

VOŠ a SPŠE Plzeň



Daniel Hajžman
hajzmand@spseplzen.cz

Semestrální práce z předmětu SZD

1 Zadání

V této semestrální práci se zabýváme základní analýzou časových řad (tj. očištění, vizualizace, tvorba modelu a predikce), konkrétně ročními daty popisující počty cestujících přepravovaných veřejnou autobusovou dopravou v období 1995–2022 v ČR. Cílem modelu je předpověď počtu cestujících, kteří budou přepraveni autobusovou dopravou na příštích několik období, tj. 1, 2 a 3 roky, analýza či úspěšnost předpovědí.

2 Data

Data byla získaná ze stránek ČSÚ, konkrétně nás zajímala data z let 1995–2019. Tyto data jsou přehledně zpracovaná v excelovském souboru *SZD_Hajžman-SP.xlsx* na listu *data*, v tomto souboru se také nacházejí další části semestrální práce. Aby byl zohledněn pohyb obyvatelstva na území ČR, byl přidán do dat i počet obyvatel ČR. Z těchto dat jsme poté spočetli počet cestujících na 100 tisíc obyvatel ČR.

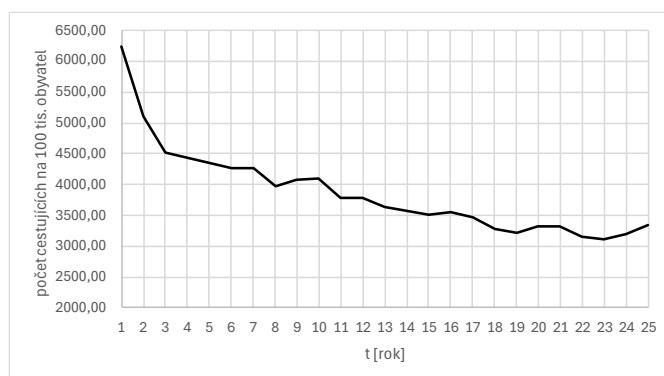
Důvodem, proč byl zvolen tento rozsah dat, je, že zbylá data, tj. 2020, 2021 a 2022 poslouží k posouzení kvality naší předpovědi. Data jsou v souboru seřazena vzestupně. Níže uvádíme krátkou ukázkou struktury našich dat ke zpracování.

t [rok]	rok	# cestujících	# obyv.	# cestujících na 100 K
1	1995	34 623	10 330 759	6235,75
2	1996	32 528	10 315 353	5112,42
.

Tabulka 1: Data ke zpracování.

3 Grafická analýza

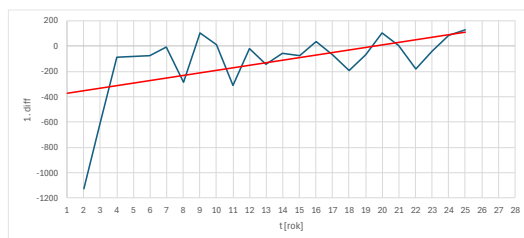
Prvotní informaci o datech získáme ze spojnicového grafu (nebo bodového s rovnými spojnicemi). Na x-ové ose se nacházejí jednotlivé roky od 1995 až do roku 2019 (t: 1 = 1995, 2 = 1996, ...). Na y-ovou osu nanášíme počet cestujících přepravovaných veřejnou autobusovou dopravou na 100 tisíc obyvatel v daném období (y_t).



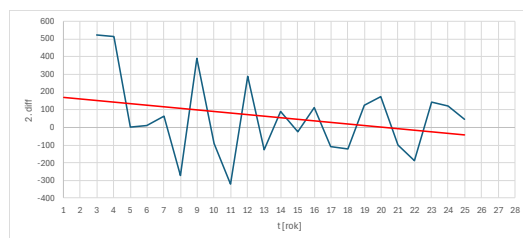
Obrázek 1: Spojnicový graf časové řady

4 Identifikace modelu a předpověď

Pro identifikaci modelu a vhodnou volbu trendové funkce použijeme jednoduchý informativní test. V níže uvedených obrázcích nalezneme 1. a 2. diferenci zkoumaných dat.



(a) 1. difference



(b) 2. difference

Obrázek 2: 1. a 2. difference zkoumané řady

Na základě výše uvedených grafů nelze jednoduše zjistit předpoklad modelu, jelikož se obě difference zdají nekonstantní. Pro přesnější výsledky jsme využili kritéria M.S.E., díky kterému jsme dostali následující výsledky

	lin. trend	kvadr. trend
M.S.E.	112646,7	55175,08

Tabulka 2: M.S.E.

Odhadnutý kvadratický trend a současně vyrovnaná časová řada mají tedy tvar

$$\hat{y}_t = 5,1668t^2 - 219,62t + 5576,8$$

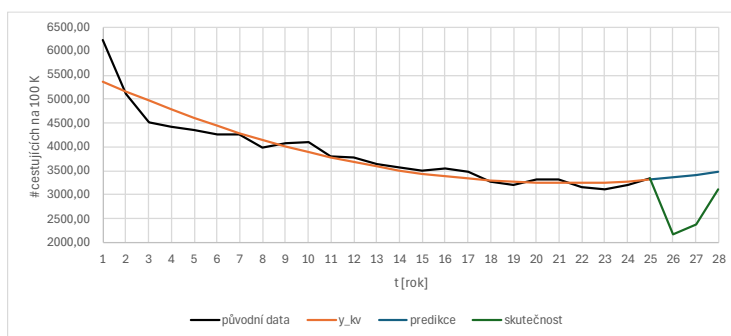
Předpovědi pro roky 2020, 2021 a 2022 jsou postupně

$$\hat{y}_{26} = 5,1168 \cdot 26^2 - 219,62 \cdot 26 + 5576,8 = 3325,64$$

$$\hat{y}_{27} = 5,1168 \cdot 27^2 - 219,62 \cdot 27 + 5576,8 = 3377,21$$

$$\hat{y}_{28} = 5,1168 \cdot 28^2 - 219,62 \cdot 28 + 5576,8 = 3439,01$$

To lze srovnat se skutečnou hodnotou 3359,44 pro rok 2020, 3413,66 pro rok 2021 a 3478,21 pro rok 2022.



Obrázek 3: Předpovědi na 3 roky dopředu