

Biomeetria

Kahemõõtmeline sagedustabel. Hii-ruut test (χ^2 -test)

Kahemõõtmelise sagedustabeli üldkuju



Näide. Kurja tõppe haigestunud 25 koera raviti samal meetodil. Tulemusena sai 14 koera terveks ja 11-l ei õnnestunud haigust välja ravida. Võttes arvesse ka ravitud koerte soo, kas võib väita, et antud haiguse ja ravimeetodi korral ravi tulemus sõltub koera soost?

Suhtelised sagedused:

Sugu	Ravi tulemus		Kokku
	Terve	Haige	
Isane	0,83	0,17	1,00
Emane	0,31	0,69	1,00
Kokku	0,56	0,44	1,00

Kahemõõtmeline sagedustabel:

Sugu	Ravi tulemus		Kokku
	Terve	Haige	
Isane	10	2	12
Emane	4	9	13
Kokku	14	11	25

Sugu	Ravi tulemus		Kokku
	Terve	Haige	
Isane	0,71	0,18	0,48
Emane	0,29	0,82	0,52
Kokku	1,00	1,00	1,00

Sugu	Ravi tulemus		Kokku
	Terve	Haige	
Isane	0,40	0,08	0,48
Emane	0,16	0,36	0,52
Kokku	0,56	0,44	1,00

Kahemõõtmelise sagedustabeli üldkuju



Kahemõõtmeline sagedustabel võimaldab uurida kahe nominaaltunnuse või diskreetse arvtunnuse vahelist seost.

Olgu vaatluse all tunnus X , millel on m erinevat väärtust x_1, x_2, \dots, x_m ja tunnus Y , millel on k erinevat väärtust y_1, y_2, \dots, y_k . Ja olgu valimi maht n , kusjuures igal valimi objektil on mõlemad tunnused mõõdetud.

$X \setminus Y$	y_1	y_2	\dots	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	\dots	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	\dots	n_{2k}	$n_{2.}$
\dots	\dots	\dots	\dots	\dots	\dots
x_m	n_{m1}	n_{m2}	\dots	n_{mk}	$n_{m.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	\dots	$n_{.k}$	n

$$n_{i.} = \sum_{j=1}^k n_{ij}, n_{.j} = \sum_{i=1}^m n_{ij}, n = \sum_{j=1}^k n_{.j} = \sum_{i=1}^m n_{i.}$$

Rea suhtelised sagedused saadakse, jagades lahtrite sagedused läbi vastava rea ääresagedusega: $u_{ij} = n_{ij}/n_{i.}$.

Veeru suhtelised sagedused saadakse, jagades lahtrite sagedused läbi vastava veeru ääresagedusega: $s_{ij} = n_{ij}/n_{.j}$.

Tabeli suhtelised sagedused saadakse, jagades lahtrite sagedused läbi valimi mahuga: $t_{ij} = n_{ij}/n$.

Hii-ruut test (χ^2 -test) kahemõõtmelise sagedustabeli korral



Võrreldakse andmete alusel konstrueeritud sagedustabelit sõltumatu juhuvee vastava teoreetilise sagedustabeliga. Viimases peaksid ridade suhtelised sagedused võrduma summaarse suhteliste sageduste reaga ja veergude suhtelised sagedused summaarse suhteliste sageduste veeruga, ehk $n_{ij} = n_{i.}n_{.j}/n$.

H_0 – tunnused on sõltumatud, st $n_{ij} = n_{i.}n_{.j}/n$,

H_1 – tunnused on sõltuvad.

Teststatistik:
$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n} \underset{H_0}{\sim} \chi_{df}^2,$$
 kus $df = (m-1)(k-1)$

$X \setminus Y$	y_1	\dots	y_k	$n_{i.}$
x_1	n_{11}	\dots	n_{1k}	$n_{1.}$
\dots	\dots	\dots	\dots	\dots
x_m	n_{m1}	\dots	n_{mk}	$n_{m.}$
$n_{.j}$	$n_{.1}$	\dots	$n_{.k}$	n

Eeldused: kõik nullhüpooteesile vastavad sagedused ≥ 5 ($n_{i.}n_{.j}/n \geq 5$, iga i, j).

Otsuse vastuvõtmine:

kui teststatistiku väärtus on suurem kui χ^2 -jaotuse vastav kriitiline väärtus

$$(\chi^2 \geq h_{1-\alpha, (m-1)(k-1)}), \text{ või kui } p \leq \alpha,$$

siis on tõestatud H_1 (tunnused on sõltuvad),

vastupidisel juhul jäädakse tunnuste sõltumatu hüpoteesi juurde (H_0).

**χ^2 -jaotuse
1- α -kvantiilide
($h_{1-\alpha, df}$)
väärtused**

*MS Excelis saab
 χ^2 -jaotuse
1- α -kvantiili
leidmiseks
kasutada
funktsiooni
CHIINV(α ;df)*

df	$\alpha = 0,05$	$\alpha = 0,01$
1	3,841	6,635
2	5,991	9,210
3	7,815	11,345
4	9,488	13,277
5	11,070	15,068
6	12,592	16,812
7	14,067	18,475
8	15,507	20,090
9	16,919	21,666
10	18,307	23,209
12	21,026	26,217
14	23,685	29,141
16	26,296	32,000
18	28,869	34,805
20	31,410	37,566
25	37,652	45,624
30	43,773	50,892
35	49,802	57,342
40	55,758	63,691
45	61,656	69,957
50	67,505	76,154
60	79,082	88,379
70	90,531	100,425
100	124,32	135,807

Hii-ruut test (χ^2 -test) kahemõõtmelise sagedustabeli korral

Näide. Sugu versus ravi tulemus?

H_0 – ravi tulemus ei sõltu koera soost,
 H_1 – ravi tulemus on soospetsiifiline.

Teststatistik: $\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - n_i n_j / n)^2}{n_i n_j / n}$

MS Excel leiab vastava p-väärtuse valemiga CHIDIST(χ^2 ;df)

n_{ij}	Ravi tulemus		Kokku
	Terve	Haige	
Isane	10	2	12
Emane	4	9	13
Kokku	14	11	25

$\frac{n_i n_j}{n}$	Terve	Haige
Isane	6,72	5,28
Emane	7,28	5,72

$\frac{(n_{ij} - n_i n_j / n)^2}{n_i n_j / n}$	Terve	Haige
Isane	1,60	2,04
Emane	1,48	1,88

$\chi^2 = 6,997$

$\Rightarrow H_1$ – ravi tulemus on soospetsiifiline ($p=0,0082$).



Hii-ruut testi (χ^2 -test) olemus

H_0 – vaatlustulemused on kooskõlas teooria poolt ennustatuga,
 H_1 – vaatlustulemused ei kinnita teooriat.

	Vaadeldud juhtude arv (empiiriline sagedus)	Teooria poolt ennustatud juhtude arv (teoreetiline sagedus)
1. võimalus	E_1	T_1
2. võimalus	E_2	T_2
3. võimalus	E_3	T_3
...
k . võimalus	E_k	T_k

Katsetulemus

Teststatistik: $\chi^2 = \sum_{i=1}^k \frac{(E_i - T_i)^2}{T_i} \underset{H_0}{\sim} \chi_{df}^2$

Vabadusastmete arv (*df, degrees of freedom*) = erinevate võimalike tulemuste (väärtuste) arv – valimi põhjal hinnatud teoreetiliste parameetrite arv



Hii-ruut test (χ^2 -test) – näiteid geneetikast

Näide. Mesilaste vastsed arenevad kärjekannudes ning neid hooldavad töomesilased. Töomesilaste ülesanne on haiged või surnud vastsed kõrvaldada, et tagada hügieeniline keskkond.

Selles tegevuses on kaks geneetiliselt kontrollitud etappi: (1) haiget vastset sisaldava kärjekannu avamine; (2) vastse eemaldamine kärjekannust.

Osa mesilasi on ebahügieenilised ning jätvavad haiged või surnud vastsed kärke.

Kui ristata ebahügieenilisi mesilasi hügieenilistega (mõeldud on nn puhtaid liine, st et nii hügieenilised kui ka ebahügieenilised mesilased on saadud paljude põlvkondade vastavate omadustega mesilaste järjestikuse ristamise tulemusel), on kõik järglased ebahügieenilised, mis näitab, et ebahügieeniline käitumine on dominantne tunnus.

Selgitamaks täpsemalt mesilaste hügieenilisuse geneetilisi tagamaid, ristati saadud järglasi uuesti hügieeniliste mesilastega.

Katse tulemused on kokku võetud tabelis järgmisel slaidil.

```

      Eba-      Hügieenilised
      hügieenilised      P2
      P1 ————— P2
                |
                F1
                |
                F2
  
```

Hii-ruut test (χ^2 -test) – näiteid geneetikast



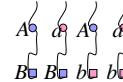
E_i	Ristandmesilaste (F_2) tüüp			Kokku
	Hügieenilised	Ebahüg. ¹	Ebahüg. ²	
	110	98	87	105
				400

¹ avavad kärjekannud, kuid ei kõrvalda sealt nakatunud vastseid,

² ei ava kärjekanne, kuid kõrvaldavad nakatunud vastseid avatud kärjekannudest,

³ ei ava kärjekanne ega kõrvalda vastseid.

Kui nüüd eeldada, et mesilaste hügieenilisus on määratud kahe dialleelse lookuse poolt, siis peaks kõiki nelja tüüpi mesilasi olema võrdselt, ehk iga grupi eeldatav suurus $T_i = 400/4 = 100$ (see on nullhüpoteesi H_0).



Kontrollimaks teooria paikapidavust, viiakse läbi χ^2 -test.

$$\text{Teststatistik: } \chi^2 = \sum_{i=1}^k \frac{(E_i - T_i)^2}{T_i} = \frac{(110 - 100)^2 + (98 - 100)^2 + (-13)^2 + 5^2}{100} = 2,98.$$

Vabadusastmete arv $df = 4 - 1 = 3$, sest nelja teoreetilise sageduse T_i arvutamiseks piisab vaid ühest parameetrist – F_2 -ristandmesilaste arvust $n = 400$.

Et χ^2 -jaotuse kriitiline väärtus $df = 3$ ja $\alpha = 0,05$ korral on $7,815 > 2,98$, siis võime jääda nullhüpoteesi juurde: üks paar alleele kontrollib kärjekannu avamist, teine aga vastsete kõrvaldamist ning need alleelid päranduvad vastavalt Mendeli III seadusele.

Sagedustabelist leitavad suurused epidemioloogias



Enamasti võetakse epidemioloogias vaatluse alla 2×2 sagedustabel (suuremate tabelite korral arvutatakse kordajad analoogselt)

	Juhud – 1 (haiged/responders/...)	Kontrollid – 0 (terved/nonresponders/...)	Kokku
Eksponeeritud	a	b	$a+b$
Mitteeksponeeritud	c	d	$c+d$
Kokku	$a+c$	$b+d$	$a+b+c+d=N$

- **Sagedused** (*observed frequencies*) a, b, c, d .
- **Oodatavad sagedused** assotsiatsiooni puudumise korral (uuritavate tunnuste sõltumatuse eeldusel, st nullhüpoteesi kehtides; *expected frequencies*);
($a+b$) \times ($a+c$)/ N , ... – võimaldavad välja selgitada sõltumatuse juhust enim erinevad väärtuste kombinatsioonid;
- **Juhtude esinemissagedus** (haigestumuskordaja, *incidence rate, rate*) iga riskifaktori taseme tarvis – $a/(a+b), c/(c+d)$.

Sagedustabelist leitavad suurused epidemioloogias



	Juhud	Kontr.	Kokku
Eksp.	<i>a</i>	<i>b</i>	<i>a+b</i>
Mitteeksp.	<i>c</i>	<i>d</i>	<i>c+d</i>
Kokku	<i>a+c</i>	<i>b+d</i>	<i>N</i>

- **Riskisuhe** (*RR*, *risk ratio*, *relative risk*):

$$RR = \frac{\text{juhu risk eksponeeritudel}}{\text{juhu risk mitte eksponeeritudel}};$$

kui võrreldavaid grupe on enam kui 2, siis leitakse *RR* tavaliselt vähima juhtude esinemissagedusega rea (grupi) suhtes; näiteks eeldades, et $a/(a+b) > c/(c+d)$: $RR_1 = [a/(a+b)]/[c/(c+d)]$, $RR_2 = 1$.

- **Šansside suhe** (*OR*, *odds ratio*):

$$OR = \frac{\text{juhu : šanss eksponeeritudel}}{\text{juhu : šanss mitte eksponeeritudel}};$$

üldjuhul leitakse sarnaselt riskisuhtele enamasti vähima juhtude esinemissagedusega grupi suhtes ja on oma olemuselt *RR*-i hinnang; näiteks $OR_1 = (a/b)/(c/d)$, $OR_2 = 1$.

Sagedustabelist leitavad suurused epidemioloogias



Näide. Kas luts on sagedamini *Diphyllobothrium latum* (hariliku laiussi) plerotserkoididega nakatunud kui haug?

	Nakatunud	Terve	Kokku
Luts	59	78	137
Haug	28	186	214
Kokku	87	264	351

	Oodatav nakatunute arv	Haigestumus-kordaja	Riskisuhe	Nakatamise šanss	Šansside suhe
Luts	34	0,431	3,29	0,756	5,03
Haug	53	0,131	1	0,151	1

Järeldused (näiteks).

- Lutsudest oli nakatunud 43,1%, haugidest 13,1%.
- Lutsude šansid nakatuda on 5,03 korda suuremad võrreldes haugidega.