6. Odhady pravděpodobností v HMŘ a testy hypotéz o nich

6.1. Popis situace: Předpokládejme, že HMŘ $\{X_n; n \in N_0\}$ s konečným počtem stavů k má vektor počátečních pravděpodobností $\mathbf{p}(0) = (p_1(0), p_2(0), \ldots, p_k(0))$ a matici přechodu $\mathbf{P} = \begin{pmatrix} p_{11} & \cdots & p_{1k} \\ \cdots & \cdots & \cdots \\ p_{k1} & \cdots & p_{kk} \end{pmatrix}$. Tyto pravděpodobnosti však neznáme, můžeme je pouze odhadnout na základě dlouhodobého pozorování systému.

Pro i, j = 1, …, k označme:

c_{ij} … počet pozorování přechodu systému ze stavu i do stavu j,

$c_i = \sum_{j=1}^{k} c_{ij}$ … celkový počet přechodů, které začínaly ve stavu i,

$c = \sum_{i=1}^{k} c_i$ … celkový počet všech pozorovaných přechodů.

Pozorované hodnoty zaznamenáme do tabulky:

	1	2	…	k	∑
1	c_{11}	c_{12}	…	c_{1k}	c_1
2	c_{21}	c_{22}	…	c_{2k}	c_2
…	…	…	…	…	…
k	c_{k1}	c_{k2}	…	c_{kk}	c_k
∑					c

6.2. Bodové odhady počátečních pravděpodobností a pravděpodobností přechodu

Bodový odhad počáteční pravděpodobnosti $p_i(0)$ získáme metodou maximální věrohodnosti. Předpokládáme, že v čase n = 0 máme celkem c pozorování přechodu řetězce, z nichž v c_i případech byl řetězec ve stavu i. Číslo c_i považujeme za realizaci náhodné veličiny Y_i. Je zřejmé, že $Y_i \sim \mathrm{Bi}(c, p_i(0))$, tedy $P(Y_i = c_i) = \binom{c}{c_i} p_i(0)^{c_i} (1 - p_i(0))^{c - c_i}$, $c_i = 0,1,\ldots,c$.

Zlogaritmováním pravděpodobnostní funkce $P(Y_i = c_i)$ získáme logaritmickou věrohodnostní funkci:

$$l(p_i(0); c_i) = \ln\binom{c}{c_i} + c_i \ln p_i(0) + (c - c_i)\ln(1 - p_i(0)).$$

Tuto funkci derivujeme podle $p_i(0)$ a derivaci položíme rovnu 0:

$$\frac{dl(p_i(0); c_i)}{dp_i(0)} = \frac{c_i}{p_i(0)} - \frac{c - c_i}{1 - p_i(0)} = 0$$

Vyřešením této rovnice získáme maximálně věrohodný odhad počáteční pravděpodobnosti $p_i(0)$: $\hat{p}_i(0) = \frac{c_i}{c}$.

Je to relativní četnost počtu přechodů, které začínaly ve stavu i.

Analogicky lze odvodit, že maximálně věrohodným odhadem pravděpodobnosti přechodu p_{ij} je $\hat{p}_{ij} = \begin{cases} \frac{c_{ij}}{c_i} \text{ pro } c_i \neq 0 \\ 0 \text{ pro } c_i = 0 \end{cases}$, tedy

$\hat{p}_{ij}$ je relativní četnost případů, kdy řetězec, nacházející se v okamžiku n ve stavu i, posunul v čase n + 1 do stavu j.

6.3. Intervalové odhady počátečních pravděpodobností a pravděpodobností přechodu

a) **Waldův interval spolehlivosti**

Při výpočtu mezí intervalu spolehlivosti využijeme **centrální limitní větu**:

Jsou-li náhodné veličiny $X_1, \ldots, X_n$ stochasticky nezávislé a všechny mají stejné rozložení se střední hodnotou μ a rozptylem σ^2, pak pro velká n ($n \geq 30$) lze rozložení standardizovaného součtu $U = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ aproximovat standardizovaným normálním rozložením N(0,1). Zkráceně píšeme $U_n \approx N(0,1)$.

Vraťme se k HMŘ. Na náhodnou veličinu $Y_i \sim Bi(c, p_i(0))$ můžeme pohlížet jako součet c nezávislých náhodných veličin, z nichž každá se řídí alternativním rozložením A($p_i(0)$). Přitom $E(Y_i) = cp_i(0)$ a $D(Y_i) = cp_i(0)[1-p_i(0)]$.

Podle CLV tedy platí: $U = \frac{Y_i - cp_i(0)}{\sqrt{cp_i(0)[1-p_i(0)]}} \approx N(0,1)$.

Čitatele a jmenovatele podělíme c a $p_i(0)$ ve jmenovateli nahradíme odhadem $\hat{p}_i(0)$: $U = \frac{Y_i/c - p_i(0)}{\sqrt{\frac{\hat{p}_i(0)[1-\hat{p}_i(0)]}{c}}} \approx N(0,1)$

100(1-α)% asymptotický Waldův interval spolehlivosti pro počáteční pravděpodobnost $p_i(0)$ má tedy meze:

$$D = \frac{Y_i}{c} - \sqrt{\frac{\hat{p}_i(0)[1-\hat{p}_i(0)]}{c}} u_{1-\alpha/2}, \quad H = \frac{Y_i}{c} + \sqrt{\frac{\hat{p}_i(0)[1-\hat{p}_i(0)]}{c}} u_{1-\alpha/2}$$

Po nahrazení náhodné veličiny Y_i její realizací c_i dostaneme empirický interval spolehlivosti:

$$\hat{p}_i(0) \pm \sqrt{\frac{\hat{p}_i(0)[1-\hat{p}_i(0)]}{c}} u_{1-\alpha/2}, \text{ kde } \hat{p}_i(0) = \frac{c_i}{c}$$

Analogicky se odvodí 100(1-α)% asymptotický Waldův interval spolehlivosti pro pravděpodobnost přechodu. Empirické meze jsou $\hat{p}_{ij} \pm \sqrt{\frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{c_i}} u_{1-\alpha/2}$, kde $\hat{p}_{ij} = \frac{c_{ij}}{c_i}$

Waldův interval spolehlivosti lze korektně použít, je-li splněna podmínka dobré aproximace: $c\hat{p}_i(0)[1-\hat{p}_i(0)] > 9$ resp. $c_i\hat{p}_{ij}(0)[1-\hat{p}_{ij}(0)] > 9$. Není-li tato podmínka splněna, doporučuje se použít skórový interval spolehlivosti.

b) **Skórový interval spolehlivosti**

Pokud ve statistice $U = \frac{Y_i/c - p_i(0)}{\sqrt{p_i(0)[1-p_i(0)]/c}}$ nenahradíme $p_i(0)$ ve jmenovateli odhadem $\hat{p}_i(0)$, můžeme sestrojit skórový interval spolehlivosti pro $p_i(0)$.

Meze tohoto intervalu splňují nerovnost: $\left|\frac{Y_i/c - p_i(0)}{\sqrt{p_i(0)[1-p_i(0)]/c}}\right| \le u_{1-\alpha/2}$.

Umocníme na druhou: $\left[\frac{Y_i/c - p_i(0)}{\sqrt{p_i(0)[1-p_i(0)]/c}}\right]^2 \le u^2_{1-\alpha/2}$.

Řešením kvadratické rovnice pro $p_i(0)$ obdržíme : $\frac{2\frac{Y_i}{c} + \frac{u^2_{1-\alpha/2}}{c} \pm \sqrt{\left(2\frac{Y_i}{c} + \frac{u^2_{1-\alpha/2}}{c}\right)^2 - 4\left(1+\frac{u^2_{1-\alpha/2}}{c}\right)\left(\frac{Y_i}{c}\right)^2}}{2\left(1+\frac{u^2_{1-\alpha/2}}{c}\right)}$.

Po nahrazení náhodné veličiny Y_i její realizací c_i dostaneme empirický interval spolehlivosti:

$\frac{2\hat{p}_i(0) + \frac{u^2_{1-\alpha/2}}{c} \pm \sqrt{\left(2\hat{p}_i(0) + \frac{u^2_{1-\alpha/2}}{c}\right)^2 - 4\left(1+\frac{u^2_{1-\alpha/2}}{c}\right)\hat{p}_i(0)^2}}{2\left(1+\frac{u^2_{1-\alpha/2}}{c}\right)}$, kde $\hat{p}_i(0) = \frac{c_i}{c}$

Analogicky se odvodí 100(1-α)% asymptotický skórový interval spolehlivosti pro pravděpodobnost přechodu.

Empirické meze jsou

$$\frac{2\hat{p}_{ij} + \frac{u^2_{1-\alpha/2}}{c_i} \pm \sqrt{\left(2\hat{p}_{ij} + \frac{u^2_{1-\alpha/2}}{c_i}\right)^2 - 4\left(1 + \frac{u^2_{1-\alpha/2}}{c_i}\right)\hat{p}_{ij}^{\,2}}}{2\left(1 + \frac{u^2_{1-\alpha/2}}{c_i}\right)}, \text{ kde } \hat{p}_{ij} = \frac{c_{ij}}{c_i}$$

6.4. Příklad: V jistém regionu bylo náhodně vybráno 2501 domácností. Bylo zjištěno, že k určitému datu 629 domácností nepředplácelo žádný deník, 750 předplácelo regionální deník a zbytek celostátní deník. Z těch domácností, které neměly žádné předplatné, hodlá v příštím měsíci 126 předplácet regionální a 63 celostátní deník. Z domácností, které předplácejí regionální deník, u něj v příštím měsíci zůstane 525 domácností a 75 začne předplácet celostátní deník. A nakonec z těch domácností, které předplácejí celostátní deník, 673 nezmění předplatné a 112 přejde na předplatné regionálního deníku.
Modelujte situaci pomocí homogenního markovského řetězce a najděte bodové a intervalové odhady (se spolehlivostí 95 %) počátečních pravděpodobností a pravděpodobností přechodu.

Řešení: Zavedeme homogenní markovský řetězec $\{X_n; n \in N_0\}$ s množinou stavů J = {1, 2, 3}, kde $X_n = 1$, když v n-tém měsíci náhodně vybraná domácnost nemá žádné předplatné, $X_n = 2$, když má předplatné na regionální deník a $X_n = 3$, když má předplatné na celostátní deník. Údaje obsažené v textu úlohy uspořádáme do tabulky:

	1	2	3	Σ
1	440	126	63	629
2	150	525	75	750
3	337	112	673	1122
Σ				2501

Nejprve odhadneme počáteční pravděpodobnosti podle vzorce $\hat{p}_i(0) = \frac{c_i}{c}, i = 1,2,\ldots,k$.

V našem případě k = 3, $c_1 = 629$, $c_2 = 750$, $c_3 = 1122$, c = 2501.

$$\hat{p}_1(0) = \frac{629}{2501} = 0{,}2515,\ \hat{p}_2(0) = \frac{750}{2501} = 0{,}2999,\ \hat{p}_3(0) = \frac{1122}{2501} = 0{,}4486$$

Odhad vektoru počátečních pravděpodobností: $\hat{p}(0) = (0{,}25; 0{,}3; 0{,}45)$.

Znamená to, že na počátku sledování 25 % domácností v daném regionu nemělo žádné předplatné, 30 % předplácelo regionální deník a 45 % celostátní deník.

Před výpočtem intervalů spolehlivosti ověříme, zda jsou splněny podmínky dobré aproximace $\hat{p}_i(0)[1-\hat{p}_i(0)]c > 9$.

Přitom $\hat{p}_1(0) = \frac{629}{2501}, \hat{p}_2(0) = \frac{750}{2501}, \hat{p}_3(0) = \frac{1122}{2501}, c = 2501$. Tedy

i = 1: $\frac{629}{2501}\left(1-\frac{629}{2501}\right)2501 = 469,$

i = 2: $\frac{750}{2501}\left(1-\frac{750}{2501}\right)2501 = 525,$

i = 3: $\frac{1122}{2501}\left(1-\frac{1122}{2501}\right)2501 = 619.$

Vidíme, že podmínky jsou splněny. Pro i = 1, 2, 3 a $\alpha = 0{,}05$ dosadíme do vzorce $\hat{p}_i(0) \pm \sqrt{\frac{\hat{p}_i(0)[1-\hat{p}_i(0)]}{c}} u_{1-\alpha/2}$.

Dostaneme meze 95% asymptotických Waldových intervalů spolehlivosti pro $p_1(0)$, $p_2(0)$, $p_3(0)$.

$p_1(0) \in (0{,}2345; 0{,}2685)$, $p_2(0) \in (0{,}2819; 0{,}3178)$, $p_3(0) \in (0{,}4291; 0{,}4681)$ vždy s pravděpodobností 95 %.

Interpretujeme např. 1. interval spolehlivosti: Ve sledovaném regionu je k danému datu s pravděpodobností 95 % 23,45 % až 26,85 % domácností, které nepředplácejí žádný deník.

Pro porovnání nyní vypočteme meze 95% skórových intervalů spolehlivosti pro $p_1(0)$, $p_2(0)$, $p_3(0)$.

Pro i = 1, 2, 3 a $\alpha = 0{,}05$ dosadíme do vzorce $\frac{2\hat{p}_i(0) + \frac{u^2_{1-\alpha/2}}{c} \pm \sqrt{\left(2\hat{p}_i(0) + \frac{u^2_{1-\alpha/2}}{c}\right)^2 - 4\left(1 + \frac{u^2_{1-\alpha/2}}{c}\right)\hat{p}_i(0)^2}}{2\left(1 + \frac{u^2_{1-\alpha/2}}{c}\right)}$.

Dostaneme meze 95% asymptotických skórových intervalů spolehlivosti pro $p_1(0)$, $p_2(0)$, $p_3(0)$.

$p_1(0) \in (0{,}2349; 0{,}2689)$, $p_2(0) \in (0{,}2822; 0{,}3191)$, $p_3(0) \in (0{,}4292; 0{,}4682)$ vždy s pravděpodobností 95 %.

Nyní se budeme věnovat odhadům pravděpodobností přechodu. Použijeme vzorec $\hat{p}_{ij} = \frac{c_{ij}}{c_i}, i, j = 1, \ldots, k$.

Znovu uvedeme tabulku se zadanými údaji:

	1	2	3	Σ
1	440	126	63	629
2	150	525	75	750
3	337	112	673	1122
Σ				2501

V našem případě k = 3,

$c_{11} = 440$, $c_{12} = 126$, $c_{13} = 63$, $c_1 = 629$,

$c_{21} = 150$, $c_{22} = 525$, $c_{23} = 75$, $c_2 = 750$,

$c_{31} = 337$, $c_{32} = 112$, $c_{33} = 673$, $c_3 = 1122$.

$$\hat{p}_{11} = \frac{440}{629} = 0{,}6995,\ \hat{p}_{12} = \frac{126}{629} = 0{,}2003,\ \hat{p}_{13} = \frac{63}{629} = 0{,}1002$$

$$\hat{p}_{21} = \frac{150}{750} = 0{,}2,\ \hat{p}_{22} = \frac{525}{750} = 0{,}7,\ \hat{p}_{23} = \frac{75}{750} = 0{,}1$$

$$\hat{p}_{31} = \frac{337}{1122} = 0{,}3004,\ \hat{p}_{32} = \frac{112}{1122} = 0{,}0998,\ \hat{p}_{33} = \frac{673}{1122} = 0{,}5998$$

$$\hat{\mathbf{P}} = \begin{pmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}2 & 0{,}7 & 0{,}1 \\ 0{,}3 & 0{,}1 & 0{,}6 \end{pmatrix}$$

Interpretujeme např. 1. řádek odhadnuté matice přechodu: Pokud v jednom měsíci náhodně vybraná domácnost neodebírala žádný deník, tak v příštím měsíci s pravděpodobností 0,7 opět nebude mít žádné předplatné, s pravděpodobností 0,2 si předplatí regionální deník a s pravděpodobností 0,1 celostátní deník.

Před výpočtem intervalů spolehlivosti ověříme splnění podmínek dobré aproximace $c_i \hat{p}_{ij}(1-\hat{p}_{ij}) > 9$.

Připomínáme, že

$c_{11} = 440$, $c_{12} = 126$, $c_{13} = 63$, $c_1 = 629$,
$c_{21} = 150$, $c_{22} = 525$, $c_{23} = 75$, $c_2 = 750$,
$c_{31} = 337$, $c_{32} = 112$, $c_{33} = 673$, $c_3 = 1122$.

i = 1: $\frac{440}{629}\left(1-\frac{440}{629}\right)629 = 132, \frac{126}{629}\left(1-\frac{126}{629}\right)629 = 101, \frac{63}{629}\left(1-\frac{63}{629}\right)629 = 57$

i = 2: $\frac{150}{750}\left(1-\frac{150}{750}\right)750 = 120, \frac{525}{750}\left(1-\frac{525}{750}\right)750 = 158, \frac{75}{750}\left(1-\frac{75}{750}\right)750 = 68$

i = 3: $\frac{337}{1122}\left(1-\frac{337}{1122}\right)1122 = 236, \frac{112}{1122}\left(1-\frac{112}{1122}\right)1122 = 109, \frac{673}{1122}\left(1-\frac{673}{1122}\right)1122 = 269$

Ve všech devíti případech jsou podmínky dobré aproximace splněny, můžeme tedy spočítat meze 95% Waldových asymptotických intervalů spolehlivosti pro pravděpodobnosti přechodu. Pro i, j = 1, 2, 3 a $\alpha = 0{,}05$ dosadíme do vzorce $\hat{p}_{ij} \pm \sqrt{\frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{c_i}} u_{1-\alpha/2}$.

$p_{11} \in (0{,}6637;0{,}7354), p_{12} \in (0{,}169;0{,}2316), p_{13} \in (0{,}0767;0{,}1236),$
$p_{21} \in (0{,}1714;0{,}2286), p_{22} \in (0{,}6672;0{,}7328), p_{23} \in (0{,}0785;0{,}1215),$
$p_{31} \in (0{,}2735;0{,}3272), p_{32} \in (0{,}0823;0{,}1174), p_{33} \in (0{,}5712;0{,}6285).$

Interpretujeme např. interval spolehlivosti pro p_{11}: Pokud v jednom měsíci náhodně vybraná domácnost neodebírala žádný deník, tak v příštím měsíci můžeme se spolehlivostí 95 % zaručit, že s pravděpodobností 66,37 % až 73,54 % opět nebude odebírat žádný deník.

Pro srovnání spočteme meze 95% skórových asymptotických intervalů spolehlivosti pro pravděpodobnosti přechodu. Pro i, j = 1, 2, 3 a $\alpha = 0{,}05$ dosadíme do vzorce

$$\frac{2\hat{p}_{ij} + \frac{u^2_{1-\alpha/2}}{c_i} \pm \sqrt{\left(2\hat{p}_{ij} + \frac{u^2_{1-\alpha/2}}{c_i}\right)^2 - 4\left(1 + \frac{u^2_{1-\alpha/2}}{c_i}\right)\hat{p}_{ij}^{\,2}}}{2\left(1 + \frac{u^2_{1-\alpha/2}}{c_i}\right)}.$$

$p_{11} \in (0{,}6626;0{,}7341), p_{12} \in (0{,}1709;0{,}2334), p_{13} \in (0{,}0791;0{,}1261),$
$p_{21} \in (0{,}1729;0{,}2301), p_{22} \in (0{,}6663;0{,}7317), p_{23} \in (0{,}0805;0{,}1236),$
$p_{31} \in (0{,}2743;0{,}3278), p_{32} \in (0{,}0836;0{,}1188), p_{33} \in (0{,}5709;0{,}6281)$

6.5. Testy o počátečních pravděpodobnostech

Máme HMŘ $\{X_n; n \in N_0\}$ s konečným počtem stavů k a vektorem počátečních pravděpodobností **p**(0) = ($p_1(0)$, $p_2(0)$, …, $p_k(0)$). Nechť c_i je počet těch případů, kdy se řetězec v okamžiku n = 0 nachází ve stavu i a počet všech výskytů řetězce v jednotlivých stavech je $c = \sum_{i=1}^{k} c_i$.

Na hladině významnosti α testujeme hypotézu H_0: $p_1(0) = p_1 \wedge \ldots \wedge p_k(0) = p_k$ ($p_1 \geq 0$, …, $p_k \geq 0$ jsou předem dané pravděpodobnosti, jejich součet je 1) proti alternativě H_1: aspoň jedna rovnost neplatí.

a) **Waldův test** (známý také jako Pearsonův chí-kvadrát test dobré shody)

Testová statistika: $T_0 = \sum_{i=1}^{k} \frac{(c_i - cp_i)^2}{cp_i} \approx \chi^2(k-1)$, když H_0 platí.

Kritický obor: $W = \langle \chi^2_{1-\alpha}(k-1), \infty)$

$T_0 \in W \Rightarrow H_0$ zamítáme na asymptotické hladině významnosti α.

b) **Test poměrem věrohodnosti**

Testová statistika: $T_0 = 2\sum_{i=1}^{k} c_i \ln \frac{c_i / c}{p_i} \approx \chi^2(k-1)$, když H_0 platí.

Kritický obor: $W = \langle \chi^2_{1-\alpha}(k-1), \infty)$

$T_0 \in W \Rightarrow H_0$ zamítáme na asymptotické hladině významnosti α.

Upozornění: Musí být splněny podmínky dobré aproximace $cp_i \geq 5$ pro všechna i = 1, …, k. Není-li pro některé stavy tato podmínka splněna, slučujeme tyto stavy se stavy jim nejbližšími.

6.6. Příklad: Vraťme se k příkladu s předplácením denního tisku v daném regionu. Připomeňme, že z 2501 náhodně vybraných domácností jich k určitému datu 629 nepředplácelo žádný deník, 750 předplácelo regionální deník a zbytek (tj. 1122) celostátní deník. V celostátním měřítku k témuž datu žádné předplatné nemá 23 % domácností, jakýkoliv regionální deník si předplácí 29 % domácností a 48 % domácností si předplácí celostátní deník. Na hladině významnosti 0,05 testujte hypotézu, že vektor počátečních pravděpodobností v daném regionu se shoduje s vektorem počátečních pravděpodobností v celostátním měřítku.

Řešení: Testujeme hypotézu H_0: $p_1(0) = 0{,}23 \wedge p_2(0) = 0{,}29 \wedge p_3(0) = 0{,}48$.

Ověříme podmínky dobré aproximace: $cp_i \geq 5$ pro i = 1, 2, 3.

i = 1: 2501.0,23 = 575,23, i = 2: 2501.0,29 = 725,29, i = 3: 2501.0,48 = 1200,48

Testová statistika Waldova testu:

$$T_0 = \sum_{i=1}^{k} \frac{(c_i - cp_i)^2}{cp_i} = \frac{(629 - 2501 \cdot 0{,}23)^2}{2501 \cdot 0{,}23} + \frac{(750 - 2501 \cdot 0{,}29)^2}{2501 \cdot 0{,}29} + \frac{(1122 - 2501 \cdot 0{,}48)^2}{2501 \cdot 0{,}48} = 10{,}9986$$

Kritický obor: $W = \langle \chi^2_{0{,}95}(2), \infty) = \langle 5{,}9915; \infty)$

$T_0 \in W \Rightarrow H_0$ zamítáme na asymptotické hladině významnosti 0,05. Pro úplnost: p = 0,0041.

Testová statistika testu poměrem věrohodnosti:

$$T_0 = 2\sum_{i=1}^{k} c_i \ln \frac{c_i/c}{p_i} = 2\left(629 \cdot \ln \frac{629/2501}{0{,}23} + 750 \cdot \ln \frac{750/2501}{0{,}29} + 1122 \cdot \ln \frac{1122/2501}{0{,}48} \right) = 10{,}9551$$

Kritický obor: $W = \langle \chi^2_{0{,}95}(2), \infty) = \langle 5{,}9915; \infty)$

$T_0 \in W \Rightarrow H_0$ zamítáme na asymptotické hladině významnosti 0,05. Pro úplnost: p = 0,0042.

6.7. Testy o pravděpodobnostech přechodu

Nechť $\{X_n; n \in N_0\}$ je HMŘ s konečným počtem stavů k a maticí přechodu **P** = $(p_{ij})_{i,j=1}^{k}$. Na hladině významnosti α testujeme hypotézu H_0: $p_{ij} = p_{ij}^{\,0}$ pro $\forall i, j = 1, \ldots, k$ ($p_{ij}^{\,0} \geq 0$ jsou předem dané pravděpodobnosti, $\sum_{j=1}^{k} p_{ij}^{\,0} = 1, i = 1, \ldots, k$) proti alternativě H_1: existují i,j taková, že $p_{ij} \neq p_{ij}^{\,0}$.

a) **Waldův test** (známý také jako Pearsonův chí-kvadrát test dobré shody)

Testová statistika: $T_0 = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{c_i \left(\hat{p}_{ij} - p_{ij}^{\,0}\right)^2}{p_{ij}^{\,0}} \approx \chi^2\left(k(k-1)\right)$, když H_0 platí.

Kritický obor: $W = \left\langle \chi^2_{1-\alpha}\left(k(k-1)\right), \infty \right)$

$T_0 \in W \Rightarrow$ H_0 zamítáme na asymptotické hladině významnosti α.

b) **Test poměrem věrohodnosti**

Testová statistika: $T_0 = 2\sum_{i=1}^{k} \sum_{j=1}^{k} c_{ij} \ln \frac{\hat{p}_{ij}}{p_{ij}^{\,0}} \approx \chi^2\left(k(k-1)\right)$, když H_0 platí.

Kritický obor: $W = \left\langle \chi^2_{1-\alpha}\left(k(k-1)\right), \infty \right)$

$T_0 \in W \Rightarrow$ H_0 zamítáme na asymptotické hladině významnosti α.

Upozornění:

Musí být splněny podmínky dobré aproximace $c_i p_{ij} \geq 5$ pro všechna i,j = 1, …, k. Není-li pro některé stavy tato podmínka splněna, slučujeme tyto stavy se stavy jim nejbližšími.

Tvrzení o rozložení testové statistiky za platnosti H_0 je pravdivé, pokud všechny pravděpodobnosti p_{ij}^0 jsou kladné. V případě, že některá p_{ij}^0 jsou nulová, pak ve vzorci pro T_0 uvažujeme pouze takové dvojice indexů (i, j), pro které $p_{ij}^0 > 0$. Pak T_0 má za platnosti H_0 asymptoticky rozložení $\chi^2(k(k-1)-s)$, kde s je počet nulových pravděpodobností.

6.8. Příklad: V příkladu s předplácením denního tisku jsme na základě údajů zjištěných v daném regionu, tj. pomocí tabulky

	1	2	3	Σ
1	440	126	63	629
2	150	525	75	750
3	337	112	673	1122
Σ				2501

odhadli matici přechodu: $\hat{\mathbf{P}} = \begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,2 & 0,7 & 0,1 \\ 0,3 & 0,1 & 0,6 \end{pmatrix}$. Je známo, že v celostátním měřítku má matice přechodu tvar $\mathbf{P}^0 = \begin{pmatrix} 0,72 & 0,19 & 0,09 \\ 0,21 & 0,68 & 0,11 \\ 0,30 & 0,12 & 0,58 \end{pmatrix}$.

Na hladině významnosti 0,05 testujte hypotézu H_0: $\mathbf{P} = \mathbf{P}^0$ proti H_1: $\mathbf{P} \neq \mathbf{P}^0$.

Řešení:

Ad a) Waldův test

Musíme ověřit splnění podmínek dobré aproximace.

i = 1: $c_{11} = 440$, $c_{12} = 126$, $c_{13} = 63$, $c_1 = 629$

$c_1 p_{11}^{\,0} = 629.0,72 = 452,88$, $c_1 p_{12}^{\,0} = 629.0,19 = 119,51$, $c_1 p_{13}^{\,0} = 629.0,09 = 56,61$

i = 2: $c_{21} = 150$, $c_{22} = 525$, $c_{23} = 75$, $c_2 = 750$

$c_2 p_{21}^{\,0} = 750.0,21 = 157,5$, $c_2 p_{22}^{\,0} = 750.0,68 = 510$, $c_2 p_{23}^{\,0} = 750.0,11 = 82,5$

i = 3: $c_{31} = 337$, $c_{32} = 112$, $c_{33} = 673$, $c_3 = 1122$

$c_3 p_{31}^{\,0} = 1122.0,3 = 336,6$, $c_3 p_{32}^{\,0} = 1122.0,12 = 134,64$, $c_3 p_{33}^{\,0} = 1122.0,58 = 650,76$

Podmínky dobré aproximace jsou splněny, můžeme vypočítat testovou statistiku.

Pro výpočet testové statistiky potřebujeme výchozí tabulku četností přechodu mezi jednotlivými stavy a matici $\mathbf{P}^0$:

	1	2	3	Σ
1	440	126	63	629
2	150	525	75	750
3	337	112	673	1122
Σ				2501

$$\mathbf{P}^0 = \begin{pmatrix} 0{,}72 & 0{,}19 & 0{,}09 \\ 0{,}21 & 0{,}68 & 0{,}11 \\ 0{,}30 & 0{,}12 & 0{,}58 \end{pmatrix}$$

$$T_0 = \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{c_i\left(\hat{p}_{ij} - p_{ij}^{\,0}\right)^2}{p_{ij}^{\,0}} = 629\left[\frac{\left(\frac{440}{629} - 0{,}72\right)^2}{0{,}72} + \frac{\left(\frac{126}{629} - 0{,}19\right)^2}{0{,}19} + \frac{\left(\frac{63}{629} - 0{,}09\right)^2}{0{,}09}\right] +$$

$$+ 750\left[\frac{\left(\frac{150}{750} - 0{,}21\right)^2}{0{,}21} + \frac{\left(\frac{525}{750} - 0{,}68\right)^2}{0{,}68} + \frac{\left(\frac{75}{750} - 0{,}11\right)^2}{0{,}11}\right] +$$

$$+ 1122\left[\frac{\left(\frac{337}{1122} - 0{,}3\right)^2}{0{,}3} + \frac{\left(\frac{112}{1122} - 0{,}12\right)^2}{0{,}68} + \frac{\left(\frac{673}{1122} - 0{,}58\right)^2}{0{,}58}\right] = 7{,}4877$$

Kritický obor: $W = \left\langle \chi^2_{1-\alpha}\left(k(k-1)\right), \infty\right) = \left\langle \chi^2_{0{,}95}(6), \infty\right) = \langle 12{,}5916; \infty)$

Testová statistika nepatří do kritického oboru, nulovou hypotézu tedy nezamítáme na asymptotické hladině významnosti 0,05.

Ad b) Test poměrem věrohodnosti

Pro výpočet testové statistiky potřebujeme výchozí tabulku četností přechodu mezi jednotlivými stavy a matici $\mathbf{P}^0$:

	1	2	3	Σ
1	440	126	63	629
2	150	525	75	750
3	337	112	673	1122
Σ				2501

$$\mathbf{P}^0 = \begin{pmatrix} 0{,}72 & 0{,}19 & 0{,}09 \\ 0{,}21 & 0{,}68 & 0{,}11 \\ 0{,}30 & 0{,}12 & 0{,}58 \end{pmatrix}$$

$$T_0 = 2\sum_{i=1}^{k}\sum_{j=1}^{k} c_{ij} \ln \frac{\hat{p}_{ij}}{p_{ij}^{\,0}} =$$

$$= 2\left(\begin{array}{l} 440 \cdot \ln \frac{\frac{440}{629}}{0{,}72} + 126 \cdot \ln \frac{\frac{126}{629}}{0{,}19} + 63 \cdot \ln \frac{\frac{63}{629}}{0{,}09} + 150 \cdot \ln \frac{\frac{150}{750}}{0{,}21} + 525 \cdot \ln \frac{\frac{525}{750}}{0{,}68} + 75 \cdot \ln \frac{\frac{75}{750}}{0{,}11} + \\ + 337 \cdot \ln \frac{\frac{337}{1122}}{0{,}3} + 112 \cdot \ln \frac{\frac{112}{1122}}{0{,}12} + 673 \cdot \ln \frac{\frac{673}{1122}}{0{,}58} \end{array} \right) = 7{,}7073$$

Kritický obor: $W = \langle \chi^2_{1-\alpha}(k(k-1)), \infty) = \langle \chi^2_{0{,}95}(6), \infty) = \langle 12{,}5916; \infty)$

Testová statistika nepatří do kritického oboru, nulovou hypotézu tedy nezamítáme na asymptotické hladině významnosti 0,05.

6.9. Test homogenity dvou vektorů počátečních pravděpodobností

Předpokládáme, že máme dva nezávislé homogenní markovské řetězce $\{X_n; n \in N_0\}$ a $\{Y_n; n \in N_0\}$, které mají stejnou konečnou množinu stavů J = {1, 2, …, k}. Označme $\mathbf{p}_X(0)$ vektor počátečních pravděpodobností 1. řetězce a $\mathbf{p}_Y(0)$ vektor počátečních pravděpodobností 2. řetězce. Na hladině významnosti α testujeme hypotézu H_0: $\mathbf{p}_X(0) = \mathbf{p}_Y(0)$ proti alternativě H_1: $\mathbf{p}_X(0) \neq \mathbf{p}_Y(0)$. Přitom máme k dispozici celkové počty přechodů, které začínaly ve stavech 1, 2, …, k:

	1	2	…	k	Σ
1. řetězec	c_1	c_2	…	c_k	c
2. řetězec	d_1	d_2	…	d_k	d
Σ	c_1+d_1	c_2+d_2	…	c_k+d_k	c+d

Testová statistika: $T_0 = \sum_{i=1}^{k} \left[\frac{\left(c_i - \frac{c(c_i+d_i)}{c+d} \right)^2}{\frac{c(c_i+d_i)}{c+d}} + \frac{\left(d_i - \frac{d(c_i+d_i)}{c+d} \right)^2}{\frac{d(c_i+d_i)}{c+d}} \right]$ se za platnosti H_0 asymptoticky řídí rozložením $\chi^2(k-1)$.

Kritický obor: $W = \langle \chi^2_{1-\alpha}(k-1), \infty)$

$T_0 \in W \Rightarrow H_0$ zamítáme na asymptotické hladině významnosti α.

Upozornění: Musí být splněny podmínky dobré aproximace, tj. teoretické četnosti $\frac{c(c_i+d_i)}{c+d}, \frac{d(c_i+d_i)}{c+d}, i = 1, \ldots, k$ musí být aspoň v 80 % případů ≥ 5 a ve zbylých 20 % nesmí klesnout pod 2.

6.10. Příklad: Budeme se zabývat průzkumem předplácení tisku ve dvou různých regionech. V regionu č. 1 z 2501 náhodně vybraných domácností jich k určitému datu nepředplácelo žádný deník 629, regionální deník předplácelo 750 a celostátní deník předplácelo 1122. V regionu č. 2 z 2793 náhodně vybraných domácností jich k témuž dat 678 nepředplácelo žádný deník, 1322 předplácelo regionální deník a 1322 celostátní deník. Na hladině významnosti 0,05 testujte hypotézu, že vektory počátečních pravděpodobností v 1. a 2. regionu se shodují.

Řešení:
Zjištěná data zapíšeme do kontingenční tabulky.

	bez předplatného	regionální deník	celostátní deník	Σ
1. region	629 (25,15 %)	750 (29,99 %)	1122 (44,86 %)	2501
2. region	678 (24,27 %)	793 (28,39 %)	1322 (47,35 %)	2793
Σ	1307	1543	2444	5294

Vypočteme teoretické četnosti.

	bez předplatného	regionální deník	celostátní deník	Σ
1. region	$\frac{2501 \cdot 1307}{5294} = 617{,}46$	$\frac{2501 \cdot 1543}{5294} = 728{,}95$	$\frac{2501 \cdot 2444}{5294} = 1154{,}6$	2501
2. region	$\frac{2793 \cdot 1307}{5294} = 689{,}55$	$\frac{2793 \cdot 1543}{5294} = 814{,}05$	$\frac{2793 \cdot 2444}{5294} = 1289{,}4$	2793
Σ	1307	1543	2444	5294

Vidíme, že podmínky dobré aproximace jsou splněny.

	bez předplatného	regionální deník	celostátní deník	Σ
1. region	629	750	1122	2501
2. region	678	793	1322	2793
Σ	1307	1543	2444	5294

	bez předplatného	regionální deník	celostátní deník	Σ
1. region	$\frac{2501\cdot 1307}{5294}=617,46$	$\frac{2501\cdot 1543}{5294}=728,95$	$\frac{2501\cdot 2444}{5294}=1154,6$	2501
2. region	$\frac{2793\cdot 1307}{5294}=689,55$	$\frac{2793\cdot 1543}{5294}=814,05$	$\frac{2793\cdot 2444}{5294}=1289,4$	2793
Σ	1307	1543	2444	5294

Dosadíme do vzorce pro testovou statistiku.

$$T_0=\sum_{i=1}^{k}\left[\frac{\left(c_i-\frac{c(c_i+d_i)}{c+d}\right)^2}{\frac{c(c_i+d_i)}{c+d}}+\frac{\left(d_i-\frac{d(c_i+d_i)}{c+d}\right)^2}{\frac{d(c_i+d_i)}{c+d}}\right]=\frac{(629-617,46)^2}{617,46}+\frac{(750-728,95)^2}{728,95}+\frac{(1122-1154,6)^2}{1154,6}+$$

$$+\frac{(678-689,55)^2}{689,55}+\frac{(793-814,05)^2}{814,05}+\frac{(1322-1289,4)^2}{1289,4}=3,3062$$

Kritický obor: $W=\langle\chi^2_{1-\alpha}(k-1),\infty)=\langle\chi^2_{0,95}(2),\infty)=\langle 5,9915;\infty)$

Protože $T_0\notin W$, hypotézu o shodě vektorů počátečních pravděpodobností nezamítáme na asymptotické hladině významnosti 0,05.

6.11. Test shody stochastických vektorů matice přechodu

Máme homogenní markovský řetězec $\{X_n; n \in N_0\}$ s množinou stavů J = {1, 2, …, k} a maticí přechodu $\mathbf{P} = \begin{pmatrix} p_{11} & \cdots & p_{1k} \\ \cdots & \cdots & \cdots \\ p_{k1} & \cdots & p_{kk} \end{pmatrix} = \begin{pmatrix} \mathbf{p_1} \\ \vdots \\ \mathbf{p_k} \end{pmatrix}$. Na hladině významnosti α testujeme hypotézu H_0: $\mathbf{p}_1 = \ldots = \mathbf{p}_k$ proti alternativě H_1: aspoň jedna dvojice vektorů se liší. Přitom máme k dispozici tabulku pozorovaných četností všech přechodů daného řetězce:

	1	2	…	k	∑
1	c_{11}	c_{12}	…	c_{1k}	c_1
2	c_{21}	c_{22}	…	c_{2k}	c_2
…	…	…	…	…	…
k	c_{k1}	c_{k2}	…	c_{kk}	c_k
∑					c

Tuto tabulku přepíšeme do jiného tvaru, který nám umožní provést požadovaný test.

stav	setrvání ve stavu	přechod jinam	∑
1	c_{11}	c_1-c_{11}	c_1
2	c_{22}	c_2-c_{22}	c_2
⋮	⋮	⋮	⋮
k	c_{kk}	c_k-c_{kk}	c_k
∑	$\sum_{i=1}^{k} c_{ii}$	$c - \sum_{i=1}^{k} c_{ii}$	c

Pro přehlednější zápis kontingenční tabulky (a následně testové statistiky) zavedeme následující označení:

stav	setrvání ve stavu	přechod jinam	$\sum$
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
⋮	⋮	⋮	⋮
k	n_{k1}	n_{k2}	$n_{k.}$
$\sum$	$n_{.1}$	$n_{.2}$	n

Testová statistika $T_0 = \sum_{i=1}^{k}\sum_{j=1}^{2} \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}}$ se za platnosti H_0 asymptoticky řídí rozložením $\chi^2(k-1)$.

Kritický obor: $W = \langle \chi^2_{1-\alpha}(k-1), \infty)$

$T_0 \in W \Rightarrow H_0$ zamítáme na asymptotické hladině významnosti α.

Upozornění: Musí být splněny podmínky dobré aproximace, tj. teoretické četnosti $\frac{n_{i.}n_{.j}}{n}$, i = 1, 2, …, k, j = 1, 2 musí být aspoň v 80 % případů ≥ 5 a ve zbylých 20 % nesmí klesnout pod 2.

6.12. Příklad: V příkladu 6.4. jsme získali údaje o četnostech přechodu domácností mezi jednotlivými stavy (1 – žádné předplatné, 2 – předplatné regionálního deníku, 3 – předplatné regionálního deníku) v měsíčním kroku:

	1	2	3	Σ
1	440	126	63	629
2	150	525	75	750
3	337	112	673	1122
Σ				2501

Na hladině významnosti 0,05 testujte hypotézu, že všechny tři stochastické vektory v matici přechodu jsou shodné.

Řešení: V příkladu 6.4. jsme vypočítali odhad matice přechodu $\hat{\mathbf{P}} = \begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,2 & 0,7 & 0,1 \\ 0,3 & 0,1 & 0,6 \end{pmatrix}$, tedy

$\hat{\mathbf{p}}_1 = (0,7 \quad 0,2 \quad 0,1), \hat{\mathbf{p}}_2 = (0,2 \quad 0,7 \quad 0,1), \hat{\mathbf{p}}_3 = (0,3 \quad 0,1 \quad 0,6)$. Na asymptotické hladině významnosti 0,05 testujeme H_0: $\mathbf{p}_1 = \mathbf{p}_2 = \mathbf{p}_3$ proti alternativě H_1: aspoň jedna dvojice stochastických vektorů se liší.

Zjištěné údaje přepíšeme do kontingenční tabulky.

stav	setrvání ve stavu	přechod jinam	∑
1	440	189	629
2	525	225	750
3	673	449	1122
∑	1638	863	2501

Vypočteme teoretické četnosti.

stav	setrvání ve stavu	přechod jinam	∑
1	$\frac{629 \cdot 1638}{2501} = 411{,}96$	$\frac{629 \cdot 863}{2501} = 217{,}04$	629
2	$\frac{750 \cdot 1638}{2501} = 491{,}2$	$\frac{750 \cdot 863}{2501} = 258{,}8$	750
3	$\frac{1122 \cdot 1638}{2501} = 734{,}84$	$\frac{1122 \cdot 863}{2501} = 387{,}16$	1122
∑	1638	863	2501

Podmínky dobré aproximace jsou splněny.
Dosadíme do vzorce pro testovou statistiku:

$$T_0 = \sum_{i=1}^{k}\sum_{j=1}^{2} \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}} = \frac{(440-411{,}96)^2}{411{,}96} + \frac{(189-217{,}04)^2}{217{,}04} + \frac{(525-491{,}2)^2}{491{,}2} + \frac{(225-258{,}8)^2}{258{,}8} +$$

$$+ \frac{(673-734{,}84)^2}{734{,}84} + \frac{(449-387{,}16)^2}{387{,}16} = 27{,}3533$$

Kritický obor: $W = \langle \chi^2_{1-\alpha}(k-1), \infty) = \langle \chi^2_{0{,}95}(2), \infty) = \langle 5{,}9915; \infty)$

Protože $T_0 \in W$, hypotézu o shodě stochastických vektorů v matici přechodu zamítáme na asymptotické hladině významnosti 0,05.

6.13. Test shody dvou matic přechodu

Máme dva nezávislé homogenní markovské řetězce $\{X_n; n \in N_0\}$ a $\{Y_n; n \in N_0\}$, které mají stejnou konečnou množinu stavů J = {1, 2, …, k}. Označme $\mathbf{P}_X$ matici přechodu 1. řetězce a $\mathbf{P}_Y$ matici přechodu 2. řetězce. Na hladině významnosti α testujeme hypotézu H_0: $\mathbf{P}_X = \mathbf{P}_Y$ proti alternativě H_1: $\mathbf{P}_X \neq \mathbf{P}_Y$. Přitom máme k dispozici tabulky četností přechodů 1. a 2. řetězce:

1. řetězec

	1	2	…	k	∑
1	c_{11}	c_{12}	…	c_{1k}	c_1
2	c_{21}	c_{22}	…	c_{2k}	c_2
…	…	…	…	…	…
k	c_{k1}	c_{k2}	…	c_{kk}	c_k
∑					c

2. řetězec

	1	2	…	k	∑
1	d_{11}	d_{12}	…	d_{1k}	d_1
2	d_{21}	d_{22}	…	d_{2k}	d_2
…	…	…	…	…	…
k	d_{k1}	d_{k2}	…	d_{kk}	d_k
∑					d

Jednotlivé řádky matice přechodu 1. řetězce označíme $\mathbf{p}_{X1, \ldots, } \mathbf{p}_{Xk}$, 2. řetězce pak $\mathbf{p}_{Y1, \ldots, } \mathbf{p}_{Yk}$.
Test hypotézy o shodě matic přechodu převedeme na k testů shody stochastických vektorů, tj. H_{0j}: $\mathbf{p}_{Xj} = \mathbf{p}_{Yj}$ proti H_{1j}: $\mathbf{p}_{Xj} \neq \mathbf{p}_{Yj}$, j = 1, 2, …, k.

Tabulky četností přechodů přepíšeme do k kontingenčních tabulek, tedy j-tá tabulka (j = 1, 2, …, k) bude mít tvar:

	1	2	…	k	Σ
1. řetězec	c_{j1}	c_{j2}	…	c_{jk}	c_j
2. řetězec	d_{j1}	d_{j2}	…	d_{jk}	d_j
Σ	$c_{j1}+d_{j1}$	$c_{j2}+d_{j2}$	…	$c_{jk}+d_{jk}$	c_j+d_j

Testová statistika pro test shody j-tého řádku matice $\mathbf{P}_X$ a j-tého řádku matice $\mathbf{P}_Y$ je dána vztahem

$$T_{0j} = \sum_{i=1}^{k} \left[\frac{\left(c_{ji} - \frac{c_j\left(c_{ji}+d_{ji}\right)}{c_j+d_j} \right)^2}{\frac{c_j\left(c_{ji}+d_{ji}\right)}{c_j+d_j}} + \frac{\left(d_{ji} - \frac{d_j\left(c_{ji}+d_{ji}\right)}{c_j+d_j} \right)^2}{\frac{d_j\left(c_{ji}+d_{ji}\right)}{c_j+d_j}} \right]$$

. Za platnosti nulové hypotézy H_{0j}: $\mathbf{p}_{Xj} = \mathbf{p}_{Yj}$ se tato statistika asymptoticky řídí rozložením $\chi^2(k-1)$.

Kritický obor: $W = \left\langle \chi^2_{1-\alpha}(k-1), \infty \right)$

$T_{0j} \in W \Rightarrow$ H_{0j} zamítáme na asymptotické hladině významnosti α.

Upozornění: Musí být splněny podmínky dobré aproximace, tj. teoretické četnosti $\frac{c_j\left(c_{ji}+d_{ji}\right)}{c_j+d_j}, \frac{d_j\left(c_{ji}+d_{ji}\right)}{c_j+d_j}, i, j = 1, \ldots, k$ musí být aspoň v 80 % případů ≥ 5 a ve zbylých 20 % nesmí klesnout pod 2.

6.14. Příklad: Budeme se opět zabývat průzkumem předplácení tisku ve dvou různých regionech. V 1. regionu bylo náhodně vybráno 2501 domácností a ve 2. regionu 2793 domácností. Počty přechodů mezi stavy 1 (žádné předplatné), 2 (předplatné regionálního deníku) a 3 (předplatné celostátního deníku) máme uvedeny ve dvou tabulkách:

1. region

	1	2	3	Σ
1	440	126	63	629
2	150	525	75	750
3	337	112	673	1122
Σ				2501

2. region

	1	2	3	Σ
1	492	111	75	678
2	136	574	83	793
3	388	183	751	1322
Σ				2793

Na hladině významnosti 0,05 testujte hypotézu o shodě matic přechodu 1. a 2. řetězce.

Řešení: Budeme testovat tři hypotézy o shodě prvních, druhých a třetích řádků matic přechodu. Podrobněji ukážeme test 1. hypotézy, tj. H_{01}: $\mathbf{p}_{X1} = \mathbf{p}_{Y1}$ proti H_{11}: $\mathbf{p}_{X1} \neq \mathbf{p}_{Y1}$.
Údaje z prvních dvou řádků vstupních tabulek přepíšeme do kontingenční tabulky:

	bez předplatného	regionální deník	celostátní deník	Σ
1. region	440	126	63	629
2. region	492	111	75	678
Σ	932	237	138	1307

Pomocí tabulky četností přechodů ze stavu „bez předplatného“ sestavíme tabulku teoretických četností:

	bez předplatného	regionální deník	celostátní deník	Σ
1. region	440	126	63	629
2. region	492	111	75	678
Σ	932	237	138	1307

	bez předplatného	regionální deník	celostátní deník	Σ
1. region	$\frac{629 \cdot 932}{1307} = 448{,}53$	$\frac{629 \cdot 237}{1307} = 114{,}06$	$\frac{629 \cdot 138}{1307} = 66{,}41$	629
2. region	$\frac{678 \cdot 932}{1307} = 483{,}47$	$\frac{678 \cdot 237}{1307} = 122{,}94$	$\frac{678 \cdot 138}{1307} = 71{,}59$	678
Σ	932	237	138	1307

Podmínky dobré aproximace jsou splněny. Vypočteme testovou statistiku:

$$T_{01} = \frac{(440-448{,}53)^2}{448{,}53} + \frac{(126-114{,}06)^2}{114{,}06} + \frac{(63-66{,}41)^2}{66{,}41} + \frac{(492-483{,}47)^2}{483{,}47} +$$

$$+ \frac{(111-122{,}94)^2}{122{,}94} + \frac{(75-71{,}59)^2}{71{,}59} = 3{,}0614$$

Kritický obor: $W = \langle \chi^2_{1-\alpha}(k-1), \infty) = \langle \chi^2_{0{,}95}(2), \infty) = \langle 5{,}9915; \infty)$

Protože $T_{01} \notin W$, hypotézu o shodě 1. řádků matic přechodu nezamítáme na asymptotické hladině významnosti 0,05.

Analogicky postupujeme při testování shody druhých a třetích řádků matic přechodu. Zjistíme, že v obou případech jsou podmínky dobré aproximace splněny. Vypočteme $T_{02} = 2{,}0784$ a $T_{03} = 8{,}6394$. Protože kritický obor je $W = \langle 5{,}9915; \infty)$, vidíme, že pouze ve 3. případ zamítáme na asymptotické hladině významnosti 0,05 hypotézu o shodě stochastických vektorů matic přechodu. S rizikem omylu nejvýše 5 % jsme tedy prokázali, že z hlediska změn předplatného se v 1. a 2. regionu liší domácnosti, které na počátku sledování odebíraly celostátní deník.