Páros binomiális próbák

Kontingencia-táblázatok (*r*x*c* táblázat) elemzése, ha sem a sor-, sem az oszlop-összegek nem rögzítettek csak *N* adott

- Szimmetria-vizsgálat (összefüggés-vizsgálat)
- Függetlenség-vizsgálat

Szimmetria-vizsgálat

7. példa

G.A.Walker: Common statistical methods for clinical research with SAS examples, Collins-Wellesley Publishing, San Diego, California, 1996

Véletlenszerűen kiválasztottak 86 pácienst, akik egy adott kezelést kaptak. Mindenkinek megmérték a bilirubin-szintjét kezelés előtt és kezelés után is.

Kérdés: A kezelésnek van-e mellékhatása a vizelet bilirubin-szintjére, vagyis hogy a kezeléstől megváltozik-e a bilirubin-szint.

		Kezelés után		
		normális	magas	Σ
Kezelés előtt	normális	60	14	74
	magas	6	6	12
	Σ	66	20	86

diszkordáns egyedek

N rögzített!

a	***b***	r_1
c	***d***	r_2
c_1	c_2	N

π_{11}	π_{12}
π_{21}	π_{22}

Nullhipotézis kétféle megfogalmazása:

kezelés után (y): magas (1)

kezelés előtt (x): normális (0)

$$H_0: P(x_i = 0, y_i = 1) = P(x_i = 1, y_i = 0) \qquad H_0: \pi_{12} = \pi_{21}$$

$$H_1: P(x_i = 0, y_i = 1) \neq P(x_i = 1, y_i = 0) \qquad H_1: \pi_{12} \neq \pi_{21}$$

azaz annak a valószínűsége, hogy valaki a táblázat *b* cellájából átkerüljön a táblázat *c* cellájába, ugyanakkora, mint annak, hogy valószínűsége, hogy a *c* cellájából a *b* cellájába kerüljön át

Átrendezve:

$$H_0 : P(x_i = 0, y_i = 1) + \boxed{P(x_i = 0, y_i = 0)} = P(x_i = 1, y_i = 0) + \boxed{P(x_i = 0, y_i = 0)}$$

$$H_0 : P(x_i = 0) = P(y_i = 0)$$

$$H_1 : P(x_i = 0) \neq P(y_i = 0)$$

azaz annak a valószínűsége, hogy valaki az egyik (x) szempont szerint az első csoportba tartozzék, ugyanakkora, mint hogy a másik (y) szempont szerint az első csoportba tartozzék

a	***b***	r_1
c	***d***	r_2
c_1	c_2	N

$n = b + c$ $\quad b < c$

Szimmetria-vizsgálat számítása:

n<20, kismintás

$$p = P(k \le b) = \sum_{k=0}^{b} \binom{n}{k} \cdot 0.5^n$$

n≥20, nagymintás

$$z_0 = \frac{b - n \cdot 0.5}{\sqrt{n \cdot 0.5^2}} \xrightarrow{n = b + c} z_0 = \frac{b - c}{\sqrt{b + c}} \xrightarrow{\wedge 2} \chi_0^2 = \frac{(b - c)^2}{b + c}$$

A folytonossági korrekcióval:

$$z_0 = \frac{|b - c| - 1}{\sqrt{b + c}} \qquad \chi_0^2 = \frac{(|b - c| - 1)^2}{b + c}$$

Statistics > Nonparametrics > 2x2 table

a	b	r_1
c	d	r_2
c_1	c_2	N

	2 x 2 Table		
	Column 1	Column 2	Row Totals
Frequencies, row 1	60	14	74
Percent of total	69.767%	16.279%	86.047%
Frequencies, row 2	6	6	12
Percent of total	6.977%	6.977%	13.953%
Column totals	66	20	86
Percent of total	76.744%	23.256%	
Chi-square (df=1)	5.59	p= .0181	
V-square (df=1)	5.52	p= .0188	
Yates corrected Chi-square	3.98	p= .0460	
Phi-square	.06499		
Fisher exact p, one-tailed		p= .0282	
two-tailed		p= .0282	
McNemar Chi-square (A/D)	42.56	p= .0000	
Chi-square (B/C)	2.45	p= .1175	

$$z_0 = \frac{b-c}{\sqrt{b+c}} = \frac{14-6}{\sqrt{14+6}} = 1.788$$

$$z_0 = \frac{b-c-1}{\sqrt{b+c}} = \frac{14-6-1}{\sqrt{14+6}} = 1.565$$

$$\chi_0^2 = \frac{(b-c)^2}{b+c} = \frac{(14-6)^2}{14+6} = 3.197$$

$$\chi_0^2 = \frac{(|b-c|-1)^2}{b+c} = \frac{(14-6-1)^2}{14+6} = 2.45$$

Függetlenség-vizsgálat

8. példa (hipotetikus)

Egy szociológiai vizsgálatnál 50 véletlenszerűen kiválasztott embert a házastársi hűséghez való viszonyáról kérdeztek. Független-e a 2 kérdésre adott válasz egymástól?

H_0: független

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$$

		fontosnak tartja-e a hűséget a házasságban			
		igen	nem		
hűséges-e	igen	π_{ij} 18	2	20	π_{i+}
	nem	6	24	30	
		π_{+j} 24	26	50	

Függetlenség-vizsgálat számításának elve:

a	***b***	r_1
c	***d***	r_2
c_1	c_2	N

$$\chi_0^2 = \sum_i \sum_j \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$\nu = (r-1)\cdot(c-1)$$

O_{ij} observed value, azaz megfigyelt (mintabeli) érték

E_{ij} expected value, azaz a **függetlenség esetén** várt érték

$$E_{ij} = N\cdot\pi_{ij} = N\cdot\pi_{i+}\cdot\pi_{+j} = N\cdot\frac{r_i}{N}\cdot\frac{c_j}{N} = \frac{r_i\cdot c_j}{N}$$

7. példa számítása:

Statistics > Basic Statistics and Tables > Tables and banners

$$E_{11} = \frac{r_1 \cdot c_1}{N} = \frac{20 \cdot 24}{50} = 9.6 \qquad E_{12} = \frac{r_1 \cdot c_2}{N} = \frac{20 \cdot 26}{50} = 10.4$$

$$\chi_0^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(18-9.6)^2}{9.6} + .. + \frac{(24-15.6)^2}{15.6} = 23.56$$

$$\nu = (r-1)(c-1) = 1 \cdot 1 = 1 \qquad \chi^2_{0.05}(1) = 3.841$$

Döntés?

Summary Frequency Table (huseg in h
Marked cells have counts > 5
Weight variable: szam

fontos igen	fontos nem	Row Totals
18	2	20
6	24	30
24	26	50

Summary Table: Expected Frequencies (huseg in huseg)
Marked cells have counts > 5
Pearson Chi-square: 23,5577, df=1, p=,000001
Weight variable: szam

fontos igen	fontos nem	Row Totals
9,60000	10,40000	20,00000
14,40000	15,60000	30,00000
24,00000	26,00000	50,00000

Statistics > Nonparametrics > 2x2 tables

	2 x 2 Table (Spreadsheet21)		
	Column 1	Column 2	Row Totals
Frequencies, row 1	18	2	20
Percent of total	36.000%	4.000%	40.000%
Frequencies, row 2	6	24	30
Percent of total	12.000%	48.000%	60.000%
Column totals	24	26	50
Percent of total	48.000%	52.000%	
Chi-square (df=1)	23.56	p= .0000	
V-square (df=1)	23.09	p= .0000	
Yates corrected Chi-square	20.84	p= .0000	
Phi-square	.47115		
Fisher exact p, one-tailed		p= .0000	
two-tailed		p= .0000	
McNemar Chi-square (A/D)	.60	p= .4404	
Chi-square (B/C)	1.13	p= .2889	

A χ^2- próbához szükséges előfordulási számok

Cochran: egyik E_{ij} sem lehet kisebb 1-nél, és a cellák legföljebb 20%-ában lehet kisebb 5-nél

Conover: ha néhány E_{ij} érték 0.5 körül van, de a többség nagyobb 1-nél, az eljárás alkalmazható

Ha túlságosan kicsinyek a várható előfordulási számok, a cellákat összevonhatjuk.

Fisher „egzakt” próbája
a sor- és oszlop-összegek is adottak

9. példa

A. Agresti: Categorical data analysis, J. Wiley, 2002, p. 444

Fisher kolléganője szerint a teát úgy kell helyesen elkészíteni, hogy először a tejet öntik a csészébe, utána a teát. Fisher kétségbe vonta, hogy a tea ízéből észre lehet venni a sorrendet.
Ezért az alábbi kísérletet végezte: Készített 4-4 csésze teát mindkét sorrend szerint és megkóstoltatta a kolléganőjével (aki azt tudta, hogy mindkét recepttel 4-4 tea készült).

		vélt sorrend		
		tej előbb	tea előbb	
tényleges sorrend	tej előbb	3	1	4
	tea előbb	1	3	4
		4	4	8

		vélt sorrend		
		tej előbb	tea előbb	
tényleges sorrend	tej előbb	3	1	4
	tea előbb	1	3	4
		4	4	8

hipergeometrikus eloszlás

Mi a valószínűsége annak, hogy a 4 „tej előbb” csésze teából éppen 3-t talál el helyesen:

$$P(k=a)=\frac{kedvező\ esetek}{összes\ eset}=\frac{\binom{r_1}{a}\binom{r_1}{r_1-a}}{\binom{N}{r_1}}$$

$$P(3)=\frac{\binom{4}{3}\cdot\binom{4}{1}}{\binom{8}{4}}=0.229$$

p annak valószínűsége, hogy a talált vagy annál szélsőségesebb eredmény álljon elő:

$$P(4) = \frac{\binom{4}{4} \cdot \binom{4}{0}}{\binom{8}{4}} = 0.0143$$

$$p = 0.229 + 0.0143 = 0.2433$$

A kis minta-elemszám miatt nagyok az ugrások,
$p<0.05$ csak akkor lenne, ha mind a 4 csésze teát jól eltalálnák.

Összefoglaló táblázat 2x2 táblázatok elemzéséhez

Típus	I.	II.	III.	IV.	V.
Rögzített	-	N	c_1, c_2	r_1, r_2	c_1,c_2, r_1, r_2
Véletlen	N, c_1, c_2, r_1, r_2	c_1, r_1	r_1	c_1	-
Eloszlás	4 független Poisson	multinomiális	két független binomiális	két független binomiális	hiper-geometrikus
H_0		$\pi_{12}=\pi_{21}$ ($\pi_b=\pi_c$) (szimmetria) $\pi_{ij}=\pi_{i+}\pi_{+j}$ (függetlenség)	$\pi_{11}=\pi_{12}$ ($\pi_a=\pi_b$)	$\pi_{11}=\pi_{21}$ $\pi_a=\pi_c$	$\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}=1$ $\frac{\pi_a\pi_d}{\pi_b\pi_c}=1$
Próba		McNemar, χ^2	χ^2, binomiális egzakt	χ^2, binomiális egzakt	Fisher egzakt, χ^2
Megj.	log-linear	cross-sectional	case-control, retrospective	clinical trial, cohort study prospective	

10. példa

A biztonsági öv használata és a halálos balesetek száma közti kapcsolatot kívánjuk vizsgálni.

Hogyan kell a vizsgálatot elvégezni, hogy az előbbi táblázat szerinti I., II., III. ill. IV. típusú táblázatot kapjuk?

rxc kontingencia-táblázatok elemzése
- homogenitás vizsgálat χ^2-próbával -

11. példa

(Box-Hunter-Hunter: Statistics for Experimenters, J. Wiley, 1978, p. 145)

5 kórházban vizsgálták egy bizonyos betegség-típusnál elért javulást.

		nincs javulás	részleges javulás	teljes gyógyulás	Σ
kórház	A	13	18	16	47
	B	5	10	16	31
	C	8	36	35	79
	D	21	56	51	128
	E	43	29	10	82
	Σ	90	149	128	367

Sorok (kórházak) független multinomiális eloszlást alkotnak.

Nullhipotézis: a multinomiális eloszlások paraméterei kórházanként megegyeznek, azaz a különböző kórházakban megegyezik a javulás esélye

$$H_0: \pi_{j|i} = \pi_j$$

1. Kezeljük egyelőre a javulási fokozatokat **névleges skálán** mért értékeknek!

Ha a nullhipotézis igaz:

$$\hat{\pi}_j = \frac{c_j}{N} \longrightarrow E_{ij} = r_i \cdot \hat{\pi}_j = \frac{r_i \cdot c_j}{N}$$

Summary Frequency Table (korhaz in Workbook1)
Marked cells have counts > 5
(Marginal summaries are not marked)
Weight variable: !~\FT1,,,\c238,gyakorisag

observed

korhaz	javulas nincs	javulas reszleges	javulas teljes	Row Totals
A	13	18	16	47
B	5	10	16	31
C	8	36	35	79
D	21	56	51	128
E	43	29	10	82
All Grps	90	149	128	367

$$\chi_0^2 = \sum_i \sum_j \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$\nu = (r-1)\cdot(c-1)$$

Summary Table: Expected Frequencies (korhaz in Workbook1)
Marked cells have counts > 5
Pearson Chi-square: 56,7050, df=8, p=,000000
Weight variable: !~\FT1,,,\c238,gyakorisag

javulas nincs	javulas reszleges	javulas teljes	Row Totals
11,526	19,08	16,39	47,00
7,602	12,59	10,81	31,00
19,373	32,07	27,55	79,00
31,390	51,97	44,64	128,00
20,109	33,29	28,60	82,00
90,000	149,00	128,00	367,00

$$E_{11} = \frac{r_1 \cdot c_1}{N} = \frac{47 \cdot 90}{367} = 11.526$$

2. **Sorrendi skálán** kezelve a javulási adatokat:

egyfaktoros ANOVA rangszámokra végezve

Kruskal-Wallis ANOVA by Ranks; javulas (korhaz.sta)
Independent (grouping) variable: korhaz
Kruskal-Wallis test: H (4, N= 367) =49.38915 p =.0000

Depend.: javulas	Code	Valid N	Sum of Ranks	Mean Rank
A	101	47	8417.50	179.0957
B	102	31	6733.50	217.2097
C	103	79	16926.50	214.2595
D	104	128	25674.00	200.5781
E	105	82	9776.50	119.2256

Mood-féle

Median Test, Overall Median = 102.000; javulas (korhaz.sta)
Independent (grouping) variable: korhaz
Chi-Square = 26.89190 df = 4 p = .0000

Dependent: javulas	A	B	C	D	E	Total
<= Median: observed	31.00000	15.00000	44.00000	77.0000	72.0000	239.0000
expected	30.60763	20.18801	51.44687	83.3569	53.4005	
obs.-exp	0.39237	-5.18801	-7.44687	-6.3569	18.5995	
> Median: observed	16.00000	16.00000	35.00000	51.0000	10.0000	128.0000
expected	16.39237	10.81199	27.55313	44.6431	28.5995	
obs.-exp	-0.39237	5.18801	7.44687	6.3569	-18.5995	
Total: observed	47.00000	31.00000	79.00000	128.0000	82.0000	367.0000

3. az „E” kórház és a többi kórház összehasonlítása

Summary Frequency Table (korhaz in Workbook1) Marked cells have counts > 5 (Marginal summaries are not marked) Weight variable: !~\FT1,,,\c238,gyakorisag				
melyik	javulas nincs	javulas reszleges	javulas teljes	Row Totals
E	43	29	10	82
nemE	47	120	118	285
All Grps	90	149	128	367

Summary Table: Expected Frequencies (korhaz in Workbook1) Marked cells have counts > 5 Pearson Chi-square: 49,8439, df=2, p=,000000 Weight variable: !~\FT1,,,\c238,gyakorisag			
javulas nincs	javulas reszleges	javulas teljes	Row Totals
20,10899	33,2916	28,5995	82,0000
69,89101	115,7084	99,4005	285,0000
90,00000	149,0000	128,0000	367,0000