faktory mohou mít velmi shodný průběh.
- ✓ Proto je třeba hledat, zda neexistuje nějaký vztah mezi nepravidelnými (náhodnými složkami) analyzovaných časových řad.
- ✓ Jestliže se nalezne určitá závislost mezi těmi náhodnými složkami, lze důvodně přepokládat, že existuje skutečná příčinná závislost mezi sledovanými časovými řadami.
- ✓ Pro hodnocení příčinného vztahu mezi proměnnými lze použít metody měření těsnosti závislosti řad náhodné složky, tj. řad očištěných od trendu, popř. také od sezónní složky (jde o korelaci náhodné složky).
- ✓ Znamená to tedy, že pro zkoumání, zda vztah mezi proměnnými je příčinný, lze použít metody měření těsnosti závislosti řad náhodné složky, tj. řad očištěných od trendu, popř. rovněž od sezónní složky. Předpokládáme pro jednoduchost, že pracujeme s časovými řadami typu $y_t = T_t + \varepsilon_t$, $t = 1, 2, \dots, n$, kde pro dané t značí y_t empirickou hodnotu ČŘ, T_t hodnotu trendové složky a ε_t hodnotu náhodné složky (náhodnou složku vyjadřujeme pomocí odchylek empirických hodnot ČŘ od teoretických – vyrovnaných hodnot).
- ✓ Předpokládejme nyní, že máme zkoumat závislost mezi dvěma časovými řadami, z nichž jednu označíme symbolem x_t a druhou symbolem y_t pro $t = 1, 2, \dots, n$.
- ✓ Odhadneme-li průběh trendu obou uvedených řad, dostaneme posloupnost odhadů trendových hodnot (T_x a T_y).
- ✓ Při hledání závislosti mezi oběma řadami pak budeme korelovat odhady reziduálních hodnot, které označíme jako $e_x = x_t - T_x$ a $e_y = y_t - T_y$, $t = 1, 2, \dots, n$.

Příklad:

- ✓ Chceme změřit sílu závislosti mezi dvěma časovými řadami.
- ✓ Určíme-li korelační koeficient přímo z hodnot časových řad y_t a x_t , dostaneme hodnotu $r_{yx} = 0,892952$, což by mohlo dokládat silný vztah mezi nimi.
- ✓ Avšak pohled na hodnoty ČŘ nás upozorní na to, že mezi ČŘ příliš systematický vývoj neexistuje.

| t | y_t | x_t | e_y | e_x |
|-----------------|-------|-------|-----------|----------|
| 1 | 230 | 2000 | -9,09091 | -17,7269 |
| 2 | 250 | 2200 | -0,18182 | 110,8791 |
| 3 | 275 | 2130 | 13,72727 | -30,5148 |
| 4 | 260 | 2250 | -12,36364 | 18,09124 |
| 5 | 285 | 2105 | 1,54545 | -198,303 |
| 6 | 300 | 2375 | 5,45454 | 0,30336 |
| 7 | 320 | 2500 | 14,36363 | 53,90942 |
| 8 | 315 | 2605 | -1,72728 | 87,51548 |
| 9 | 320 | 2620 | -7,81819 | 31,12154 |
| 10 | 335 | 2605 | -3,9091 | -55,2724 |
| Σ | 2890 | 23390 | -5E-05 | 0,0033 |
| Durbin – Watson | | | 2,093483 | 2,062744 |

- ✓ Obě ČŘ zbavíme trendové složky. Pro odhad trendu nalezneme rovnice $T_y = 228 + 1,09091 \cdot t$ a $T_x = 1946,333 + 71,39394 \cdot t$, na jejichž základě spočteme rezidua. Jejich korelace je však již velmi nízká, a sice $r_{ey,ex} = -0,02968$, což značí slabou intenzitu závislosti. Vysoká úroveň koeficientu korelace původních hodnot byla tedy velmi klamná.

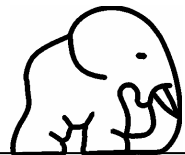


- ✓ Uvedený příklad popisuje problém tzv. **zdánlivé korelace**.
- ✓ Zdánlivá korelace spočívá v tom, že je někdy možné pozorovat silnou závislost mezi proměnnými i v případě, kdy mezi proměnnými ve skutečnosti závislost buď skoro nebo vůbec neexistuje. Dochází k ní proto, že obě proměnné vykazují stejný lineární trend.
- ✓ Je potřeba mít na zřeteli, že volba a použití určité trendové funkce je do určité míry ovlivněna subjektivismem a zvolená funkce nemusí být vždy vhodná.
- ✓ Hodnověrnost vypočteného koeficientu korelace je totiž podmíněna správnou volbou typu trendových funkcí, od nichž počítáme odchylky, které pak korelujeme.
- ✓ Vyjádření trendu nevhodnou funkcí se může potom projevit také v tom, že odchylky empirických hodnot od teoretických nebudou správně vystihovat náhodnou složku ČŘ, což bude mít za následek, že odchylky nebudou v čase náhodně uspořádány a bude se mezi nimi projevovat **autokorelace**, tj. korelace mezi sousedními odchylkami od trendu.
- ✓ Vhodným prostředkem ověření náhodnosti uspořádání je např. Durbin – Watsonův test autokorelace.
- ✓ V analýze ČŘ se někdy setkáváme s případem, kdy vliv určitého jevu na jiný jev se neprojevuje ve stejných obdobích, ale často až po určité době, tj. po uplynutí jednoho, dvou nebo více období. Pak mluvíme o tzv. **opožděné korelaci**.
- ✓ Zkoumáme ji stejnými metodami jako korelaci mezi dvěma stejnými obdobími pouze s tím rozdílem, že posunujeme jednu ČŘ (závisle proměnná) o jedno, dvě nebo více období dále. To znamená, že sledujeme vliv hodnot např. i-tého roku na hodnoty závisle proměnné v obdobích $i + 1$, $i + 2$,
- ✓ Tak zjistíme, zda vliv vysvětlující proměnné na závisle proměnnou je omezen pouze na dané období nebo zda působí i po následující období.

Zkoumání vzájemných vztahů mezi hodnotami jedné časové řady:

- ✓ V souvislosti se zkoumáním korelace ČŘ přichází v úvahu analýza závislostí mezi sousedními členy jedné ČŘ.
- ✓ Mírou těsnosti této závislosti je koeficient autokorelace prvního řádu, který můžeme vyjádřit ve formě

$$r_1 = \frac{\frac{1}{n-1} \sum_{t=1}^{n-1} \left\{ \left(y_t - \frac{1}{n-1} \sum_{t=1}^{n-1} y_t \right) \left(y_{t+1} - \frac{1}{n-1} \sum_{t=1}^{n-1} y_{t+1} \right) \right\}}{\sqrt{\frac{1}{n-1} \sum_{t=1}^{n-1} \left(y_t - \frac{1}{n-1} \sum_{t=1}^{n-1} y_t \right)^2 \frac{1}{n-1} \sum_{t=1}^{n-1} \left(y_{t+1} - \frac{1}{n-1} \sum_{t=1}^{n-1} y_{t+1} \right)^2}} =$$
$$= \frac{(n-1) \sum_{t=1}^{n-1} y_t y_{t+1} - \left(\sum_{t=1}^{n-1} y_t \right) \left(\sum_{t=1}^{n-1} y_{t+1} \right)}{\sqrt{\left\{ (n-1) \sum_{t=1}^{n-1} y_t^2 - \left(\sum_{t=1}^{n-1} y_t \right)^2 \right\} \left\{ (n-1) \sum_{t=1}^{n-1} y_{t+1}^2 - \left(\sum_{t=1}^{n-1} y_{t+1} \right)^2 \right\}}}$$



- ✓ Obdobně koeficient korelace, jímž je možné měřit těsnost závislosti dvou členů časové řady, mezi nimiž je $k-1$ jiných pozorování, se nazývá **koeficient autokorelace k -tého řádu**. Vyjadřuje se vzorcem

$$r_k = \frac{\frac{1}{n-k} \sum_{t=1}^{n-k} \left\{ \left(y_t - \frac{1}{n-k} \sum_{t=1}^{n-k} y_t \right) \left(y_{t+k} - \frac{1}{n-k} \sum_{t=1}^{n-k} y_{t+k} \right) \right\}}{\sqrt{\frac{1}{n-k} \sum_{t=1}^{n-k} \left(y_t - \frac{1}{n-k} \sum_{t=1}^{n-k} y_t \right)^2 \cdot \frac{1}{n-k} \sum_{t=1}^{n-k} \left(y_{t+k} - \frac{1}{n-k} \sum_{t=1}^{n-k} y_{t+k} \right)^2}}$$

- ✓ Různé hodnoty r_k (tedy autokorelací s posunem k) je možné graficky znázornit pomocí grafu, který se nazývá korelogram.

Příklad:

- ✓ Máme údaje o těžbě kaolínu v tis. t v letech 1985 – 1995:

| Rok | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|-------|------|------|------|------|------|------|------|------|------|------|------|
| y_t | 28 | 30 | 35 | 40 | 42 | 44 | 50 | 54 | 58 | 65 | 68 |

- ✓ Z údajů vypočítáme koeficient autokorelace prvního řádu.

| y_t | y_{t+1} | y_t^2 | y_{t+1}^2 | $y_t y_{t+1}$ |
|-------|-----------|---------|-------------|---------------|
| 28 | 30 | 784 | 900 | 840 |
| 30 | 35 | 900 | 1225 | 1050 |
| 35 | 40 | 1225 | 1600 | 1400 |
| 40 | 42 | 1600 | 1764 | 1680 |
| 42 | 44 | 1764 | 1936 | 1848 |
| 44 | 50 | 1936 | 2500 | 2200 |
| 50 | 54 | 2500 | 2916 | 2700 |
| 54 | 58 | 2916 | 3364 | 3132 |
| 58 | 65 | 3364 | 4225 | 3770 |
| 65 | 68 | 4225 | 4624 | 4420 |
| 446 | 486 | 21214 | 25054 | 23040 |

$$r_1 = \frac{10 \cdot 23040 - 446 \cdot 486}{\sqrt{(10 \cdot 21214 - 446^2)(10 \cdot 25054 - 486^2)}} = 0,995$$

- ✓ Získaný výsledek svědčí o vysoké autokorelovanosti dané ČŘ.