



# KONTINGENČNÍ TABULKA. STANDARDIZOVANÝ REZIDUÁL.

ZUR357 Statistická analýza dat --

22. listopadu 2017

# DVOUROZMĚRNÁ ANALÝZA – SROVNÁVÁNÍ PRŮMĚRŮ

= bivariační analýza = třídění druhého stupně

Typicky:

- Srovnání průměrů pro dvě kategorie (např. průměrný příjem mužů a žen, porovnání míry religiozity u osob se základním a středoškolským vzděláním).
- Porovnávání s nějakou známou hodnotou (např. pochází náš výběr z populace s určitým rozdělením?)

Závislá proměnná – kardinální (věk, počet slov, délka promluvy, počet zmínek, indexy a škály,...), příp. ordinální s dlouhou škálou (např. míra religiozity, ...)

Nezávislá proměnná – kategorizovaná (nominální, ordinální)

Procedury: MEANS, T-TEST, ANOVA

# PROCEDURA MEANS

<b>Nezávislá proměnná - deník</b>	<b>Závislá proměnná - Počet slov v textech o uprchlické problematice</b>
Lidové noviny	375
Mladá Fronta DNES	128
Právo	220
Hospodářské noviny	540

# T-TEST

- Pouze 2 proměnné
- Závisle proměnná – kardinální
- Nezávisle proměnná: kategorizovaná pouze se dvěma kategoriemi
- Test signifikance

Kategorie nezávisle prom.	Průměry z.p.
Muži	1,7
Ženy	2,2

# ANOVA

- Analýza 2 proměnných
- Závisle proměnná – kardinální
- Nezávisle proměnná: kategorizovaná s více než dvěma kategoriemi
- Test signifikance vzájemných rozdílů mezi průměry

Kategorie nezávisle prom.	Průměry z.p.
20-29 let	3,4
30-39 let	4,7
40-49 let	6,6
50 a více let	8,1

# CO JE TESTOVÁNÍ STATISTICKÉ SIGNIFIKANCE?

- Oblast statistického usuzování (statistické inference)
- TESTOVÁNÍ směřuje k zobecnění dat výběrového souboru na základní soubor.
- Děje se prostřednictvím testů signifikance
- Test signifikance: test nulové hypotézy, zdali je možné, nebo ji není možné zamítnout

Co je nulová hypotéza?

# TESTOVÁNÍ HYPOTÉZ

Výzkumná hypotéza:

*Barva vlasů ženy má vliv na míru její inteligence*

Statistická hypotéza jako operacionalizovaná výzkumná hypotéza:

*Blondýnky se budou lišit od černovlásek ve výsledcích v testu inteligence*

**Nulová hypotéza** jako specifický případ statistické hypotézy:

*Mezi blondýnkami a černovláskami nebude v testech inteligence žádný rozdíl*

# TESTOVÁNÍ HYPOTÉZ

Nulová hypotéza jako specifický případ statistické hypotézy

*Mezi blondýnkami a černovláskami nebude v testech inteligence žádný rozdíl (rozdíl bude 0 nebo blízký 0)*

Alternativní hypotéza k hypotéze nulové: nesměrová

*Blondýnky a černovlásky se budou v testech inteligence odlišovat (průměry budou odlišné)*

Alternativní hypotéza k hypotéze nulové: směrová

*Blondýnky budou mít v testech inteligence nižší skóre než černovlásky (průměr B < průměr C)*

Alternativní hypotéza k hypotéze nulové: přesná

*Černovlásky budou mít v testech inteligence skóre vyšší o 20 bodů než blondýnky (Č - B*





# TESTY STATISTICKÉ VÝZNAMNOSTI

Na základě čeho rozhodují o významnosti?



## Jednostranný (*one-tailed*) a dvoustranný test (two-tailed)

### Kritika testů významnosti

- Nadužívání testů významnosti (reprezentativita)
- Rigidita 95% hranice významnosti
- Nešťastný výraz „významnost“

### Parametrické a neparametrické testy

### Independent Samples Test

		q65a_8 Homosexualita	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	,065	
	<b>Sig.</b>	,798	
t-test for Equality of Means	t	9,787	9,796
	df	1162	839,091
	<b>Sig. (2-tailed)</b>	<b>,000</b>	,000
	Mean Difference	2,088	2,088
	Std. Error Difference	,213	,213
95% Confidence Interval of the Difference	Lower	1,669	1,669
	Upper	2,506	2,506

# KONTINGENČNÍ TABULKY

- Dvourozměrná analýza kategorizovaných proměnných s menším počtem kategorií
- Česky: křížová tabulka

VEK\_KAT1 Vekové kategorie \* Q8 Důvěra v lidi Crosstabulation

Count		Q8 Důvěra v lidi		Total
		1 lidem je možné důvěřovat	2 člověk musí být opatrný	
VEK_KAT1	1 18-29	87	327	414
Vekové kategorie	2 30-49	158	508	666
	3 50+	199	585	784
Total		444	1420	1864

VEK\_KAT1 Vekové kategorie \* Q8 Důvěra v lidi Crosstabulation

				Q8 Důvěra v lidi		Total
				1 lidem je možné důvěřovat	2 člověk musí být opatrný	
VEK_KAT1 Vekové kategorie	1 18-29	Count	87	327	414	
		Row %	21,0%	79,0%	100,0%	
		Column %	19,6%	23,0%	22,2%	
		Total %	4,7%	17,5%	22,2%	
	2 30-49	Count	158	508	666	
		Row %	23,7%	76,3%	100,0%	
		Column %	35,6%	35,8%	35,7%	
		Total %	8,5%	27,3%	35,7%	
	3 50+	Count	199	585	784	
		Row %	<b>25,4%</b>	74,6%	<b>100,0%</b>	
		Column %	<b>44,8%</b>	41,2%	42,1%	
		Total %	<b>10,7%</b>	31,4%	42,1%	
Total	Count	444	1420	1864		
	Row %	23,8%	76,2%	100,0%		
	Column %	<b>100,0%</b>	100,0%	100,0%		
	Total %	23,8%	76,2%	<b>100,0%</b>		

- Řádková %,
- Sloupcová %,
- Celková %

VEK\_KAT1 Vekové kategorie \* Q8 Důvěra v lidi Crosstabulation

				Q8 Důvěra v lidi		Total
				1 lidem je možné důvěřovat	2 člověk musí být opatrný	
VEK_KAT1	1 18-29	Count		87	327	414
Vekové kategorie		Row %		<b>21,0%</b>	79,0%	100,0%
	2 30-49	Count		158	508	666
		Row %		<b>23,7%</b>	76,3%	100,0%
	3 50+	Count		199	585	784
		Row %		<b>25,4%</b>	74,6%	100,0%
Total		Count		444	1420	1864
		Row %		23,8%	76,2%	100,0%

### Rámeček 7.1

Při publikaci výsledků ovšem tabulku v takového podobě, jako jsou tab. 7.2 nebo 7.3, nikdy nezveřejňujeme. Nejsou totiž přehledné. Proto je musíme upravit. Zásady jsou následující:

1. Každá tabulka musí mít číslo a název.
2. Všechny popisky tabulky musí být česky.
3. Názvy proměnných jsou ve sloupcích a řádcích jasně vyjádřeny.
4. Nezávisle proměnnou obvykle umístíme do sloupců, takže počítáme sloupcová procenta.
5. Závisle proměnná, která je v řádku, by měla mít varianty uspořádané od nejvyšší po nejnižší (pokud je měřená na ordinální nebo intervalové úrovni). Tento požadavek se ale nemusí dodržovat příliš striktně.
6. Poslední řádek uvádí celková procenta (obvykle tedy 100 %) a současně i absolutní počty případů.
7. V poznámce pod tabulkou se uvádí zdroj dat.

Tabulka 7.3 by tedy podle těchto zásad musel být pro případnou publikaci upravena takto:

**Tabulka 7.3: Důvěra k lidem podle věku (sloupcová %)**

Důvěra k lidem	Věkové kategorie		
	18-29	30-49	50+
Lidem je možné důvěřovat	21	24	25
Člověk musí být ve styku s ostatními lidmi opatrný	79	76	75
Celkem	100 % (414)	100 % (666)	100 % (784)

Pramen: EVS ČR 1999

## VZDĚLÁNÍ \* Q8 Důvěra v lidi Crosstabulation

			Q8 Důvěra v lidi		Total
			1 lidem je možné důvěřovat	2 člověk musí být opatrný	
VZDĚLÁNÍ kategorizace q94	1 základní	Count	71	292	363
		Expected Count	86,6	276,4	363,0
		Row %	19,6%	80,4%	100,0%
		Residual	-15,6	15,6	
	2 vyučen	Count	145	618	763
		Expected Count	182,0	581,0	763,0
		Row %	19,0%	81,0%	100,0%
		Residual	-37,0	37,0	
	3 SŠ	Count	151	395	546
		Expected Count	130,2	415,8	546,0
		Row %	27,7%	72,3%	100,0%
		Residual	20,8	-20,8	
	4 VŠ	Count	78	116	194
		Expected Count	46,3	147,7	194,0
		Row %	40,2%	59,8%	100,0%
		Residual	31,7	-31,7	
Total	Count	445	1421	1866	
	Expected Count	445,0	1421,0	1866,0	
	Row %	23,8%	76,2%	100,0%	



ROD\_STAV rodinný stav (sloučeno rozvedený + odloučení) \* PREFEREN volební preference 99 Crosstabulation

			PREFEREN volební preference 99						Total
			1 KSČM	2 ČSSD	3 KDU	4 US	5 ODS	90 nevolili by	
ROD_STAV rodinný stav (sloučeno rozvedený + odloučení)	1 ženatý/vdaná	Count	101	196	58	97	253	198	903
		Row %	11,2%	21,7%	6,4%	10,7%	28,0%	21,9%	100,0%
		Adjusted Residual	-,4	1,9	-1,1	-1,2	1,6	-1,6	
	2 vdovec/vdova	Count	36	25	15	9	29	35	149
		Row %	24,2%	16,8%	10,1%	6,0%	19,5%	23,5%	100,0%
		Adjusted Residual	5,2	-1,1	1,6	-2,2	-2,1	,1	
	3 rozvedený/á	Count	13	14	4	11	27	38	107
		Row %	12,1%	13,1%	3,7%	10,3%	25,2%	35,5%	100,0%
		Adjusted Residual	,2	-1,9	-1,4	-,4	-,3	3,1	
	4 nikdy neoženěn/nepro vdána	Count	11	49	21	45	66	56	248
		Row %	4,4%	19,8%	8,5%	18,1%	26,6%	22,6%	100,0%
		Adjusted Residual	-3,8	-,2	1,0	3,6	,0	-,3	
Total	Count	161	284	98	162	375	327	1407	
	Row %	11,4%	20,2%	7,0%	11,5%	26,7%	23,2%	100,0%	