STATISTILISTE JAOTUSTE KASUTAMINE

Binoomjaotus

(dihhotoomilistele tunnustele)

N=3

3 kulli	2 kulli	1 kull	0 kulli
(0 kirja)	(1 kirja)	(2 kirja)	(3 kirja)
1	3	3	1
0,125	0,375	0,375	0,125

N=4

4 kulli	3 kulli	2 kulli	1 kull	0 kulli
(0 kirja)	(1 kirja)	(2 kirja)	(3 kirja)	(4 kirja)
1	4	6	4	1
0,06	0,25	0,38	0,25	0,06

Kullide arv	Võimalusi	Tõenäosus
12	1	0,0002
11	12	0,0029
10	66	0,0161
9	220	0,0537
8	495	0,1208
7	792	0,1934
6	924	0,2256
5	792	0,1934
4	495	0.1208
3	220	0,0537
2	66	0,0161
1	12	0,0029
0	1	0,0002
Σ	4096	1,0

Näide.

$P=Q=0{,}5$ $\quad p = 0{,}05$ $\quad N=12$

Kas on tõenäone, et tulemus $X \leq 3$ kulli on saadud 12 mündi viskamisel?

$$P_{X \leq 3} = P_{X=3} + P_{X=2} + P_{X=1} + P_{X=0} =$$

$$= 0{,}0537 + 0{,}0161 + 0{,}0029 + 0{,}0002 = 0{,}0729 > 0{,}05$$

=>Jah

Suunatud ja mittesuunatud hüpotees

Näide. Olgu meil 2 rühma inimesi:

X1 - haiged, kellele antakse ravimit,

X2 - kontrollrühm (ei anta ravimit).

Mittesuunatud hüpotees

$H_0 : P_{X1}=P_{X2} = 0{,}5$=> haigete osa, kellel tekkisid negatiivsed kõrvalnähud, ei erine rühmades oluliselt;

$H_1 : P_{X1} \neq P_{X2} \neq 0{,}5$=> haigete osa, kellel tekkisid negatiivsed kõrvalnähud, erinevad rühmades oluliselt.

Suunatud hüpotees

$H_0 : P_{X1} > P_{X2}$ => ravimit saanutel kõrvalnähtude osa suurem kui kontrollrühmas;

$H_1 : P_{X1} \leq P_{X2}$ => ravimit saanutel kõrvalnähtude osa ei ületa kontrollrühma kõrvalnähtude osa.

***NB!* Suunatud hüpoteesi korral kasutatakse ühepoolset, mittesuunatud hüpoteesi korral aga kahepoolset kriteeriumit.**

Näide.

Firmale esitati süüdistus naiste diskrimineerimises tööle võtmisel. Selgus, et rea aastate kestel oli laekunud töölevõtu avaldusi naistelt ja meestelt ligikaudu ühepalju, tööle oli võetud 19 meest ja 6 naist.

$H_0 : P_{X1}=P_{X2} = 0,5$ (kas võib eeldada, et see väljavõtt (19, 6) on tehtud üldkogumist, milles soo eelistuse tõenäosused ei erine oluliselt?)

$H_1 : P_{X1} \neq P_{X2} \neq 0,5$.

N = 25 p = 0,05 Y = 19

Tabelist: $Y_{KR} = 18$. Kuna $Y > Y_{KR}$, hüljatakse H_0.

⇨Meeste eelistamine naistele polnud juhuslik.

Näide.

Otsustati streikida. A/Ü delegaatidest poolt hääletanud jagunesid soo järgi: 11 meest ja 3 naist.

$H_0 : P_{X1}=P_{X2} = 0,5$ (nais- ja meesdelegaatide osa poolthääletanute korral oluliselt ei erine)

$H_1 : P_{X1} \neq P_{X2} \neq 0,5$.

N = 14 p = 0,05

a) Y = 3.

$P_{Y \leq 3}$=0,0287; 2*0,0287 = 0,0574 > 0,05 => H_0.

b) Y= 11. Tabelist $Y_{KR} = 11$. $Y = Y_{KR}$ => H_1.

$P_Y = N!/(Y!(N-Y)!)*P^Y*Q^{N-Y}$, kus

Y - kulli,
N-Y - kirja,
N - katsete arv,
P - kulli tõenäosus ühes katses,
Q - kirja tõenäosus ühes katses.

- *Üherühma* χ^2 *(hii-ruut) kriteerium*

***Näide*.**

Uuritakse transpordivahendite kasutamist linnas. Küsitletakse 132 töölist. *Eesmärk*: teha kindlaks, kas mõni transpordiliik on eelistatud.

H_0 : transpordiliigi eelistustes oluline erinevus puudub.

T_j

Jalgsi 1	Buss 2	Rong 3	Auto 4	Jalgratas 5	Mootorratas 6	Kokku
50	19	17	30	9	7	132

$O_j(H_0)$

22	22	22	22	22	22	132

$\chi^2 = {}_j\Sigma\ (T_j - O_j)^2/O_j j=1,\dots,6(K)$ $p=0{,}05$

$\chi^2_{ARV} = 57{,}9$ $df = 6-1 = 5$ $\chi^2_{KR} = 11{,}07$

$\chi^2_{ARV} > \chi^2_{KR}$ **=> hülgame H_0.**

***NB!* Kui K = 2, siis oht, et χ^2 väärtus üle hinnatud.** $\chi^2 = {}_j\Sigma\ (|T_j - O_j| - 0{,}5)^2/O_j$.

Kui K > 2 ja T_j > 5, siis soovitav väärtuskategooriaid ühendada - suureneb χ^2 usaldatavus.

***Näide*.**

A/Ü delegaadid…. p =0,05

H_0 : meeste ja naiste osa oluliselt ei erine.

	Mehi	naisi
T_j	11	3
O_j	7	7

$\chi^2_{ARV} = (|11-7|-0{,}5)^2/7 + (|3-7|-0{,}5)^2/7 =$

$= 3{,}5^2/7 + 3{,}5^2/7 = 3{,}5.$ $df = 2-1 = 1$

Tabelist: $\chi^2_{KR} = 3{,}841.$ $\chi^2_{ARV} < \chi^2_{KR} \Rightarrow H_0$

χ^2 arvutamine K*N tabelites

***Näide*.**

Uuritakse 1., 2.,3. kursuse tudengite reageeringut väitele "Ei tohi keelata Üliõpilaste nõukogu koosolekutest". *Eesmärgiks* on kindlaks teha, kas üliõpilaste arvamused sõltuvad ülikoolis õppimise kestusest.

H_0 : üliõpilaste arvamuse ja õppimise kestuse vahel oluline erinevus puudub.

T_{ij}

Kursus	Täiesti nõus	Nõus	Ei tea	Pole nõus	Üldse pole nõus	Kokku
1.	6	12	40	15	10	83
2.	14	22	50	11	9	106
3.	30	50	10	6	2	98
Σ	50	84	100	32	21	287

O_{ij}

1.	14,5	24,3	28,9	9,3	6,1
2.	18,5	31,0	36,9	11,8	7,8
3.	18,0	28,7	34,2	10,9	7,1

P(A,B)=P(A)*P(B) $\quad$ **K_A=3;** $\quad$ **K_B=5**

$\chi^2_{ARV} = (6-14{,}5)^2/14{,}5 + (12-24{,}3)^2/24{,}3 + \ldots = 78{,}7$

$df = (K_A - 1)(K_B - 1) = (3-1)(5-1) = 2*4 = 8$

p = 0,05 $\quad$ **Tabelist $\chi^2_{KR} = 20{,}09$**

$\chi^2_{ARV} > \chi^2_{KR}$ **=> hülgame H_0**

Üherühma märgitest

Oletame, et viskame münti 10 korda (H-kiri, T-kull):
1) H T H T H T H T H T
2) H H H H H T T T T T
Häirib juhuslikkuse puudumine:
1) regulaarse vaheldumise liiasus,
2) regulaarse vaheldumise vähesus.

Märgitest - et määrata, kas tegemist juhuvälja-võtuga või mitte.
"Märk" on ühesuguste tähiste (näiteks H) rida, millele eelneb või järgneb erinev tähis või mitte-mingisugust tähist.

1)	**H T H T H T H T H T**	
	1 2 3 4 5 6 7 8 9 10	**märki**
2)	$\underbrace{\text{H H H H H}}_{1}$ $\underbrace{\text{T T T T T}}_{2}$	**märki**

***Näide*.**
Tööbüroos fikseeritakse etnilises ülekaalus (1) ja vähemuses (0) oleva rassi töötute meeste järjekord $\underbrace{0\,0\,0\,0\,0}_{1}\ \underbrace{1}_{2}\ \underbrace{0\,0\,0\,0\,0}_{3}\ \underbrace{1\,1\,1\,1\,1}_{4}\ \underbrace{0\,0\,0\,0\,0}_{5}$

H_0 : etnilise enamuse ja vähemuse töötute meeste järjekord on juhuslik.
$N_1 = 6$ $N_2 = 15$ $R_{ARV} = 5$ $p = 0{,}05$
1) Tabelist $R_{KR} = 5$. $R_{ARV} = R_{KR}$ => hülgame H_0.
2) kasutatakse Z-kriteeriumit

$$Z_{ARV} = [R+1- \{2N_1N_2/(N_1+N_2)\}]/\sqrt{A}, \quad \text{kus}$$

$$A = 2N_1N_2(2N_1N_2-N_1-N_2)/\{(N_1+N_2)^2(N_1+N_2 -1)\}$$

Tabelist Z_{KR} .   Kui $Z_{ARV} > Z_{KR}$, hülgame H_0.

Fisheri täpne kriteerium

A \ B	***1***	***2***	**Kokku**
1	f_{11}	f_{12}	f_{10}
2	f_{21}	f_{22}	f_{20}
Kokku	f_{01}	f_{02}	f_{00}

$$P = (f_{10}!\ f_{20}!\ f_{01}!\ f_{02}!)/(\ f_{11}!\ f_{12}!\ f_{21}!\ f_{22}!\ f_{00}!)$$

$n! = 0*1*2*...*(n-1)*n;$ $0! = 1$

Näide.

Olgu antud jaotus X

A \ B	***1***	***2***	**Kokku**
1	**10**	**20**	**30**
2	**5**	**25**	**30**
Kokku	**15**	**45**	**60**

H_0 : tunnused A ja B on sõltumatud

H_1 : tunnused A ja B ei ole sõltumatud

p = 0,05

10	20		11	19		12	18
5	25		4	26		3	27

P=0,0805 P=0,0281 P=0,0066

13	17		15	16		15	15
2	28		1	29		0	30

P=0,001 P=0,0001 P=0,0000

Jaotuse $\leq$ X tõenäosus:

($\leq$ X , s.t. veel äärmuslikumad jaotused kui X)

P=0,0805+0,0281+0,0066+0,001+0,0001+0=0,1163

Kuna 0,1163 > 0,05, võtame vastu H_0.