

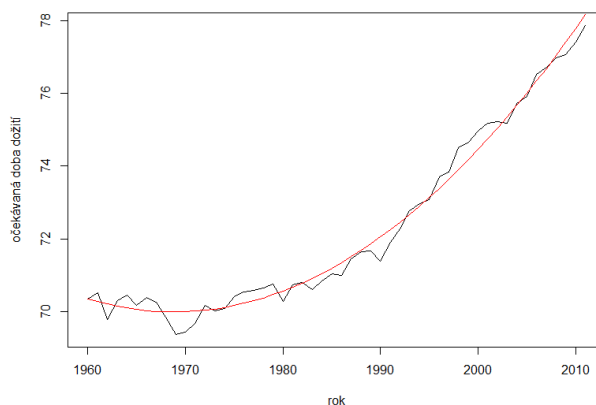
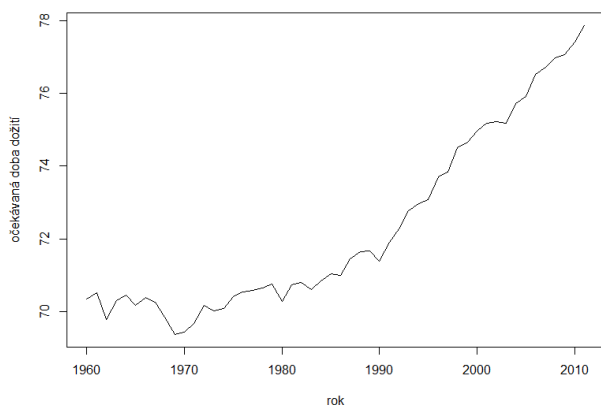
Příklady užití časových řad k predikci rizikových jevů

1 Očekávaná doba dožití v ČR

Máme k dispozici časovou řadu udávající očekávanou dobu dožití v České republice od roku 1960 do roku 2011 (datový soubor life_expectancy_CR.txt). Budeme predikovat vývoj očekávané doby dožití pomocí

- regresní analýzy,
- exponenciálního, resp. Holt-Wintersova vyrovňávání,
- ARIMA modelu.

Lineární regrese



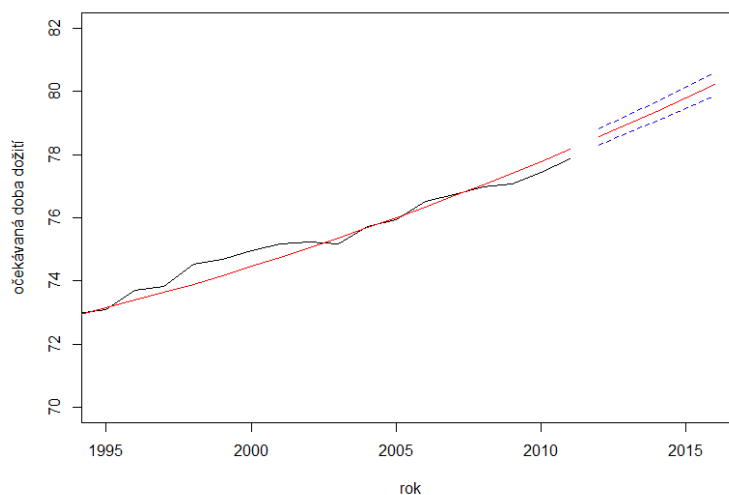
Odhady parametrů kvadratické regresní funkce (pro hodnoty časové proměnné $t = \text{rok} - 1959 = 1, \dots, 52$)

	Odhad	Sm. chyba	t -test	p -hodnota
konst.	70.4349	0.1312	536.70	0.0000
t	-0.0907	0.0114	-7.94	0.0000
t^2	0.0046	0.0002	22.03	0.0000

Predikce na 5 let dopředu spolu s 95% intervaly spolehlivosti jsou:

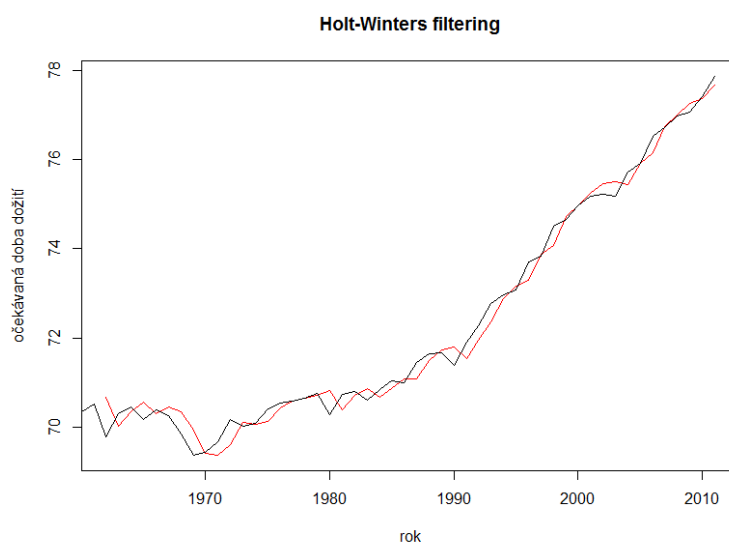
	predikce	dolní	horní
2012	78.56	78.30	78.82
2013	78.96	78.68	79.25
2014	79.37	79.07	79.68
2015	79.79	79.46	80.12
2016	80.22	79.87	80.57

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



Holt-Wintersovo vyrovňování

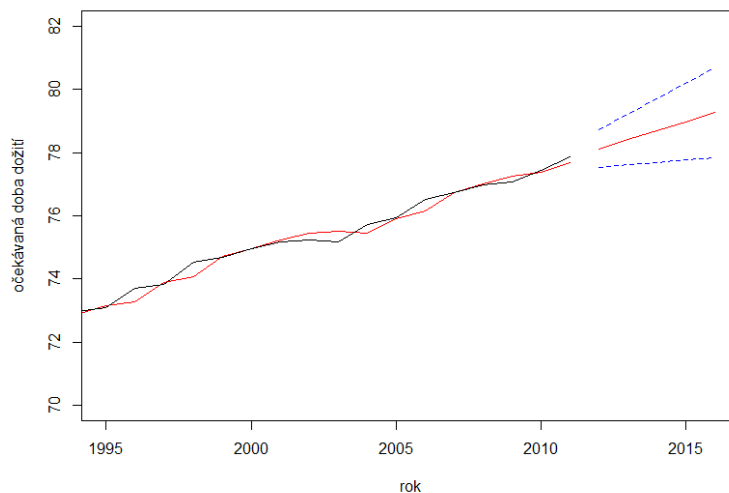
Výsledky Holt-Wintersova vyrovňování jsou zachyceny na následující obrázku (červená čára)



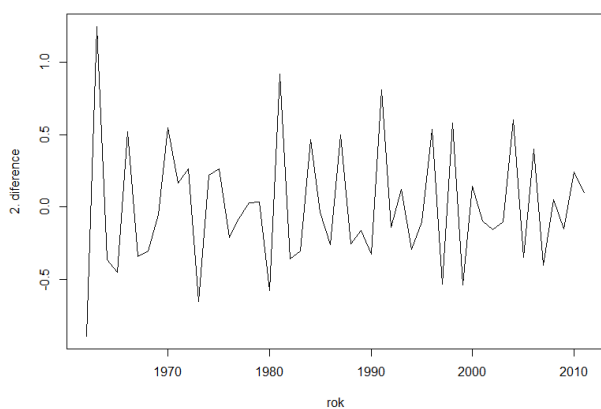
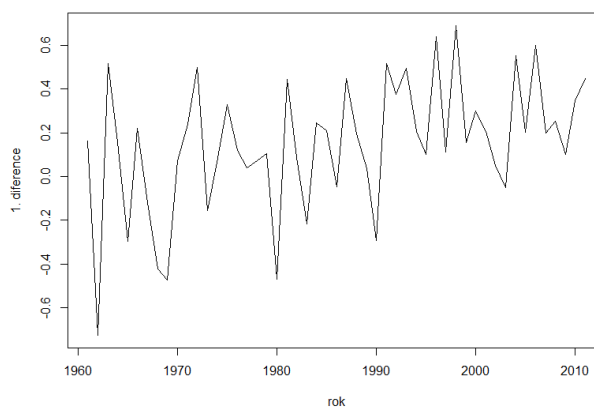
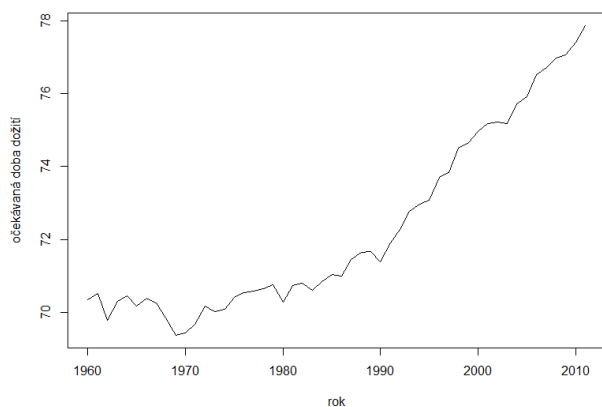
	predikce	dolní	horní
2012	78.12	78.71	77.54
2013	78.41	79.21	77.61
2014	78.69	79.70	77.68
2015	78.98	80.19	77.76
2016	79.26	80.69	77.84

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Holt-Winters filtering



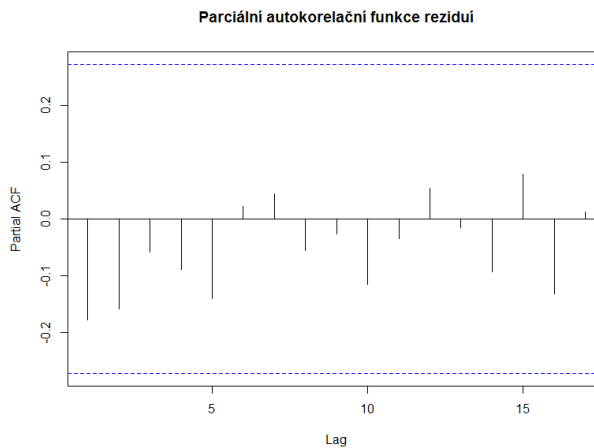
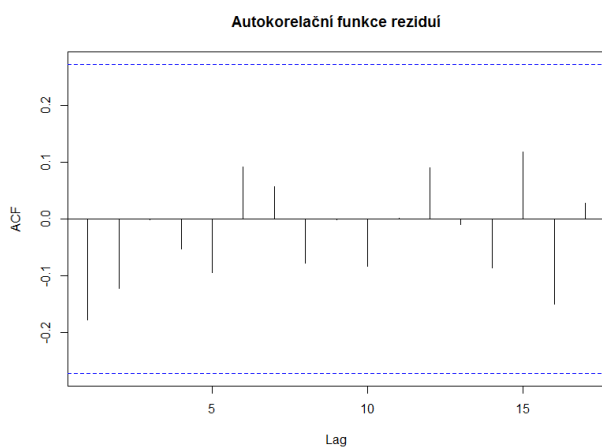
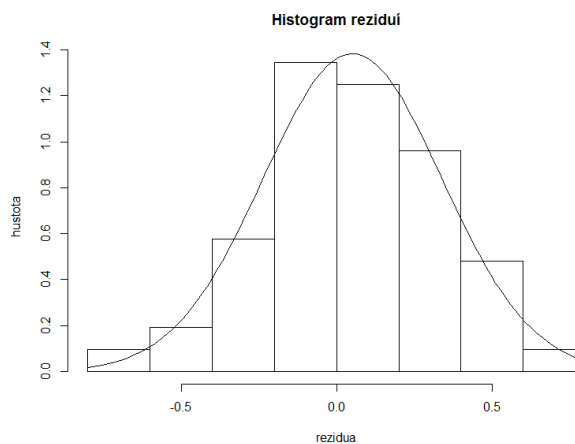
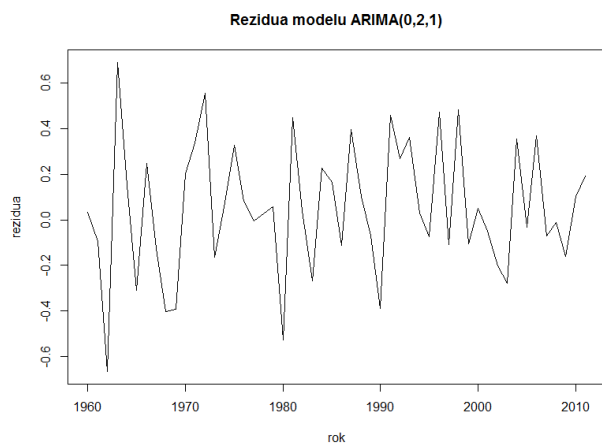
ARIMA model



Grafy znázorňují původní časovou řadu a časové řady 1. a 2. diferencí. Z jejich tvaru plyne, že druhé diference jsou již stacionární.

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

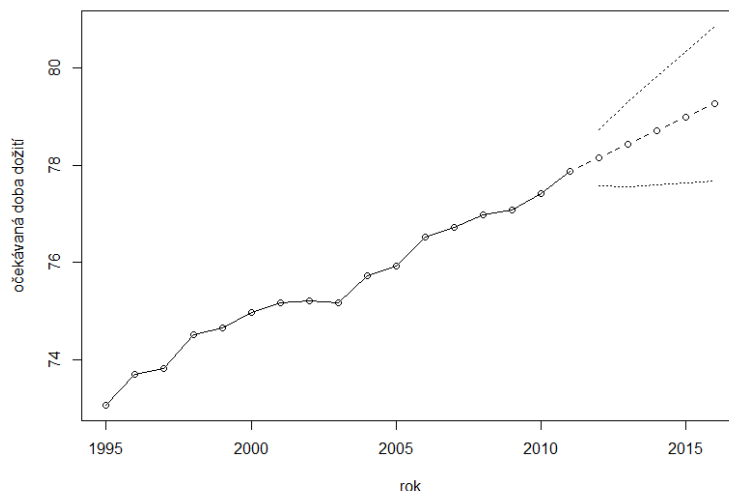
Pro modelování očekávané doby dožití použijeme model ARIMA(0,2,1). Analýza reziduí je shrnuta v následujících grafech.



	predikce	dolní	horní
2012	78.15	77.57	78.73
2013	78.43	77.56	79.29
2014	78.70	77.59	79.82
2015	78.98	77.63	80.34
2016	79.26	77.67	80.85

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Predikce na 5 kroků dopředu



2 Oxid uhličitý

Máme k dispozici časovou řadu koncentrací CO_2 v atmosféře. Jedná se o měsíční časovou řadu od roku 1959 do konce roku 1997, což je 468 pozorování.

(data(co2, package="datasets")).

Budeme predikovat vývoj koncentrace

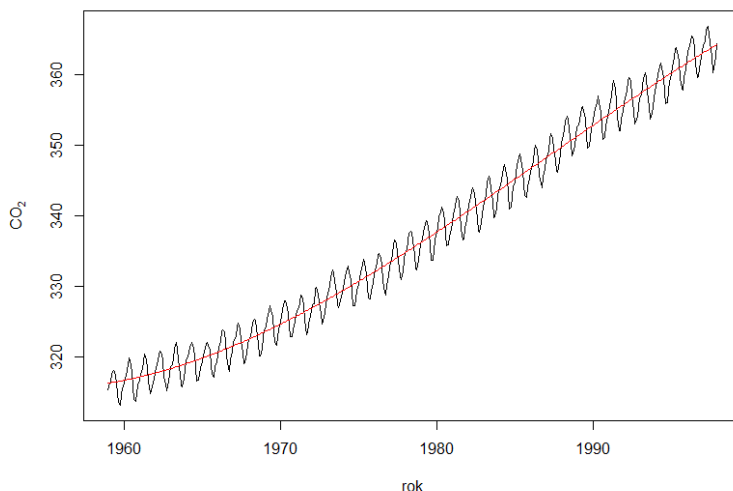
- regresní analýzy,
- exponenciálního, resp. Holt-Wintersova vyrovňování,
- SARIMA modelu.

Regresní analýza

Pro účely regresní analýzy bude používat časovou proměnnou $t = 1, 1 + 1/12, 1 + 2/12, \dots$. Trend lze popsat pomocí polynomu 3. stupně.

	Odhad	Sm. chyba	t -test	p -hodnota
konst.	315.9816	0.4679	675.38	0.0000
t	0.2701	0.0965	2.80	0.0053
t^2	0.0435	0.0054	8.05	0.0000
t^3	-0.0005	0.0001	-5.76	0.0000

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



Periodickou složku nejprve popíšeme pomocí „dummy“ proměnných. Jelikož se jedná o časovou řadu měsíčních údajů, použijeme v regresním modelu celkem 12 „dummy“ proměnných, konstantní člen již součástí modelu nebude.

$m1 = \text{rep}(c(1,0,0,0,0,0,0,0,0,0,0,0),39)$

$m2 = \text{rep}(c(0,1,0,0,0,0,0,0,0,0,0,0),39)$

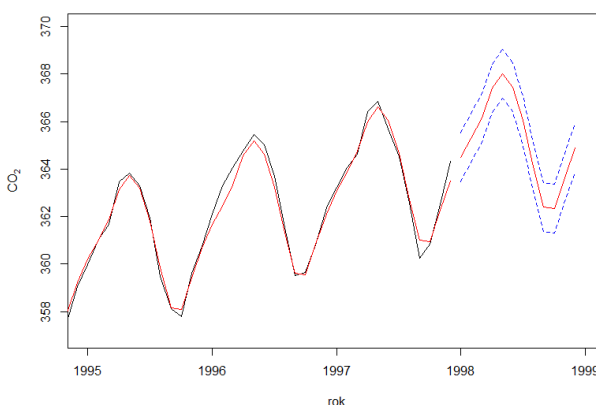
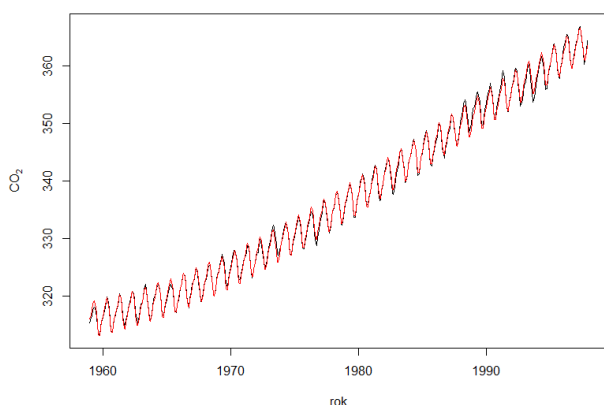
$m3 = \text{rep}(c(0,0,1,0,0,0,0,0,0,0,0,0),39)$

$m4 = \text{rep}(c(0,0,0,1,0,0,0,0,0,0,0,0),39)$

...

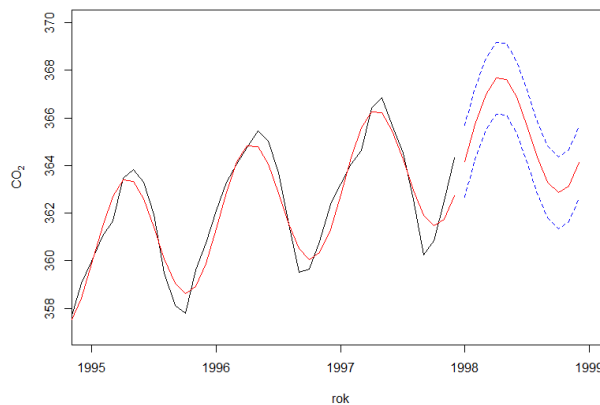
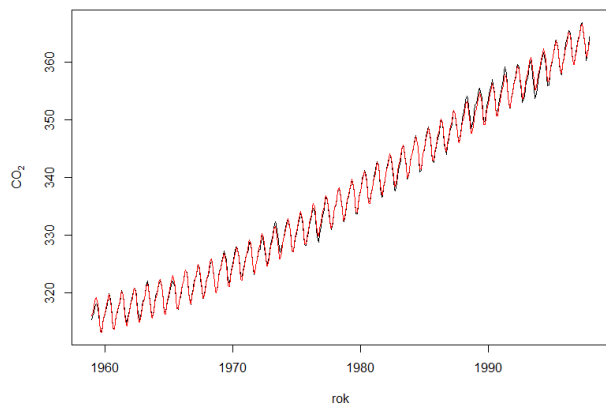
$m12 = \text{rep}(c(0,0,0,0,0,0,0,0,0,0,0,1),39)$

Regresní model se bude skládat z polynomu 3. stupně bez konstanty a z proměnných $m1, \dots, m12$. Výsledný fit a predikce na 12 hodnot dopředu jsou zachyceny na následujících obrázcích.



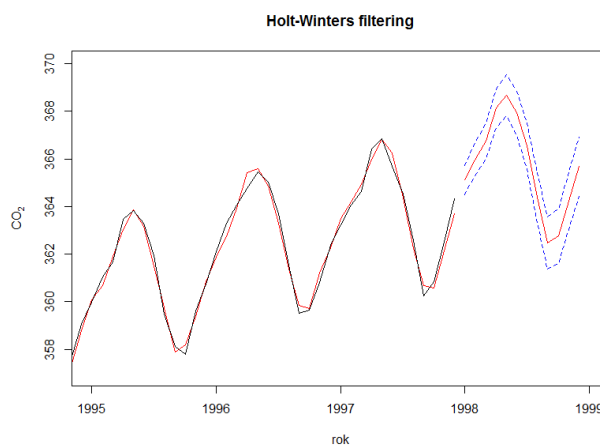
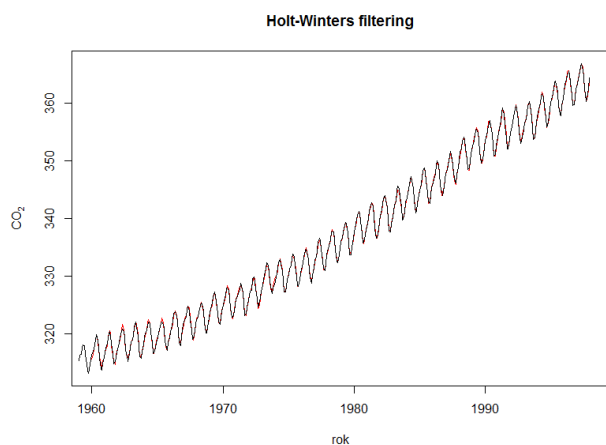
Uvedená časová řada vykazuje roční periodicitu. Použijeme periodické funkce sinus a cosinus pro konstrukci alternativního regresního modelu. Ten bude obsahovat polynom 3. stupně periodické funkce $\cos(2\pi t) + \sin(2\pi t)$. Výsledný fit a predikce na 12 hodnot dopředu jsou zachyceny na následujících obrázcích.

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



Holt-Wintersovo vyrovňování

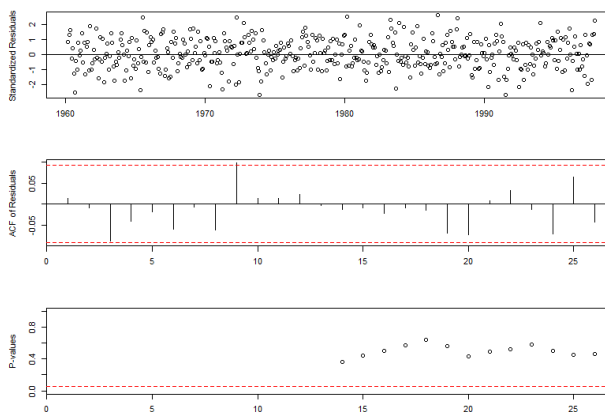
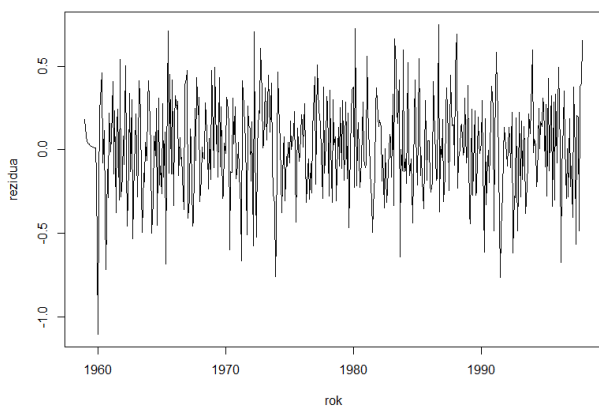
Výsledky Holt-Wintersova vyrovňování spolu s predikcemi na následujících 12 hodnot (jeden rok) jsou zachyceny na následujících obrázcích.



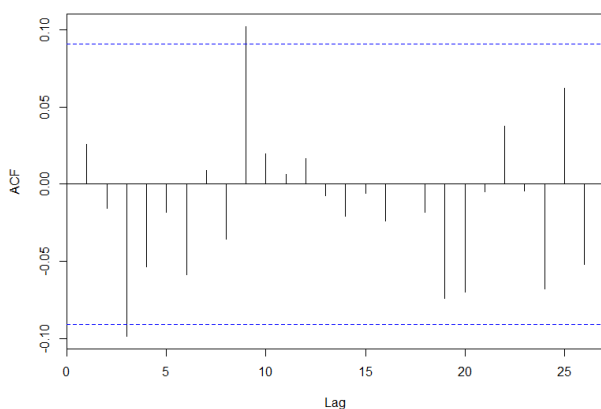
SARIMA model

Analyzovanou řadu budeme popisovat pomocí modelu SARIMA(0,1,1)(0,1,1) s periodou 12. Výsledky analýzy reziduí a následná predikce budoucích 12 hodnot jsou zachyceny v následujících grafech.

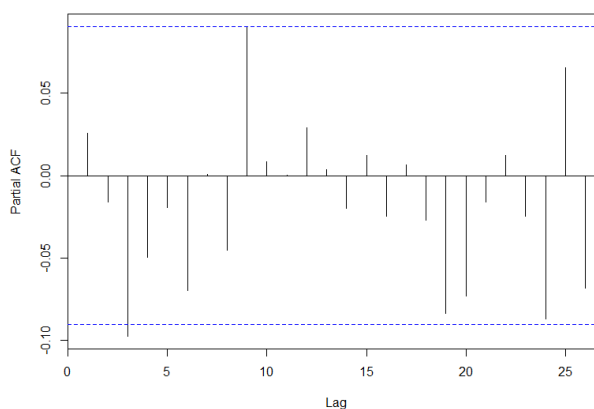
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



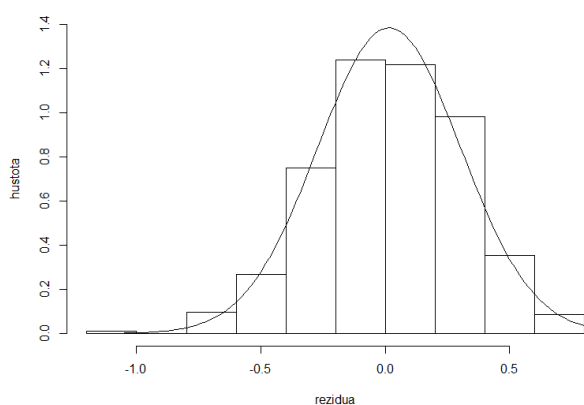
Rezidua modelu SARIMA(0,1,1)(0,1,1) s periodou 12



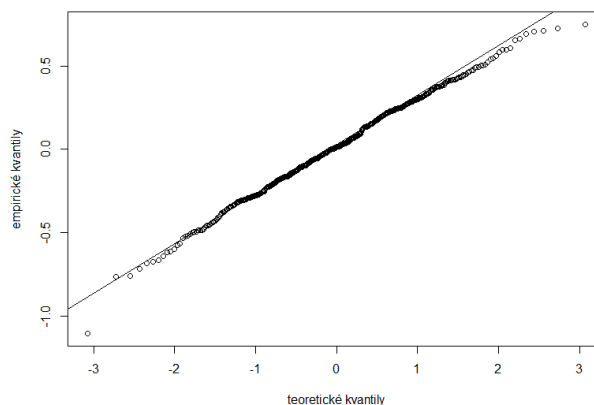
Rezidua modelu SARIMA(0,1,1)(0,1,1) s periodou 12

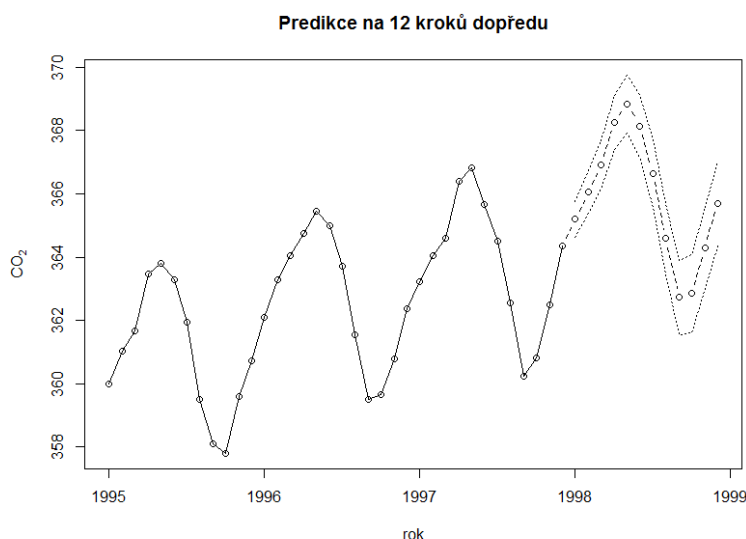


Histogram reziduí



QQ-plot reziduí





Příklady k procvičení

1. Datový soubor `population_USA.txt` zachycuje vývoj počtu obyvatel v USA v období 1790 až 1970. Pomocí vhodného regresního modelu predikujte vývoj populace v následujících desetiletích. Pro predikci použijte také metody exponenciálního vyrovnávání. [Datový soubor: `population_USA.txt`]
2. Časová řada Nile z balíčku `datasets` obsahuje hodnoty měření ročního průtoku řeky Nilu v Asuánu (1871–1970) v miliardách kubických metrů. Díky výstavbě přehrady na této řece došlo ke změně úrovně průtoku okolo roku 1998. Pomocí vhodného regresního modelu odhadněte velikost této změny. Pokuste se identifikovat významné periody v této řadě. [Příkaz v R: `data(Nile)`]
3. Časová řada `sunspots` z balíčku `datasets` obsahuje průměrné měsíční počty slunečních skvrn od roku 1749 do roku 1983. Nalezněte významné periody v této časové řadě. [Příkaz v R: `data(sunspots)`]
4. Data v souboru `prumerna_mzda_CR2000_2012.txt` obsahuje vývoj hrubé měsíční mzdy v ČR v období 2000–2012, jedná se o čtvrtletní data. Popište tuto časovou řadu vhodným SARIMA modelem a spočítejte predikce na jeden rok dopředu. Výsledky srovnajte s předpověďmi získanými pomocí lineární regrese s „dummy“ proměnnými a s Holt-Wintersovým vyrovnáváním. [Datový soubor: `prumerna_mzda_CR2000_2012.txt`]