

Časové rady: domáca úloha 4

Termín odovzdania: 16. 12. 2024

Odovzdávanie:

- Úlohu rieši každý samostatne. Za odpísané úlohy je nula bodov - pre tých, ktorí úlohu odpísali aj pre tých, ktorí ju dali odpísať. Rovnako bude nula bodmi hodnotený aj príklad, kde sa dá vidieť spolupráca - v zadaní je povedané, že úlohu rieši každý samostatne.
- Odovzdávanie: mailom na adresu **beata.ulohy@gmail.com**, s predmetom **CR 2024 - DU4 - meno** - v rovnakej forme ako v prvej domácej úlohe
- Ak použijete nejaký kód, výpočet, tvrdenie a pod., ktoré nájdete na internete, treba ho citovať - uviesť na konci riešenia odkaz a odvolať sa naň na príslušnom mieste v texte. V prípade použitia umelej inteligencie treba v maili okrem úlohy uviesť aj kompletný priebeh jej použitia - všetky vaše otázky a všetky odpovede UI (v pdf alebo txt formáte) a informáciu o tom, akú konkrétnu UI ste použili.
- Rezevácia dát: <http://bit.ly/4en5NZW>

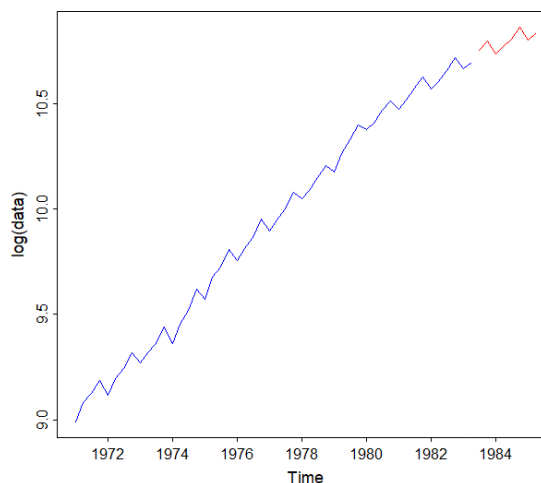
Príklad 1 [10 bodov]. Budeme pracovať s dátami z knižnice **Ecdat** zobrazenými na obrázku vľavo. Posledné dva roky použijeme na porovnanie s predikciami. Modelovať a robiť predikcie budeme pre **logaritmy týchto dát**. Na obrázku vpravo sú modrou farbou vyznačené hodnoty, pomocou ktorých budeme hľadať model a červenou tie, ktoré budeme porovnávať s predikciami.

```
> library(Ecdat)
> data("IncomeUK")
> IncomeUK[, 2]
```

	Qtr1	Qtr2	Qtr3	Qtr4
1971	8016	8798	9184	9755
1972	9113	9837	10337	11135
1973	10632	11144	11650	12586
1974	11643	12750	13608	15068
1975	14369	15944	16758	18140
1976	17212	18221	19217	21025
1977	19812	20901	21984	23781
1978	23068	24028	25496	27056
1979	26244	28610	30593	32709
1980	32130	33027	35077	36761
1981	35314	36924	39002	41099
1982	38815	40771	42486	45090
1983	42795	43955	46665	48792
1984	45938	47553	49119	52063
1985	49095	50481		

dáta na modelovanie

na zhodnotenie predikcií



- (a) [1 bod] Čo vyjadrujú dáta, s ktorými pracujeme?
- (b) [3 body] V dátach je trend, preto je jasné, že ich budeme aspoň raz diferencovať, klasicky alebo sezónne. Určte rád a typ diferencovania a popíšte všetky testy, ktoré k tomuto záveru viedli. Uveďte vždy, na aké dáta ste ho použili, parametre testu a ich zdôvodnenie a záver testu. Treba pritom myslieť na to, že jednotkový koreň môže byť aj sezónny a váš test musí byť schopný odhaliť ho. Výsledkom je časový rad bez trendu a bez jednotkového koreňa.

- (c) [2 body] Nájdite vhodný SARIMA model. Nemusíte popisovať postup, ako ste sa k nemu dopracovali, stačí uviesť typ modelu, zaradenie konštanty a jej zdôvodnenie (iba v prípade, že sú dáta diferencované len raz) a testovanie rezíduí.
- (d) [3 body] Zapište rovnicu pre získaný model. Ukážte, že je stacionárny a invertovateľný.
- (e) [1 bod] Spravte predikcie na dva roky a porovnajte so skutočnými hodnotami.

Alternatíva za dva bonusové body. Spravte hľadanie SARIMA modelu podľa jednotlivých bodov na vlastných aktuálnych dátach - takých, ktoré končia v roku 2023 alebo 2024. Požiadavky na dáta:

- Sezónne dáta - teda model obsahuje sezónne diferencie, sezónny AR alebo sezónny MA člen (môže mať aj viaceré z týchto vlastností) a nie sú do modelu zaradené zbytočne. Výrazná sezónnosť je v dátach viditeľná - ak pre ne nájdete model s dobrými rezíduami, môžete ich použiť.
- Originálnosť - nemôžu byť z tej istej R-kovskej knižnice alebo tej istej webstránky, ako už použil niekto pred vami (rozhoduje rezervácia dát v tabuľke).
- Nemôže ísť o dáta z prednášok alebo z cvičení.

Príklad 2 [15 bodov]. Nájdite si vlastné dáta pre ARIMA modelovanie. Požiadavky na dáta:

- Dáta bez sezónnosti.
 - Nemôže ísť o dáta z prednášok alebo z cvičení.
 - Každý má iný časový rad.
 - **Nepovinné - za 2 bonusové body:** Aktuálne dáta, ktoré končia v roku 2023 alebo 2024. V tomto prípade je znovu ďalšou požiadavkou originálnosť - nemôžu byť z tej istej R-kovskej knižnice alebo tej istej webstránky, ako už použil niekto pred vami (rozhoduje rezervácia dát v tabuľke).
- (a) [3 body] Vysvetlite, aké dáta modelujete, ak treba, zdôvodnite logitnickú transformáciu a zobrazte priebeh dát. Uveďte zdroj dát - slovný popis (z akej sú databázy a pod.) a odkaz na webstránku, resp. balík pre R a názov dát v tomto balíku. Pred ďalším modelovaním vynechajte niekoľko posledných pozorovaní, aby ste ich mohli použiť na porovnanie predikcií so skutočnosťou. Napíšte, aké dáta ste si takto odložili.
- (b) [3 body] Rovnaké zadanie ako v predchádzajúcej domácej úlohe - tento prípad pošlem opravený v najbližších dňoch, aby ste túto úlohu robili s už skontrolovanými cvičnými dátami.

Podrobne rozpíšte postup pri testovaní, či a koľkokrát treba dáta diferencovať. Štruktúra bude nasledovná:

Pre každý uvažovaný časový rad (teda pôvodný časový rad, v prípade potreby diferencie, druhé diferencie) uveďte:

- Ktorá z nasledovných možností nastáva:
 - Časový rad má lineárny trend, treba ho diferencovať a ideme testovať jeho diferencie.
 - Časový rad nemá lineárny trend, ideme testovať prítomnosť jednotkového koreňa.
- Napíšte a zdôvodnite typ ADF testu jednotkového koreňa.
- Skontrolujte, že použitý počet lagov v odhadnutej regresii nie je maximálny možný, ktorý ste povolili.

- Porovnajte testovaciu štatistiku s kritickou hodnotou a napíšte, čo tento výsledok znamená pre modelovanie.

Záver musí byť, že testovaný časový rad (pôvodný alebo diferencie) nemá ani trend, ani jednotkový koreň. To znamená, že sa preň bude hľadať ARMA model.

- (c) **[3 body]** Podrobne rozpíšte posledný ADF test, ktorý ste robili.
- Na základe výstupu napíšte regresiu, ktorá sa odhadla.
 - Aká hypotéza o parametroch tejto regresie sa testuje?
 - Odvodte, že táto hypotéza je ekvivalentná s hypotézou o jednotkovom koreni.
- (d) **[3 body]** Nájdite vhodný model pre vaše dáta. Ku grafu s p-hodnotami z R-ka doplňte tabuľku, v ktorej budú tieto p-hodnoty vypísané. Zapište odhadnutú rovnicu a ukážte, že je model stacionárny a invertovateľný. Korene zodpovedajúce tomuto overovaniu znázorníte graficky. Ak sú dáta jedenkrát diferencované, zdôvodnite zaradenie, resp. nezaradenie konštantného člena do modelu.
- (e) **[3 body]** Spravte predikcie a porovnajte ich so skutočnými hodnotami. Zmestili sa do intervalov spoľahlivosti?