

STATISTIKA

9. PŘEDNÁŠKA



**SILESIAN
UNIVERSITY**
SCHOOL OF BUSINESS
ADMINISTRATION IN KARVINA

Téma přednášky:

Chi-kvadrát test

- a) test nezávislosti,*
- b) test dobré shody.*

Mgr. Radmila Krkošková, Ph.D.



Chi – kvadrát test

- Data mohou být **nominální** (nejslabší požadavek)!
- Testuje se (nulová) hypotéza

H_0 : výběr pochází z populace s daným rozdělením

- Zadané rozdělení je obvykle:
 - diskretní rozdělení se stejnými pr- stmi
 - (tzv. **test nezávislosti**)
 - diskretní rozdělení s rozdílnými pr- stmi
 - (tzv. **test dobré shody**)

Příklad – test nezávislosti - limonády

Nová limonáda se prodávala za stejnou cenu jeden týden ve 3 různých typech obalu: A, B, C, počet prodaných limonád viz tabulka:

Typ obalu	Prodané kusy
A	135
B	130
C	155
Celkem	420

Ovlivňuje styl designu obalu počet prodaných limonád?

Jinak: Závisí prodej na obalu?

Příklad – test nezávislosti - limonády

Krok 1. Nulová hypotéza H_0 :

Počet prodaných kusů **nezávisí** na typu obalu (rozdíly v prodeji u vzorku jsou pouze dílem náhody).

Očekávané četnosti (Expected): $E_1 = E_2 = E_3 = 420/3 = 140$

Pozorované četnosti (Observed): $O_1 = 135, O_2 = 130, O_3 = 155$

Krok 2. Testové kritérium:
$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

k - počet kategorií ($k = 3$)

Příklad – test nezávislosti - limonády

Krok 3. Porovnání hodnoty vypočítaného kritéria

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} = 2,5 \quad \text{CHIINV}(0,05;2) = 6,0$$

s tabulkovou **kritickou hodnotou** rozdělení $\chi^2_{\alpha}(2) = 6,0$

kde $\alpha (= 0,05)$ je zadaná hladina významnosti

V každé kategorii: O_i alespoň 5 !

Jestliže $X^2 = 2,5 < \chi^2_{0,05}(2) = 6,0$

potom H_0 nezamítáme! (jinak **zamítáme**)

p -hodnota (signifikance) = 0,287 > 0,05 (Nezamítáme)

Příklad – test nezávislosti – limonády – nové zadání – domácí úkol

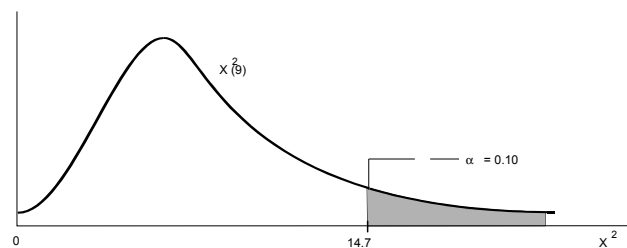
Nová limonáda se prodávala za stejnou cenu jeden týden ve fakultním bufetu ve 3 různých typech obalu: A, B, C, počet prodaných limonád viz tabulka:

NOVÉ ZADÁNÍ:

Typ obalu	Prodané kusy	
A	135	120
B	130	130
C	155	170
Celkem	420	420

Ovlivňuje styl designu obalu počet prodaných limonád?

Kritické hodnoty
rozdělení Chi-kvadrát $\chi^2_{\alpha}(n)$



**SILESIAN
UNIVERSITY**
SCHOOL OF BUSINESS
ADMINISTRATION IN KARVINA

n \ α	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01	0,005
1	0,00	0,00	0,00	0,00	0,02	2,7	3,8	5,0	6,6	7,9
2	0,01	0,02	0,05	0,10	0,21	4,6	6,0	7,4	9,2	10,6
3	0,07	0,12	0,22	0,35	0,58	6,3	7,8	9,4	11,3	12,8
4	0,21	0,30	0,48	0,71	1,06	7,8	9,5	11,1	13,3	14,9
5	0,41	0,55	0,83	1,15	1,61	9,2	11,1	12,8	15,1	16,7
6	0,68	0,87	1,24	1,64	2,20	10,6	12,6	14,4	16,8	18,5
7	0,99	1,24	1,69	2,17	2,83	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	13,4	15,5	17,5	20,1	22,0
9	1,74	2,09	2,70	3,33	4,17	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	21,0	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,09	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,86	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,12	11,65	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,85	12,44	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,28	11,59	13,24	29,6	32,7	35,5	38,9	41,4
22	8,64	9,51	10,98	12,34	14,04	30,8	33,9	36,8	40,3	42,8
23	9,26	10,20	11,69	13,09	14,58	32,0	35,2	38,1	41,6	42,2
24	9,89	10,86	12,40	13,85	15,66	33,2	36,4	39,4	43,0	45,6
25	10,52	11,52	13,12	14,61	16,47	34,4	37,7	40,6	44,3	46,9
26	11,16	12,20	13,84	15,38	17,29	35,6	38,9	41,9	45,6	48,6
27	11,81	12,88	14,57	16,15	18,11	36,7	40,1	43,2	47,0	49,6
28	12,46	13,56	15,31	16,93	18,94	37,9	41,3	44,5	48,3	51,0
29	13,12	14,26	16,05	17,71	19,77	39,1	42,6	45,7	49,6	52,3
30	13,79	14,95	16,79	18,49	20,60	40,3	43,8	47,0	50,9	53,7

Příklad – test dobré shody – barvy automobilů

Automobil Škoda - Felicia se prodává ve čtyřech barvách:

- 40% zákazníků požaduje zelenou barvu automobilu
- 25% červenou barvu,
- 25% modrou barvu a
- 10% bílou barvu.

K ověření správnosti předpokladu o struktuře poptávky podle barev použijte záznamy o nákupech v dané prodejně v jistém měsíci.

Příklad – test dobré shody – barvy automobilů

Vstupní údaje obsahuje následující tabulka:

j	Barva	$p_{0,j}$	n_j
1	zelená	0,40	201
2	červená	0,25	105
3	modrá	0,25	144
4	bílá	0,10	30
součet		1,00	480

Na hladině významnosti $\alpha = 0,05$ testujte hypotézu, že uvedené pravděpodobnostní odhady odpovídají zjištěným hodnotám prodeje.

Příklad – test dobré shody – barvy automobilů

Krok 1. Nulová hypotéza H_0 :

$$H_0 : p_{0,1} = 0,4, p_{0,2} = p_{0,3} = 0,25, p_{0,4} = 0,1$$

Očekávané četnosti:

$$E_1 = 192, E_2 = 120, E_3 = 120, E_4 = 48$$

Pozorované četnosti:

$$O_1 = 201, O_2 = 105, O_3 = 144, O_4 = 30$$

Krok 2. Testové kritérium:
$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

k - počet kategorií ($k = 4$)

Příklad – test dobré shody – barvy automobilů



SILESIAN
UNIVERSITY
SCHOOL OF BUSINESS
ADMINISTRATION IN KARVINA

$$\text{Očekáv_čet}_i = \text{Pravděp}_i \times \text{celk_čet}$$

Příklad:

$$i = \text{zelená}, \text{Pravděp}_i = 0,40, \text{celk_čet} = 480$$

$$E_1 = \text{Očekáv_čet}_i = 0,4 * 480 = 192$$

atd.

Příklad – test dobré shody – barvy automobilů

Krok 3. Porovnání hodnoty vypočítaného kritéria

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} = 13,85$$

s tabulkovou **kritickou hodnotou** rozdělení $\chi_{0,05}^2(3) = 7,81$

V každé kategorii: O_i je alespoň 5 (>30)

Platí $\chi^2 = 13,85 > \chi_{0,05}^2(3) = 7,81 = \text{CHIINV}(0,05;3)$

proto H_0 zamítáme!

Alternativně: $\text{Sig} = \text{CHIDIST}(13,85; 3) = 0,003 < 0,05$

Testování nezávislosti kvalitativních znaků



SILESIAN
UNIVERSITY
SCHOOL OF BUSINESS
ADMINISTRATION IN KARVINA

V jednom vzorku (výběru) můžeme současně sledovat dva nebo i více (kvalitativních) znaků

Příklad:

Při kontrole jakosti výrobku sledujeme přítomnost nebo nepřítomnost vady A (znak A), nebo přítomnost nebo nepřítomnost vady B (znak B).

A i B nabývají pouze dvě alternativní hodnoty –

kategorie: Ano, Ne

(Přítomnost, Nepřítomnost, apod.).

Testování nezávislosti kvalitativních znaků

Uvažujte soubor se dvěma **kvalitativními** znaky A a B

Znak A má r možných kategorií hodnot

označených: A_1, A_2, \dots, A_r

znak B má s možných kategorií hodnot: B_1, B_2, \dots, B_s

Výsledek celého složeného experimentu lze shrnout do
kontingenční tabulky:

Závěr přednášky



**SILESIAN
UNIVERSITY**
SCHOOL OF BUSINESS
ADMINISTRATION IN KARVINA

Děkuji Vám za pozornost !!!