Quantitative methods

Lesson 13

Daróczi Gergely

Corvinus University of Budapest, Hungary

2011 May 3



◆□▶ ◆□▶ ◆三▶ ◆三▶ ○□ ○○○





4 Standardization and decomposition



글 🕨 🖌 글

Real association?



Wind (miles per hour)

Daróczi Gergely (BCE)

Real association?



Daróczi Gergely (BCE)

Real association?



Daróczi Gergely (BCE)

Real association?



Daróczi Gergely (BCE)

Simpson's paradox

Berkeley sex bias case



Simpson's paradox

Berkeley sex bias case

	Admitted	Deny	Σ
Female	1494	2827	4321
Male	3738	4704	8442
Σ	5232	7531	12763

Table: Observed values

	Admitted	Deny	Σ
Female	34.6 %	65.4 %	100 %
Male	44.3 %	55.7 %	100 %
Σ	41 %	59 %	100 %

Table: Row percentages

$$\chi^2 = 110.8489; d.f. = 1; p = 6.385628e - 26$$

Daróczi Gergely (BCE)

Berkeley sex bias case

	Applicants	Admitted
Men	8442	44%
Women	4321	35%

	Me	en	Won	nen
Departement	Applicants	Admitted	Applicants	Admitted
A	825	62%	108	82%
В	560	63%	25	68%
С	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

ヘロト 人間 とくほ とくほとう

Batting averages in professional baseball

	1995		1996		Combined	
	Runs/Outs	%	Runs/Outs	%	Runs/Outs	%
Derek Jeter	12/48	25 %	183/582	31.4 %	195/630	31 %
David Justice	104/411	25.3 %	45/140	32.1 %	149/551	27 %

Who is the better player?

Discrete (qualitative) variables

_

ID	gender	color
1	Female	pink
2	Female	pink
3	Female	pink
4	Female	pink
5	Female	pink
6	Female	pink
95	Male	yellow
96	Male	yellow
97	Male	yellow
98	Male	yellow
99	Male	yellow
100	Male	yellow

<ロ> (四) (四) (三) (三) (三)

Discrete (qualitative) variables



Discrete (qualitative) variables



Discrete (qualitative) variables

	green	pink	yellow
Female	17	30	13
Male	18	10	12

< ロ > < 四 > < 回 > < 回 > < 回 > 、

Discrete (qualitative) variables

	green	pink	yellow	
Female	17	30	13	Marginals
Male	18	10	12	warymais
	Marginals			Ν

Discrete (qualitative) variables

	green	pink	yellow	Σ
Female	17	30	13	60
Male	18	10	12	40
Σ	35	40	25	100

≣▶ ≣ •⁄০৭ে 5/3/2011 12/44

Percentages

	green	pink	yellow	Σ
Female	17	30	13	60
Male	18	10	12	40
Σ	35	40	25	100

Table: Counted values

	green	pink	yellow	Σ
Female	17 %	30 %	13 %	60 %
Male	18 %	10 %	12 %	40 %
Σ	35 %	40 %	25 %	100 %

Table: Total percentages

< D > < B

▶ < 글 ▶ < 글 ▶</p>

Row percentages

	green	pink	yellow	Σ
Female	17	30	13	60
Male	18	10	12	40
Σ	35	40	25	100

Table: Counted values

	green	pink	yellow	Σ
Female	28.3 %	50 %	21.7 %	100 %
Male	45 %	25 %	30 %	100 %
Σ	35 %	40 %	25 %	100 %

Table: Row percentages

> < E > < E >

< □ > < 🗇

Column percentages

	green	pink	yellow	Σ
Female	17	30	13	60
Male	18	10	12	40
Σ	35	40	25	100

Table: Counted values

	green	pink	yellow	Σ
Female	48.63 %	75 %	52 %	60 %
Male	51.4 %	25 %	48 %	40 %
Σ	100 %	100 %	100 %	100 %

Table: Column percentages

> < E > < E >

< □ > < 🗇

Expected values

	green	pink	yellow	Σ
Female	17	30	13	60
Male	18	10	12	40
Σ	35	40	25	100

Table: Counted values

	green	pink	yellow	Σ
Female	21	24	15	60
Male	14	16	10	40
Σ	35	40	25	100

Table: Expected values

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

where:

- χ^2 : Pearson's cumulative test statistic,
- O_i: an observed (counted) frequency,
- E_i: an expected (theoretical) frequency,
- *n*: the number of cells in the table.

 H_0 : observed and expected values are all the same

Requirements!

Computed chi-square

	green	pink	yellow	Σ
Female	$\frac{(17-21)^2}{21}$	$\frac{(30-24)^2}{24}$	<u>(13–15)²</u> 15	-
Male	$\frac{(18-14)^2}{14}$	<u>(10–16)²</u> 16	$\frac{(12-10)^2}{10}$	-
Σ	-	-	-	-

Table: Computed distances between observed and expected values

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = 6.321429$$

degrees of freedom: (3-1)(2-1) = 2

Computed chi-square



 $\Rightarrow p = 0.04239545$

A basic example



Henderson & Velleman (1981): Building multiple regression models interactively

ъ

A basic example



Henderson & Velleman (1981): Building multiple regression models interactively

Daróczi Gergely (BCE)

5/3/2011 20 / 44

Basic theory of normalization

Standard score (z-values, z-scores, normal scores, standardized variables) indicates how many standard deviations an observation is above or below the mean:

$$z = \frac{x-\mu}{\sigma}$$



Decomposition

	Miami				Alaska			U.S.		
Age	Pop.	Deaths	Rate*	Pop.	Deaths	Rate*	Pop.+	Deaths ⁺	Rate*	
< 15	114,350	136	1.19	37,164	59	1.59	23,961	32	1.34	
15-24	80,259	57	0.71	20,036	18	0.90	15,420	9	0.58	
25-44	133,440	208	1.56	32,693	37	1.13	21,353	30	1.40	
45-64	142,670	1,016	7.12	14,947	90	6.02	19,609	140	7.14	
65+	92,168	3,605	39.11	2,077	81	39.00	10,685	529	49.51	
	562,887	5,022		106,917	285		91,028	740		
Crude death rate*			8.92			2.67			8.13	

Population and Deaths by Age in 1970 for White Females in Miami, Alaska, and the U.S.

* Deaths per 1,000 population

+ in thousands

Daróczi Gergely (BCE)

ъ

< □ > < 🗇

→ 4 Ξ

Direct standardization

Definition

In direct standardization the stratum-specific rates of study populations are applied to the age distribution of a standard population.

Directly standardized rate =
$$\frac{\sum stratum specific rates \times standard weights}{\sum standard weights}$$

$$\begin{aligned} \text{Miami} &= \frac{(1.19x23,961) + \dots + (39.11x10,685)}{91,208} = 6.92 \ \text{deaths/thousand} \\ \text{Alaska} &= \frac{(1.59x23,961) + \dots + (39x10,685)}{91,208} = 6.71 \ \text{deaths/thousand} \end{aligned}$$

Indirect standardization

Definition

In indirect standardization, the standard population provides the rates and the study population provides the weights.

Indirectly standardized rate = $\frac{\sum observed \ values}{\sum expected \ values}$

Expected values = Stratum specific rates from the study population \times stratum sizes from the study population

		Study population	Standard popula	ation	
Directly-standardized rate		Rates	Weights		_
Indirectly-standardized rate		Weights	Rates		
		4		₹ 9	٩
Daróczi Gergely (BCE)	C	auantitative methods, 13/14	5/3/201	1 24	/4



5/3/2011 25 / 44



i



Visits from search engines

Daróczi Gergely (BCE)

≣▶ ≣ ∽Ω< 5/3/2011 26/44

Graphs Line



Daróczi Gergely (BCE)

≣▶ ≣ ዏ�? 5/3/2011 27/44



Industrial Growth Rate (Country)



5/3/2011 28 / 44

Graphs

Area



Daróczi Gergely (BCE)

Quantitative methods, 13/14

5/3/2011 29 / 44





≣। ≣ √००. 5/3/2011 30/44

◆□▶ ◆圖▶ ◆厘▶ ◆厘≯

Graphs Combo



5/3/2011 31 / 44





Daróczi Gergely (BCE)

Graphs Heatmap



Daróczi Gergely (BCE)

≣। ≣ √००. 5/3/2011 33/44

・ロト ・四ト ・ヨト ・ヨト



Calendar Heat Map of MSFT Adjusted Close

Daróczi Gergely (BCE)





5/3/2011 35 / 44

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ ○臣

Graphs Dot plot

Gas Milage for Car Models grouped by cylinder

4						
÷.	Tovota Corolla					•
	Fiat 128					••••••
	Lotus Europa					•
	Honda Civic					•
	Fiat X1-9				••••••	
	Porsche 914-2				•••••	
	Merc 240D				• •	
	Merc 230			•••••		
	Datsun 710			•••••		
	Toyota Corona			• • • • • • • • • • • • • • • • • • • •		
	Volvo 142E			• • • • • • • • • • • • • • • • • • • •		
6						
	Hornet 4 Drive			• • • • • • • • • • • • • • • • • • • •		
	Mazda RX4 Wag			•••••		
	Mazda RX4			•••••		
	Ferrari Dino			•		
	Merc 280			•		
	Valiant		• • •			
	Merc 280C		• • • • • • • • • • • • • • • • • • • •			
8						
	Pontiac Firebird			•		
	Hornet Sportabout		•			
	Merc 450SL		• • • • • • • • • • • • • • • • • • • •			
	Merc 450SE					
	Ford Pantera L		•			
	Dodge Challenger		• • • • • • • • • • • • • • • • • • • •			
	AMC Javelin		• • • • • • • • • • • • • • • • • • • •			
	Merc 450SLC		••••			
	Maserati Bora					
	Chrysler Imperial					
	Duster 360		•			
	Camaro ∠28					
	Lincoln Continental					
	Cadillac Fleetwood					
				1		
		10	15	20	25	30
				Miles Per Gallo	0	

Daróczi Gergely (BCE)

≣। ≣ √००. 5/3/2011 36/44

・ロト ・四ト ・ヨト ・ヨト







How Firefox Users Use Bookmarks

< D > < B

5/3/2011 38 / 44

표 🖌 🔺 표



≣▶ ≣ ዏ�? 5/3/2011 39/44

・ロト ・ 理 ト ・ ヨ ト ・ ヨ ト

Graphs Mosaic chart



Daróczi Gergely (BCE)

≣▶ ≣ ∽Ω< 5/3/2011 40/44

・ロト ・ 理 ト ・ ヨ ト ・ ヨ ト



Graphs "Crayola Color Chart, 1903-2010"



DATA POINTED datapointed.n

イロト イポト イヨト イヨト

Daróczi Gergely (BCE)

≣▶ ≣ •⁄০৭ে 5/3/2011 42/44

- http://www.visual-literacy.org/periodic_table/periodic_table.html
- http://www.edwardtufte.com/tufte/
- http://www.perceptualedge.com/
- http://www.visualcomplexity.com/vc/
- http://flowingdata.com/
- http://infosthetics.com/
- http://chartsgraphs.wordpress.com/
- http://www.informationisbeautiful.net/
- http://chartporn.org/

It was a pleasure!

Daróczi Gergely daroczi.gergely@btk.ppke.hu

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへで