

STATISTIKA

10. PŘEDNÁŠKA



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

*Téma přednášky:
testování nezávislosti kvalitativních
znaků*

Mgr. Radmila Krkošková, Ph.D.

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:

Testování nezávislosti kvalitativních znaků



Kategorie znaku A/B	B_1	B_2	B_3	B_s	Součet
A_1	n_{11}	n_{12}	n_{13}	n_{1s}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	n_{23}	n_{2s}	$n_{2\cdot}$
A_3	n_{31}	n_{32}	n_{33}	n_{3s}	$n_{3\cdot}$
.....
A_r	n_{r1}	n_{r2}	n_{r3}	n_{rs}	$n_{r\cdot}$
Součet	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot s}$	n

Testování nezávislosti kvalitativních znaků

Čtyřpolní kontingenční tabulka



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

Vzhled / Hmotnost výrobků	Vyhovující hmotnost	Nevyhovující hmotnost	Součet- Marg. četnost
Vyhovující vzhled	239	60	299
Nevyhovující vzhled	14	7	21
Součet - Marg. četnost	253	67	320

Testování nezávislosti kvalitativních znaků

Krok 1. Nulová hypotéza H_0 :

Vzhled výrobku nezávisí na hmotnosti
(rozdíly u vzorku jsou pouze dílem náhody).

Očekávané četnosti: $E_{11} = 253 \cdot 299 / 320 = 236,4$
 $E_{21} = 253 \cdot 21 / 320 = 16,6$
 $E_{12} = 67 \cdot 299 / 320 = 62,6$
 $E_{22} = 67 \cdot 21 / 320 = 4,4$

Pozorované četnosti: $O_{11} = 239$, $O_{12} = 14$, $O_{21} = 60$, $O_{22} = 7$

Krok 2. Testové kritérium X^2 :
$$X^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 2,086$$

$df = (r-1)(s-1)$ počet stupňů volnosti ($k = (2-1)(2-1) = 1$)

Testování nezávislosti kvalitativních znaků



SILESIAN
UNIVERSITY
SCHOOL OF BUSINESS
ADMINISTRATION IN KARVINA

$$\frac{\text{Očekáv}_{\check{c}}_{i,j}}{\text{celk.c.}} = \frac{\text{Marg}_{\check{c}}_{i}}{\text{celk.}\check{c}} \times \frac{\text{Marg}_{\check{c}}_{j}}{\text{celk.}\check{c}}$$

$$\text{Očekáv}_{\check{c}}_{i,j} = \text{Marg}_{\check{c}}_{i} \times \text{Marg}_{\check{c}}_{j} / \text{celk}_{\check{c}}$$

Příklad:

$i = 1$: Hmotnost-Nevyhovující

$j = 2$: Vzhled-Vyhovující

$\text{celk}_{\check{c}} = 320$

$E_{12} = \text{Očekáv}_{\check{c}}_{1,2} = 299 * 67 / 320 = 62,6$

atd.

Testování nezávislosti kvalitativních znaků

Krok 3. Porovnání hodnoty vypočítaného kritéria s tabulkovou kritickou hodnotou rozdělení, kde $\alpha = 0,10$ je zadaná hladina významnosti.

V každé kategorii má být alespoň 5 hodnot!

Jestliže $X^2 = 2,1 < \chi_{0,1}^2(1) = 2,7$ potom H_0 nezamítáme!

Alternativně:

Pro hodnotu X^2 zjistíme p -hodnotu (tj. signifikanci -
- má být menší než 0,1)

$p = \text{CHIDIST}(2,1;1) = 0,147$ - tedy H_0 nezamítáme!

Čtyřpolní tabulka – kontingenční tabulka 2 x 2:

Znak2			Součet
Znak1	h_1	h_2	
h_1	A	B	$A+B$
h_2	C	D	$C+D$
Součet	$A+C$	$B+D$	n

Kritérium:
$$X^2 = \frac{n(AD - BC)^2}{(A+B)(C+D)(A+C)(B+D)}$$

Jestliže $X^2 > \chi_{\alpha}^2(1)$, pak H_0 zamítáme, jinak ji nezamítáme!

Příklad: VZHLED X HMOTNOST



SILESIAN
UNIVERSITY
SCHOOL OF BUSINESS
ADMINISTRATION IN KARVINA

$$A = 239, B = 60, C = 14, D = 7$$

$$X^2 > \chi_{0,1}^2(1) = 2,7$$

$$X^2 = \frac{n(AD - BC)^2}{(A + B)(C + D)(A + C)(B + D)} = 2,1$$

Vliv kouření na úmrtnost v Karviné

Kontingenční tabulka pro 2917 zemřelých v Karviné
Kouření versus Počet zemřelých na rakovinu plic

Pozorované čet.	zemřel RP	zemřel JINAK
kouření ANO	137	817
kouření NE	198	1765

Analyzujte, zda kouření respondentů ovlivnilo úmrtnost na rakovinu plic (*RP*).

Použijte Chi-kvadrát test.

Vliv kouření na úmrtnost v Karviné



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

Pozorované čet.	zemřel RP	zemřel JINAK	Suma
kouření ANO	137	817	954
kouření NE	198	1765	1963
Suma	335	2582	2917

Očekávané čet.	zemřel RP	zemřel JINAK	$(E_{ij}-O_{ij})^2/E_{ij}$	
kouření ANO	109,56	844,44	6,87	0,89
kouření NE	225,44	1737,56	3,34	0,43

CHI-SQUARE	11,54
alfa	0,05
df	1
CHIINV	3,8415
CHIDIST=Sig	0,0007

Vliv kouření na úmrtnost v Karviné

Nulovou hypotézu o **nezávislosti** znaků zamítáme!

(Úmrtnost na rakovinu plic závisí na kouření respondentů)

$$X^2 = \frac{n(AD - BC)^2}{(A + B)(C + D)(A + C)(B + D)} = 11,54$$

Závěr přednášky



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

Děkuji Vám za pozornost !!!