



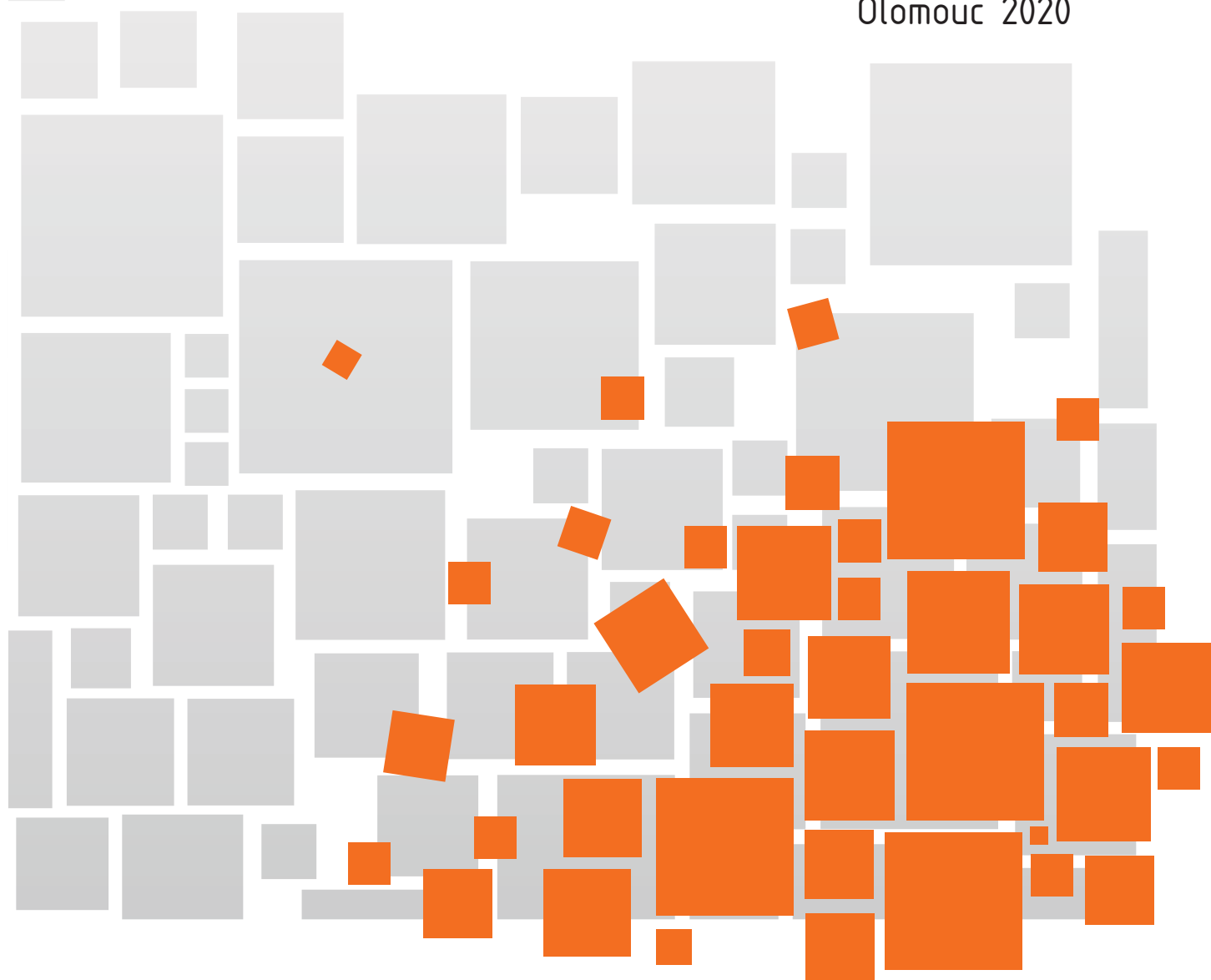
Přírodovědecká  
fakulta

Univerzita Palackého  
v Olomouci

# ČASOVÉ ŘADY ČASNĚ A ŘÁDNĚ

Ondřej Vencálek

Olomouc 2020



Univerzita Palackého v Olomouci  
Přírodovědecká fakulta

# Časové řady časně a řádně

Ondřej Vencálek

Olomouc 2020



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



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

Publikace vznikla za podpory projektu OP VVV s názvem Univerzita Palackého jako komplexní vzdělávací instituce, reg. č. CZ.02.2.69/0.0/0.0/16\_015/0002337

Oponenti:

RNDr. Tomáš Füst, Ph.D.

Ing. Ondřej Vozár, Ph.D.

Neoprávněné užití tohoto díla je porušením autorských práv a může zakládat občanskoprávní, správněprávní, popř. trestněprávní odpovědnost.

1. vydání

©Ondřej Vencálek, 2020

©Univerzita Palackého v Olomouci, 2020

ISBN 978-80-244-5839-7

# Obsah

Slovo úvodem	4
Cvičení 1: práce s daty v Excelu (Google doodles)	6
Cvičení 2: očišťování od kalendářních variací a popis změn pomocí relativních přírůstků (produkce mléka a ceny másla)	8
Cvičení 3: lineární trend I (zalednění Arktidy)	10
Cvičení 4: lineární trend II a kvadratický trend (zalednění Arktidy)	12
Cvičení 5 – verze A: exponenciální trend (šíření „koronaviru“ SARS-CoV-2 v České republice v roce 2020)	14
Cvičení 5 – verze B: exponenciální trend (Moorův zákon)	16
Cvičení 6 – verze A: logistický trend (růst slunečnice)	18
Cvičení 6 – verze B: modifikovaný exponenciální trend (vývoj počtu obyvatel Japonska)	20
Cvičení 7: klouzavé průměry (kurz koruny vůči jenu)	22
Cvičení 8: exponenciální vyrovnávání (kurz koruny vůči jenu)	24
Cvičení 9: modelování sezónnosti (návštěvnost webových stránek)	26
Cvičení 10: model skrytých period (návštěvnost webových stránek)	28

## Slovo úvodem

Cvičení k předmetům věnovaným analýze časových řad vedu s kratšími přestávkami od roku 2009, kdy jsem nastoupil jako odborný asistent na Katedře matematické analýzy a aplikací matematiky Přírodovědecké fakulty Univerzity Palackého v Olomouci. Jejich dnešní podoba je určena převážně následujícími čtyřmi faktory:

- Práce v prostředí Excel.
- Samostatná příprava studentů dopředu na každé cvičení a následná diskuse o řešení v rámci cvičení.
- Analýza reálných dat, o kterých se domnívám, že jsou pro studenty zajímavá.
- Komunikace se studenty prostřednictvím uzavřené facebookové skupiny.

Uvádím tyto faktory v pořadí, v jakém jsem je do výuky zařazoval. Myslím, že každý z těchto faktorů si zaslouží alespoň krátký komentář.

Když jsem v září 2009 jako začínající odborný asistent dostal na starost cvičení k předmětu Pravděpodobnost a statistika 3, ve kterém se studenti měli seznámit se základy analýzy časových řad, dal jsem studentům na výběr, v jakém softwaru chtějí pracovat. Drtivá většina volila Excel. Dnes se studenti mnohem více setkávají s R-kem, ale tehdejší třetíáci na oborech matematika-ekonomie tento software prakticky vůbec neznali. S Excelem jsem zas neměl velké zkušenosti já. Zkusit v Excelu udělat to, co bych snadno zvládl v R-ku, byla pro mě výzva. Postupně jsem objevoval možnosti, o kterých jsem dříve nevěděl. A mnohdy mi pomohli a rozšířili obzory studenti. Stále jsem přesvědčen, že se s Excelem bude v budoucnu ve větší či menší míře potkávat naprostá většina dnešních studentů. Na současných studentech vidím, že jsou často překvapeni, kolik toho Excel umí; stejně, jako jsem byl překvapen já.

Samostatná příprava studentů dopředu na každé cvičení a následná diskuse o řešení v rámci cvičení je z mého pohledu velmi efektivním způsobem, jak studenty něco naučit. Zavést tento způsob práce pro mě přitom byla čirá nutnost, protože cvičení byla vždy pouze 45minutová. V tak krátkém čase je prakticky vyloučeno teprve hledat řešení zadávaných úloh. Studenti si vše mohou v klidu dopředu rozmyslet a vyzkoušet. Já během 45 minut, které mám k dispozici, stihnu pouze ukázat své „řešení“ a dát prostor pro několik krátkých dotazů. Skripta, která se vám nyní dostávají do rukou, nemají ambici tato společná setkání se studenty nahradit. Nicméně jejich součástí jsou i řešení úloh zde zadaných. Tato řešení mají podobu souborů Excelu. Nejdůležitějšív komentáře, které obvykle zazní v průběhu cvičení, jsem se snažil zaznamenat přímo do těchto dokumentů. Studenti je mohou zobrazit volbou *Zobrazit všechny komentáře* v nabídce Excelu *Revize*.

I když jsem se nikdy nevěnoval rozvoji metodologie analýzy časových řad, poměrně často jsem se musel vypořádávat s požadavkem analyzovat nějakou časovou řadu. A byly to většinou velmi zajímavé časové řady: týkaly se například sledovanosti televize, počtu onemocnění chřipkou nebo množství zakoupeného zboží určitého typu. Bohužel jsem tato data nemohl použít pro výuku, protože šlo o data neveřejná. Hledání veřejně dostupných časových řad, které by byly aktuální a přitom studenty zaujaly, jsem věnoval

značné úsilí. Aktuálnost časových řad je přitom klíčová. V mnoha cvičeních proto pracujeme s řadami, které každý rok aktualizujeme a výsledky se proto rok od roku mírně liší. Sami můžete porovnat své výsledky s výsledky, které získali studenti loni či předloni. Je třeba zde ještě upozornit, že odkazy na datové zdroje, které v této práci uvádím, se v budoucnu mohou měnit, a stávající odkazy budou nefunkční. V případě, že najdete v práci nefungující odkaz, zkuste dohledat aktuální umístění dat, případně pracujte s daty v souborech, které tvoří přílohu k této publikaci.

Na podzim 2016 jsem zadání cvičení začal studentům předávat prostřednictvím uzavřené facebookové skupiny, kterou jsem za tím účelem založil. Od té doby to tak dělám až dosud. Troufám si tvrdit, že rok 2020 znamenal přelom v „online“ komunikaci vyučujících se studenty. Je možné, že facebookovou skupinu brzy nahradí jiná forma komunikace. Uvidíme.

Skripta vznikla v rámci projektu Operačního programu Výzkum, vývoj a vzdělávání *Univerzita Palackého jako komplexní vzdělávací instituce* (CZ.02.2.69/0.0/0.0/16\_015/0002337). Rád bych toto skromné dílko věnoval Tomáši Ciprovi, u něž jsem se poprvé seznamoval s problematikou analýzy časových řad, a milému kolegovi Ivo Müllerovi, který kurz analýzy časových řad na Univerzitě Palackého v Olomouci přednášel. Právě on do značné míry určil obsah tohoto kurzu.

Při hledání příkladů, psaní skript i tvorbě řešení jsem především myslel na vás, studenty. Doufám, že se vám cvičení budou líbit a že se při řešení zadaných úloh hodně naučíte. Nabízené úlohy vám k tomu vytváří prostor, záleží však především na vaší aktivitě.

V Olomouci, v říjnu 2020  
Ondřej Vencálek

# Cvičení 1: práce s daty v Excelu (Google doodles)

## Cíl cvičení

Cílem cvičení je připomenout si (případně naučit se ovládat) základní nástroje prostředí Microsoft Excel pro práci s daty.

## Příběh a data

Možná jste si povšimli, že se čas od času objeví na stránce vyhledávače Google svátečně upravené logo této společnosti. Google na svých stránkách uvádí: *Sváteční loga (anglicky doodles) jsou zábavné, překvapující a někdy spontánní změny loga Google, které provádíme, abychom oslavili svátky, výročí a životy známých umělců, inovátorů či vědců.* Může nás zajímat, jak umístění takovéto „upoutávky“ na stránce Google zvýší zájem o určitou osobu či událost. Zájem přitom můžeme „měřit“ například počtem zobrazení stránky věnované dané osobě či události na Wikipedii.

## Zadání úkolu

1. Vyberte si osobnost, které byl v nedávné době věnován nějaký doodle.
2. Na Wikipedii vyhledejte data o počtu zobrazení stránky věnované této osobnosti (např. za posledních 90 dnů).
3. Data vykreslete do grafu. Zvažte použití logaritmického měřítko. Co vám může přinést?
4. Graf komentujte.

## Nápověda k jednotlivým krokům

1. Kliknutím na tlačítko „zkusím štěstí“ na stránce Google (bez předchozího vyplnění hledaného hesla) se dostaneme na stránku Google doodles. Výběr doodlu je ponechán vašemu zájmu, vybírejte však raději osobnost či událost, ke které existuje heslo na Wikipedii. Na stránku Google doodles se můžete dostat přímo: <https://www.google.com/doodles>.
2. Zadejme na stránkách Wikipedie heslo, které se v poslední době objevilo mezi doodles. V pravé horní části stránky pak zvolme „View history (Zobrazit historii)“ a poté „Pageviews (Statistika návštěvnosti)“. Zobrazme údaje za posledních 90 dnů. Data můžeme stáhnout např. ve formátu .csv, který lze otevřít v Excelu. Setkáváte-li se s formátem csv (Comma-Separated Values) poprvé, přečtěte si o něm například na Wikipedii. Narozdíl od formátu běžného pro Excel (.xlsx) jsou v tomto formátu uchovávána pouze samotná data (v souborech s příponou .xlsx je uloženo mnoho dalších informací – třeba velikost a barva buněk tabulky).
3. V Excelu je třeba data nejprve upravit. Protože střeoevropské nastavení formátu csv používá jako oddělovač středník, zatímco jinde se používá čárka, jsou data

v jednom sloupci místo ve dvou. Toto napravíme následovně: volme *Data – text do sloupců* (oddělovač čárka).

4. Pokud se rozhodnete pro logaritmické měřítko, zaveďte novou proměnnou (dekadický logaritmus původního počtu). Vykreslete do grafu. Rozmyslete si, co znamená změna logaritmu o jednotku pro původní veličinu.
5. V komentáři zkuste zodpovědět výše uvedenou otázku, jak se umístění doodlu na stránce Google projevilo na zájmu o vámi vybranou osobnost – jak moc se zvýšila návštěvnost jejích stránek na Wikipedii a jak dlouho zvýšený zájem trval.



## Cvičení 2: očišťování od kalendářních variací a popis změn pomocí relativních přírůstků (produkce mléka a ceny másla)

### Cíl cvičení

Toto cvičení má dvě témata a dva cíle. Cílem první části je vyzkoušet si očišťování časové řady od kalendářních variací. Cílem druhé části cvičení je provést výpočet relativních změn a rozmyslet si jejich interpretaci.

### Příběh a data

Dvě krátké kapitoly z teorie analýzy časových řad (očišťování od kalendářních variací a popis změn pomocí relativních přírůstků) propojíme v tomto cvičení tématicky – obě datové sady, které budeme analyzovat, se týkají mléka a mléčných produktů. V první části tohoto cvičení budeme analyzovat data týkající se produkce mléka v USA. Tato data každý měsíc zveřejňuje Ministerstvo zemědělství USA na stránce <https://www.ers.usda.gov/data-products/dairy-data/>.

V druhé části budeme analyzovat časovou řadu cen másla v České republice. O tuto časovou řadu jsem se začal zajímat v říjnu 2016, kdy ceny másla už čtvrtým měsícem rostly a, i když oficiální čísla udávají průměrnou cenu másla v tomto měsíci 161 Kč za kilogram, koupit čtvrtkilové balení v supermarketu pod 50 Kč bylo téměř nemožné.

### Zadání úkolu

#### Produkce mléka v USA

1. Aktualizujte přiloženou datovou sadu v souboru `ProdukceMlekaUSA.csv` obsahující údaje o produkci mléka v USA o nejnovější údaje podle výše uvedené webové stránky, případně zde po zadání hesla „milk production“.
2. Vykreslete data nejprve celá a poté jejich část začínající rokem 2010, dále pracujte jen s touto zkrácenou časovou řadou.
3. Všimněte si jednotek a převed'te na jednotky běžně používané v naší zemi.
4. Rozmyslete si, že jde o intervalovou časovou řadu a proved'te očištění od kalendářních variací. Volte přitom délku standardního měsíce 30 dní a poté  $365/12$  dní. Porovnejte graficky s původními daty.

#### Cena másla v ČR

1. Ze stránek Českého statistického úřadu <https://vdb.czso.cz/vdbvo2/> stáhněte data o vývoji cen másla v ČR (sekce Statistiky, Ceny a inflace, Spotřebitelské ceny vybraných druhů zboží a služeb).
2. Kdy během posledních let docházelo k nejprudšímu růstu/poklesu cen másla?

3. V médiích se vyskytlo tvrzení, že cena másla v létě vždy dočasně vzroste, protože krávy nadojí méně mléka, navíc s nižším obsahem tuku. Zkuste z dat posoudit, zda opravdu cena másla pravidelně v létě roste. Navrhněte, jaký graf by tuto skutečnost (pokud je tento efekt skutečný) dobře vystihl.

# Cvičení 3: lineární trend I (zalednění Arktidy)

## Cíl cvičení

Cílem cvičení je odhadnout parametry lineárního trendu a provést predikci v časové řadě s lineárním trendem.

## Příběh a data

O změnách globálního klimatu se hovoří již dlouho. V posledních desetiletích sledujeme například zmenšující se plochu zaledněných oblastí Arktidy. Naopak v poměrně nedávné době jsme mohli téměř pravidelně číst o rekordním zalednění Antarktidy (2012, 2013, 2014).

V rámci tohoto cvičení prozkoumáme data o zalednění polárních oblastí. Jsou volně k dispozici na stránkách National Snow & Ice Data Center. Na stránce věnované rozsahu zalednění najdeme mimo popis dat také odkaz na FTP, kde jsou na adrese `ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/north/monthly/data/N_10_extent_v3.0.csv` uložená data. Budeme zpracovávat data o zalednění Arktidy za měsíc říjen z let 1979-2019.

## Zadání úkolu

1. Stáhněte si data do Excelu. Zamyslete se nad tím, proč bereme data jen za určitý měsíc v roce.
2. Přečtěte si, jaký je rozdíl mezi sloupci `extent` a `area`. Kterou proměnnou byste si vybrali pro analýzu časové řady?
3. Budeme pracovat se sloupcem `extent`. Vykreslete data.
4. Odhadněte parametry lineárního trendu – vyzkoušejte alespoň dva různé postupy uvedené níže a zkontrolujte, zda dostanete stejné výsledky (doporučuji začít prvním a třetím z níže uvedených postupů):

- pomocí vzorců  $\hat{\beta}_1 = \frac{\sum ty_t - n\bar{t}\bar{y}}{\sum t^2 - n\bar{t}^2}$ ,  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{t}$ ,
- pomocí maticového násobení, kdy použijeme vzorec  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ ,
- pomocí funkce Excelu `LINREGRESE`,
- pomocí tzv. **Analytických nástrojů** v Excelu. Návod k jejich aktivaci naleznete při zadání hesla **Analytické nástroje** v Excelu. Po aktivaci **Analytických nástrojů** zvolte v nabídce **Data** možnost **Analýza dat** a vyberte položku **Regrese**.

5. Interpretujte odhadnuté parametry (řekněte, co znamenají). Interpretace absolutního členu bude záležet na tom, zda roky číslujete 1, 2, ... nebo 1979, 1980, ... Zkuste obě možnosti a parametry porovnejte. Co platí pro směrnici při změně číslování let?

6. Do grafu s daty vykreslete přímkou znázorňující trend. Zkuste to nejprve bez použití speciálních funkcí, poté pomocí toho, že pravým tlačítkem kliknete na body časové řady v grafu a zvolíte „Přidat spojnicí trendu“.
7. Vypočtete bodovou predikci rozsahu zalednění pro říjen 2020.

## Nápověda

U maticového násobení, stejně jako při použití funkce LINREGRESE bude výstup v několika buňkách. Ty je nutno předem označit a „vzorečky“ pak potvrzovat nikoliv pouhým stisknutím klávesy Enter, ale kombinací kláves Ctrl+Shift+Enter. Pokud vám maticové násobení nepůjde, netrapte se tím, ukážeme si celý postup na cvičení.

## Pokračování příště

K lineárnímu trendu a analýze dat o zalednění se budeme vracet ještě v dalším cvičení, zkonstruujeme predikční interval a budeme se zabývat otázkou, zda je pozorovaný pokles plochy zalednění statisticky významný. Budeme rovněž hledat další modely pro tuto časovou řadu, např. kvadratický trend, a posuzovat, který z modelů je vhodnější.

# Cvičení 4: lineární trend II a kvadratický trend (zalednění Arktidy)

## Cíl cvičení

Předmětem cvičení je diagnostika vhodnosti modelu pomocí residuí, testování hypotéz o parametrech trendu a konstrukce predikčního intervalu pro časovou řadu s lineárním trendem, dále pak odhad parametrů kvadratického trendu, predikce v tomto modelu a porovnání vhodnosti kvadratického a lineárního trendu pro analyzovanou časovou řadu.

## Zadání úkolu

Pokračování analýzy předpokládající lineární trend:

1. Vypočítejte residua v modelu s lineárním trendem z předešlého cvičení. Vykreslete residua do grafu. Graf komentujte, tj. vyjádřete se ke splnění předpokladů lineárního regresního modelu.
2. Všimněte si variability odchylek od přímky (velikosti residuí). Jak souvisí tato variabilita s parametrem  $\sigma$ ? Odhadněte tento parametr.
3. V minulém cvičení jsme odhadli, že rozsah zalednění v průběhu sledovaného období klesal. Odpovězme na otázku, zda je pozorovaný pokles plochy zalednění statisticky významný.
4. Zkonstruuje intervalovou predikci (při lineárním trendu) rozsahu zalednění pro říjen 2020.

Analýza předpokládající kvadratický trend:

1. Na časové řadě zalednění severního pólu odhadněte parametry kvadratického trendu. K odhadu parametrů můžete použít maticové násobení nebo některou z funkcí Excelu.
2. Trend zakreslete společně s daty do grafu a porovnejte s lineárním trendem.
3. Vypočítejte hodnoty residuí v tomto modelu a posuďte na základě jejich grafu splnění předpokladů modelu.
4. Pro kvadratický trend zkonstruuje bodovou predikci na rok 2020 a srovnajte ji s predikcí při použití lineárního trendu. Zkonstruuje také intervalovou predikci.
5. Srovnajte modely lineárního a kvadratického trendu: vypočtete koeficient determinace pro oba modely a porovnejte je. Zkuste hodnotu koeficientu determinace vypočítat samostatně a „zkontrolovat si ji“ porovnáním s hodnotou, kterou vrací Excel (např. při tzv. přidání spojnice trendu do grafu je možno zaškrtnout položku pro vrácení hodnoty  $R^2$ ). Jak byste rozhodli, který z modelů (lineární trend nebo kvadratický trend) je pro data o zalednění lepší? Je koeficient determinace k tomuto rozhodnutí vhodný? Zdůvodněte.

## Nápověda

- V prvním úkolu analyzujeme residua modelu. Residua jsou vlastně odhady náhodné složky. Proto na jejich základě můžeme posuzovat předpoklady činěné o náhodné složce. Zejména předpoklad nezávislosti jednotlivých náhodných odchylek by mohl být porušen. To by nás vedlo k úpravě předpokládaného trendu. Cílem tohoto cvičení je uvědomit si, že analýza residuí není jen formalitou, ale může nám pomoci při hledání vhodného modelu.
- Připomeňte si:  $\text{var}\epsilon = E(\epsilon - E\epsilon)^2 = E(\epsilon^2) - (E\epsilon)^2 = \sigma^2$ . Nepřekvapí nás proto, že parametr  $\sigma^2$  můžeme odhadnout jako „průměr“ druhých mocnin residuí  $e_1, \dots, e_n$ :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n e_i^2.$$

- Při posuzování signifikance (statistické významnosti) poklesu testujeme hypotézu nulovosti směrnice přímky trendu  $\beta_1 = 0$ . Testová statistika samozřejmě vychází z odhadu tohoto parametru, tedy je-li  $|\hat{\beta}_1| > k$ , kde  $k$  je vhodně zvolená konstanta, budeme hypotézu nulovosti zamítat. K určení konstanty  $k$  využijeme, co víme o vlastnostech odhadu  $\hat{\beta}$ . Jelikož

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}),$$

bude mít testová statistika podobu

$$t = \frac{\hat{\beta}_1}{s\sqrt{v_{2,2}}},$$

kde  $v_{22}$  je prvek na 2. řádku ve 2. sloupci matice  $(\mathbf{X}'\mathbf{X})^{-1}$ . Tato testová statistika má za platnosti nulové hypotézy t-rozdělení o  $n-2$  stupních volnosti. Vypočítejte tuto statistiku. Určete  $p$ -hodnotu testu nulovosti parametru  $\beta_1$  pomocí příkazu T.DIST a rozhodněte o platnosti nulové hypotézy.

- Při konstrukci predikčních intervalů bude třeba kvantilů t-rozdělení; využijte příkaz T.INV.

# Cvičení 5 – verze A: exponenciální trend (šíření „koronaviru“ SARS-CoV-2 v České republice v roce 2020)

## Cíl cvičení

Cílem cvičení je pochopit vlastnosti exponenciálního růstu. Odhadneme parametry exponenciálního trendu, interpretujeme jeho parametry a provedeme predikce v časové řadě s exponenciálním trendem.

## Příběh a data

V říjnu 2020 vyšla kniha Vojtěcha Gibiše a Michala Kubala *Pandemie*, která mimo jiné odhalila identitu „muže, který přiměl české úřady k rázné akci“ – člověka, jehož predikce počtu nakažených virem SARS-CoV-2 v březnu 2020 vedla k přísným omezením společenského života v České republice (11. března 2020 uzavření škol, o den později vyhlášení nouzového stavu, viz Wikipedia).

Přehlednou excelovou tabulku shrnující modely vývoje počtu infikovaných v ČR, na jejichž základě tehdy vláda ČR přistoupila k výše zmíněným restrikcím, naleznete zde. Predikce byla učiněna na základě počtu potvrzených případů v prvních sedmi dnech března 2020. V těchto dnech byly, dle této tabulky, zjištěny následující počty infikovaných:

3, 4, 5, 8, 12, 19, 26, 32.

## Zadání úkolu

1. Přečtěte si výše citovaný článek o „tajemném muži, který na jaře 'zachránil Česko'“.
2. Uvedenou časovou řadu vykreslete do grafu. Zkuste také graf v logaritmickém měřítku. Porovnejte údaje z tabulky s oficiálními údaji ÚZIS ČR dostupnými v sekci *otevřená data* na stránce COVID-19: Přehled aktuální situace v ČR.
3. Odhadněte parametry exponenciálního trendu. Vyzkoušejte různé metody odhadu:
  - metodu logaritmické transformace,
  - metodu vybraných bodů pro různé volby vybraných bodů,
  - některou z metod nabízených Excelem, např. pomocí přidání spojnice trendu v grafu.
4. Zamyslete se nad interpretací parametrů.
5. Představte si, že je 8. března 2020 ráno, a proved'te predikci počtu nakažených po 7. březnu 2020.
6. Za jak dlouho se dle modelu počet nakažených zdvojnásobí? Za jak dlouho se zdesetinásobí?

## Poznámky

- Doplňte si představu o exponenciálním růstu četbou příběhu (pohádky) o počátcích šachu.
- Predikce, kterou viděli vládní činitelé po 8. březnu 2020, končila 1. dubna 2020. Protáhněte tuto predikci dále. Zamyslete se nad tím, jak dlouhá predikce má smysl.
- Porovnejte predikci na březen 2020 se skutečně zjištěnými počty (viz data ÚZIS ČR). Byly predikce na období do 20.3. správné? (Vliv restrikcí není okamžitý a projeví se údajně až po dvou týdnech.)
- Zkuste se zamyslet, zda záleží na tom, je-li pro logaritmickou transformaci použit logaritmus přirozený, dekadický, dvojkový či jiný. Možná, že použití logaritmu o základu 2 (resp. 10) usnadní zodpovězení poslední otázky (zdvojnásobení, resp. zdesetinásobení, počtu nakažených).
- Uvědomte si, že počty zveřejňované ÚZIS ČR, ze kterých v tomto cvičení vycházíme, jsou počty pozitivně testovaných jedinců. Skutečný počet nakažených může být dosti vyšší, jak nasvědčují tzv. séroprevalenční studie.
- V říjnu 2020 v ČR exponenciálně narůstá počet úmrtí na COVID-19, jak vyplývá z grafu publikovaného Jaroslavem Flégrem, viz tento článek. Zájemci mohou na aktuálních datech o počtu zemřelých odhadnout parametry trendu v této řadě a zamyslet se nad predikcí.



# Cvičení 5 – verze B: exponenciální trend (Moorův zákon)

## Cíl cvičení

Cílem cvičení je pochopit vlastnosti exponenciálního růstu. Odhadneme parametry exponenciálního trendu a interpretujeme jeho parametry.

## Příběh a data

Moorův zákon je empirické pravidlo, které roku 1965 vyslovil chemik a spoluzakladatel firmy Intel Gordon Moore. Původní znění bylo: „počet tranzistorů, které mohou být umístěny na integrovaný obvod se při zachování stejné ceny zhruba každých 18 měsíců zdvojnásobí.“ Takovýto růst se nazývá exponenciální.

Složitost dnešních procesorů se poměruje především počtem tranzistorů v nich zapojených. Rychlost růstu počtu tranzistorů na plošné jednotce se časem zpomalila a nyní se jejich počet zdvojnásobuje přibližně jednou za dva roky.

[https://cs.wikipedia.org/wiki/Moore%C5%AFv\\_z%C3%A1kon](https://cs.wikipedia.org/wiki/Moore%C5%AFv_z%C3%A1kon)

## Zadání úkolu

1. Stáhněte si data o počtu tranzistorů v mikroprocesorech ze stránky [https://en.wikipedia.org/wiki/Transistor\\_count](https://en.wikipedia.org/wiki/Transistor_count) (jde o první tabulku na této stránce).
2. Uvědomte si, že takto získaná data tvoří časovou řadu v pravém slova smyslu, protože pro jeden časový údaj máme k dispozici několik pozorování. Není vlastně zřejmé, co je sledovanou veličinou (jaký je klíč pro zařazení mikroprocesoru do tabulky na Wikipedii). Jde o nejvýkonnější procesor uvedený na trh v nějakém období? Vezměte za každý rok maximální hodnotu.
3. Časovou řadu vykreslete do grafu. Zkuste také graf v logaritmickém měřítku (původní hodnoty zlogaritmujte a vykreslete graf těchto hodnot).
4. Odhadněte parametry exponenciálního trendu a zamyslete se nad jejich interpretací.
5. Platí Moorův zákon? Odpovězte na základě provedených analýz.

## Nápověda

- Metoda odhadu parametrů vychází z logaritmické transformace a následného použití metody nejmenších čtverců.
- Zamyslete se, zda záleží na tom, je-li použit logaritmus přirozený, dekadický, dvojkový či jiný. Možná, že použití logaritmu o základu 2 usnadní zodpovězení poslední otázky, zda platí Moorův zákon.

- Na Wikipedii si přečtete vyjádření Gordona Moora z roku 1995, kdy o platnosti zákona, který je po něm pojmenován, uvedl: „Tak to nemůže pokračovat napořád. Povaha exponenciál je taková, že je tlačíme mimo limity a následně nastane pohroma“. Exponenciální růst málokdy trvá dlouho.

# Cvičení 6 – verze A: logistický trend (růst slunečnice)

## Cíl cvičení

Cílem cvičení je osvojit si model logistického trendu a vyzkoušet si odhad parametrů tohoto trendu a interpretaci jeho parametrů.

## Příběh a data

Před sto lety byla zveřejněna data obsahující záznamy o růstu slunečnice, která najdete v souboru `Slunecnice.csv`:

Reed, H.S. and Holland, R. H., *The Growth of an Annual Plant Helianthus*, *Proceedings of the National Academy of Sciences (USA)*, 5, 135-144 (1919).

Růst bývá dosti často zprvu pomalý, postupně se zrychluje, ale po dosažení jisté rychlosti se už rychlost růstu nezvětšuje, ba naopak, růst je čím dál tím pomalejší, až pozvolna ustává úplně. Tomuto popisu odpovídá logistický trend.

## Zadání úkolu

1. Odhadněte parametry logistického trendu

$$T_t = \frac{k}{1 + \alpha\beta^t}$$

- (a) metodou vybraných bodů (vybrané body první, šestý a jedenáctý:  $y_t, y_{t+m}, y_{t+2m}$ ); rozmyslete si, čemu se rovnají hodnoty  $t$  a  $m$ ,
  - (b) metodou vybraných bodů (pro jinou trojici vybraných bodů),
  - (c) metodou inverzní transformace následovanou odhadem parametrů modifikovaného exponenciálního trendu buďto opět metodou vybraných bodů nebo metodou postupných součtů.
2. Který ze tří předešlých odhadů byste vybrali jako „nejlepší“? Proč?
  3. Kdy rostla slunečnice nejrychleji? Odhad tohoto času dle modelu porovnejte s pozorovanými přírůstky.



## Cvičení 6 – verze B: modifikovaný exponenciální trend (vývoj počtu obyvatel Japonska)

### Cíl cvičení

Cílem cvičení je osvojit si model modifikovaného exponenciálního trendu a vyzkoušet si odhad jeho parametrů.

### Příběh a data

Budeme analyzovat údaje o počtu obyvatel Japonska za roky 1971 až 2000, které můžete najít na Wikipedii.

### Zadání úkolu

- Ukažte, že se tato časová řada řídí *modifikovaným (posunutým) exponenciálním trendem*:

$$T_t = k + \alpha\beta^t.$$

Nápověda: vyšetřete chování prvních diferencí, tj. hodnot  $y_t - y_{t-1}$ ; vykreslete graf prvních diferencí. Rozmyslete si, co platí pro první difference při modifikovaném exponenciálním trendu a vizuálně ověřte, zda studovaná řada tuto vlastnost má.

- Než se dáte do odhadu parametrů modifikovaného exponenciálního trendu, rozmyslete si, jaké hodnoty parametrů očekáváte.
- Odhadněte parametry modifikovaného exponenciálního trendu a interpretujte je. Můžete zkusit porovnat odhady získané různými metodami – metodou vybraných bodů (třeba i pro různé vybrané body) a metodou postupných součtů.
- Vykreslete residua a graf residuí interpretujte.



# Cvičení 7: klouzavé průměry (kurz koruny vůči jenu)

## Cíl cvičení

Cílem cvičení je provést vyhlazení datové řady kurzu dvou měn pomocí klouzavých průměrů a predikce v modelu lokálních trendů.

## Příběh a data

Kurz české měny k cizím měnám vyhlašuje Česká národní banka (ČNB), a to ve formě *kurzů devizového trhu* (pro běžně obchodované měny) a ve formě *kurzů ostatních měn*. V řešení úlože budeme analyzovat vývoj kurzu české koruny a japonského jenu za rok 2019, můžete však analyzovat kurz k jiné měně, která vás zajímá více, a za jiný (např. aktuálně probíhající) rok.

## Zadání úkolu

1. Ze stránky ČNB stáhněte data o vývoji kurzu koruny a japonského jenu (JPY) za rok 2019.
2. Proveďte vyhlazení časové řady metodou prostých klouzavých průměrů při různé délce vyhlazovacího okna (volte např. délku 5 a 15). Podívejte se, jak se s měnící se délkou vyhlazovacího okna mění „hladkost“ vyhlazení.
3. Porovnejte vámi získaný výsledek s tím, který získáte při použití funkce Excelu „Přidat spojnicí trendu – klouzavý průměr“.
4. Při délce vyhlazovacího okna 5 vypočtete také vážené klouzavé průměry a porovnejte vyhlazení s vyhlazením prostými klouzavými průměry při stejné délce vyhlazovacího okna.
5. Proveďte (pomocí klouzavých průměrů) předpověď na týden 22. 7.–26. 7. (tj. představte si, že je 21. 7. 2019 a vy chcete predikovat vývoj kurzu v následujícím týdnu na základě dat dostupných k 21. 7. 2019). Podobně vyzkoušejte predikce na týden 29. 7.–2. 8. (učiněné dne 28. 7.). Zhodnoťte kvalitu svých předpovědí.

## Bonusová úloha

Prozkoumejte kurz koruny vůči euru za rok **2016**. Ten byl poněkud nezajímavý, neboť se příliš neodchyloval od úrovně 27 Kč za 1 euro (kvůli regulaci ČNB). Zamyslete se, jak byste charakterizovali (číselně vyjádřili) variabilitu jednotlivých řad (řady kurzu koruny vůči jenu a řadu kurzu koruny vůči euru, případně spočtete i pro jiné měny, např. americký dolar).





## Cvičení 8: exponenciální vyrovnávání (kurz koruny vůči jenu)

### Cíl cvičení

Cílem cvičení je provést vyhlazení datové řady kurzu dvou měn pomocí jednoduchého exponenciálního vyrovnávání a predikce v tomto modelu.

### Příběh a data

Opět budeme pracovat s daty ze stránky ČNB týkající se vývoje kurzu koruny a japonského jenu (JPY) za rok 2019.

### Zadání úkolu

1. Proveďte vyhlazení časové řady metodou jednoduchého exponenciálního vyrovnávání při volbě vyhlazovací konstanty  $\alpha = 0,7$ .
2. Proveďte dále vyhlazení pro další volby vyhlazovací konstanty, např. 0,72, 0,74, ..., 0,98. Podívejte se, jak se s měnící se hodnotou  $\alpha$  mění „hladkost“ vyhlazení.
3. Jakou hodnotu  $\alpha$  byste pro tato data zvolili a proč?
4. Proveďte předpověď na týden 22. 7. – 26. 7. (tj. představte si, že je 21. 7. 2019 a vy chcete predikovat vývoj kurzu v následujícím týdnu na základě dat dostupných k 21. 7. 2019). Podobně vyzkoušejte predikce na týden 29. 7. – 2. 8. (z času 28. 7.). Zhodnoťte kvalitu svých předpovědí.



# Cvičení 9: modelování sezónnosti I (návštěvnost webových stránek)

## Cíl cvičení

Cílem cvičení je analyzovat časovou řadu týkající se návštěvnosti webových stránek, která vykazuje značnou periodicitu.

## Příběh a data

Soubor `NavstevnostMK.csv` obsahuje denní záznamy o počtech zobrazení francouzské, anglické a české stránky Wikipedie věnované spisovateli Milanu Kunderovi za období 2. ledna až 15. března 2015. V tomto cvičení budeme používat data o návštěvnosti anglické verze za období pěti týdnů 9. 2.–15. 3. 2015. Můžete zkusit analyzovat také aktuální data o návštěvnosti vaší oblíbené stránky na Wikipedii (viz první cvičení).

## Zadání úkolu

1. Vykreslete data a povšimněte si týdenní periodicity.
2. Týdenní periodicitu můžete jasně vizualizovat grafem, v němž na x-ové ose je pořadové číslo dne v týdnu. Použijte funkci `DENTÝDNE`.
3. Rozmyslete si, zda použít model konstantní či proporcionální sezónnosti. Rozmyslete, jaký trend použít, případně vyzkoušejte různé trendové funkce.
4. Odhadněte parametry zvoleného modelu a interpretujte je.
5. Spočítejte a vykreslete modelové hodnoty (fitted values) a zakreslete je do grafu s daty.
6. Spočítejte a vykreslete residua, graficky ověřujte vhodnost modelu.

Další úkoly pro zájemce:

- Zkuste předešlý postup použít na celou časovou řadu. Jak moc se změní odhady parametrů?
- Změňte typ grafu zkonstruovaného v druhém kroku na „krabicový graf“ a vyhodnoťte přítomnost odlehlých pozorování (outlierů).
- Odhadněte znovu parametry modelu na všech datech po vynechání odlehlých pozorování a srovnajte opět odhady parametrů.

## Nápověda

- Uvažujte model konstantní sezónnosti s konstantním trendem.
- K vytvoření sloupců s pomocnými vysvětlujícími proměnnými můžete využít příkazu `KDYŽ`. Rozmyslete si, jak budou tyto pomocné proměnné vypadat a kolik jich bude.
- Sloupce s pomocnými proměnnými lze vytvořit několika způsoby. Můžeme například uvažovat omezení na efekty jednotlivých dnů v podobě

$$\sum_{j=1}^7 S_{i,j} = 0.$$

V tom případě můžeme efekt sedmého dne  $S_{i,7}$  vyjádřit pomocí předchozích šesti efektů, neboť platí  $S_{i,7} = -S_{i,1} - \dots - S_{i,6}$ . Omezení používané např. v R-ku má podobu  $S_{i,1} = 0$ .

## Cvičení 10: modelování sezónnosti II – model skrytých period (návštěvnost webových stránek)

### Cíl cvičení

Cílem cvičení je vyzkoušet si modelování periodicity pomocí modelu skrytých period a porovnat tento přístup s modelem konstantní sezónnosti použitým v minulém cvičení.

### Příběh a data

Stejně jako v minulém cvičení, i nyní budeme používat data o návštěvnosti anglické verze stránky Wikipedie věnované Milanu Kunderovi za období pěti týdnů 9. 2.–15. 3. 2015. Také tentokrát můžete zkusit analyzovat aktuální data o návštěvnosti vaší oblíbené stránky na Wikipedii.

### Zadání úkolu

Najděte vhodný model popisující chování výše uvedené časové řady. Zkuste najít co nejlepší model s co nejméně parametry. Jaký model byste volili, budete-li mít k dispozici pouze 3 parametry? A jaký s 5 parametry?

1. Zkuste uvažovat model skrytých period.
2. Nakreslete periodogram a vyberte významné periody.
3. Dosazením odhadnutých hodnot spočtete modelové hodnoty a vykreslete je do grafu s daty.
4. Vypočtete residua a graficky ověřujte vhodnost modelu.

**Mgr. Ondřej Vencálek, Ph.D.**

**Časové řady časně a řádně**

Výkonný redaktor prof. RNDr. Karel Lemr, Ph.D

Odpovědný redaktor Bc. Otakar Loutocký

Technická redakce Mgr. Ondřej Vencálek, Ph.D.

Návrh a grafické zpracování obálky Bc. Karina Pavlíková, Ivana Perůtková

Publikace neprošla redakční úpravou ve VUP

Vydala Univerzita Palackého v Olomouci

Křížkovského 8, 771 47 Olomouc

[www.vydavatelstvi.upol.cz](http://www.vydavatelstvi.upol.cz)

[www.e-shop.upol.cz](http://www.e-shop.upol.cz)

[vup@upol.cz](mailto:vup@upol.cz)

1. vydání

Olomouc 2020

Neprodejná publikace

ISBN 978-80-244-5839-7 (online: iPDF)

VUP 2020/0441