

MODELOVÁNÍ V EPIDEMIOLOGII

Radmila Stoklasová

Klíčová slova:

Epidemiologie, modelování, klasický epidemiologický model, analýza časových řad, sezónní dekompozice, Boxův – Jenkinsovův model časové řady

Key words:

Epidemiology, modeling, classical epidemiological model, time series analysis, seasonal decomposition, Box–Jenkinson time series model.

Abstrakt

Článek se zabývá studiem vhodných matematických modelů pro popis šíření epidemie chřipky. Modely jsou aplikovány na data (počet nakažených jedinců akutním respiračním onemocněním včetně chřipky) Zdravotního ústavu se sídlem v Ostravě získaných v rámci Systému monitorování zdravotního stavu a životního prostředí ČR za období 2001 – 2010. Cílem práce je konstrukce vhodného matematického modelu šíření chřipky. Článek uvádí standardní deterministický model šíření epidemie a dvě metody analýzy časové řady.

Abstract

This paper deals with mathematical models suitable to describe spread of the epidemic of influenza and validates the application of these models to the data of the Health Institute in Ostrava obtained under System of health and environment monitoring in the Czech Republic for the period 2001 – 2010. Aim is to design a suitable mathematical model of the spread of flu, so such a model to be valid.

Úvod

Termíny používané v simulaci a modelování mohou mít jiný význam v běžném jazyku, resp. v odborných terminologiích jiných profesí. Matematické modelování má pak význam při studiu reálných procesů. [13] Článek se zabývá modelováním epidemie chřipky.

Článek prezentuje klasický model šíření epidemie: Kermackův – Mc Kendrickův model a simulační experiment v závislosti na počátečních podmínkách. Z analýzy časových řad jsou v článku uvedeny dva přístupy: sezónní dekompozice časové řady a Boxova – Jenkinsova metoda

V závěru práce se pak ověřují výsledky těchto modelů na experimentálních datech pro vybrané části města Ostrava. Vliv na počet nemocných akutním respiračním onemocněním včetně chřipky má jistě také lokalita, ve které se epidemie chřipky zkoumá. V Ostravě je hlavním problémem množství emisí v ovzduší, které v současných letech klesá. Celkově lze říci, že počet nakažených jedinců akutním respiračním onemocněním včetně chřipky má klesající trend. [11]

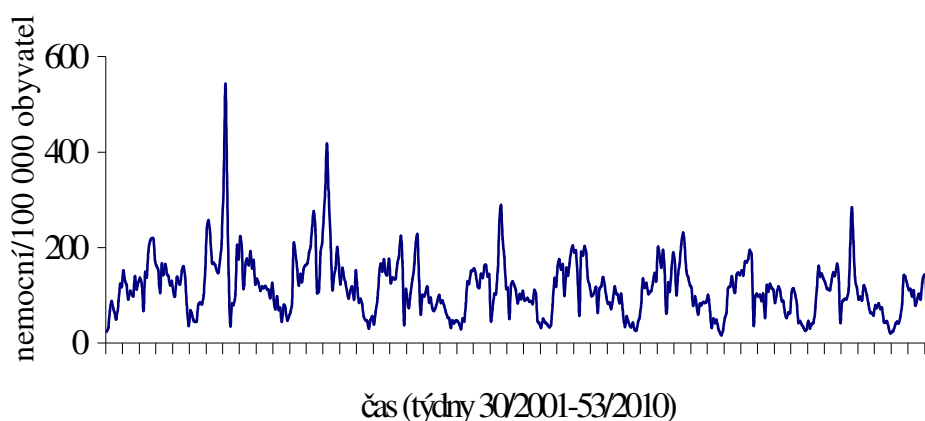
1. Základní charakteristiky epidemiologických dat

Data, která jsou použita v této práci, pocházejí z databáze sledování MONARO za roky 2001–2010 a byla poskytnuta Zdravotním ústavem se sídlem v Ostravě. Data byla získána v rámci Systému monitorování zdravotního stavu a životního prostředí ČR. Databáze obsahuje kód

obvodu, den vyšetření, rok narození, pohlaví, diagnózu v rozlišení akutní respirační onemocnění a ostatní. Zdrojem dat je Zdravotní ústav se sídlem v Ostravě a data jsou získaná v rámci Systému monitorování zdravotního stavu a životního prostředí ČR za období 2001–2010. [4]

Jedná se o denní údaje spolupracujících lékařů s ordinacemi v obvodech Slezská Ostrava, Moravská Ostrava a Přívoz, Mariánské Hory a Hulváky, Poruba, Ostrava-Jih. Chřipková epidemie se zkoumá od 30. týdne jednoho roku do 29. týdne dalšího roku.

Veškerá poskytnutá data Zdravotním ústavem se sídlem v Ostravě zachycuje Obr. 1. Denní údaje o počtu nemocných byly upraveny pro lepší interpretaci na týdenní počty nemocných akutním respiračním onemocněním. V období od 30. týdne roku 2001 do 53. týdne roku 2010 byla průměrná týdenní nemocnost akutních respiračních onemocnění včetně chřipky na 100000 obyvatel 116 nemocných.



Obr. 1 Týdenní nemocnost v jednotlivých týdnech (30. týden/2001 – 53. týden/2010)

2. Kermackův – Mc Kendrickův model

Tento model je při své jednoduchosti obecně přijímán pro aproximaci průběhu epidemie nakažlivým onemocněním. Dále vycházím z prací [5,7,8,15].

Předpokládá splnění těchto podmínek:

- Nemoc je přenášena kontaktem mezi infikovaným a zdravým jedincem, který není vůči této nemoci imunní. Tento jedinec se nazývá vnímavým jedincem.
- Všichni infikovaní jedinci mohou se stejnou pravděpodobností přenést nemoc, stejně jako všichni vnímaví jedinci mohou tuto nemoc dostat se stejnou pravděpodobností.
- U každé nemoci je důležitá doba, která je potřebná k tomu, aby se vnímavý jedinec po setkání s nakažlivým stál sám nakažlivým. Předpokládá se, že tato latentní doba je tak malá, že ji lze považovat za nulovou. Proto se vnímavý jedinec okamžitě po styku s nemocí stává infikovaným jedincem.
- Zkoumaná populace je uzavřená, tzn., že nedochází k narození žádných nových jedinců, neuvažujeme žádný možný pohyb obyvatelstva, mrtví jedinci nejsou z populace vyřazeni. Celkový počet jedinců v populaci je tedy konstantní.

Uvažuje se rozdělení populace do tří skupin:

- Skupina infikovaných jedinců s počtem $I(t)$ v čase t .
- Skupina vnímavých jedinců s počtem $V(t)$ v čase t .

- Skupina imunních jedinců, tedy přesněji těch, kteří prodělali nemoc a nejsou dále infekční ani vnímaví, s počtem $R(t)$ v čase t . Do této skupiny jsou zahrnuti i zemřelí jedinci. Podmínku uzavřenosti a konečnosti populace lze vyjádřit vztahem:

$$I(t) + V(t) + R(t) = N = \text{konst.} \quad (1)$$

Platí: $I'(t) = \alpha I(t)V(t) - \beta I(t), \quad (2)$

kde α je koeficient šíření nákazy a koeficient β představuje poměrnou část infikovaných jedinců, kteří přejdou do imunní skupiny. Pro její časovou změnu platí rovnost

$$R'(t) = \beta I(t). \quad (3)$$

Třetí diferenciální rovnice, která vyjadřuje počet vnímavých jedinců, je derivací podmínky

(1). Tedy $I'(t) + V'(t) + R'(t) = \frac{dN}{dt} = 0$ a lze psát následující rovnosti:

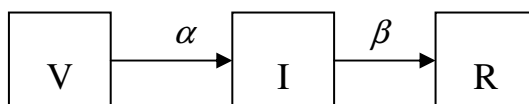
$$\begin{aligned} V'(t) &= -I'(t) - R'(t) = -\alpha I(t)V(t) + \beta I(t) - \beta I(t) \\ V'(t) &= -\alpha I(t)V(t) \end{aligned} \quad (4)$$

Dostáváme soustavu tří diferenciálních rovnic (2) – (4) tedy:

$$\begin{aligned} I'(t) &= \alpha I(t)V(t) - \beta I(t), \\ R'(t) &= \beta I(t), \\ V'(t) &= -\alpha I(t)V(t), \end{aligned}$$

kteřá je známá pod názvem Kermackův – Mc Kendrickův model.

Tab. 1 Schématické znázornění Kermackova – Mc Kendrickova modelu



Tento model bývá např. v [8] uváděn jako model SIR (*Susceptible* – náchylný, vystavený určitému vlivu, *Infective* – infekční, *Removed* – odstraněný). Uvažuje se zde model bez vitální dynamiky, tzn., že tento model nebere v úvahu přírůstek populace ani úmrtnost.

Vzhledem k podmínce uzavřenosti populace (1) lze počet diferenciálních rovnic zmenšit. Například je $V(t) = N - I(t) - R(t)$. Dosazením $V(t)$ do rovnice (2) vede k soustavě dvou diferenciálních rovnic

$$I'(t) = I(t)(\alpha N - \beta - \alpha I(t) - \alpha R(t)), \quad (5)$$

$$R'(t) = \beta I(t). \quad (6)$$

Koeficienty α , β jsou kladné. Předpokládá se, že $V(0) = V_0 > 0$, $I(0) = I_0 > 0$, $R(0) = 0$, $V_0 + I_0 = N$. Každá počáteční úloha pro (5), (6) má jediné úplné řešení. Z rovnice

$I'(t) = \alpha I(t)V(t) - \beta I(t)$ plyne, že pro $V_0 \leq \frac{\beta}{\alpha}$ by epidemie vůbec nezačala, protože, že pokud

$I'(t) \leq 0$ je funkce $I(t)$ klesající. Proto předpokládáme $V_0 > \frac{\beta}{\alpha}$. Epidemie tedy začíná až při V_0 větším než jistá prahová hodnota $\rho = \frac{\beta}{\alpha}$. Tomuto jevu se říká prahový efekt. [12], [14]
V literatuře [5] je dokázána Kermackova – Mc Kendrickova prahová věta:

Pokud populace měla na počátku epidemie $\rho + \varepsilon$ vnímavých jedinců, pak se jejich počet na konci epidemie sníží na $\rho - \varepsilon$.

Epidemická křivka

Důležitým pojmem je epidemická křivka $W(t)$, která popisuje změny v počtu infikovaných. Tato křivka je dána rovnicí $W(t) = \frac{dI(t)}{dt} = \alpha I(t)(N - I(t))$, kde α je počet vnímavých jedinců, na které přeneše nemoc jeden infikovaný jedinec za jednotku času.

Řešením diferenciální rovnice dostáváme:

$$W(t) = \frac{I_0 \alpha (N - I_0) e^{\alpha N t}}{\left(1 - \frac{I_0}{N} (1 - e^{\alpha N t})\right)^2}. \quad (7)$$

Monotónnost $W(t)$ se určí na základě vlastností její derivace.

$$\begin{aligned} W'(t) &= \frac{dI'(t)}{dt} = \frac{d}{dt}(\alpha I(t)(N - I(t))) = \alpha I'(t)(N - I(t)) - \alpha I(t)I'(t) = \\ &= \alpha I'(t)(N - 2I(t)) = \alpha^2 I(t)(N - I(t))(N - 2I(t)) \\ W'(t) &= 0 \Leftrightarrow I(t) = 0 \vee I(t) = \frac{N}{2} \vee I(t) = N \end{aligned}$$

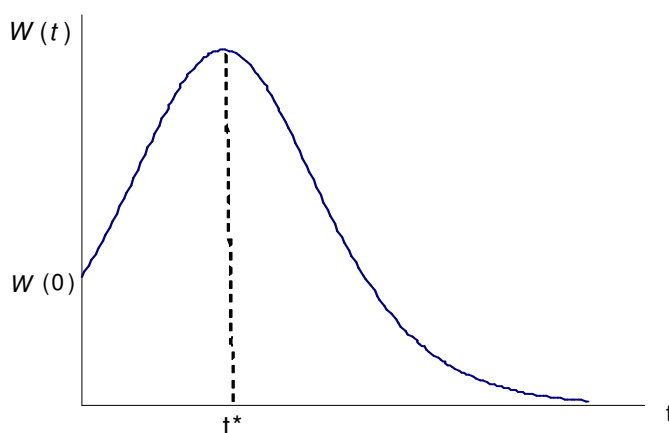
Potom, je-li $I(t) \in \left(0, \frac{N}{2}\right)$, pak $W'(t) > 0$ a tedy epidemická křivka $W(t)$ je na tomto intervalu rostoucí. Je-li $I(t) \in \left(\frac{N}{2}, N\right)$, pak $W'(t) < 0$ a epidemická křivka $W(t)$ je na tomto intervalu klesající. Epidemická křivka $W(t)$ nabývá svého maxima pro $I(t) = \frac{N}{2}$.

Z toho také vyplývá, že řešení rovnice $I'(t) = \alpha I(t)[N - I(t)]$ pro $I(t) \in \left(0, \frac{N}{2}\right)$ je na intervalu $\left(0, \frac{N}{2}\right)$ konvexní, na intervalu $\left(\frac{N}{2}, N\right)$ konkávní a v bodě t^* , pro který platí $I(t^*) = \frac{N}{2}$ má inflexní bod. Tento inflexní bod musí být v bodech, pro které platí $W'(t) = 0$. Z předchozích závěrů vyplývá, že $W'(t) = 0$ právě, když $I(t^*) = \frac{N}{2}$, což je $\frac{NI_0 e^{\alpha N t}}{N - I_0 + I_0 e^{\alpha N t}} = \frac{N}{2}$.

To znamená

$$t^* = \frac{1}{\alpha N} \ln \frac{N - I_0}{I_0}.$$

Z výpočtů tedy vyplývá, že pokud $I_0 < \frac{N}{2}$, pak epidemická křivka nabývá svého maxima v bodě $t^* = \frac{1}{\alpha N} \ln \frac{N - I_0}{I_0}$, tedy v tomto čase se epidemie šíří nejrychleji a $W(t^*) = \frac{\alpha N^2}{4}$.



Pokud $I_0 > \frac{N}{2}$, pak $t^* < 0$, pak funkce $W(t)$ v čase $t \in (0, \infty)$ svého maxima nenabývá.

Obr. 2 Epidemická křivka

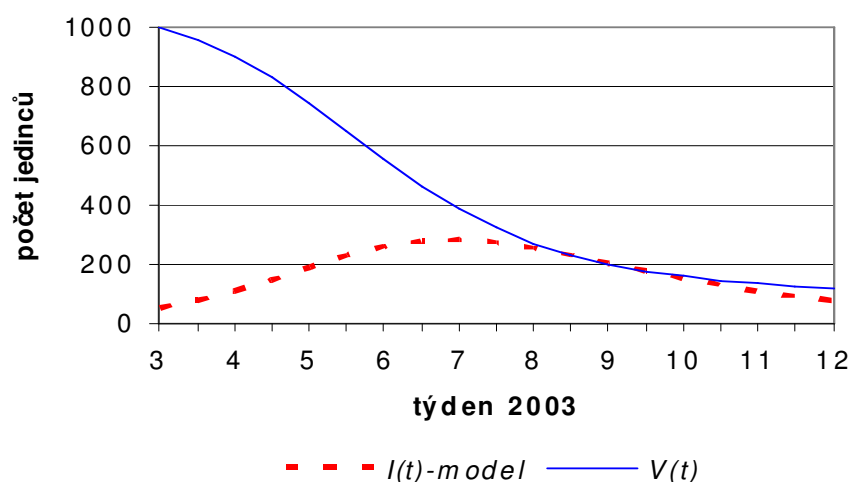
2.1 Aplikace klasického epidemiologického modelu

Pro Kermackův – Mc Kendrickův model byl vytvořen simulační program v jazyku C. [9, 10, 13]. Pomocí tohoto programu je simulováno šíření epidemie chřipky v sezóně 2002-2003 u dětí ve věku od 6 – 14 let ve vybraných částech města Ostravy, konkrétně Mariánské Hory a Hulváky. V těchto částech Ostravy žije 12 982 obyvatel, z toho děti ve věku od 6 do 14 let tvoří asi 7,8% [16]. Počáteční hodnota počtu vnímavých jedinců byla zvolena 1000.

Největším problémem bylo určit koeficienty α , β . Zde vycházím z publikací [2, 3], kde jsou uvedeny intervaly pro odhady těchto koeficientů. Na základě pozorovaných dat program identifikoval hodnoty parametrů $\alpha = 6,5 \cdot 10^{-4} \text{ týden}^{-1}$, $\beta = 0,262 \text{ týden}^{-1}$.

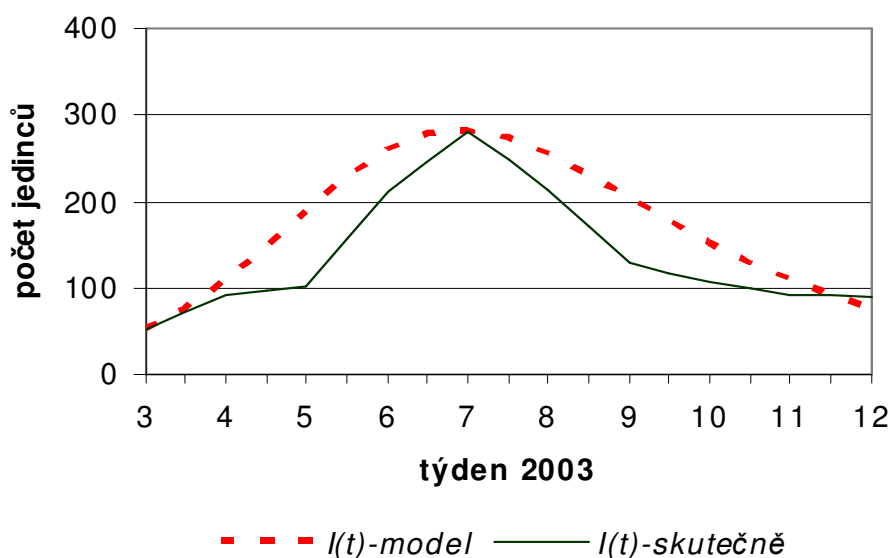
Pro simulační pokus byly stanoveny tyto hodnoty:

$$N = 1053, V(0) = 1000, R(0) = 0, I(0) = 53, \alpha = 6,5 \cdot 10^{-4} \text{ týden}^{-1}, \beta = 0,262 \text{ týden}^{-1}.$$



Obr. 3 Vývoj chřipkové epidemie věkové skupiny 6 – 14 let v sezóně 2002-2003

Následující Obr. 4 ukazuje, jak se simulovaná data shodují s daty skutečnými. Na počátku, na konci a ve vrcholu epidemie chřipky v 3. – 12. týdnu roku 2003 se hodnoty shodují.



Obr. 4 Simulované a skutečné počty infikovaných jedinců

Tab. 2 Simulované a skutečné počty infikovaných jedinců

Týden 2003	Simulované hodnoty $I(t)$	Skutečný počet infikovaných jedinců
3	53	53
4	108,8	91
5	189,2	101
6	260	211
7	283,3	282
8	255,8	215
9	205,1	130
10	153,1	107
11	109,5	92
12	76,4	90

V sezóně 2002 – 2003 byla nejvyšší hodnota počtu infikovaných dětí ve věku od 6 – 14 let 282 a to v 7. týdnu roku 2003. Hodnota vypočítaná pomocí simulačního modelu byla ve stejném týdnu 283 infikovaných dětí. Od tohoto týdne začíná počet infikovaných dětí ve věku od 6 – 14 let klesat a např. v 12. týdnu roku 2003 je skutečná hodnota 90 infikovaných jedinců a odhadnutá hodnota je 76 infikovaných jedinců.

Tímto modelem byly dosaženy dobré výsledky v případech, kdy se jednalo o uzavřenou skupinu jedinců.

3. Aplikace metod analýzy časových řad

Tato kapitola je věnována dvěma metodám analýzy časové řady. Nejprve je provedena sezónní dekompozice časové řady a na jejím základě odhadnut vývoj časové řady do konce roku 2012. V druhé části se pak modeluje časová řada pomocí ARIMA modelů.

Pro tyto účely musela být data upravena do měsíčních hodnot počtu nakažených akutním respiračním onemocněním včetně chřipky, protože program SPSS nemá ve své nabídce volbu datové proměnné „rok-týden“. [15]

3.1 Sezónní dekompozice časové řady

Program SPSS nabízí dvě varianty sezónní dekompozice časové řady: aditivní nebo multiplikativní. Pro tuto časovou řadu byl zvolen multiplikativní model, protože amplituda časové řady se s rostoucím časem zmenšuje. [1] V následující tabulce jsou uvedeny sezónní faktory, které říkají, že například nejvyšší počet nakažených akutním respiračním onemocněním včetně chřipky bývá v únoru a to o 43,6 % více než je předpokládaný počet nakažených jedinců. Nejnižší hodnotu dostáváme pro měsíc srpen (37,7 %), tedy v srpnu bývá o 62,3 % nakažených méně než je měsíční předpokládaný počet nakažených jedinců akutním respiračním onemocněním.

Tab. 3 Sezónní faktory – výstup programu SPSS

Měsíc	Sezónní faktory	Měsíc	Sezónní faktory
Leden	131,5	Červenec	41,5
Únor	143,6	Srpen	37,7
Březen	104,2	Září	98,1
Duben	94,4	Říjen	128,1
Květen	75,8	Listopad	134,8
Červen	76,9	Prosinec	133,4

Rovnice odhadnutá regresní analýzou má tvar $y_t = 1618 - 6,3t \quad t = 1,2,\dots$

Teoretické hodnoty počtu nakažených jedinců na 100 000 obyvatel od ledna 2002 do prosince 2011 jsou uvedeny níže v Tab. 5.

3.2 Boxův – Jenkinsův model časové řady

V této metodě se neklade důraz na systematickou složku časové řady y_t , ale na složku nesystematickou, tedy reziduální složku a_t . Předpokladem pro použití této metody je, že modelovaná časová řada musí být stacionární. V případě, že stacionární není, časovou řadu diferencujeme. Při analýze časové řady počtu jedinců nakažených akutním respiračním onemocněním byla provedena difference časové řady tak i sezónní difference. Kde pod pojmem difference rozumíme $\Delta y_t = y_t - y_{t-1}$, $t = 2, \dots, n$ a sezónní difference $\Delta y_{ts} = y_{ts} - y_{t-s}$, kde s je délka sezónní periody (u měsíčních časových řad 12, u čtvrtletních 4). Na základě identifikace modelu byla do modelu zahrnuta složka klouzavých průměrů (obyčejných i sezónních). Jako nejlepší z hlediska kvality modelu vychází model SARIMA(0,1,1)(0,1,1)₁₂. [1]

Tab. 4 Hodnoty koeficientů v SARIMA modelu – výstup programu SPSS

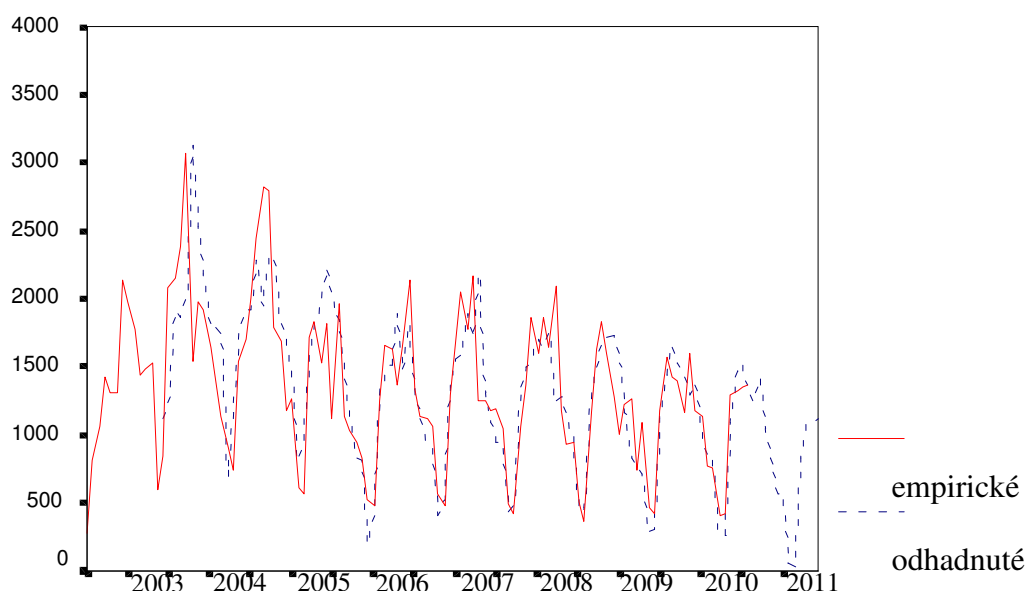
	B	SEB	T-RATIO	PROB.
MA1	0,7378	0,0687	10,739	<u>0,0000</u>
SMA1	0,7873	0,1528	5,152	<u>0,00000134</u>
CONSTANT	-3,2	3,6	-0,88	0,381

Oba koeficienty $MA1 = 0,737$; $SMA1 = 0,787$ jsou statisticky významné na hladině významnosti 1% .

Dostáváme tedy následující vztah, který popisuje tento model:

$$y_t = -3,2 + y_{t-1} + y_{t-12} + y_{t-13} + a_t - 0,737a_{t-1} - 0,787a_{t-12} + 0,58a_{t-13}.$$

Následující graf opět ilustruje skutečné hodnoty počtu nakažených jedinců akutním respiračním onemocněním a hodnoty odhadnuté pomocí SARIMA modelu. V grafu jsou znázorněny i hodnoty predikované až do prosince roku 2011.



Obr.5 Porovnání skutečných a teoretických hodnot počtu nakažených jedinců na 100 000 obyvatel od ledna 2001 do prosince 2011

Tab.5 Srovnání predikovaných hodnot získaných analýzou časové řady (sezónní dekompozice časové řady, Boxův – Jenkinsův model časové řady) a skutečných hodnot počtu nakažených jedinců na 100 000 obyvatel

Rok 2011	Odhadnutý počet nakažených jedinců na 100000 obyvatel modelem sezónní dekompozice (Ostrava)	Odhadnutý počet nakažených jedinců na 100000 obyvatel Boxovým – Jenkinsovým modelem (Ostrava)	Skutečný počet nakažených jedinců na 100000 obyvatel v Moravskoslezském kraji
Leden	1321,69	1246,8	1692
Únor	1435,96	1420,6	2115
Březen	1034,42	1049,4	1543
Duben	931,57	830,5	1120
Květen	743,41	563,2	neuvedeno
Červen	749,80	569,4	neuvedeno
Červenec	402,54	64,3	neuvedeno
Srpen	362,47	33,2	neuvedeno
Září	937,40	829,4	1042
Říjen	1216,29	1084	neuvedeno
Listopad	1270,51	1085,6	1046
Prosinec	1247,83	1121,8	1356

V období chřipkové epidemie, tj. 4. – 12. kalendářní týden dosahují obě časové řady svých maximálních hodnot. Predikované hodnoty počtu nakažených jedinců se výrazně odlišují v letních měsících. V tomto období se chřipková epidemie nepředpokládá, takže tento nedostatek nepovažuji za významný.

Ve třetím sloupci tabulky jsou uvedeny skutečné počty nakažených jedinců na 100 000 obyvatel v Moravskoslezském kraji získané z Krajské hygienické stanice Moravskoslezského kraje se sídlem v Ostravě [11]. Skutečné počty nakažených jedinců pro vybrané části města Ostravy nejsou dostupné.

Závěr

V této práci jsou uvedeny matematické modely pro popis šíření epidemie infekčních onemocnění. Epidemiologie je vědní obor, který se podle definice WHO zabývá studiem rozdělení a příčin nemocí a událostí spjatých se zdravotním stavem lidské populace a aplikací těchto poznatků při řešení zdravotních problémů.

Dále jsou v této práci charakterizována epidemiologická data. Jsou zde aplikovány dva přístupy k časovým řadám (sezónní dekompozice časové řady a Boxova – Jenkinsova metodologie). Dále je zde uveden klasický model šíření epidemie a simulační experiment v závislosti na počátečních podmínkách.

Cílem práce bylo navrhnout vhodné matematické modely pro popis šíření epidemií chřipky a ověřit tyto modely na experimentálních datech pro město Ostrava, resp. vybrané části Ostravy.

Literatura:

- [1] ARLT, J., ARLTOVÁ, M. *Příklady z analýzy ekonomických časových řad*. Praha: VŠE, 1997. ISBN 80-7079-056-3.
- [2] BAILEY, N.T.J. *The Elements of Stochastic Processes: with Applications to the Natural Sciences*. New York: John Wiley&Sons, 1990. ISBN 0-471-52368-2.
- [3] BIČÍK, V., CRHÁK, L. *Modely a modelování v biologii*. Olomouc:UP, 1985.
- [4] *Bulletin Krajské hygienické stanice Ostrava*. Ostrava:KHS, leden 2000 – prosinec 2009.
- [5] BRAUNER, F., CASTILLO-CHÁVEZ, C. *Mathematical Models in Population Biology and Epidemiology*. New York: Springer, 2001. ISBN 0-387-98902-1.
- [6] GÖPFERTO VÁ, D. *Epidemiologie*. Praha: TRITON, 1999. ISBN 80-7254-037-8.
- [7] HOLČÍK, J., FOJT, O. *Modelování biologických systémů*. Brno: VUT, 2001. ISBN 80-214-2023-5.
- [8] KALAS, J., POSPÍŠIL, Z. *Spojité modely v biologii*. Brno: Masarykova univerzita, 2001. ISBN 80-210-2626-X.
- [9] KERNIGHAN, B.W., RITCHE, D.M. *Učebnice programovacího jazyka C*. Bratislava: Alfa, 1998.
- [10] KINDLER, E. *Simulační programovací jazyky*. Praha: SNTL, 1980.
- [11] ODBOR PROTIEPIDEMICKÝ. *Krajská hygienická stanice Moravskoslezského kraje se sídlem v Ostravě* [online]. 2011 [cit. 2012-06-20]. Dostupné z: <http://www.khsova.cz/01_aktuality/pe.php>.
- [12] KŘIVÝ, I. *Modely v populační biologii a ekologii*. Ostrava: Aleko, 1991. ISBN 80-7042-047-2.
- [13] KŘIVÝ, I., KINDLER, E. *Simulace a modelování*. Ostrava: OU, 2001. ISBN 80-7042-809-0.
- [14] KŘIVÝ, I. *A Branching Process Model in Epidemiology*. *Biocybernetics and Biomedical Engineering*, vol. 17, no. 1, pp. 169-179. ISSN 0208-5216.
- [15] MAREK, L. *Statistika v SPSS. Časové řady*. Praha: VŠE, 1995. ISBN 80-7079-642-1.
- [16] OSTRAVA. *Faktografické listy*. Statutární město Ostrava – magistrát: odbor ekonomického rozvoje, 2010.

JEL C22, C15

Mgr. Radmila Stoklasová, Ph.D.

Katedra matematických metod v ekonomii

SU OPF Karviná

Univerzitní nám. 1934/3

733 40 Karviná

stoklasova@opf.slu.cz