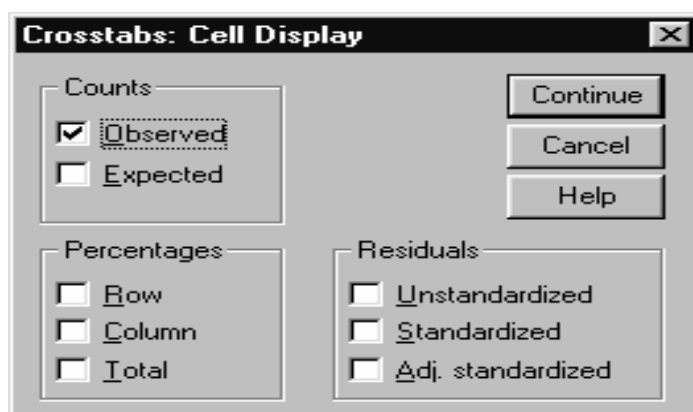
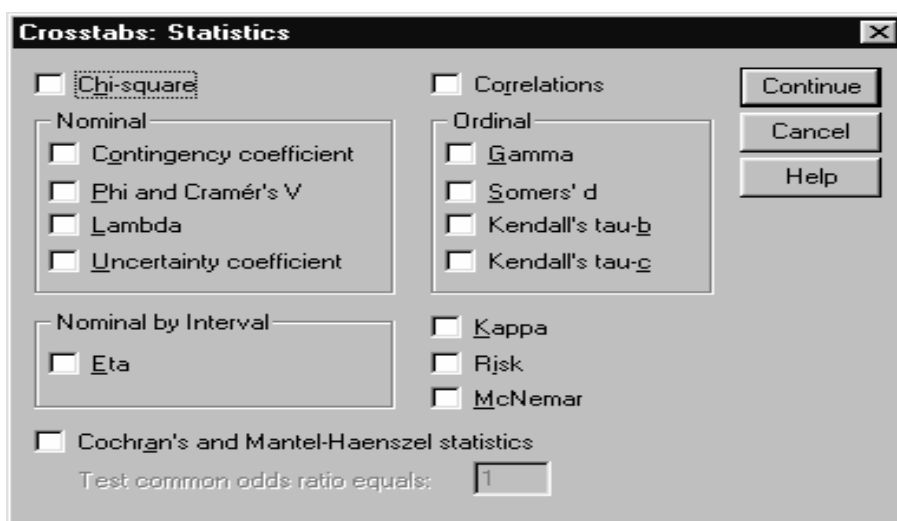
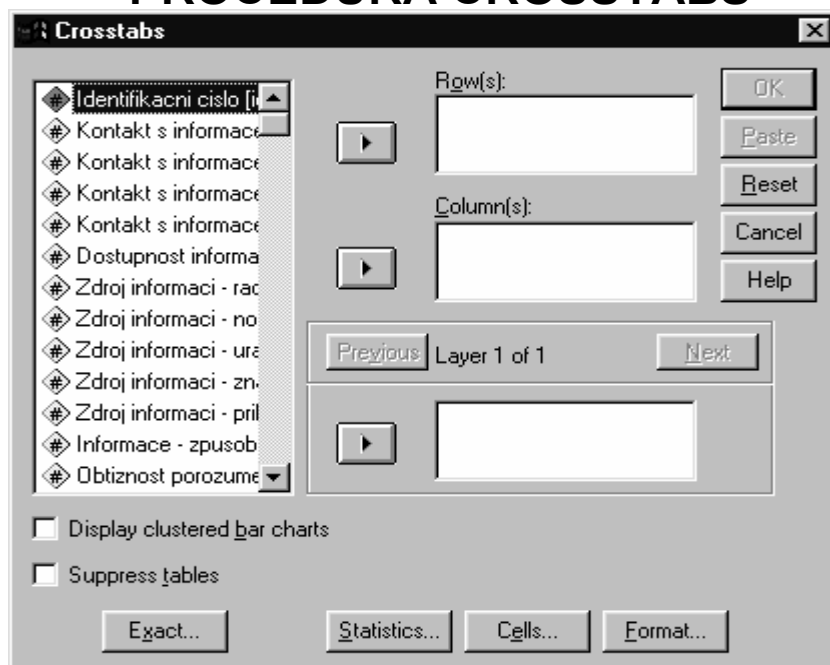


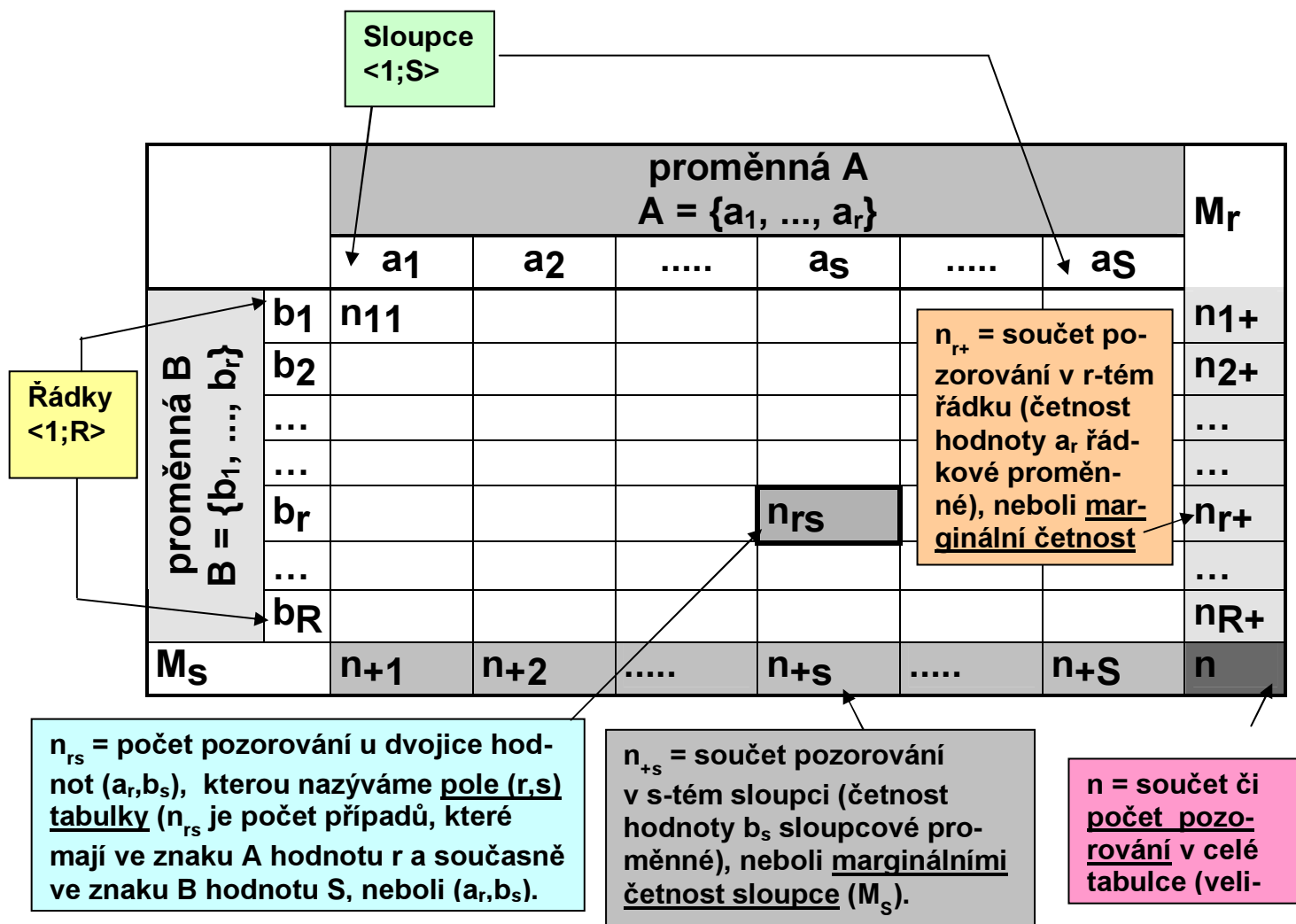
LEKCE07 ZÁKLADY BIVARIAČNÍ ANALÝZY: ROZLOŽENÍ DAT V KONTINGENČNÍ TABULCE

PROCEDURA CROSSTABS



KONTINGENČNÍ TABULKA

(absolutní četnosti)



KONTINGENČNÍ TABULKA (relativní četnosti)

		proměnná A $A = \{a_1, \dots, a_s\}$						M_r
		a_1	a_2	a_s	a_S	
proměnná B $B = \{b_1, \dots, b_r\}$	b_1	f_{11}						f_{1+}
	b_2							f_{2+}

	b_r							f_{r+}

	b_R							f_{R+}
M_s		f_{+1}	f_{+2}	f_{+s}	f_{+S}	n

Sloupce
<1;S>

Řádky
<1;R>

$f_{r+} = n_{r+}/n =$ marginální relativní četnost hodnoty a_r (četnost a_r v jednorozměrném rozložení proměnné A); odpovídá též sumě f_{rs} od $s=1$ do $s=S$.

celkové relativní četnosti
řádkové relativní četnosti
sloupcové relativní četnosti

$f_{+s} = n_{+s}/n =$ marginální relativní četnost hodnoty b_s (četnost b_s v jednorozměrném rozložení proměnné B); odpovídá též sumě f_{rs} od $r=1$ do $r=R$.

$n =$ součet či počet pozorování v celé tabulce (velikost souboru).

- $f_{rs} = n_{rs}/n =$ relativní četnost dvourozměrného rozložení pro dvojici (a_r, b_s) ; Násobeno 100 jde o procento pole tabulky z celkového počtu pozorování n (CELKOVÉ neboli TOTAL %).
- $f_{s/r} = n_{rs}/n_{r+} = f_{rs}/f_{r+} =$ relativní četnost hodnoty b_s v r -tém řádku tabulky (podmíněná četnost pro r -tý řádek, řádková relativní četnost); $100 f_{s/r} \%$ je ŘÁDKOVÉ PROCENTO.
- $f_{r/s} = n_{rs}/n_{+s} = f_{rs}/f_{+s} =$ relativní četnost hodnoty a_r v s -tém sloupci tabulky (podmíněná četnost pro s -tý sloupec, sloupcová relativní četnost); $100 f_{r/s} \%$ je SLOUPCOVÉ PROCENTO.

Poznámka: Hodnoty f_{r+} a f_{+s} se mohou lišit od původních jednorozměrných empirických rozložení díky vlivu vynechaných dat (missing date).

JEN ABSOLUTNÍ ČETNOSTI?

vztah mezi významem „porozumění a snášenlivosti (Q40_8)“ a „vzájemné úcty a uznání (Q40_4)“ pro úspěšnost manželství

Count

		Q40_8 Porozumění manželů			Total
		1 velmi důležité	2 spíše důležité	3 nepříliš důležité	
Q40_4 Úcta a uznání manželů	1 velmi důležité	1483	168	6	1657
	2 spíše důležité	132	93	1	226
	3 nepříliš důležité	5	4	2	11
Total		1620	265	9	1894

RELATIVNÍ PROCENTA SLOUPCŮ NEBO ŘÁDKŮ?

Vztah mezi významem „porozumění a snášenlivosti (Q40_8)“ a „vzájemné úcty a uznání (Q40_4)“ pro úspěšnost manželství:

OBOJÍ: Nelze jednoznačně určit závisle proměnnou a nezávisle proměnnou! SYMETRICKÝ VZTAH.

			Q40_8 Porozumění manželů			Total
			1 velmi důležité	2 spíše důležité	3 nepříliš důležité	
Q40_4 Úcta a uznání manželů	1 velmi důležité	Count	1483	168	6	1657
		% within Q40_4 Úcta a uznání manželů	89.5%	10.1%	.4%	100.0%
		% within Q40_8 Porozumění manželů	91.5%	63.4%	66.7%	87.5%
	2 spíše důležité	Count	132	93	1	226
		% within Q40_4 Úcta a uznání manželů	58.4%	41.2%	.4%	100.0%
		% within Q40_8 Porozumění manželů	8.1%	35.1%	11.1%	11.9%
3 nepříliš důležité	Count	5	4	2	11	
	% within Q40_4 Úcta a uznání manželů	45.5%	36.4%	18.2%	100.0%	
	% within Q40_8 Porozumění manželů	.3%	1.5%	22.2%	.6%	
Total		Count	1620	265	9	1894
		% within Q40_4 Úcta a uznání manželů	85.5%	14.0%	.5%	100.0%
		% within Q40_8 Porozumění manželů	100.0%	100.0%	100.0%	100.0%

LEKCE 7: ZÁKLADY BIVARIAČNÍ ANALÝZY: ROZLOŽENÍ DAT V KONTINGENČNÍ TABULCE

„Co by měla společnost zajišťovat, aby byla považována za spravedlivou (Q76_1)“ versus „společenská skupina (Q110C):

ŘÁDKOVÉ. Určíme závisle proměnnou a nezávisle proměnnou (tou je řádková proměnná Q110C)! ASYMETRICKÝ VZTAH.

			Q76_1 Zabránit velkým nerovnostem					Total
			1 velmi důležité	2	3	4	5 ani trochu důležité	
Q110C Společenská skupina	1 nižší	Count	142	79	32	15	9	277
		% z Q110C	51.3%	28.5%	11.6%	5.4%	3.2%	100.0%
	2 nižší střední	Count	173	182	144	78	33	610
		% z Q110C	28.4%	29.8%	23.6%	12.8%	5.4%	100.0%
	3 střední	Count	221	215	240	120	55	851
		% z Q110C	26.0%	25.3%	28.2%	14.1%	6.5%	100.0%
	4 vyšší střední	Count	22	18	42	36	16	134
		% z Q110C	16.4%	13.4%	31.3%	26.9%	11.9%	100.0%
Total		Count	558	494	458	249	113	1872
		% z Q110C	29.8%	26.4%	24.5%	13.3%	6.0%	100.0%

PAMATUJME SI! pro asymetrické vztahy

Spočítáme-li řádková procenta (nezávislá proměnná v řádcích jako v tomto případě), interpretujeme je ve sloupcích.

Spočítáme-li sloupcová procenta (nezávislá proměnná ve sloupcích), interpretujeme je v řádcích.

MÁ SMYSL POUŽÍVAT CELKOVÝCH PROCENT?

KOLIK JE V SOUBORU PŘÍPADŮ URČITÉHO TYPU? KONKRÉTNĚ (kolik manželských párů pochází z úplných rodin)?

			N7REC typ rodiny nevesty - recod			Total
			1 neuplna	2 uplna opakovana	3 uplna puvodni	
Z2REC typ rodiny	1 neuplna	Count	7		20	27
		Total %	2,7%		7,8%	10,5%
zenicha-recod	2 uplna opakovana	Count	4	2	14	20
		Total %	1,6%	,8%	5,4%	7,8%
	3 uplna puvodni	Count	21	29	161	211
Total %		8,1%	11,2%	62,4%	81,8%	
Total		Count	32	31	195	258
		Total %	12,4%	12,0%	75,6%	100,0%

Párů, kde ženich a nevěsta pocházejí oba z úplných původních rodin, je v našem souboru 62,4% našeho souboru (párů, kde oba pocházejí z neúplných rodin je zde

LEKCE 7: ZÁKLADY BIVARIAČNÍ ANALÝZY: ROZLOŽENÍ DAT V KONTINGENČNÍ TABULCE

2,7% ...). Párů, ve kterých pochází ženich z úplné původní rodiny je zde 10,5% případů (co se týče nevěst, je to 12,4% případů

MÁ SMYSL POUŽÍVAT CELKOVÝCH PROCENT?

KOLIK JE V SOUBORU PŘÍPADŮ URČITÉHO TYPU? KONKRÉTNĚ (kolik, respektive jaký podíl, mužů či žen si myslí, že „žena musí mít děti, aby se naplnilo její poslání“?)

			Q84 Pohlaví		Total
			1 muž	2 žena	
Q42 Žena musí mít děti, aby splnila poslání	1 ano	Count	363	432	795
		% of Total	20.1%	24.0%	44.1%
	2 není to nutné	Count	500	508	1008
		% of Total	27.7%	28.2%	55.9%
Total		Count	863	940	1803
		% of Total	47.9%	52.1%	100.0%

Muži, kteří si to myslí že „ŽENA MUSÍ MÍT DĚTI, ABY SE NAPLNILO JEJÍ POSLÁNÍ“ představují 20% souboru a ženy 24% souboru. POZOR: Nikoliv 20% mužů či 24% žen (abychom zjistili podíl tohoto názoru mezi muži a ženami, museli bychom vypočítat sloupcové četnosti), proto nemůžeme tato procenta srovnávat. To, že je v souboru více žen než mužů s tímto názorem může být dáno i tím, že je v něm vůbec více žen! CELKEM SI TO V NAŠEM SOUBORU MYSLÍ 44% OSOB.

POZOROVANÉ A OČEKÁVANÉ HODNOTY A RESIDUÁLY

		proměnná A						M_r
		a_1	a_2	a_s	a_s	
proměnná B	b_1	f_{11}						f_{1+}
	b_2							f_{2+}

	b_r							f_{r+}

	b_R							f_{R+}
M_s		$f_{+1}=100$	$f_{+2}=100$ =100	$f_{+s}=100$ =100	$f_{+S}=100$	$n=100$

- **EXPECTED COUNT** = očekávané četnosti, počet jednotek, který by byl v tomto políčku při nezávislosti obou znaků (náhodné rozložení).
- **RESIDUAL** = rozdíl mezi pozorovaným počtem jednotek, které mají příslušnou empirickou kombinaci hodnot obou znaků a očekávanou četností. Residuály se dále standardizují a používají se v adjustované (na velikost tabulky) podobě.
- **STD. RESIDUAL** = Standardizované χ^2 residuály, neboli residuály vydělené druhou odmocninou očekávaných hodnot.
- **ADJUSTED RESIDUAL** = Adjustované residuály (tak, aby měly přibližně normální rozložení s průměrem = 0 a standardní odchylkou rovnou 1).

OČEKÁVANÉ HODNOTY PRO JEDNOTLIVÁ POLÍČKA

celková četnost v řádku
(marginální řádková
četnost)

celková četnost ve sloupci
(marginální sloupcová
četnost)

očekávané hodnoty

$$= \frac{(R) \cdot (C)}{N}$$

celkový počet případů
v souboru

N

POZOROVANÉ A OČEKÁVANÉ HODNOTY A RESIDUÁLY

			Q84 Pohlaví		Total
			1 muž	2 žena	
Q30_1 Bůh existuje	1 ano	Count	253	379	632
		Expected Count	309.0	323.0	632.0
		Residual	-56.0	56.0	
		Std. Residual	-3.2	3.1	
		Adjusted Residual	-5.7	5.7	
	2 ne	Count	541	451	992
		Expected Count	485.0	507.0	992.0
		Residual	56.0	-56.0	
		Std. Residual	2.5	-2.5	
		Adjusted Residual	5.7	-5.7	
Total	Count	794	830	1624	
	Expected Count	794.0	830.0	1624.0	

INTERPRETACE KONTINGENČNÍ TABULKY Z ROZLOŽENÍ ČETNOSTÍ

1. JE OBVYKLE PROVÁDĚNA Z ŘÁDKOVÝCH ČI SLOUPCOVÝCH RELATIVNÍCH ČETNOSTÍ (PROCENT). Na rozdíl od absolutních hodnot nám totiž umožňují vzít v úvahu počet případů v souboru.

Příklad:

Údaj o tom, že 20 žen v souboru má vysokoškolské vzdělání, zatímco mezi muži je vysokoškoláků jen 6 (ženy s vysokoškolským vzděláním se zde vyskytují častěji) může dostat zcela jiný výklad, uvědomíme-li si, že žen je v souboru 100 a mužů jen 10 (vysokoškolské vzdělání má v souboru 60% mužů, zatímco jen 20% žen).

2. URČÍME, KTEROU PROMĚNNOU BUDEME POVAŽOVAT ZA NEZÁVISLOU. Případy, jako je vztah pohlaví a spokojenosti s platem jsou jednoduché: na pohlaví závisí spokojenost s platem, spokojenost s platem pohlaví ovlivnit nemůže. V jiných případech je to složitější (nezávisle proměnnou určujeme arbitrárně).

Příklad:

Preference soukromého vlastnictví před státním některými lidmi může vést k tomu, že tito lidé volí pravicové strany a ne levicové (a naopak).

Vede však tato preference k volební preferenci, nebo je to naopak volební preference, která je příčinou preference určitého typu vlastnictví?

- Pro analýzu VOLÍME PŘIMĚŘENÉ RELATIVNÍ ČETNOSTI (sloupcové, jsou-li varianty nezávisle proměnných ve sloupcích a řádkových, jsou-li v řádcích).
- POROVNÁME PROCENTA pro jednotlivé varianty nezávisle proměnné pro každou variantu závisle proměnné.
 - Je-li nezávisle proměnná ve sloupcích, porovnááme tyto sloupcová % v každém řádku.
 - Je-li nezávislá proměnná v řádcích, porovnááme řádková % v každém sloupci.

Příklad: Spokojenost s platem u mužů a žen (vztah mezi pohlavím a spokojeností s platem)

	absolutní četn. řádkové % sloupcové % celkové %	PROMĚNNÁ A pohlaví		CELKEM
		muži	ženy	
PROMĚNNÁ B - spokojenost	spokojen/a	30	28	58
		52%	48%	100%
		25%	40%	
		16%	14%	
	střed	30	30	60
		50%	50%	100%
		25%	43%	
		16%	16%	
	nespokojen/a	60	12	72
		83%	17%	100%
		50%	17%	
		32%	6%	
CELKEM	120	70	190	
	100%	100%		

Muži jsou méně spokojeni s platem než ženy, jak lze zjistit porovnáním sloupcových četností v jednotlivých řádcích (spokojených je mezi muži 25%, zatímco mezi ženami 40%).

Pokud bychom chtěli interpretovat řádkové četnosti, dozvíme se jen, že mezi spokojenými je 52% mužů a 48% žen, což nedává moc smyslu, protože počet žen a mužů je v souboru rozdílný. Museli bychom jedinečně porovnat podíl spokojených mužů (resp. žen) s jejich podílem v celém souboru.

Všimněte si, že skutečně potřebujeme nikoliv absolutní, ale RELATIVNÍ ČETNOSTI (spokojených mužů je v absolutním počtu prakticky stejně jako spokojených žen).