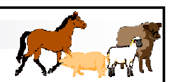
Biomeetria

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?

Kahemõõtmeline sagedustabel:

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

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

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 $x_1, x_2, \ldots, x_m$ ja tunnus Y, millel on k erinevat väärtust $y_1, y_2, \ldots, y_k$. Ja olgu valimi maht n, kusjuures igal valimi objektil on mõlemad tunnused mõõdetud.

$X \backslash Y$	y_1	y_2	$\ldots$	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	$\ldots$	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	$\ldots$	n_{2k}	$n_{2.}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_m	n_{m1}	n_{m2}	$\ldots$	n_{mk}	$n_{m.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\ldots$	$n_{.k}$	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 nn ideaalse, sõltumatuse juhule vastava 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.

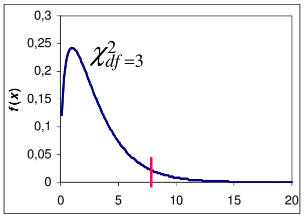
$X \backslash Y$	y_1	$\ldots$	y_k	$n_{i.}$
x_1	n_{11}	$\ldots$	n_{1k}	$n_{1.}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_m	n_{m1}	$\ldots$	n_{mk}	$n_{m.}$
$n_{.j}$	$n_{.1}$	$\ldots$	$n_{.k}$	n

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

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

Otsuse vastuvõtmine:
¤ kui teststatistiku väärtus on suurem kui χ^2-jaotuse vastav kriitiline väärtus $(\chi^2 \geq h_{1-\alpha, (m-1)(k-1)})$, või kui $\boldsymbol{p \leq \alpha}$, siis on tõestatud H_1 (tunnused on sõltuvad),
¤ vastupidisel juhul jäädakse tunnuste sõltumatuse 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	α = 0,05	α = 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: $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n}$

n_{ij}

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

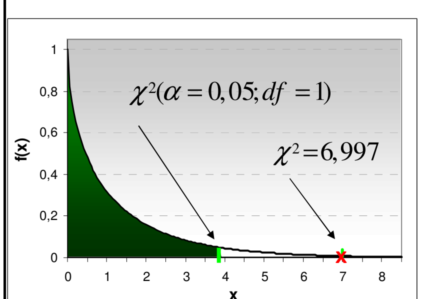
$\frac{n_{i\cdot}n_{\cdot j}}{n}$

	Terve	Haige
Isane	6,72	5,28
Emane	7,28	5,72

$\frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n}$

	Terve	Haige
Isane	1,60	2,04
Emane	1,48	1,88

$\chi^2 = 6,997$



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

=> 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)
Katsetulemus	1. võimalus	E_1	T_1
	2. võimalus	E_2	T_2
	3. võimalus	E_3	T_3
	…	…	…
	k. võimalus	E_k	T_k

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

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 (1)

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

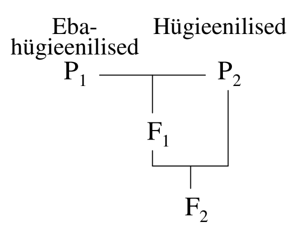
Selles tegevuses on kaks geneetiliselt kontrollitud etappi: (1) haiget vastset sisaldava kärjekannu avamine; (2) vastse eemaldamine kärjekannust.

Osa mesilasi on ebahügieenilised ning jätavad haiged või surnud vastsed kärge.

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.

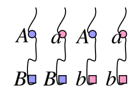
Katse tulemused on kokku võetud tabelis järgmisel leheküljel.



Hii-ruut test (χ^2-test) – näiteid geneetikast (1)

	Ristandmesilaste (F_2) tüüp				Kokku
	Hügieenilised	Ebahüg.[1]	Ebahüg.[2]	Ebahüg.[3]	
E_i	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 nullhüpotees H_0).

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

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 $df = 3$, sest teoreetiliste sageduste 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.

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

Hardy-Weinbergi seadus. See populatsioonigeneetika põhiseadus (populatsiooni geneetilise tasakaalu seadus) väidab, et kui populatsioon on piisavalt suur, paarumine on juhuslik ning puuduvad looduslik ja kunstlik valik, migratsioon jmt, siis püsivad geeni- ja genotüübisagedused põlvkonniti konstantsed.

Lihtsaim viis H-W seaduse matemaatiliseks formuleerimiseks on võtta vaatluse alla üks kahe esinemisvormiga (dialleelne) geen (alleelide tähisteks traditsiooniliselt *a* ja *A*) ning eeldada, et alleeli *A* esinemissagedus populatsioonis on p (tõenäosus, et populatsioonist juhuslikult valitud geen on *A*, on p).

Siis juhul, kui populatsioon on H-W tasakaalus, peaks genotüüpide jaotus olema järgmine:

genotüüp	esinemistõenäosus
AA	p^2
Aa	$2p(1-p)$
aa	$(1-p)^2$

Näide. Selgitamaks, kas paljude populatsioonigeneetikas (ja loomade aretuses) rakendatavate meetodite eelduseks olev Hardy-Weinbergi seadus kehtib ka tänapäevastes aretuspopulatsioonides, viidi läbi veiste veregruppide uuring.

Järgnevas tabelis on näitena toodud 40 eesti punast tõugu lehma dialleelse veregrupilookuse, tähisega EAF ning alleelidega vastavalt ’01’ ja ’02’, genotüübisagedused.

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

01/01-tüüpi isendeid: 13
01/02-tüüpi isendeid: 23
02/02-tüüpi isendeid: 4

Kontrollimaks hüpoteesi leitud genotüübisageduste vastavusest Hardy-Weinbergi seadusele (H_0), tuleb leida alleeli '01' esinemistõenäosus:

$$p = \mathrm{P}('01') = (2*13+23)/(2*40) = 0,6125.$$

Eelmise slaidi valemite alusel saab leida, kui palju ühe või teise genotüübiga isendeid pidanuks valimisse sattuma Hardy-Weinbergi seaduse kehtides, ning viia läbi χ^2-test:

	tegelik	oodatav prop.	oodatav arv	erinevus
01/01-tüüpi isendeid:	13	0,375	15	-2
01/02-tüüpi isendeid:	23	0,475	19	4
02/02-tüüpi isendeid:	4	0,150	6	-2

Teststatistik: $$\chi^2 = \sum_{i=1}^{k} \frac{(E_i - T_i)^2}{T_i} = \frac{(-2)^2}{15} + \frac{4^2}{19} + \frac{(-2)^2}{6} = 1,786.$$

Vabadusastmete arv $df = 3 - 2 = 1$, sest andmetest (3 genotüübisagedust) peame hindama 2 parameetrit: valimi suuruse n ja alleeli '01' esinemissageduse p.

Et χ^2-jaotuse kriitiline väärtus $df = 1$ ja $\alpha = 0,05$ korral on 3,841 > 1,786, siis järeldame, et erinevus Hardy-Weinbergi tasakaalus oleva populatsiooni ja eesti punast tõugu veiste populatsiooni vahel on väike ning jääme nullhüpoteesi juurde.

Sagedustabelist leitavaid suuruseid

	Juhud	Kontr.	Kokku
Eksp.	*a*	*b*	*a*+*b*
Mitteeksp.	*c*	*d*	*c*+*d*
Kokku	*a*+*c*	*b*+*d*	*N*

- **Sagedused** (*observed frequences*): *a*, *b*, *c*, *d*.
- **Oodatavad sagedused** (*expected frequences*): $(a+b)\times(a+c)/N$, ….
- **Juhtude esinemissagedus** (haigestumuskordaja, *IR*, *incidence rate*, *rate*) iga riskifaktori taseme tarvis: $a/(a+b)$, $c/(c+d)$.
- **Riskisuhe** (*RR*, *risk ratio*, *relative risk*):

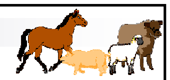
$$RR = \frac{\text{juhu risk eksponeeritutel}}{\text{juhu risk mitte eksponeeritutel}};$$

kui võrreldavaid gruppe on enam kui 2, siis leitakse *RR* tavaliselt vähima juhtude esinemissagedusega rea (grupi) suhtes; näiteks eeldades, et $a/(a+b) > c/(c+d)$:

$$RR_1 = [a/(a+b)]/[c/(c+d)],\ RR_2 = 1.$$

Näide.

	Tervete arv	Haigete arv	Oodatav haigete arv	*IR*	*RR*	95% CI_{RR}	või	*RR*	95% CI_{RR}
Isane	10	2	5,28	0,167	0,24	(0,05; 1,11)		1	
Emane	4	9	5,82	0,692	1			4,15	(0,90; 19,23)



Šansid, šansside suhe [*Odds, odds ratio*]

Sündmuse toimumise **šansid** [*odds*] näitavad, mitmel juhul sündmus toimub võrreldes sellega, mitmel juhul ta ei toimu.

Šansside suhe (OR – *odds ratio*) näitab, kui mitu korda erineb uuritava sündmuse toimumise šanss ühes grupis võrreldes teis(te)ga:

$$OR = \frac{\text{juhu shanss eksponeeritutel}}{\text{juhu shanss mitte eksponeeritutel}}.$$

Näiteks 2x2-tabeli korral võib leida $OR_1=(a/b)/(c/d)$, $OR_2=1$.

Näide.

	Tervete arv	Haigete arv	Terve olemise šanss	Haige olemise šanss	Šansside suhe (OR_{Haige})	95% CI_{OR}
Isane	10	2	5,00	0,20	0,089	(0,013; 0,607)
Emane	4	9	0,44	2,25	1	

Šansid isastel *vs* emastel:

OR_{Haige} = (2/10) / (9/4) = 8/90 = 0,089

OR_{Terve} = (10/2) / (4/9) = 90/8 = 11,25

	Šansside suhe (OR_{Haige})	95% CI_{OR}
Isane	1	
Emane	11,25	(1,64; 76,85)