Cvičení 4 – Statistické úlohy v HMŘ

Příklad: Máme k dispozici kategorizované údaje o míře nezaměstnanosti ve 196 náhodně vybraných obcích jistého kraje ve dvou po sobě následujících letech. V roce 2008 mělo 101 obcí nízkou nezaměstnanost (do 4 %), 42 střední (mezi 4 % a 6 %) a 53 vysokou (nad 8 %). V následujícím roce s ukázalo, že ze 101 obcí s nízkou nezaměstnaností jich 36 zůstalo v téže kategorii, 43 přešlo do kategorie střední nezaměstnanost a 22 do kategorie vysoká nezaměstnanost. Ze 42 obcí se střední nezaměstnaností jich 18 zůstalo ve stejné kategorii, 12 přešlo do kategorie nízká nezaměstnanost a 12 do kategorie vysoká nezaměstnanost. Z 53 obcí s vysokou nezaměstnaností jich 40 zůstalo v téže kategorii, 5 přešlo do kategorie nízká nezaměstnanost a 8 přešlo do kategorie střední nezaměstnanost.
a) Modelujte situaci pomocí HMŘ.
b) Metodou maximální věrohodnosti odhadněte vektor počátečních pravděpodobností a matici přechodu.
c) Sestrojte 95% Waldův a skórový interval spolehlivosti pro počáteční pravděpodobnosti a pravděpodobnosti přechodu.
d) V celostátním měřítku mělo v roce 2008 56 % obcí nízkou nezaměstnanost, 23 % střední a 21 % vysokou. Na hladině významnosti 0,05 testujte hypotézu, že v daném kraji se vektor počátečních pravděpodobností shoduje s celostátním vektorem počátečních pravděpodobností.

Výsledky:
Ad a) Zavedeme HMŘ $\{X_n; n \in N_0\}$ s množinou stavů J = {1, 2, 3}, kde $X_n = 1$, když v n-tém roce náhodně vybraná obec v daném regionu má nízkou nezaměstnanost, $X_n = 2$, když má střední nezaměstnanost a $X_n = 3$, když má vysokou nezaměstnanost.

Ad b) Odhad vektoru počátečních pravděpodobností:

$$\hat{\mathbf{p}}(0) = \begin{pmatrix} \frac{101}{196} & \frac{42}{196} & \frac{53}{196} \end{pmatrix} = \begin{pmatrix} 0{,}5153 & 0{,}2143 & 0{,}2704 \end{pmatrix}$$

Znamená to, že na počátku sledování 51,53 % obcí mělo nízkou nezaměstnanost, 21,43 % obcí mělo střední nezaměstnanost a 27,04 % obcí mělo vysokou nezaměstnanost.

Odhad matice přechodu:

$$\hat{\mathbf{P}} = \begin{pmatrix} \frac{36}{101} & \frac{43}{101} & \frac{22}{101} \\ \frac{12}{42} & \frac{18}{42} & \frac{12}{42} \\ \frac{5}{53} & \frac{8}{53} & \frac{40}{53} \end{pmatrix} = \begin{pmatrix} 0{,}3564 & 0{,}4257 & 0{,}2178 \\ 0{,}2857 & 0{,}4286 & 0{,}2857 \\ 0{,}0943 & 0{,}1509 & 0{,}7547 \end{pmatrix}$$

Interpretace 1. řádku: Pokud v určitém roce měla náhodně vybraná obec nízkou nezaměstnanost, tak v příštím roce bude mít s pravděpodobností 35,64 % opět nízkou nezaměstnanost, s pravděpodobností 42,57 % střední nezaměstnanost a s pravděpodobností 21,78 % bude mít vysokou nezaměstnanost.

Ad c) Před konstrukcí 95% Waldových intervalů spolehlivosti pro $p_1(0)$, $p_2(0)$, $p_3(0)$: ověříme splnění podmínek dobré aproximace $\hat{p}_i(0)[1-\hat{p}_i(0)]c > 9$, i = 1, 2, 3. Postupně dostaneme $48{,}9541 \quad 33{,}0000 \quad 38{,}6684$, tedy podmínky jsou splněny.
Meze 95% asymptotických Waldových intervalů spolehlivosti pro $p_1(0)$, $p_2(0)$, $p_3(0)$:

$p_1(0) \in (0,4453; 0,5853)$, $p_2(0) \in (0,1568; 0,2717)$, $p_3(0) \in (0,2082; 0,3326)$ vždy

s pravděpodobností 95 %.

Meze 95% asymptotických skórových intervalů spolehlivosti pro $p_1(0)$, $p_2(0)$, $p_3(0)$:

$p_1(0) \in (0,4457; 0,5843)$, $p_2(0) \in (0,1626; 0,2769)$, $p_3(0) \in (0,2131; 0,3366)$ vždy

s pravděpodobností 95 %.

Ověření podmínek dobré aproximace pro pravděpodobnosti přechodu:

23,1683	24,6931	17,2079
8,5714	10,2857	8,5714
4,5283	6,7925	9,8113

Podmínky dobré aproximace nejsou splněny ve čtyřech případech.

Dolní a horní meze 95% Waldových intervalů spolehlivosti pro pravděpodobnosti přechodu:

dW =

0.2630	0.3293	0.1373
0.1491	0.2789	0.1491
0.0156	0.0546	0.6389

hW =

0.4498	0.5222	0.2983
0.4223	0.5782	0.4223
0.1730	0.2473	0.8706

Dolní a horní meze 95% skórových intervalů spolehlivosti pro pravděpodobnosti přechodu:

dS =

0.2699	0.3338	0.1485
0.1717	0.2912	0.1717
0.0410	0.0785	0.6243

hS =

0.4535	0.5231	0.3078
0.4357	0.5779	0.4357
0.2025	0.2705	0.8507

Ad d) Testujeme hypotézu H_0: $p_1(0) = 0,56 \wedge p_2(0) = 0,23 \wedge p_3(0) = 0,21$.
Ověříme podmínky dobré aproximace: $cp_i \geq 5$ pro i = 1, 2, 3.
i = 1: 196.0,56 = 109,76, i = 2: 196.0,23 = 45,08, i = 3: 196.0,21 = 41,16

Testová statistika Waldova testu: 4,3154
Kritický obor: $W = \langle \chi^2_{0,95}(2), \infty) = \langle 5,9915; \infty)$

$T_0 \notin W \Rightarrow H_0$ nezamítáme na asymptotické hladině významnosti 0,05.
Testová statistika testu poměrem věrohodnosti: 4,0537
Kritický obor: $W = \langle \chi^2_{0,95}(2), \infty) = \langle 5,9915; \infty)$

$T_0 \in W \Rightarrow H_0$ nezamítáme na asymptotické hladině významnosti 0,05.

Návod na řešení pomocí MATLABu:
Ad b) a c) Bodové a intervalové odhady počátečních pravděpodobností a pravděpodobností přechodu poskytne funkce odhady.m.
Ad d) Pro provedení testu dobré shody slouží funkce test_shody.m.