

Zpracování náhodného vektoru

Zpracování náhodného vektoru

- **Př. 1:** Cestující na vybraném spoji linky MHD byli dotazováni za účelem zjištění spokojenosti s kvalitou MHD.

Legenda	
1	Velmi spokojen
2	Spokojen
3	Nespokojen
4	Velmi nespokojen

Průzkum spokojenosti cestujících s MHD	
Hodnocení ceny jízdného	Hodnocení kvality cestování
1	1
1	2
1	3
2	2
2	3
2	3
3	2
3	2
3	3
3	3
3	3
3	3
3	3
3	3
3	4
3	4
4	3
4	3
4	4
4	4
4	4

Zpracování náhodného vektoru

Stanovte:

1. Sdruženou pravděpodobnostní funkci.
2. Marginální pravděpodobnostní funkce.
3. Určete, zda jsou složky náhodného vektoru nezávislé.
4. Určete střední hodnoty a rozptyly jednotlivých složek.
5. Určete kovarianční matici náhodného vektoru.
6. Určete jednoduchý korelační koeficient.
7. Určete podmíněnou pravděpodobnost $P(X = 3|Y = 3)$.

Zpracování náhodného vektoru

Zavedme složky náhodného vektoru: X – hodnocení ceny jízdného, Y – hodnocení kvality cestování.

Tabulka sdružených a marginálních četností						
		Y				Σ
		1	2	3	4	
X	1	1	1	1	0	3
	2	0	1	2	0	3
	3	0	2	5	2	9
	4	0	0	2	3	5
Σ		1	4	10	5	20

Zpracování náhodného vektoru

Korelační tabulka						
		Y				$P_X(x)$
		1	2	3	4	
X	1	0,05	0,05	0,05	0	0,15
	2	0	0,05	0,1	0	0,15
	3	0	0,1	0,25	0,1	0,45
	4	0	0	0,1	0,15	0,25
$P_Y(y)$		0,05	0,2	0,5	0,25	1

Např. $P(X = 1, Y = 1) = 0,05$, $P(X = 3, Y = 3) = 0,25$

$$P_X(X = 1) = 0,15, P_Y(Y = 2) = 0,2$$

Zpracování náhodného vektoru

- Aby byly složky X a Y nezávislé, pak musí platit (v případě diskrétního náhodného vektoru):

$$P(X = x_i, Y = y_j) = P_X(x_i) \cdot P_Y(y_j).$$

Tento předpoklad splněn není, např.

$$P(X = 1, Y = 1) = 0,05, P_X(X = 1) \cdot P_Y(Y = 1) = 0,15 \cdot 0,05 = 0,0075 \\ \Rightarrow P(X = 1, Y = 1) \neq P_X(X = 1) \cdot P_Y(Y = 1).$$

Zpracování náhodného vektoru

- Stanovení střední hodnoty a rozptylu složky X:

$$EX = \sum_{i=1}^4 x_i \cdot P_X(X = x_i) = 1 \cdot 0,15 + 2 \cdot 0,15 + 3 \cdot 0,45 + 4 \cdot 0,25 = 2,8,$$

$$EX^2 = \sum_{i=1}^4 x_i^2 \cdot P_X(X = x_i) = 1^2 \cdot 0,15 + 2^2 \cdot 0,15 + 3^2 \cdot 0,45 + 4^2 \cdot 0,25 = 8,8,$$

$$DX = EX^2 - (EX)^2 = 8,8 - 2,8^2 = 0,96.$$

Zpracování náhodného vektoru

- Stanovení střední hodnoty a rozptylu složky Y:

$$EY = \sum_{j=1}^4 y_j \cdot P_Y(Y = y_j) = 1 \cdot 0,05 + 2 \cdot 0,2 + 3 \cdot 0,5 + 4 \cdot 0,25 = 2,95,$$

$$EY^2 = \sum_{j=1}^4 y_j^2 \cdot P_Y(Y = y_j) = 1^2 \cdot 0,05 + 2^2 \cdot 0,2 + 3^2 \cdot 0,5 + 4^2 \cdot 0,25 = 9,35,$$

$$DY = EY^2 - (EY)^2 = 9,35 - 2,95^2 = 0,6475.$$

Zpracování náhodného vektoru

- Kovarianční matice má tvar $\begin{vmatrix} DX & Cov(X,Y) \\ Cov(X,Y) & DY \end{vmatrix}$.

$$\begin{aligned} Cov(X,Y) &= \sum_{i=1}^4 \sum_{j=1}^4 (x_i - EX) \cdot (y_j - EY) \cdot P(X = x_i, Y = y_j) = \\ &= (1 - 2,8) \cdot (1 - 2,95) \cdot 0,05 + (1 - 2,8) \cdot (2 - 2,95) \cdot 0,05 + \\ &+ (1 - 2,8) \cdot (3 - 2,95) \cdot 0,05 + (2 - 2,8) \cdot (2 - 2,95) \cdot 0,05 + \\ &+ (2 - 2,8) \cdot (3 - 2,95) \cdot 0,1 + (3 - 2,8) \cdot (2 - 2,95) \cdot 0,1 + \\ &+ (3 - 2,8) \cdot (3 - 2,95) \cdot 0,25 + (3 - 2,8) \cdot (4 - 2,95) \cdot 0,1 + \\ &+ (4 - 2,8) \cdot (3 - 2,95) \cdot 0,1 + (4 - 2,8) \cdot (4 - 2,95) \cdot 0,15 = 0,49. \end{aligned}$$

Zpracování náhodného vektoru

- Výpočet kovariance můžeme rovněž provést podle vztahu:

$$\text{Cov}(X, Y) = E(X \cdot Y) - EX \cdot EY = 8,75 - 2,8 \cdot 2,95 = 0,49.$$

$$\begin{aligned} E(X \cdot Y) &= 1 \cdot 1 \cdot 0,05 + 1 \cdot 2 \cdot 0,05 + 1 \cdot 3 \cdot 0,05 + 2 \cdot 2 \cdot 0,05 + \\ &+ 2 \cdot 3 \cdot 0,1 + 3 \cdot 2 \cdot 0,1 + 3 \cdot 3 \cdot 0,25 + 3 \cdot 4 \cdot 0,1 + 4 \cdot 3 \cdot 0,1 + \\ &+ 4 \cdot 4 \cdot 0,15 = 8,75. \end{aligned}$$

- Vidíme, že výsledky se shodují.

Zpracování náhodného vektoru

- Výpočet kovariance v Excelu lze provést pomocí funkce COVAR.
- Kovarianční matice má tedy tvar $\begin{vmatrix} 0,96 & 0,49 \\ 0,49 & 0,6475 \end{vmatrix}$.

Zpracování náhodného vektoru

- Jednoduchý korelační koeficient je definován vztahem:

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{DX \cdot DY}} \text{ pro } DX, DY \neq 0 .$$

- Po dosazení dostaneme:

$$\rho_{X,Y} = \frac{0,49}{\sqrt{0,96 \cdot 0,6475}} \doteq 0,62 .$$

Zpracování náhodného vektoru

- Výpočet jednoduchého korelačního koeficientu lze v Excelu provést pomocí funkce CORREL.
- Podmíněná pravděpodobnost pro diskrétní náhodný vektor je definována:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P_Y(Y = y)} \text{ pro } P_Y(Y = y) \neq 0.$$

$$\text{Tedy } P(X = 3|Y = 3) = \frac{P(X = 3, Y = 3)}{P_Y(Y = 3)} = \frac{0,25}{0,5} = 0,5.$$

Zpracování náhodného vektoru

- **Př. 2:** V kontingenční tabulce jsou uvedeny výsledky celkového průzkumu spokojenosti cestujících s MHD. Ověřte na hladině významnosti $\alpha = 0,05$ hypotézu, že hodnocení ceny jízdného a kvality cestování jsou nezávislé proměnné. Dále spočítejte koeficienty kontingence.

Zpracování náhodného vektoru

Označení: proměnná X – hodnocení ceny jízdného,
proměnná Y – hodnocení kvality cestování.

Kombinační tabulka						
		Y				Σ
		1	2	3	4	
X	1	21	9	19	11	60
	2	18	18	16	8	60
	3	10	16	9	9	44
	4	10	8	12	6	36
Σ		59	51	56	34	200

Zpracování náhodného vektoru

1. Formulace nulové a alternativní hypotézy:
 H_0 : Hodnocení ceny jízdného a hodnocení kvality cestování jsou nezávislé proměnné.
 H_1 : Hodnocení ceny jízdného závisí a hodnocení kvality cestování jsou závislé proměnné.
2. Testové kritérium je ve tvaru:

$$K = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} \rightarrow \chi_{(k-1) \cdot (m-1)}^2$$

Zpracování náhodného vektoru

3. Sestavení kritického oboru a oboru přijetí:

$$x_{krit} = \chi_{(1-\alpha);(k-1)\cdot(m-1)}^2 = CHIINV[\alpha; (k-1)\cdot(m-1)]$$

– V našem případě $\alpha = 0,05; k = 4; m = 4$, tedy:

$$x_{krit} = CHIINV[0,05; (4-1)\cdot(4-1)] = CHIINV(0,05;9) = 16,92.$$

Zpracování náhodného vektoru

4. Výpočet pozorované hodnoty testové statistiky x_{obs} :

– Výpočet teoretických četností $n_{ij}^* = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}$.

$$n_{11}^* = \frac{n_{1\bullet} \cdot n_{\bullet 1}}{n} = \frac{60 \cdot 59}{200} = 17,7$$

$$n_{12}^* = \frac{n_{1\bullet} \cdot n_{\bullet 2}}{n} = \frac{60 \cdot 51}{200} = 15,3$$

$$n_{13}^* = \frac{n_{1\bullet} \cdot n_{\bullet 3}}{n} = \frac{60 \cdot 56}{200} = 16,8$$

⋮

Zpracování náhodného vektoru

Tabulka teoretických četností						
		Y				Σ
		1	2	3	4	
X	1	17,7	15,3	16,8	10,2	60
	2	17,7	15,3	16,8	10,2	60
	3	12,98	11,22	12,32	7,48	44
	4	10,62	9,18	10,08	6,12	36
Σ		59	51	56	34	200

- Jelikož jsou všechny teoretické četnosti větší než 5, máme splněny předpoklady pro použití testu.

Zpracování náhodného vektoru

$$\begin{aligned}x_{obs} &= \sum_{i=1}^m \sum_{j=1}^n \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} = \frac{(21 - 17,7)^2}{17,7} + \frac{(9 - 15,3)^2}{15,3} + \\ &+ \frac{(19 - 16,8)^2}{16,8} + \frac{(11 - 10,2)^2}{10,2} + \frac{(18 - 17,7)^2}{17,7} + \dots + \\ &+ \frac{(6 - 6,12)^2}{6,12} \doteq 9,03.\end{aligned}$$

Zpracování náhodného vektoru

		Y				Σ
		1	2	3	4	
X	1	0,615254	2,594118	0,288095	0,062745	3,560212
	2	0,005085	0,476471	0,038095	0,47451	0,99416
	3	0,68416	2,036399	0,894675	0,308877	3,924112
	4	0,036196	0,151678	0,365714	0,002353	0,555941
Σ		1,340695	5,258665	1,58658	0,848485	9,034425

X_{obs}



Zpracování náhodného vektoru

5. Formulace závěru testu:

Jelikož platí, že $x_{\text{obs}} < x_{\text{krit}}$, nezamítáme nulovou hypotézu, tedy hodnocení ceny jízdného nezávisí na hodnocení kvality cestování.

Zpracování náhodného vektoru

- Jelikož máme čtvercovou kontingenční tabulku, pro koeficient kontingence dostáváme:

$$CC = \sqrt{\frac{K}{K+n}} = \sqrt{\frac{9,03}{9,03+200}} \doteq 0,21.$$

- Pro Cramerův koeficient dostaneme:

$$V = \sqrt{\frac{K}{n \cdot [\min(k, m) - 1]}} = \sqrt{\frac{9,03}{200 \cdot [\min(4,4) - 1]}} \doteq 0,12.$$

Zpracování náhodného vektoru

- **Př. 3:** Pro potřeby výpočtu propustnosti traťového úseku byla provedena analýza platného grafikonu vlakové dopravy. V rámci analýzy byly získány informace o jízdních dobách vlaků a jejich pořadí. Tyto informace jsou uvedeny v tabulce.

Zpracování náhodného vektoru

Pořadí vlaku	Jízdní doba	Pořadí vlaku	Jízdní doba	Pořadí vlaku	Jízdní doba	Pořadí vlaku	Jízdní doba
1	8	21	5	41	5	61	12
2	12	22	5	42	8	62	5
3	12	23	12	43	5	63	8
4	8	24	5	44	12	64	8
5	12	25	12	45	8	65	12
6	5	26	5	46	8	66	8
7	12	27	5	47	5	67	8
8	12	28	8	48	5	68	5
9	5	29	12	49	8	69	12
10	5	30	5	50	8	70	5
11	5	31	5	51	5	71	8
12	5	32	5	52	5	72	12
13	8	33	8	53	8	73	8
14	8	34	5	54	5	74	12
15	5	35	8	55	5	75	5
16	12	36	5	56	5	76	8
17	12	37	5	57	8	77	5
18	5	38	8	58	8	78	8
19	12	39	12	59	5	79	8
20	5	40	5	60	5	80	5

Zpracování náhodného vektoru

- Na základě získaných dat otestujte, zda lze považovat sledy různých druhů vlaků (lišících se v jízdních dobách) za náhodné.

Zpracování náhodného vektoru

- Z tabulky plyne, že v grafikonu jsou vedeny trasy 3 vlaků – hodnoty jízdních dob činí 5, 8 a 12 minut. Označme tyto vlaky jako rychlík (jízdní doba 5 minut), osobní (jízdní doba 8 minut) a nákladní (jízdní doba 12 minut).

Zpracování náhodného vektoru

- V další fázi musíme získat kontingenční tabulku, pomocí které znázorníme četnosti jednotlivých sledů (rychlík \rightarrow rychlík, rychlík \rightarrow osobní,...).
- Budeme-li pracovat v Excelu, lze kontingenční tabulku získat poměrně snadno. Nejdříve si vytvoříme pomocnou tabulku jednotlivých sledů.

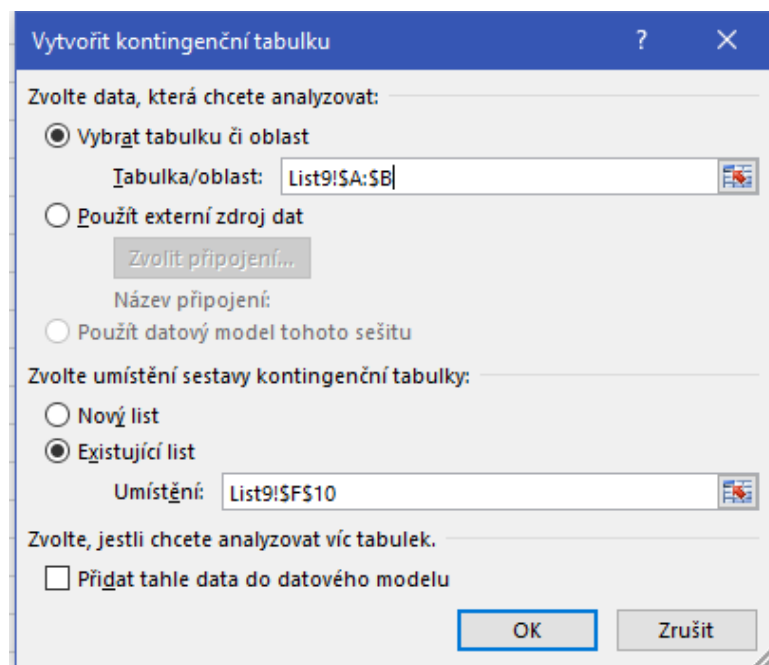
Zpracování náhodného vektoru

Pořadí vlaku	Jízdní doba		1. vlak	2. vlak
1	8	→	8	12
2	12	→	12	12
3	12	→	12	8
4	8	→	8	12
5	12		12	5
6	5		5	12
7	12		12	12
8	12		12	5
9	5		5	5
10	5			
11	5			

Levá tabulka představuje výřez tabulky s jednotlivými vlaky seřazenými dle jejich pořadí v grafikonu a jejich jízdní doby. Pravá tabulka potom představuje výřez tabulky sledů jednotlivých druhů vlaků. Šipkami je potom naznačen způsob, jakým pravou tabulku získáme. V prvním řádku pravé tabulky je sled osobní → nákladní (1. vlak a 2. vlak dle pořadí z levé tabulky), v druhém řádku je sled nákladní → nákladní (2. a 3. vlak dle pořadí z levé tabulky) atd. Poslední řádek pravé tabulky (není znázorněno) bude obsahovat sled rychlík → osobní (80. a 1. vlak dle pořadí z levé tabulky).

Zpracování náhodného vektoru

- Nyní můžeme vytvořit kontingenční tabulku.
- Pokud jste nikdy nepracovali s kontingenční tabulkou, je vhodné se podívat do nápovědy.



Jízdní_doby_test - Excel Nástroje kontingenční tabulky Přihlásit se

Soubor Domů Vlození Rozložení stránky Vzorce Data Revize Zobrazení **Analýza** Návrh Řekněte mi, co chcete udělat. Sdílet

Kontingenční tabulka Aktivní pole: Počet z 1. vlaku Nastavení pole Procházet k podrobnostem Procházet k souhrnným datům Seskupit výběr Vložit průřez Vložit časovou osu Připojení filtru Aktualizovat Změnit zdroj dat Akce Pole, položky a sady Nástroje OLAP Relace Kontingenční graf Doporučené kontingenční tabulky Zobrazit

F10 Počet z 1. vlaku

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1. vlak	2. vlak												
2	8	12												
3	12	12												
4	12	8												
5	8	12												
6	12	5												
7	5	12												
8	12	12												
9	12	5												
10	5	5				Počet z 1. vlaku	Popisky sloupců							
11	5	5				Popisky řádků		5	8	12	(prázdné)	Celkový součet		
12	5	5				5		15	14	8			37	
13	5	8				8		11	7	7			25	
14	8	8				12		11	4	3			18	
15	8	5				(prázdné)								
16	5	12				Celkový součet		37	25	18			80	
17	12	12												
18	12	5												
19	5	12												
20	12	5												
21	5	5												
22	5	5												
23	5	12												
24	12	5												
25	5	12												
26	12	5												
27	5	5												
28	5	8												
29	8	12												
30	12	5												
31	5	5												
32	5	5												
33	5	8												
34	8	5												

Pole kontingenční ..

Vyberte pole, která chcete přidat do sestavy:

Hledat

- 1. vlak
- 2. vlak

Další tabulky...

Přetáhněte pole do jedné z následujících oblastí:

<p>Filtry</p> <div style="border: 1px solid gray; height: 40px; width: 100%;"></div>	<p>Sloupce</p> <p>2. vlak</p>
<p>Řádky</p> <p>1. vlak</p>	<p>Hodnoty</p> <p>Počet z 1. vlaku</p>

Odložit aktualizaci rozlož... **Aktualizovat**

Vyberte cíl a stiskněte klávesu Enter nebo zvolte příkaz Vložit.

Zpracování náhodného vektoru

- Získali jsme kontingenční tabulku, tedy tabulku sdružených a marginálních četností jednotlivých sledů vlaků.

1. vlak / 2. vlak	Rychlík	Osobní	Nákladní	Součet
Rychlík	15	14	8	37
Osobní	11	7	7	25
Nákladní	11	4	3	18
Součet	37	25	18	80

Zpracování náhodného vektoru

- Nyní můžeme přistoupit k vlastnímu testování.
- Definujme si obě hypotézy:
 - H_0 : Sledy jednotlivých druhů vlaků jsou v grafikonu vzájemně nezávislé.
 - H_1 : Sledy jednotlivých druhů vlaků jsou v grafikonu vzájemně závislé.

Zpracování náhodného vektoru

- Abychom mohli stanovit hodnotu testového kritéria, musíme vypočítat teoretické četnosti podle nám již známého vztahu $n_{ij}^* = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$.

1. vlak / 2. vlak	Rychlík	Osobní	Nákladní	Součet
Rychlík	17,11	11,56	8,33	37,00
Osobní	11,56	7,81	5,63	25,00
Nákladní	8,33	5,63	4,05	18,00
Součet	37,00	25,00	18,00	80,00

Zpracování náhodného vektoru

- Z tabulky teoretických četností vidíme, že 8 teoretických četností je větších než 5, 1 četnost je menší než 5, ale větší než 2.
- Máme tedy splněny předpoklady pro použití testu, nemusíme používat korigované testové kritérium (Yatesovu korekci).

Zpracování náhodného vektoru

- Známe-li teoretické četnosti, lze přistoupit k výpočtu hodnoty testového kritéria:

$$x_{obs} = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}.$$

1. vlak / 2. vlak	Rychlík	Osobní	Nákladní	Součet
Rychlík	0,26	0,51	0,01	0,79
Osobní	0,03	0,08	0,34	0,45
Nákladní	0,86	0,47	0,27	1,60
Součet	1,15	1,07	0,62	2,84

Zpracování náhodného vektoru

- Výpočtem jsme stanovili, že pozorovaná hodnota testového kritéria je rovna 2,84.
- Tuto hodnotu musíme porovnat s kritickou hodnotou testu, pro kterou platí:

$$x_{krit} = \chi^2_{(1-\alpha);(k-1)\cdot(m-1)} = \chi^2_{(1-0,05);(3-1)\cdot(3-1)} = CHIINV(0,05;4) \doteq 9,49.$$

Zpracování náhodného vektoru

- Srovnáním obou hodnot zjistíme, že pozorovaná hodnota testového kritéria leží v oboru přijetí ($x_{obs} < x_{krit}$), tudíž **nezamítáme nulovou hypotézu o nezávislosti sledů jednotlivých druhů vlaků.**