



Štatistické metódy v medicínskom výskume I. a II.

Štatistika v kocke

alebo

Práca s výskumnými údajmi

Košice 31.1.2020

Mgr. Peter Kolarčík, PhD.
Ústav psychológie zdravia a
metodológie výskumu
Lekárska Fakulta UPJŠ

„Definícia štatistiky: Je to veda, ktorá vyrába nespoľahlivé fakty zo spoľahlivých údajov.“

Evan Esar

Je dokázané, že častejšie oslavovanie narodenín je **zdravé**.

Štatistiky dokazujú, že ľudia, ktorí oslávili najviac narodenín, žili najdlhšie.

S. den Hartog

Ph D. Thesis University of Groningen

Štatistika

- Je len doplnok vedeckého skúmania
- Je to nástroj a pomôcka, nie základ vedeckého skúmania
- V psychológii ma za úlohu dotvárať dojem „vedeckosti“ disciplíny :D

Pre spoľahlivý vedecký výskum je dôležité:

- Oblasť výskumu
- Výskumná otázka alebo hypotéza (nie „ad hoc“)
- Design výskumu
- Metódy
- Populácia vs. vzorka (reprezentatívnosť)

- Štatistické analýzy a ich korektnosť prichádzajú na rad až keď sú splnené predchádzajúce podmienky

Sofistikovaná statistika na „zlých“
dátach nám nepomôže dosiahnuť
„dobré“ výsledky.

Dôležité informácie

Úrovne merania

- nominálna, ordinálna , kardinálna (intervalová a pomerová)
- **Nominálne premenné:**
 - Nedajú sa zoradiť podľa veľkosti, platí pravidlo patrí/nepatrí
 - (pohlavie, typ diagnózy, miesto bydliska, ...)
- **Ordinálne (poradové) premenné:**
 - Majú nepresné/nerovnaké vzdialenosti medzi po sebe nasledujúcimi hodnotami, ale ich poradie odráža postupný nárast/pokles parametra
 - (Lickertova škála: *miera súhlasu*, frekvencia: *nikdy, niekedy, občas, často, vždy*)
- **Intervalové premenné:**
 - majú rovnaké vzdialenosti medzi jednotlivými hodnotami, ale nulové hodnoty sú len dohodnuté (arbitrárne stanovené)
 - (napr. teplota v stupňoch Ceslia/Farenheita, IQ,...).
- **Pomerové premenné:**
 - majú reálnu nulovú hodnotu a definované rovnaké vzdialenosti medzi jednotlivými hodnotami
 - umožňujú najväčšiu flexibilitu štatistických metód a môžu byť použité pre analýzy
 - (reakčný čas, hmotnosť, dĺžka, ...)
- **Kategorické premenné:** nominálne a ordinálne
- **Kontinuálne premenné:** pomerové a intervalové premenné

Typ škály	Matematická štruktúra	Prípustné štatistické operácie	Prijateľné transformácie
Nominálna Kategorická	Nezoriadená množina	modus, Chi-kvadrát	One to One (equality (=))
Ordinálna (poradová)	Zoriadená množina	medián and modus, percentily, poradová korelácia (Spearman's rho), neparametrická analýza variancie,	Monotonic increasing (order (<))
Intervalová	Nekonečné kontinuum	priemer, standardná odchýlka, korelácia (Pearson's r), regresia, parametrická analýza variancie, faktorová analýza	Positive linear (affine)
Pomerová	Jednorozmerný vektorový priestor	All statistics permitted for interval scales plus the following: geometric mean, harmonic mean, coefficient of variation, logarithms	Positive similarities (multiplication)

Deskripcia vs. Inferencia?

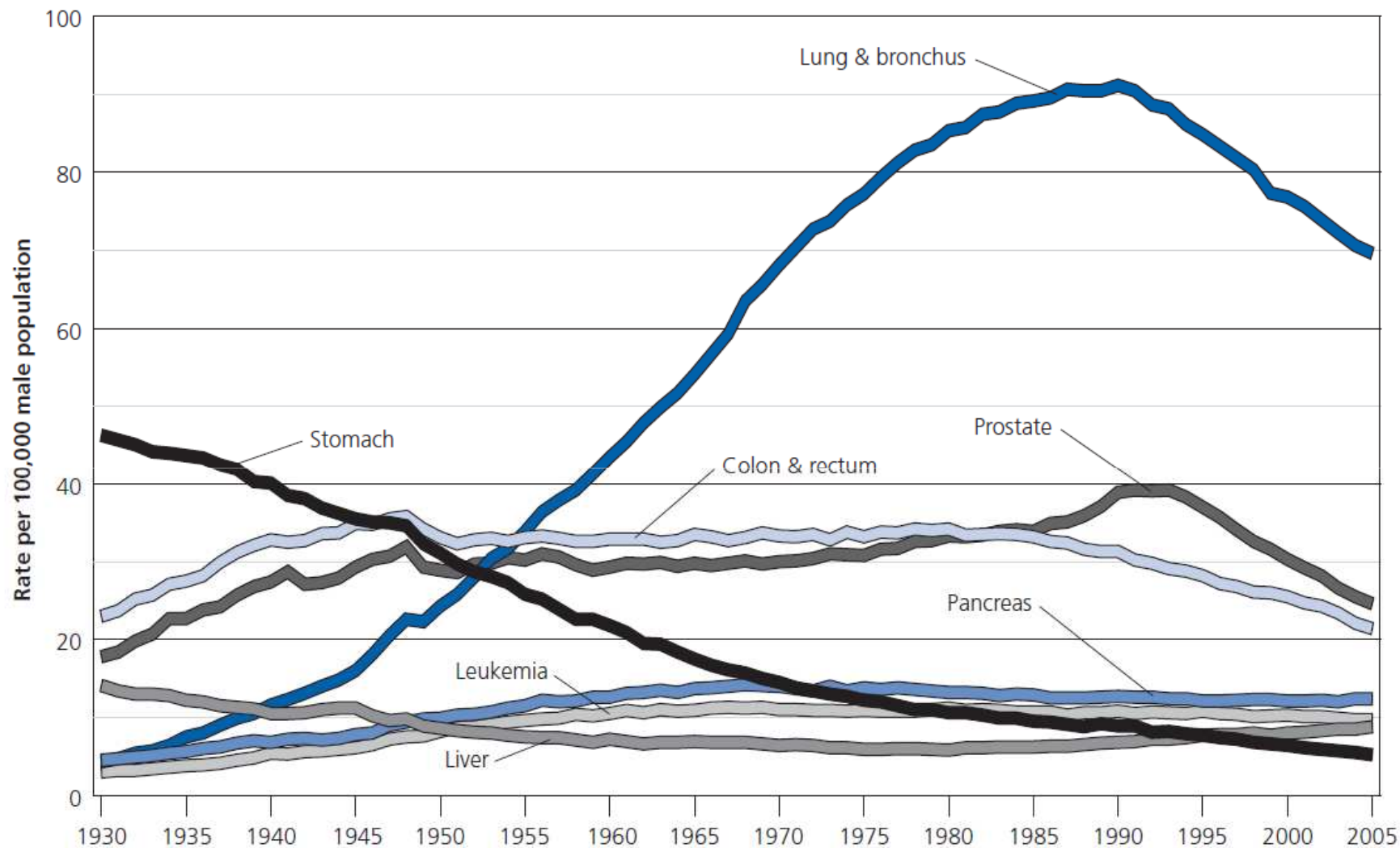
- **Deskriptívna štatistika:**

- Číselný a grafický opis súboru
- prevalencie, proporcie, priemery, kategórie,... (grafy, tabuľky)

- **Inferenčná štatistika:**

- vyvodzovanie vzťahov a súvislostí, hľadanie rozdielov (testy, korelácie, ...)
- Testovanie hypotéz (áno/nie)
- Popis vzťahov (korelácia)
- Modelovanie vzťahov (regresná analýza)

Age-adjusted Cancer Death Rates,* Males by Site, US, 1930-2005



*Per 100,000, age adjusted to the 2000 US standard population.

Note: Due to changes in ICD coding, numerator information has changed over time. Rates for cancer of the liver, lung and bronchus, and colon and rectum are affected by these coding changes.

Source: US Mortality Data, 1960 to 2005, US Mortality Volumes, 1930 to 1959, National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.

Signifikancia

- Štatistická významnosť
 - udáva s akým rizikom (pravdepodobnosťou) zamietame nulovú hypotézu v prípade, keby bola pravdivá.
 - Nehovorí o závažnosti zisteného rozdielu/vzťahu
 - Vyjadruje sa ako *p-hodnota* (nehovorí i miere efektu)
 - vysoko signifikantný výsledok nemusí mať žiaden praktický význam
 - Súvisí s veľkosťou vzorky (*sample size*)

Chyby štatistických záverov

- Ak výsledok testu zodpovedá realite, potom je rozhodnutie správne, ak nie potom došlo k chybe.
- Kvôli štatistickej povahe testu, výsledok, až na pár zriedkavých výnimiek, nie je nikdy bez chyby
- Rozlišujeme 2 druhy chýb:
 - **Chyba I. typu (type I error)**
 - **Chyba II. typu (type II error)**

Chyba I. typu

- chybné zamietnutie nulovej hypotézy
- nastane, keď nulová hypotéza (H_0) je pravdivá, ale je zamietnutá
- **Chyba falošnej positivity** (planý poplach/false alarm) je výsledok, ktorý naznačuje, že daná podmienka bola splnená, aj keď v skutočnosti splnená nebola.
- Uvádza sa ako signifikancia testu (α) = pravdepodobnosť výskytu *chyby I. typu*

Chyba II. typu

- nesprávne prijatie nulovej hypotézy
- Nulová hypotéza je nesprávna, ale na základe testu je nemožno zamietnuť
- **Chyba falošnej negativity**
- Označuje sa ako Beta (β), súvisí so „silou“ testu = $(1 - \beta)$
- Štatistický test umožní prijať/zamietnuť H_0 , ale nikdy nedokáže jej pravdivosť/neppravdivosť

	H_0 je pravdivá	H_0 je nepravdivá
Zamietnutie H_0	Chyba I. typu (Falošná pozitivita)	Správny záver (skutočná pozitivita)
Nezamietnutie H_0	Správny záver (skutočná negativita)	Chyba II. Typu (Falošná negativita)

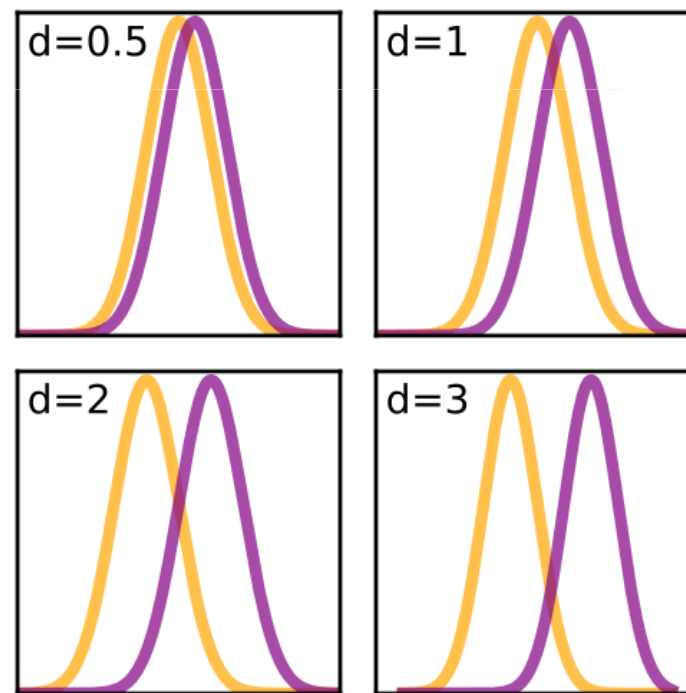
Miera efektu

- Vecná signifikancia – **effekt size** (miera efektu)
 - Cohenovo d , odds ratio (OR)

$d=0,2 - 0,3$ malý efekt

$d=\text{okolo } 0,5$ stredne veľký efekt

$d > 0,8$ veľký efekt



Miera efektu

- **Power analyses** (G*Power – nie je v SPSS)
 - Resp. udáva akú veľkú mieru efektu vieme odhaliť na danej vzorke
 - Výpočet minimálnej veľkosti vzorky potrebnej pre spoľahlivé odhalenie/detekciu stanovenej miery efektu
 - Umožňuje správne zamietnutie/prijatie nulovej hypotézy
 - Štandardne by nemala byť nižšia ako 0,80

Veľkosť vzorky v závislosti na pomere Signal/Noise
(5% sig., 2-sided)

Signal/Noise ratio	90% power	80% power
0.2	526	393
0.4	132	99
0.6	59	45
0.8	34	26
1.0	22	17
1.2	16	12
1.4	12	9
1.6	9	7
1.8	8	6
2.0	6	5
2.2	6	4
2.4	5	4
2.6	4	4
2.8	4	3
3.0	4	3

Interpretácia výsledkov

- Nie len prijatie/zamietnutie nulovej hypotézy

Je potrebné zvážiť:

- Úroveň signifikancie
- Zhodnotenie reálnych rozdielov/vzťahov
- Zohľadniť charakteristiky testovaného súboru (kontext) – súčasť diskusie

“Štatistika nikdy nemôže nahradiť
zdravý úsudok“

Henry Clay

Deskriptívna analýza



Popis súboru

- Frekvenčné tabuľky

Statistics

Věkové kategorie		
N	Valid	2182
	Missing	14
Mean		2,02
Median		2,00
Mode		3
Std. Deviation		,818

Věkové kategorie

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	11	709	32,3	32,5	32,5
	13	720	32,8	33,0	65,5
	15	753	34,3	34,5	100,0
Total		2182	99,4	100,0	
Missing	System	14	,6		
Total		2196	100,0		

Ako na to v SPSS:

Analyze -> Descriptive statistics -> Frequencies: ✓ Display frequency tables

- Statistics: priemer, medián, modus, std. odchýlka, rozptyl, rozsah, šikmosť/strmosť distribúcie
- Charts: stĺpcový graf, koláčový graf, histogram

Věkové kategorie

Pohlaví			Frequency	Percent	Valid Percent	Cumulative Percent
chlapec	Valid	11	355	34,2	34,5	34,5
		13	331	31,9	32,1	66,6
		15	344	33,2	33,4	100,0
		Total	1030	99,3	100,0	
	Missing	System	7	,7		
Total			1037	100,0		
dívka	Valid	11	354	30,5	30,7	30,7
		13	389	33,6	33,8	64,5
		15	409	35,3	35,5	100,0
		Total	1152	99,4	100,0	
	Missing	System	7	,6		
Total			1159	100,0		

Pohlaví

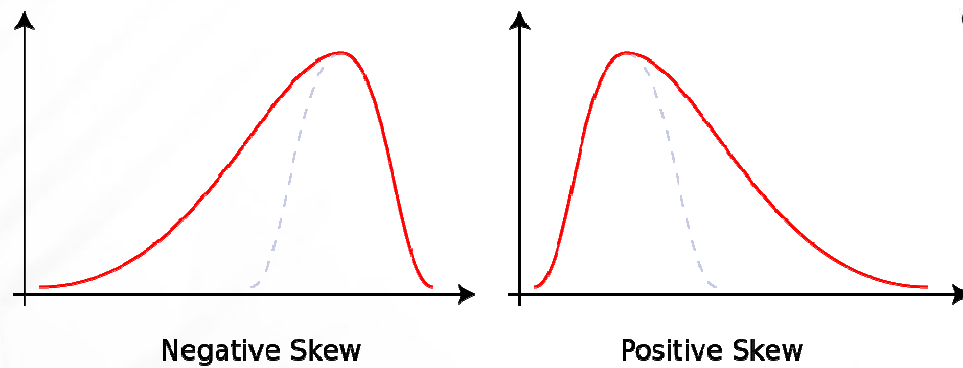
Věkové kategorie			Frequency	Percent	Valid Percent	Cumulative Percent
0	Valid	chlapec	7	50,0	50,0	50,0
		dívka	7	50,0	50,0	100,0
		Total	14	100,0	100,0	
11	Valid	chlapec	355	50,1	50,1	50,1
		dívka	354	49,9	49,9	100,0
		Total	709	100,0	100,0	
13	Valid	chlapec	331	46,0	46,0	46,0
		dívka	389	54,0	54,0	100,0
		Total	720	100,0	100,0	
15	Valid	chlapec	344	45,7	45,7	45,7
		dívka	409	54,3	54,3	100,0
		Total	753	100,0	100,0	

Popis súboru

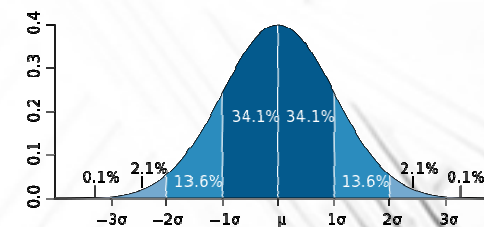
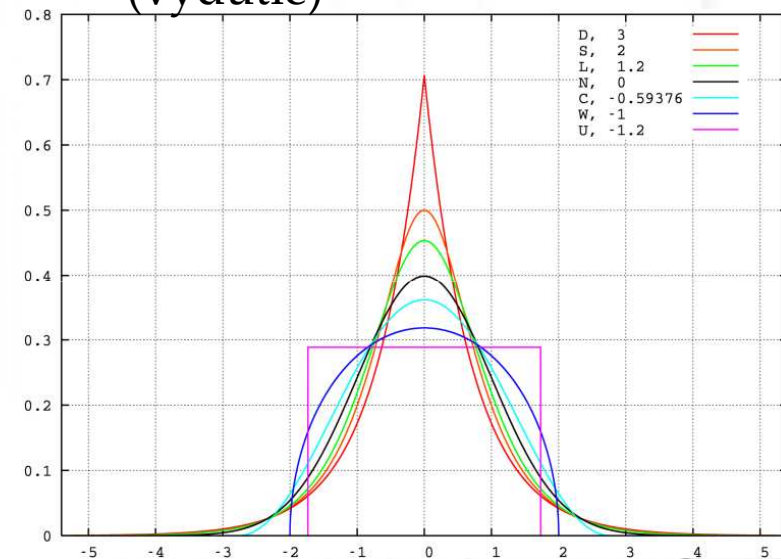
- Miery centrálnej tendencie:
 - **Modus** – najčastejšie sa vyskytujúca hodnota
 - **Medián** – polohový stred, 50-ty percentil
 - **Priemer** – aritmetický priemer
- Nominálne premenné: modus
- Ordinálne premenné: modus, medián
- Kardinálne premenné: modus, medián, priemer

Popis súboru

- Distribúcia
 - Štandardná odchýlka (sigma, σ ; SD) – rozptyl hodnôt okolo priemeru
 - Šikmosť (skewness) distribúcie



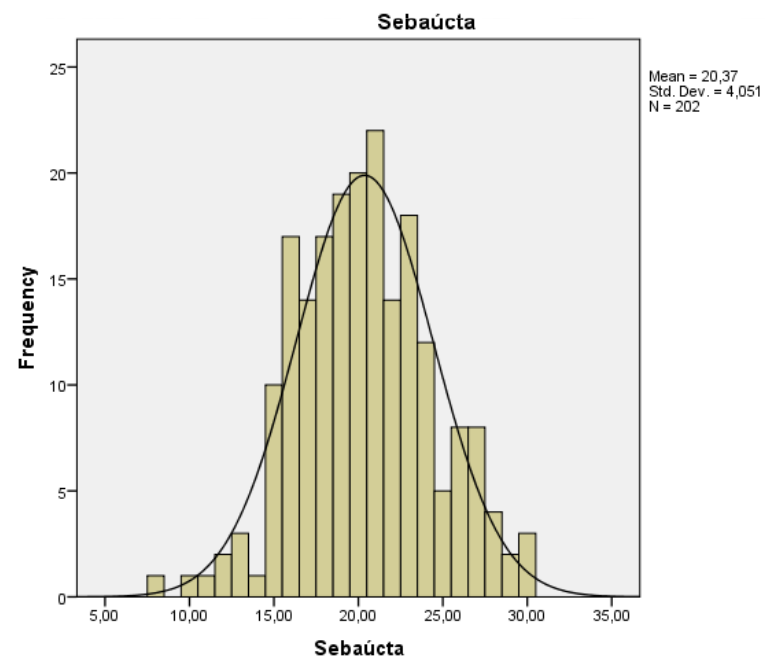
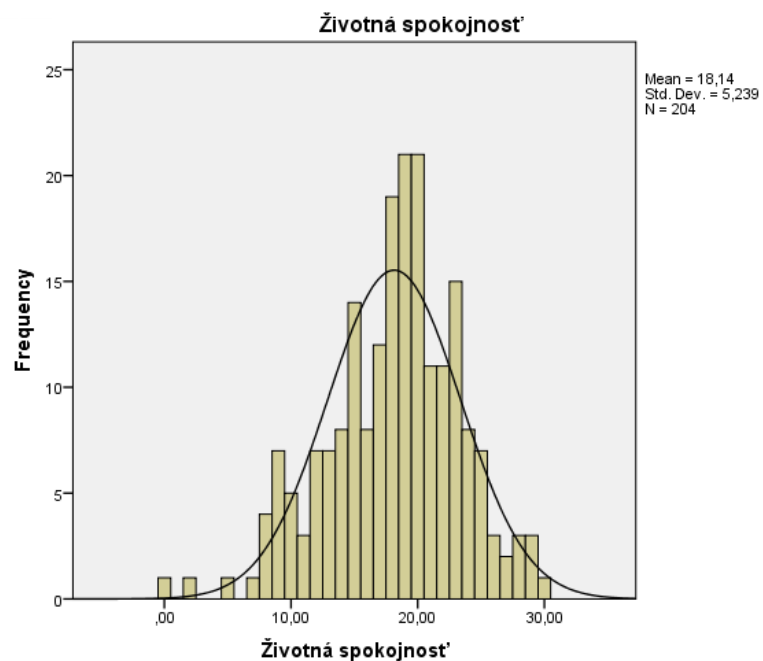
- Strmosť (kurtosis) distribúcie (vydutie)



Statistics

		Životná spokojnosť	Sebaúcta
N	Valid	204	202
	Missing	1	3
Mean		18,1373	20,3663
Median		19,0000	20,0000
Mode		19,00 ^a	21,00
Std. Deviation		5,23918	4,05122
Skewness		-,446	,057
Std. Error of Skewness		,170	,171
Kurtosis		,398	-,015
Std. Error of Kurtosis		,339	,341
Minimum		,00	8,00
Maximum		30,00	30,00

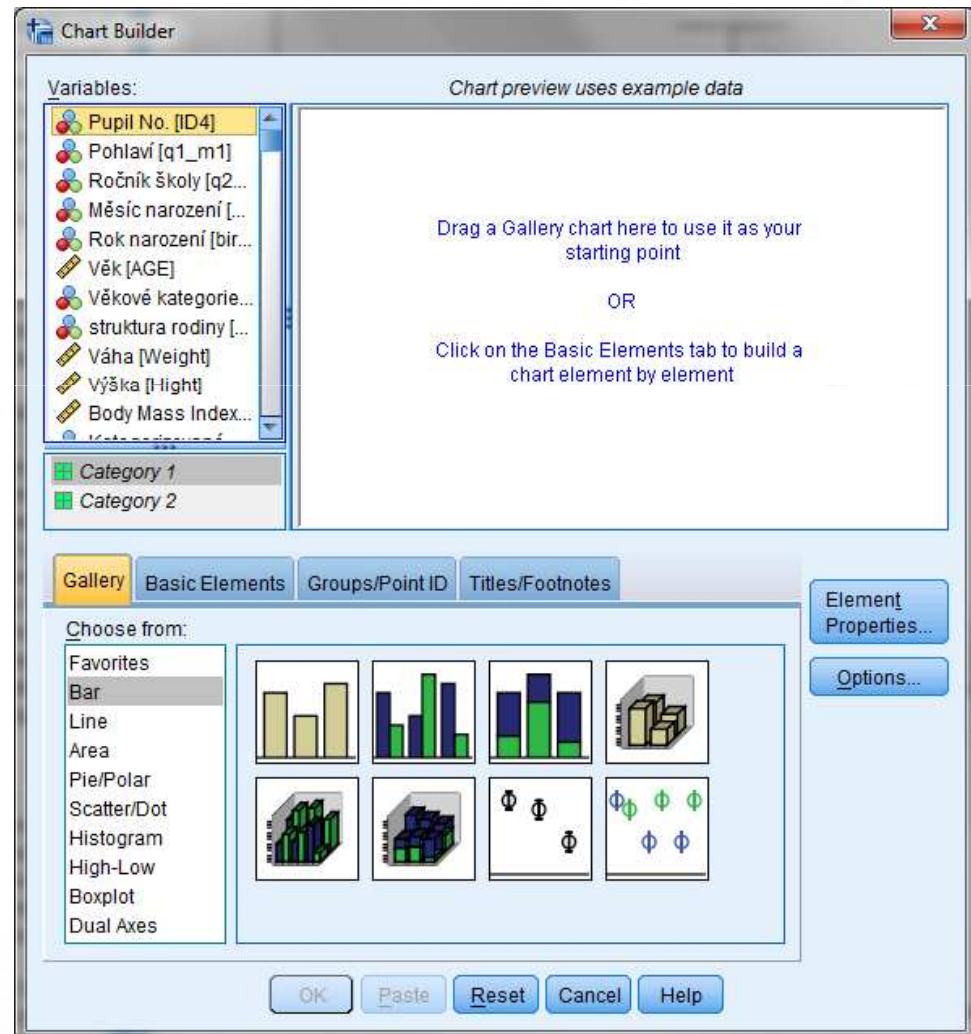
a. Multiple modes exist. The smallest value is shown



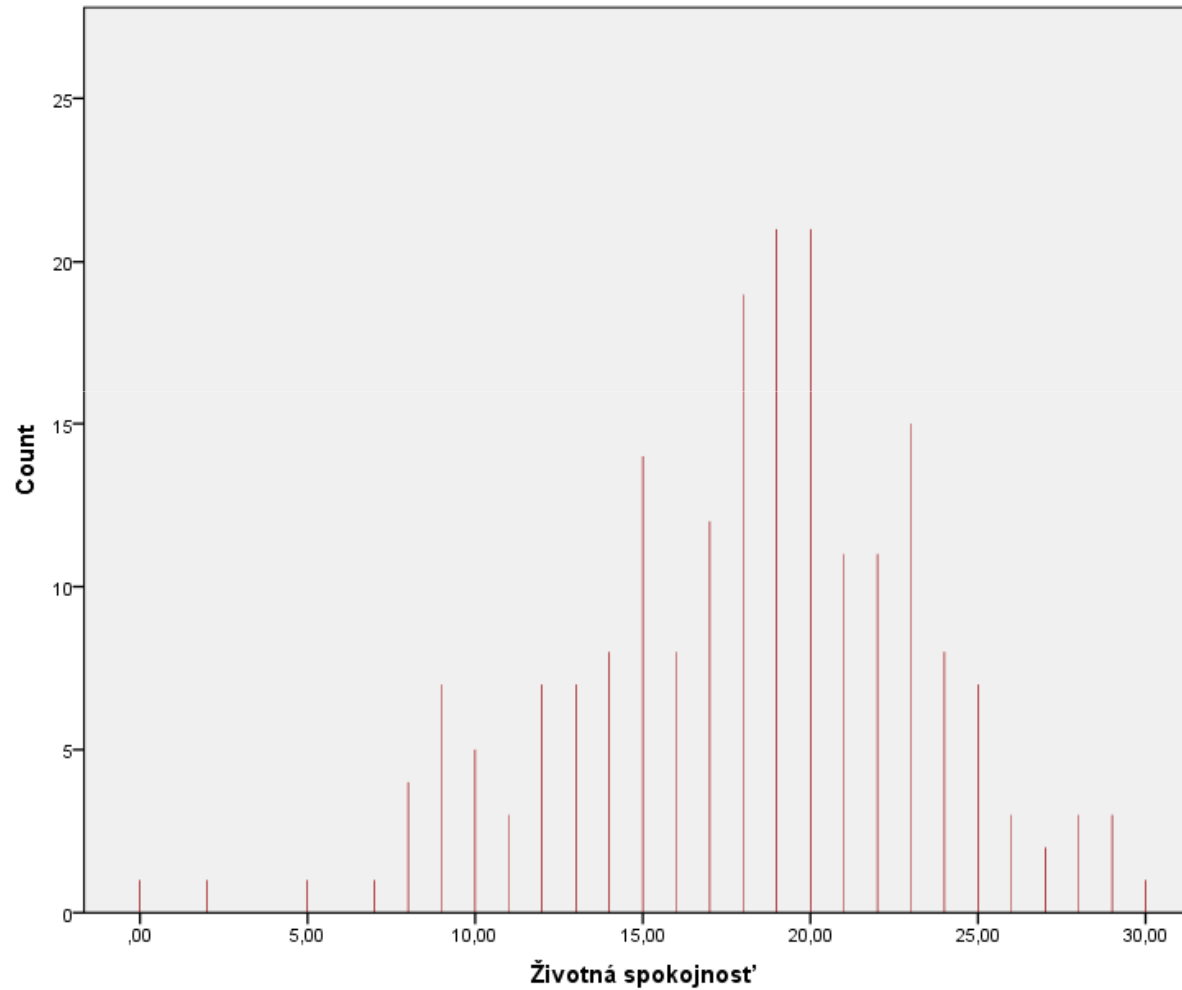
Grafické zobrazenie údajov

- Stĺpcový graf
- Čiarový graf
- Histogram
- Koláčový graf
- Bodový graf (scatter plot)
- Boxplot

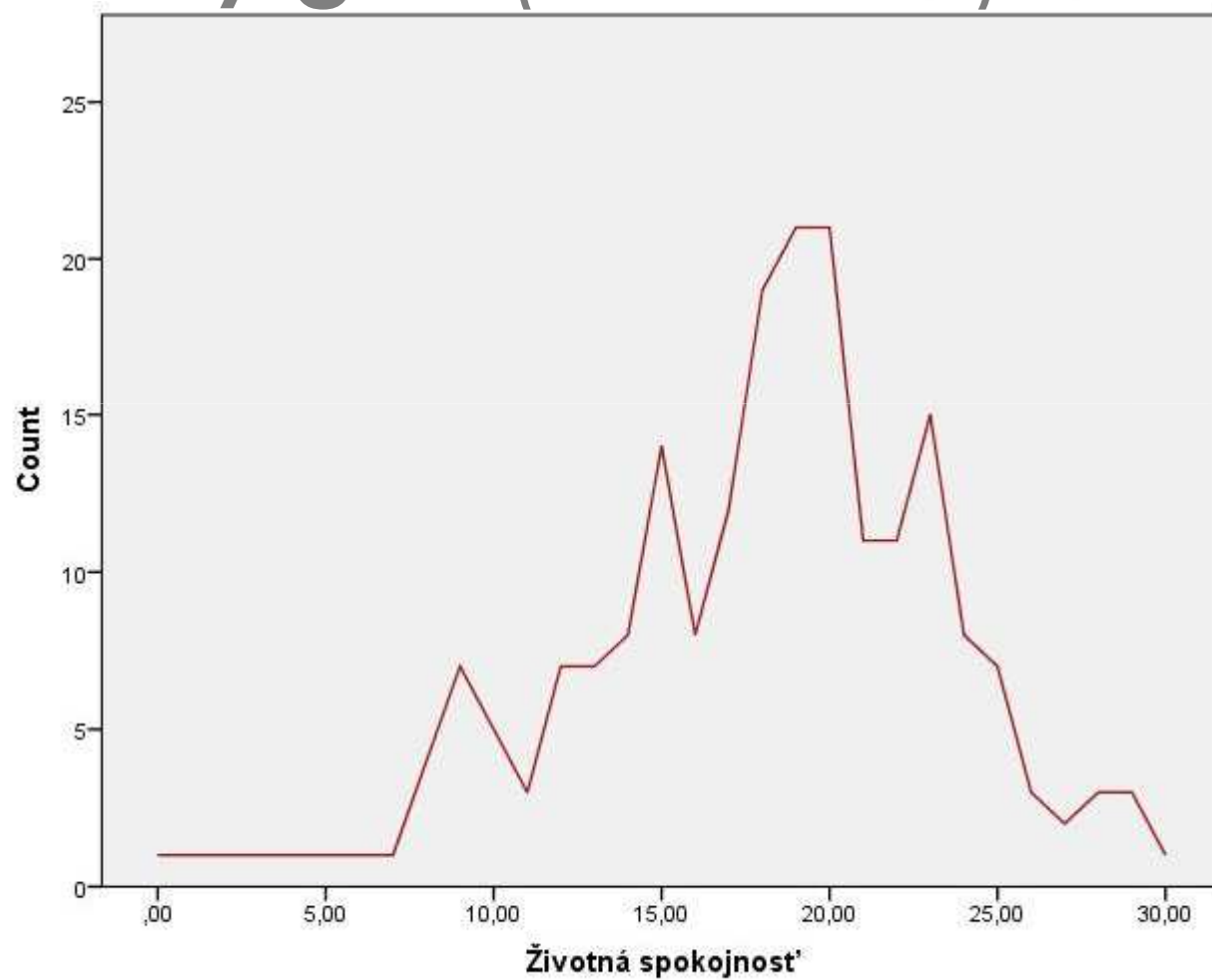
Ako na to v SPSS:
Graphs -> Chart Builder:



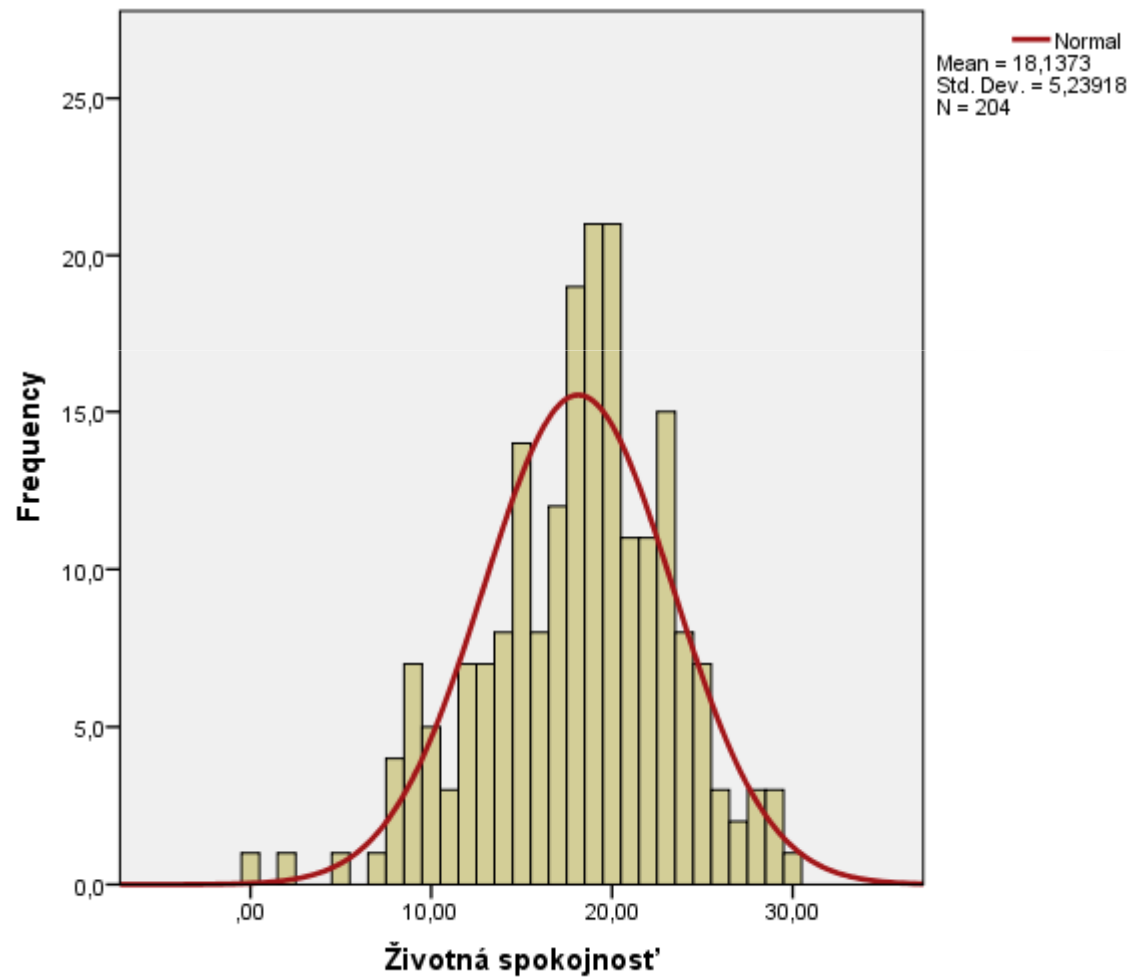
Stípcový graf (bar chart)



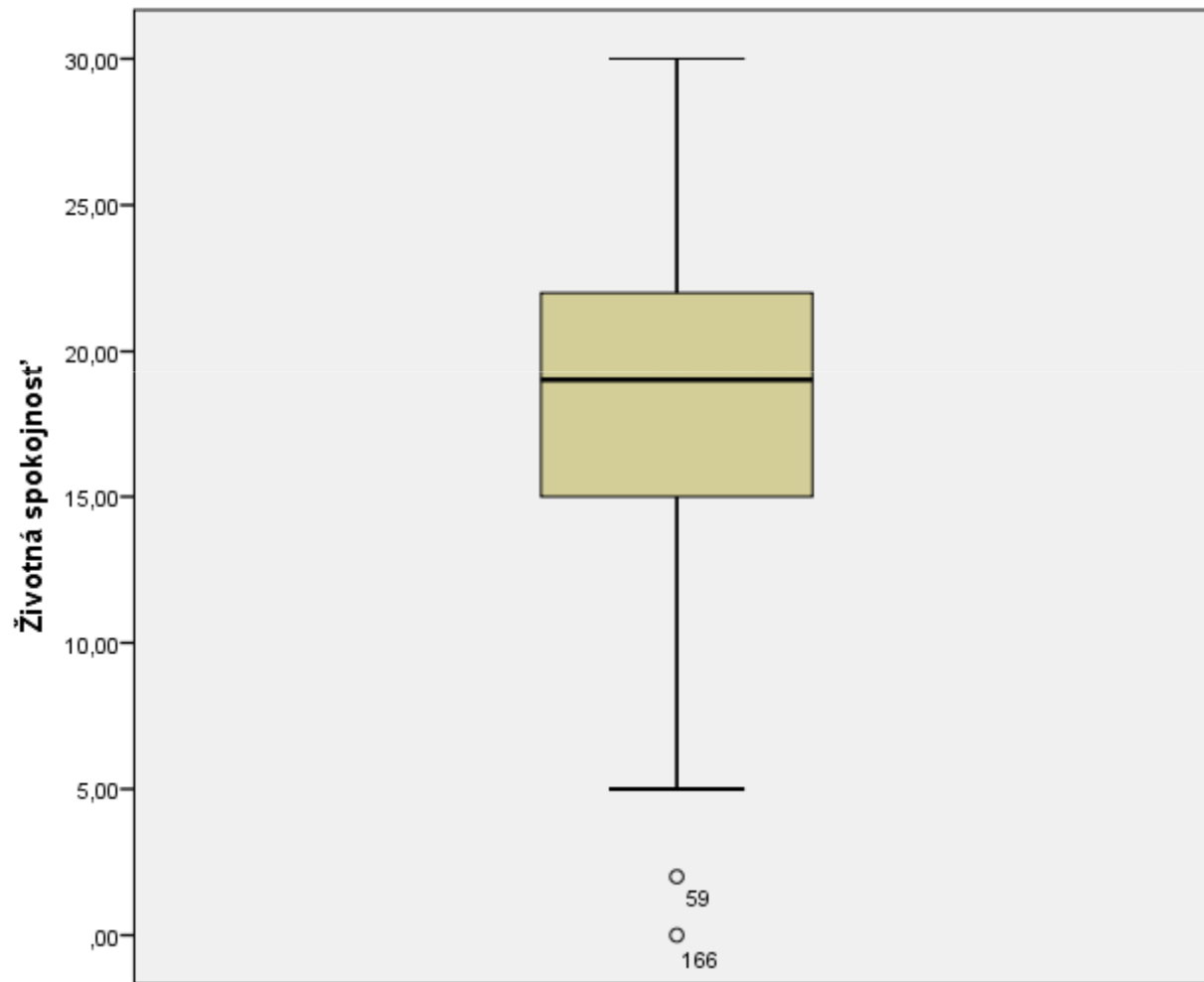
Čiarový graf (line chart)



Histogram



Boxplot



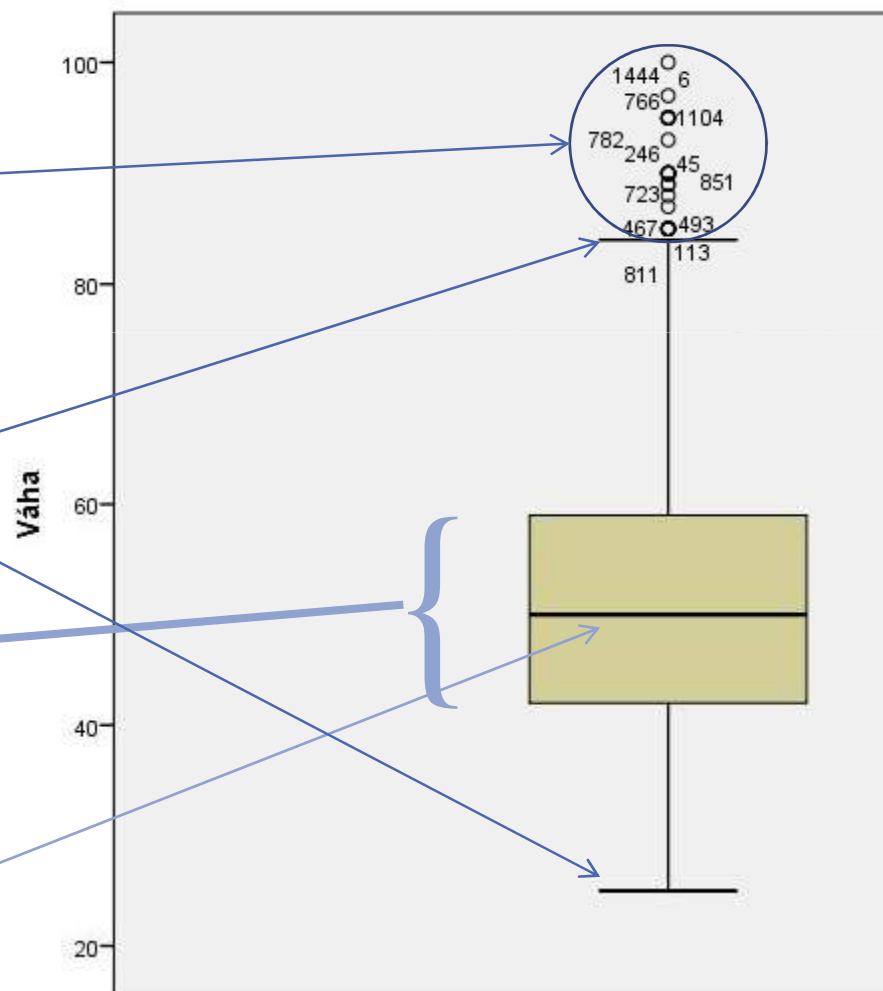
Boxplot - vysvetlenie

**Vybočujúce hodnoty
(outliers)**
($\pm 1,5 - 3$ násobok IQR)

Fúzy (whiskers)
do $\pm 1,5$ násobok IQR

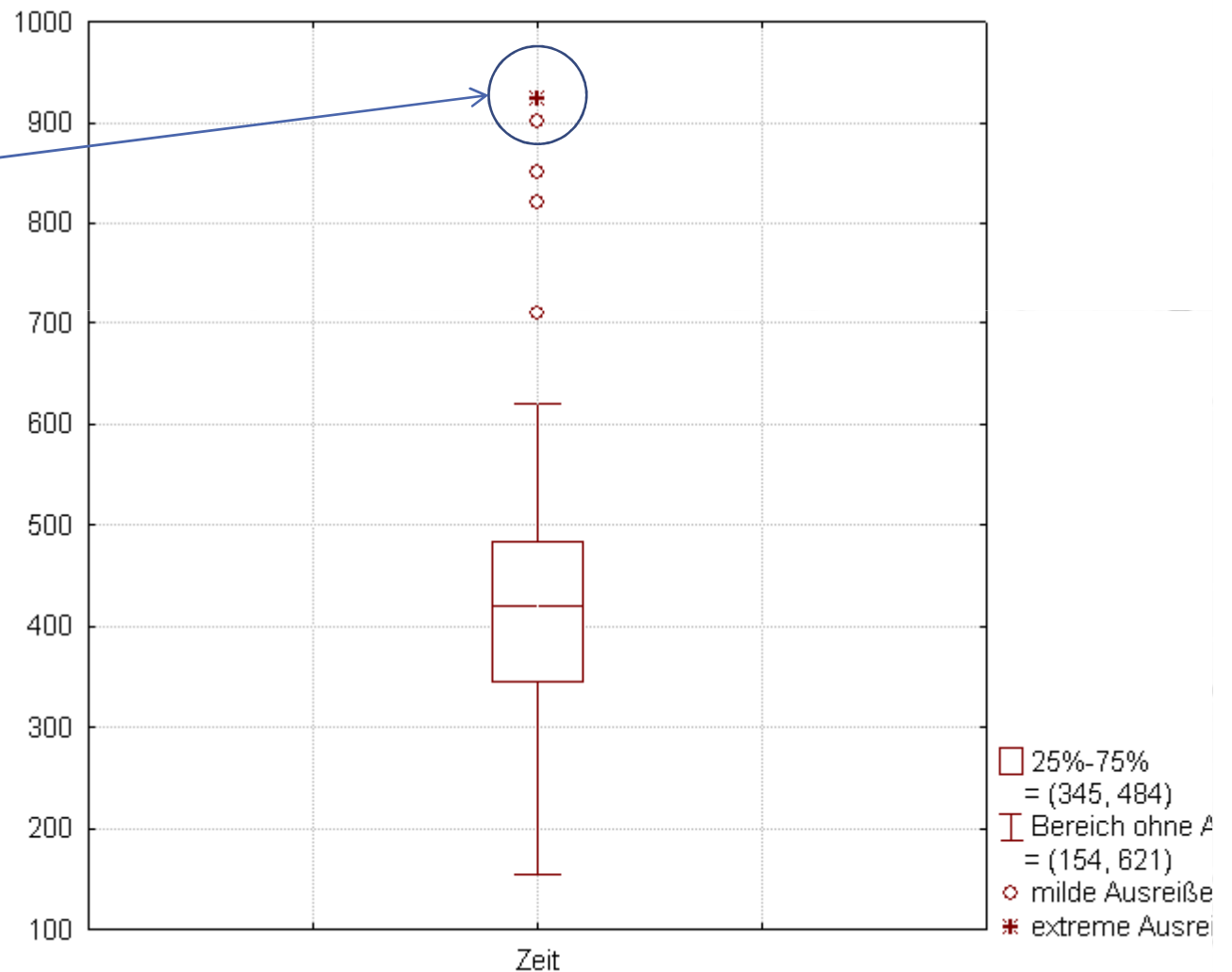
**Medzikvartilové
rozpätie (IQR)**

Medián
(polohový stred)

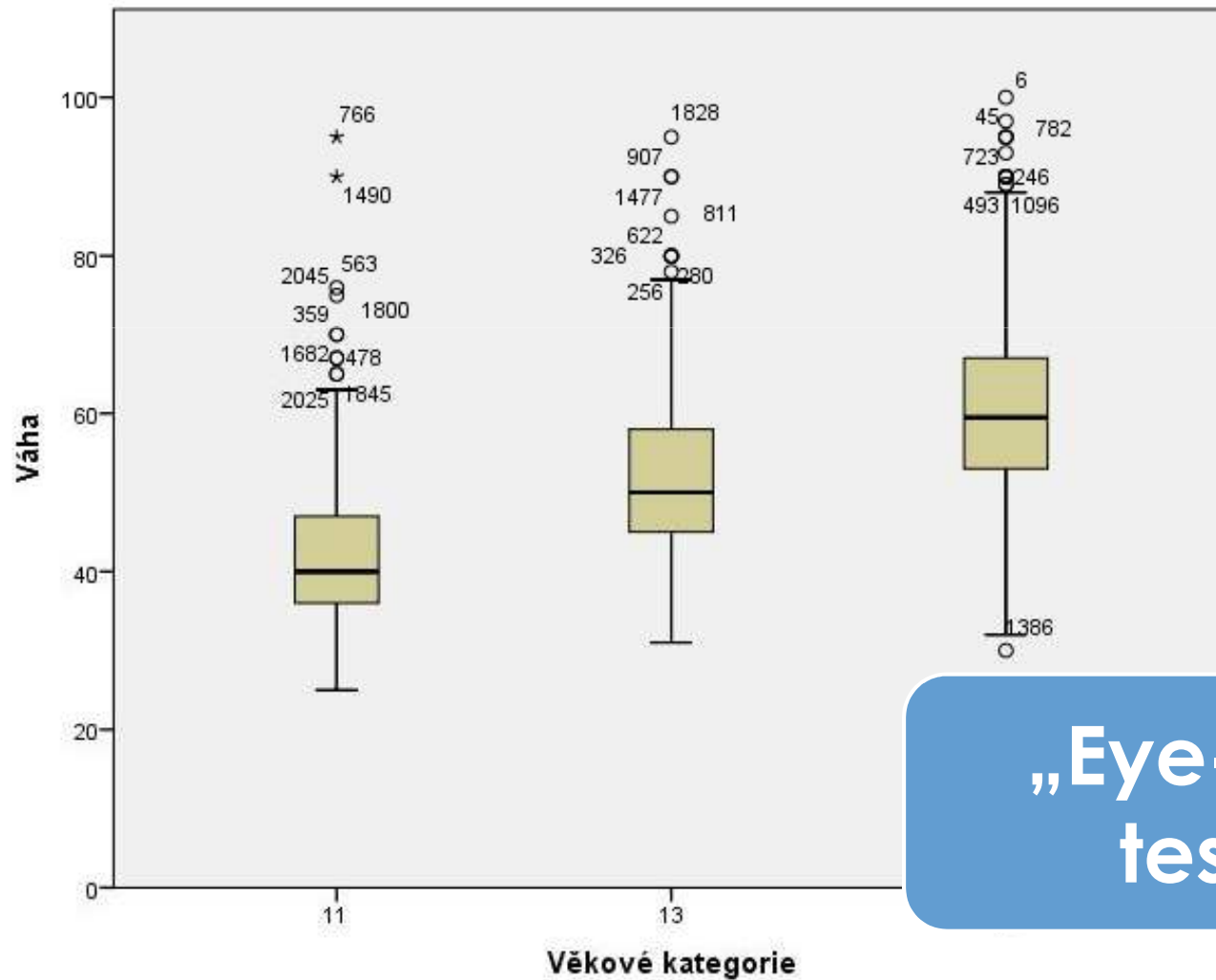


Boxplot - vysvetlenie

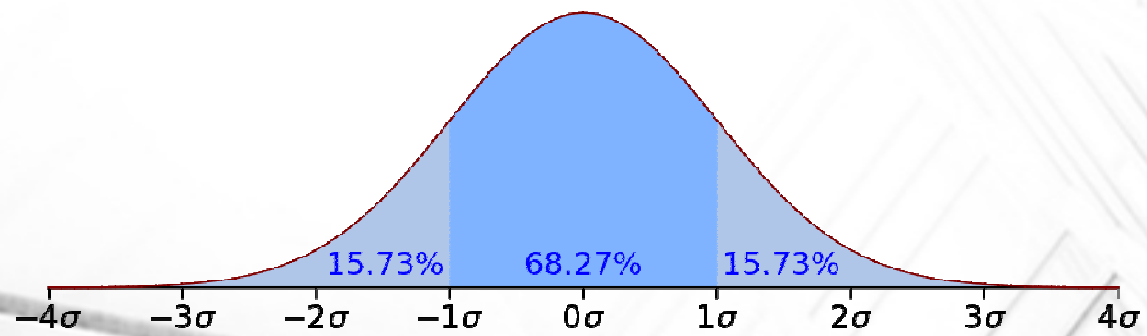
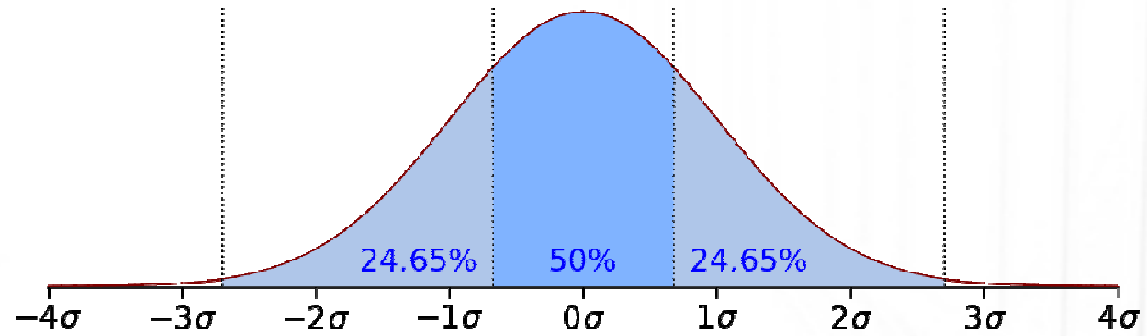
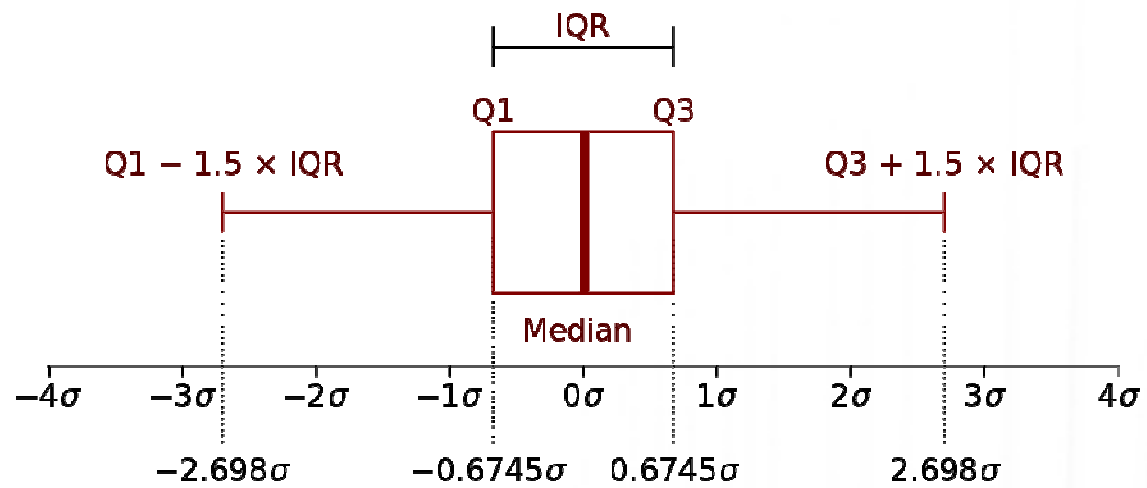
**Extrémne
hodnoty
nad +/- 3 násobok
IQR**



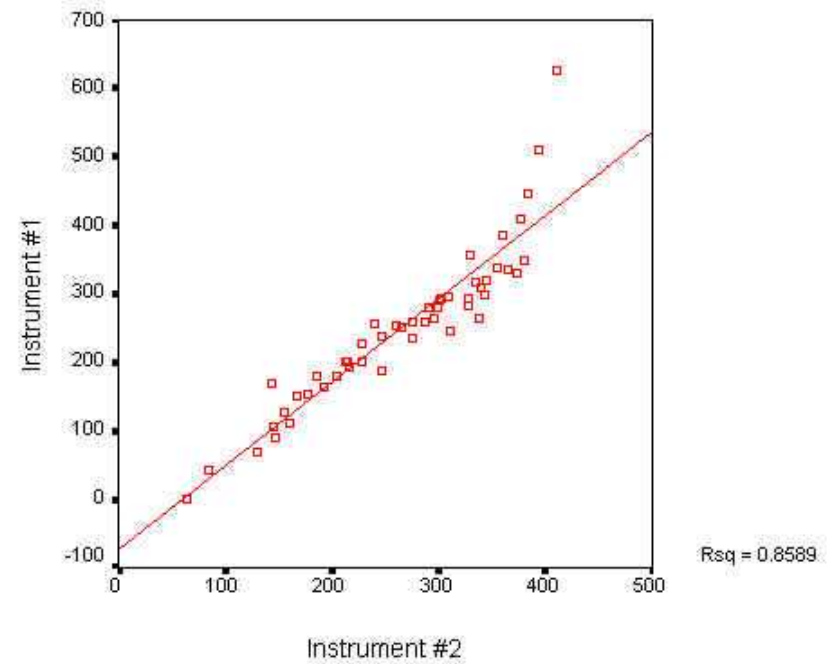
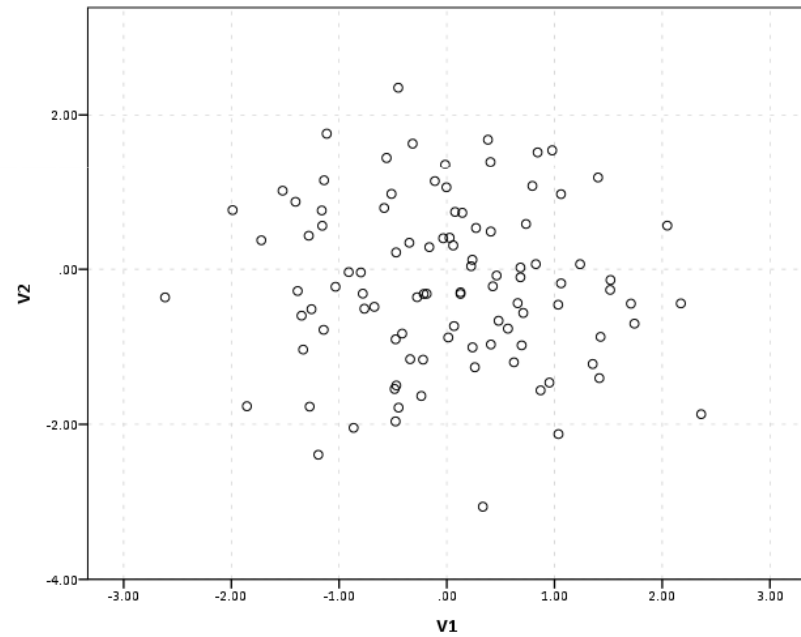
Združený boxplot



„Eye-ball test“



Scatter plot





Inferenčná štatistická analýza



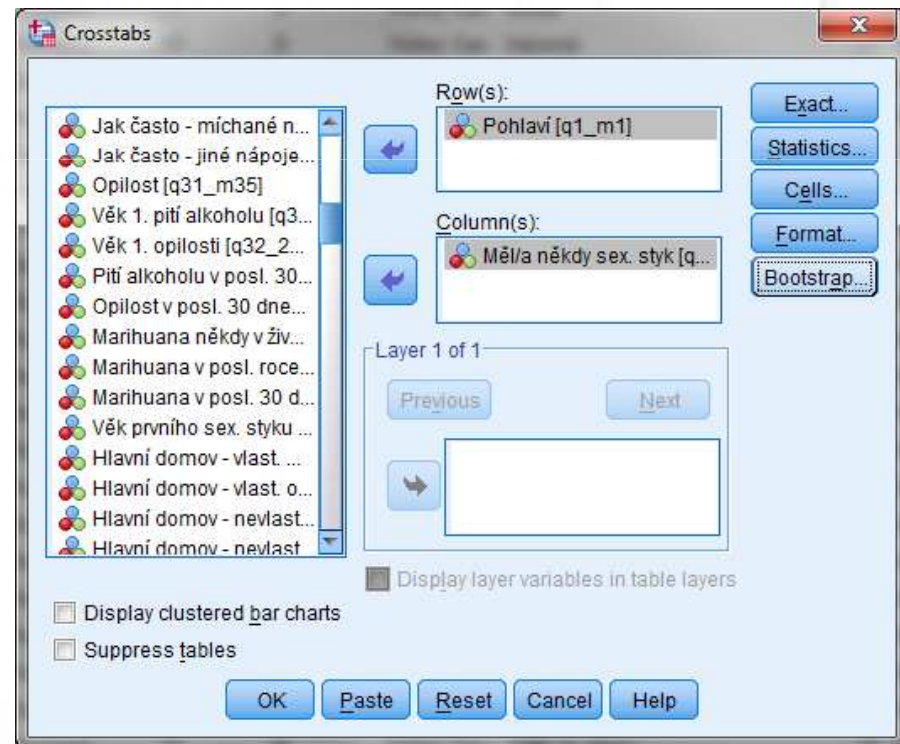
Porovnávanie 2 a viacerých súborov

Rozdiely početností (nominálna úroveň)

- **Pearsonov Chi kvadrát** – porovnáva početnosti v kategóriách a skupinách a testuje nulovú hypotézu o rovnosti početností.

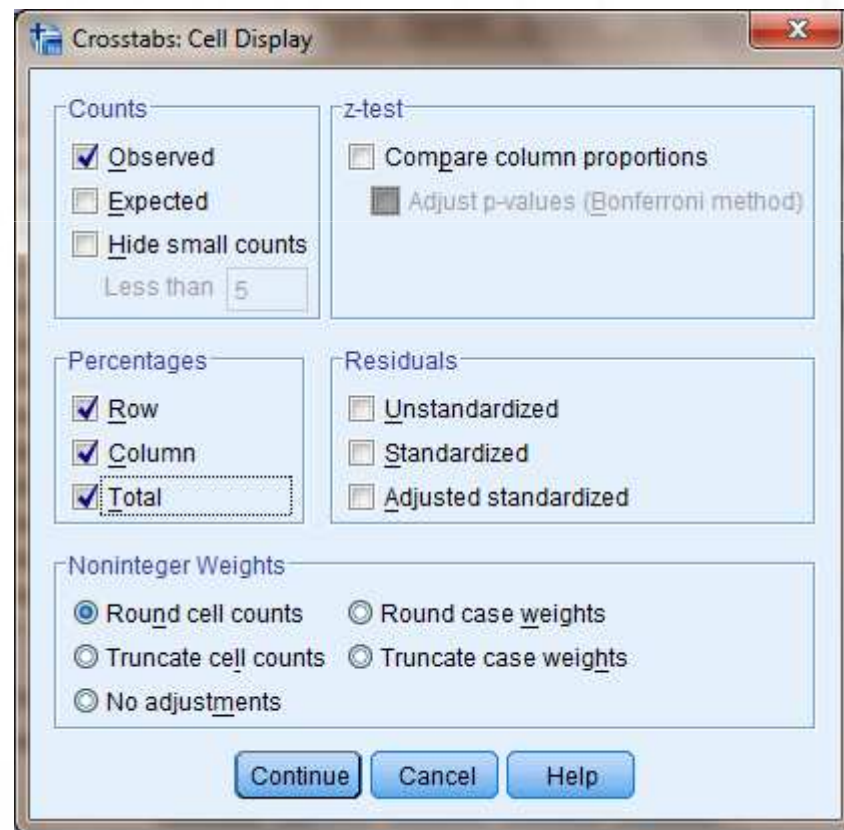
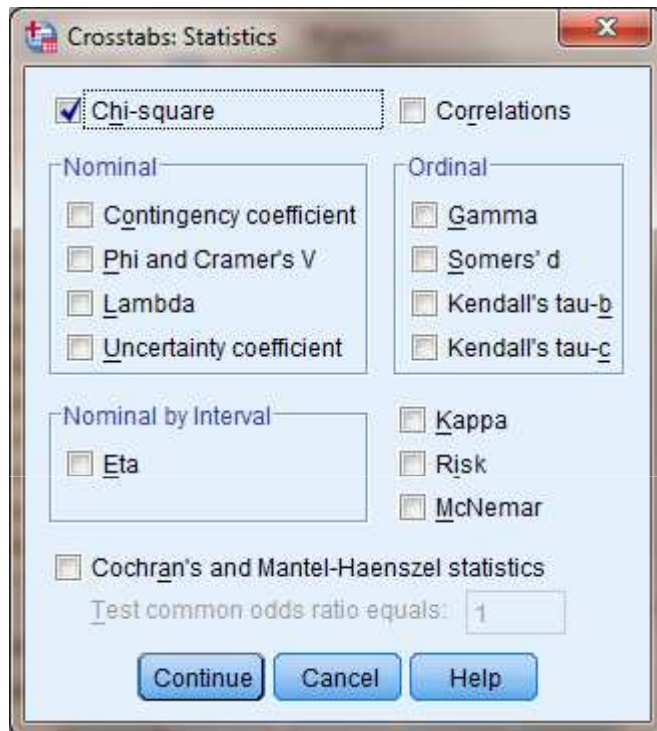
Výskumná otázka:

Majú viac sexuálnych skúseností dievčatá alebo chlapci?



Ako na to v SPSS:

Analyze -> Descriptive statistics-> Crosstabs: Statistics Chi square - OK



- Pearsonov Chi kvadrát

Pohlaví * Měl/a někdy sex. styk Crosstabulation

			Měl/a někdy sex. styk		Total
			ano	ne	
Pohlaví	chlapec	Count	54	181	235
		% within Pohlaví	23,0%	77,0%	100,0%
		% within Měl/a někdy sex. styk	40,0%	42,9%	42,2%
		% of Total	9,7%	32,5%	42,2%
dívka		Count	81	241	322
		% within Pohlaví	25,2%	74,8%	100,0%
		% within Měl/a někdy sex. styk	60,0%	57,1%	57,8%
		% of Total	14,5%	43,3%	57,8%
Total		Count	135	422	557
		% within Pohlaví	24,2%	75,8%	100,0%
		% within Měl/a někdy sex. styk	100,0%	100,0%	100,0%
		% of Total	24,2%	75,8%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	,350 ^a	1	,554
Continuity Correction ^b	,242	1	,623
Likelihood Ratio	,352	1	,553
Fisher's Exact Test			
Linear-by-Linear Association	,350	1	,554
N of Valid Cases	557		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 1,00.

b. Computed only for a 2x2 table

Záver: Chlapci a dievčatá sa nelíšia v sexuálnych skúsenostiach



WWW.GUINNESS.DE



Studentov t-test (kardinálna úroveň)

- Porovnáva priemery premenných v skupine A a B.
- Zohľadňuje distribúciu dát a veľkosť skupín
- Vhodný na súbory nad 50 respondentov



Studentov t-test (kardinálna úroveň)

- 1908 – Guinness brewery (William Sealy Gosset)
- Lacný spôsob monitorovania kvality piva
- t-test pre 2 nezávislé výbery, závislé výbery, ...
- Porovnáva vždy len 2 súbory
- *Parametrický test*

Parametre:

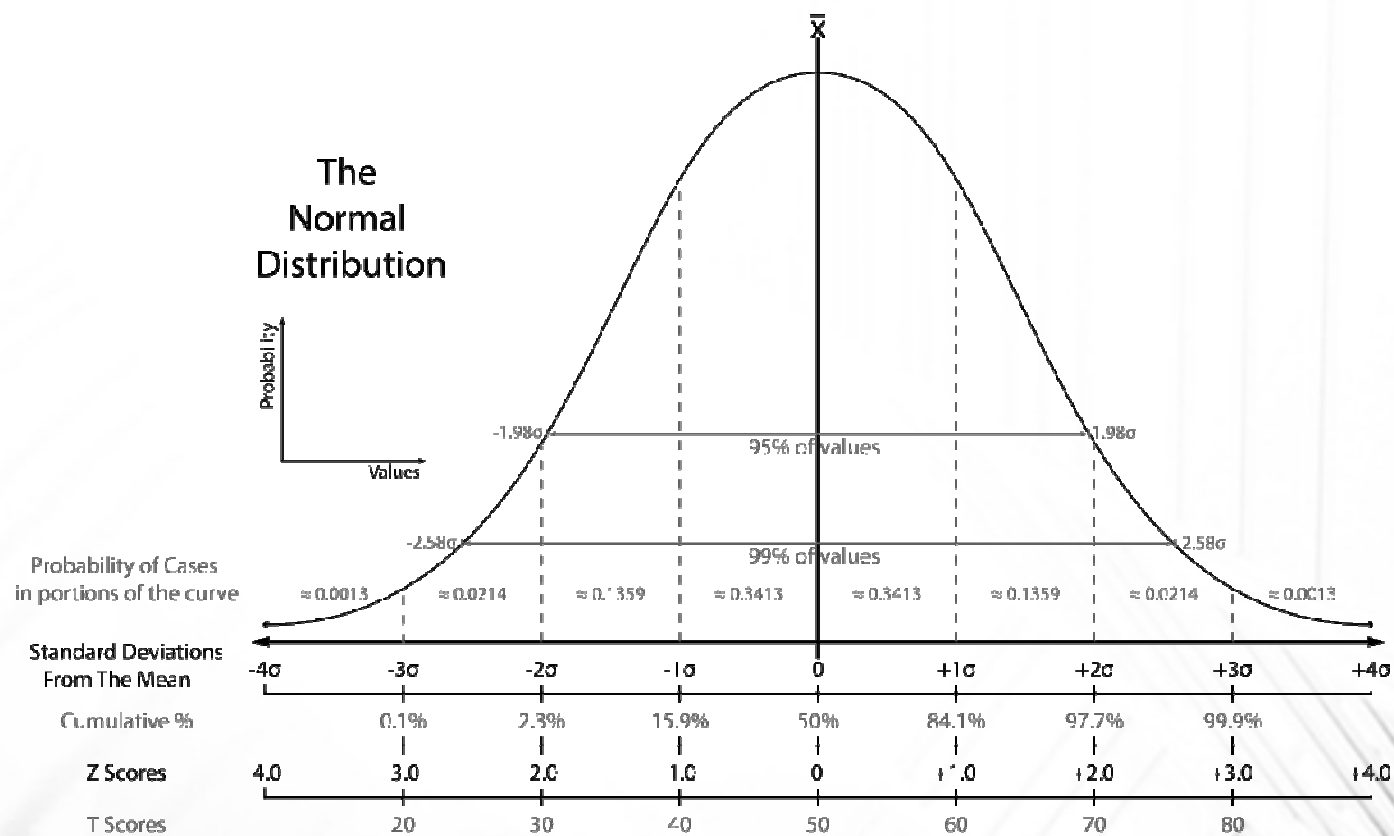
- **Normalita distribúcie** (Gaussová krivka)
- **Homoscedasticita** – homogenita, rovnosť rozptylov

Ako na to v SPSS:

Analyze -> Compare Means -> Independent sample t-test: Variables,
Define groups - OK

Studentov t-test

- Normalita distribúcie



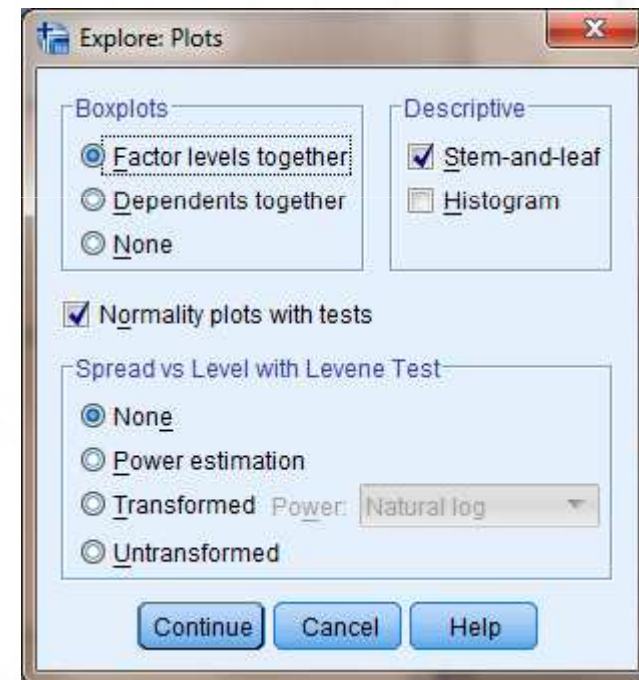
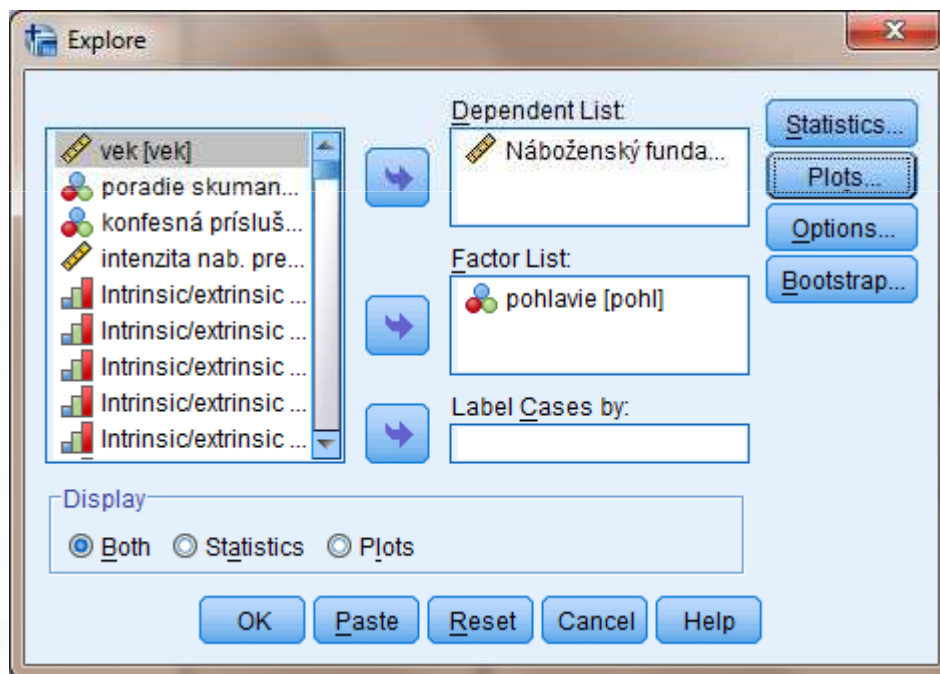
Studentov t-test

- Overujeme ju:
 - Kolmogorov-Smirnov test – neparametrický test
 - Shapiro-Wilk test
 - H_0 – populácia je normálne rozložená
- Ak zamietneme H_0 , nemali by sme pokračovať v t-teste

Ako na to v SPSS:

Analyze -> Descriptive stat. -> Explore – Plots: Normality plots with tests -
OK

Overenie normality distribúcie



Ako na to v SPSS:

Analyze -> Descriptive stat. -> Explore – Plots: ✓ Normality plots with tests - OK

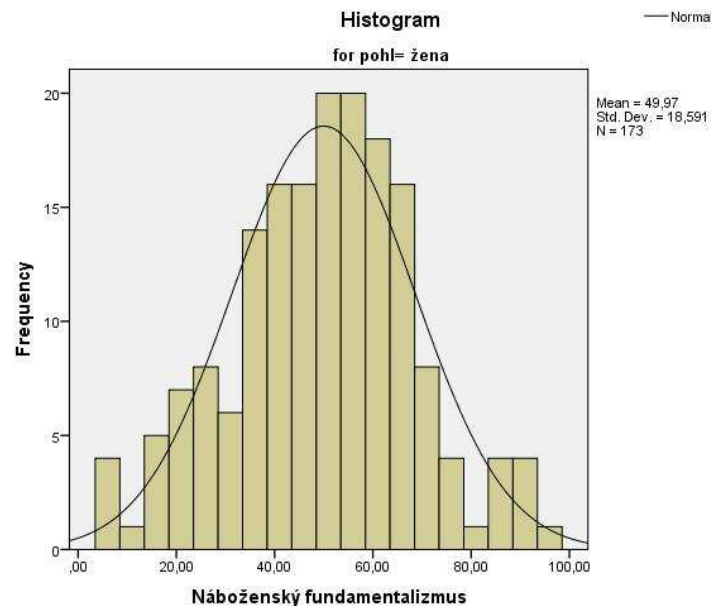
Overenie normality distribúcie

Tests of Normality

pohlavie	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Náboženský muž	,102	29	,200*	,942	29	,111
fundamentalizmus žena	,054	173	,200*	,988	173	,171

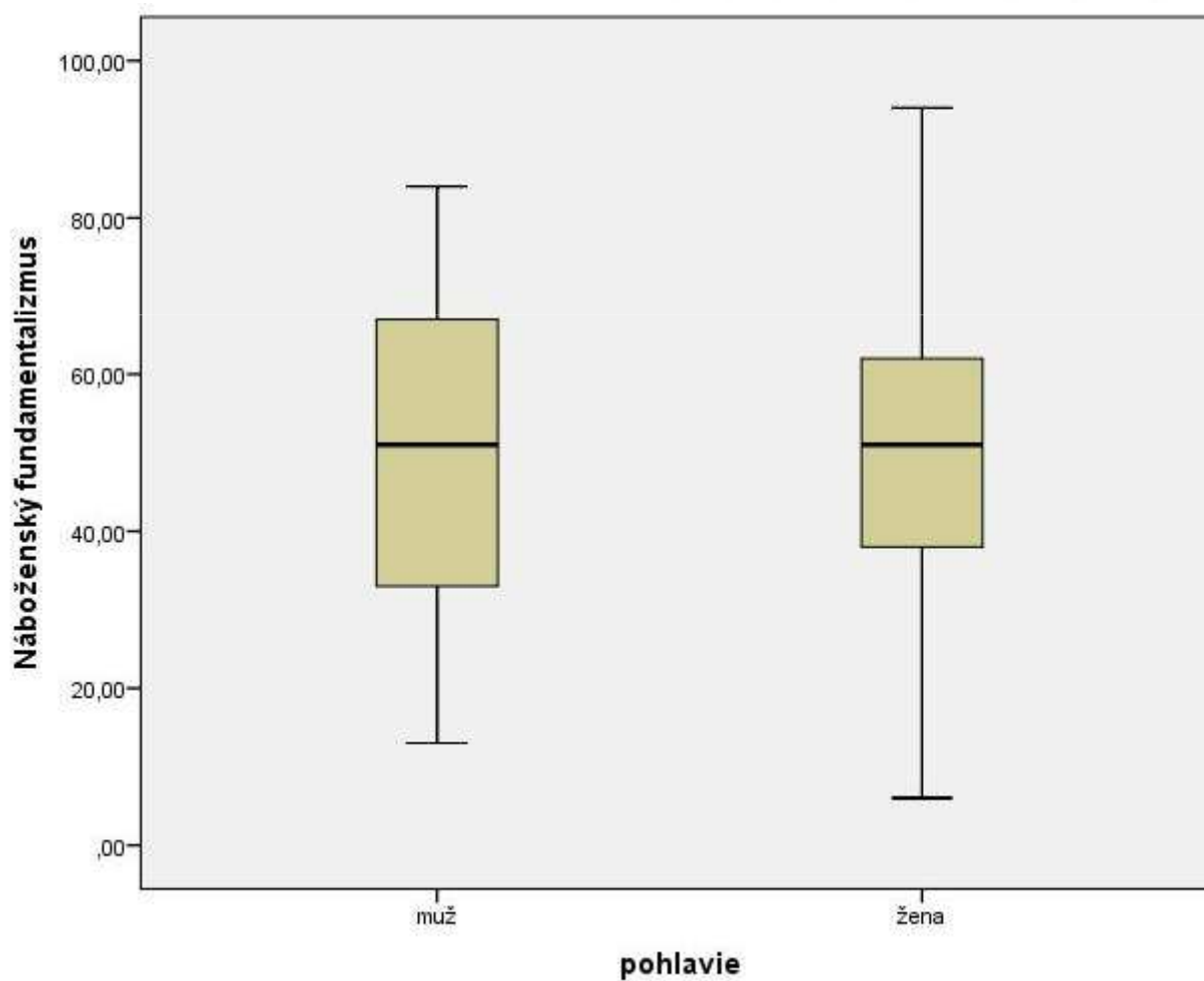
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



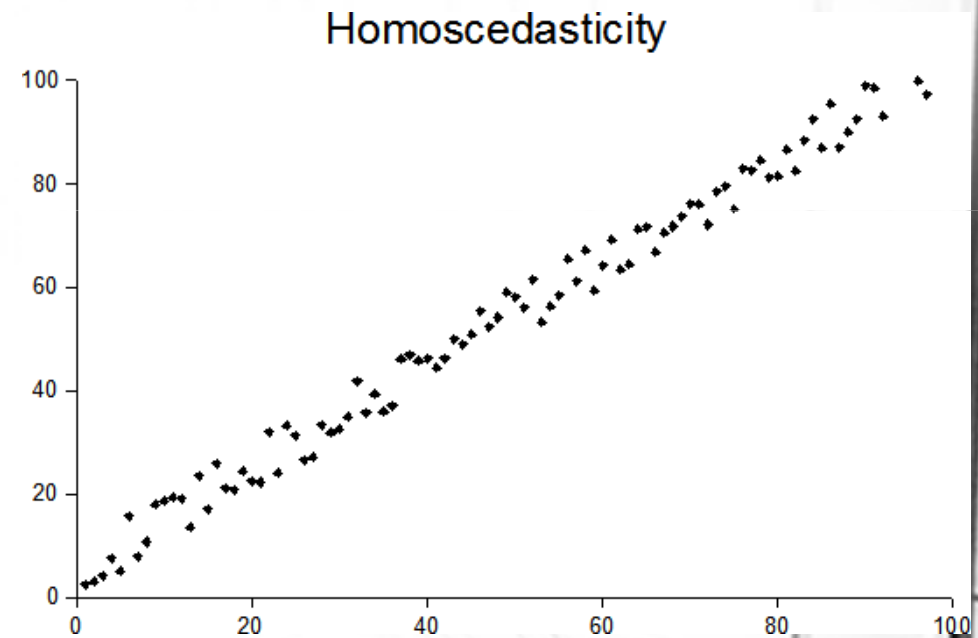
Záver: Prijímame H_0 o normálnej distribúcii hodnôt v populácii (aj u mužov aj u žien)

Overenie normality distribúcie

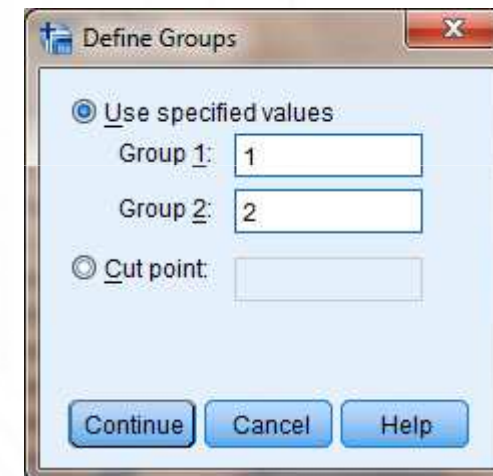
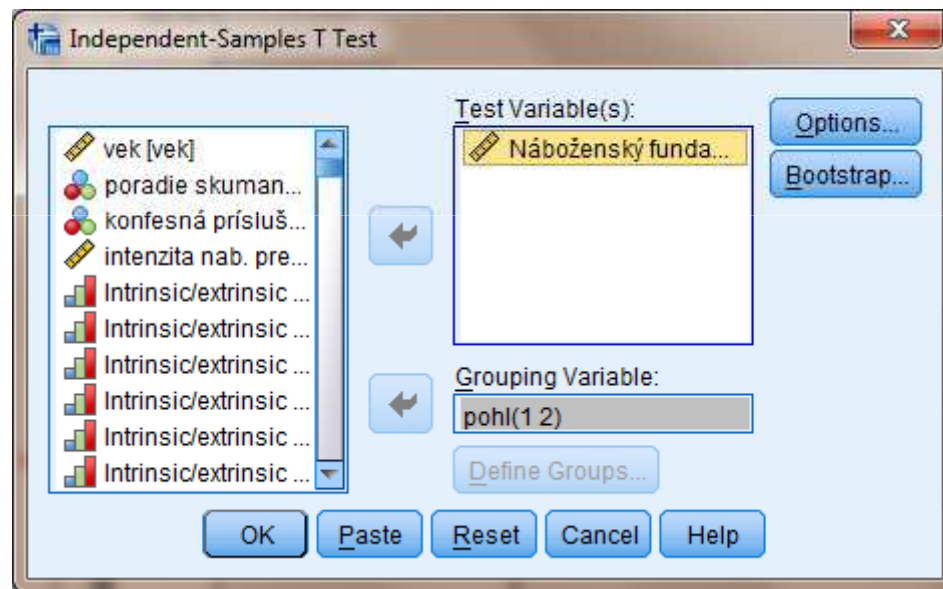


Studentov t-test

- Homoskedasticita
- Overujeme ju:
 - **Levenov test rovnosti rozptylov** (súčasť t-testu v SPSS)
 - H_0 – rozptyly
- t-test produkuje výsledky pre prípad prijatia aj zamietnutia H_0



Studentov t-test



Ako na to v SPSS:

Analyze -> Compare Means -> Independent sample t-test: Variables,
Define groups - OK

Studentov t-test

Group Statistics

	pohlavie	N	Mean	Std. Deviation	Std. Error Mean
Náboženský fundamentalizmus	muž	29	50,2759	21,89569	4,06593
	žena	173	49,9653	18,59073	1,41343

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Náboženský fundamentalizmus	Equal variances assumed	1,746	,188	,081	200	,935	,31054	3,83011	-7,2420	7,86313
	Equal variances not assumed			,072	35,09	,943	,31054	4,30460	-8,4274	9,04851

Záver: Muži a ženy sa nelíšia v úrovni náboženského fundamentalizmu.

Studentov t-test

- V prípade nenormálne distribuovaných dát môžeme rozdiely medzi skupinami testovať neparametrickou alternatívou – U-test

Rozdiely medzi skupinami (ordinálna úroveň)

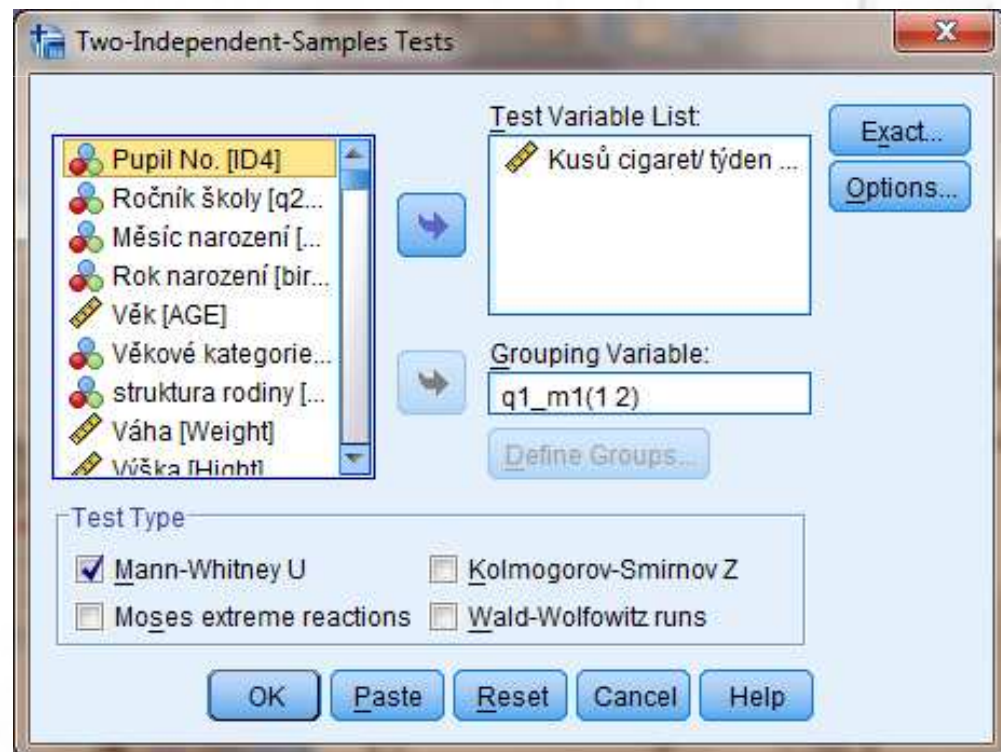
- Mann-Whitneyho U-test

Ako na to v SPSS:

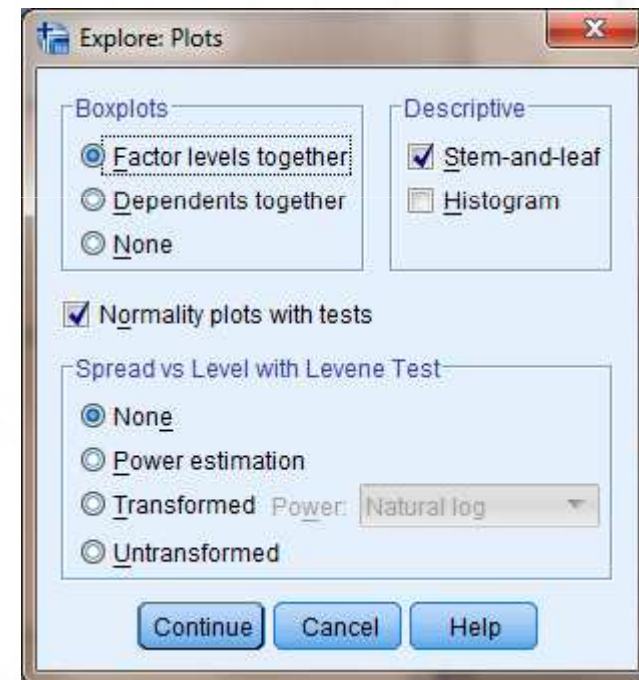
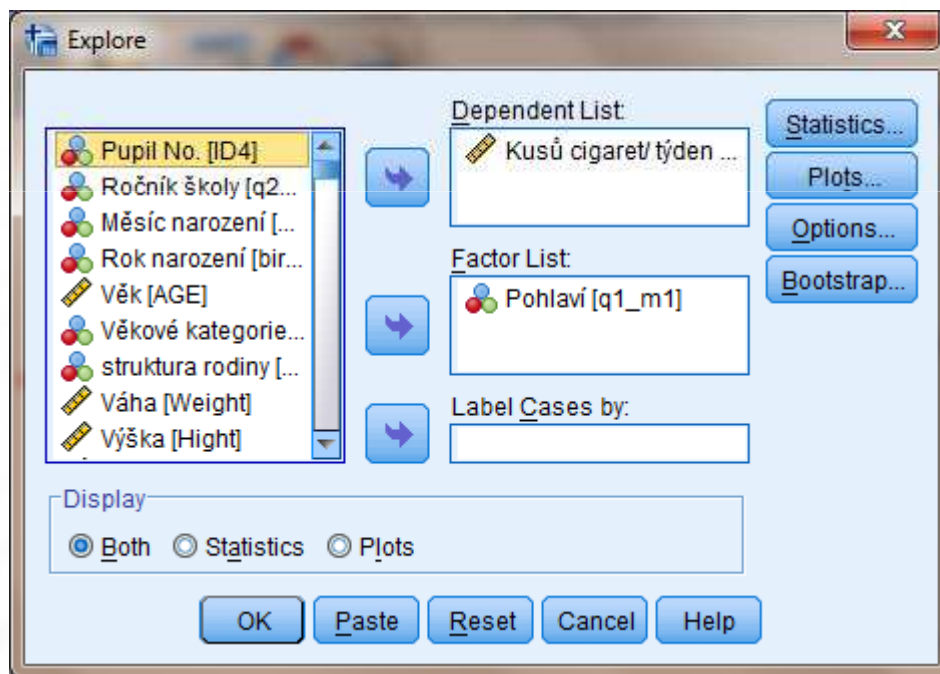
Analyze -> Nonparametric tests -> Legacy dialogs -> 2 independent samples ✓
Mann-Whitne U - OK

Výskumná otázka:

Líšia sa dievčatá a chlapci v počte vyfajčených cigariet za týždeň?



Overenie normality distribúcie



Ako na to v SPSS:

Analyze -> Descriptive stat. -> Explore - Plots: Normality plots with tests - OK

Overenie normality distribúcie

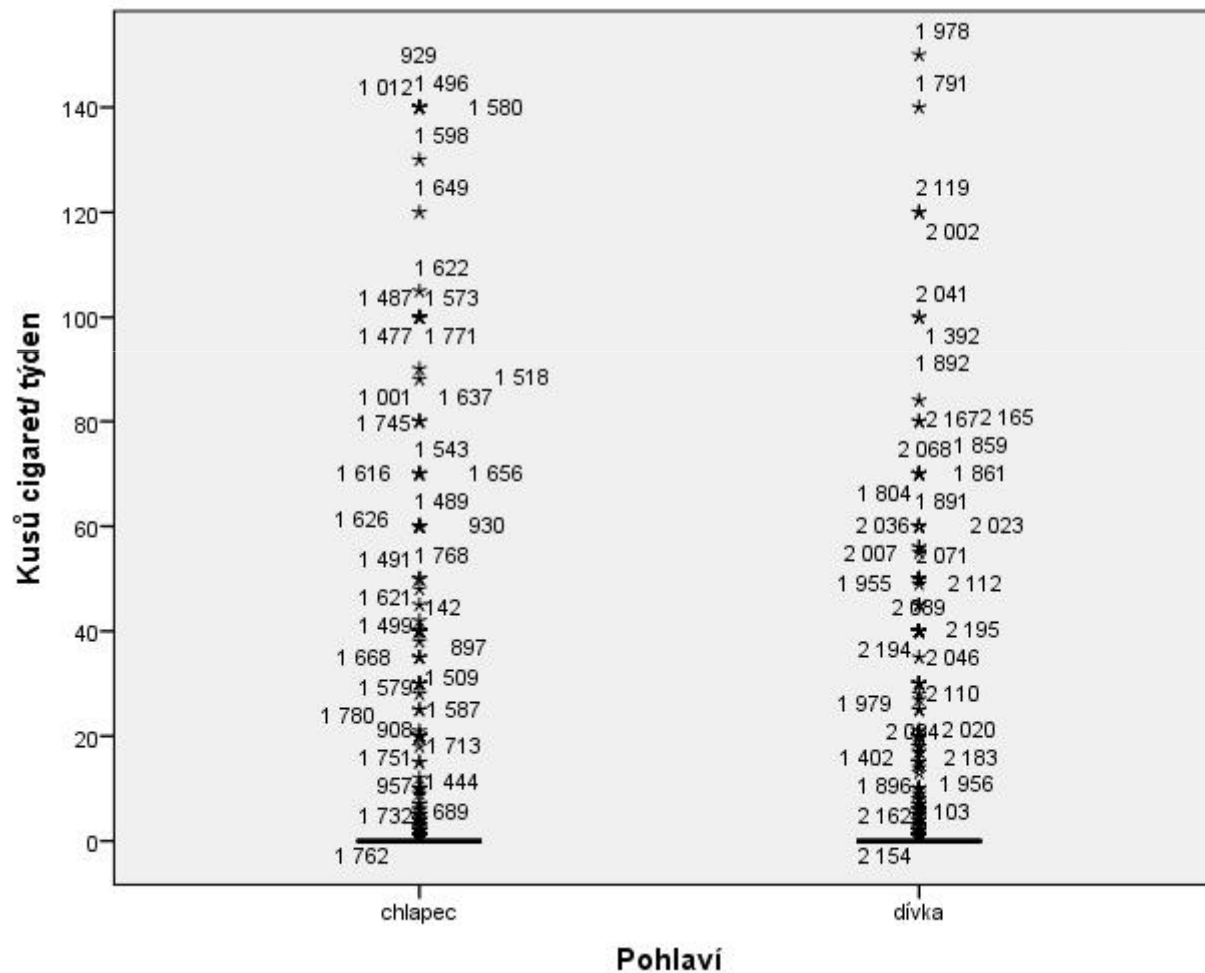
Tests of Normality

	Pohlaví	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Kusů cigaret/ týden	chlapec	,454	982	,000	,274	982	,000
	dívka	,434	1081	,000	,302	1081	,000

a. Lilliefors Significance Correction

Záver: Zamietame H_0 o normálnej distribúcii hodnôt v populácii

Overenie normality distribúcie



Rozdiely (ordinálna úroveň)

Ranks

	Pohlaví	N	Mean Rank	Sum of Ranks
Kusů cigaret/ týden	chlapec	982	1020,66	1002289,00
	dívka	1081	1042,30	1126727,00
	Total	2063		

Test Statistics^a

	Kusů cigaret/ týden
Mann-Whitney U	519636,000
Wilcoxon W	1002289,000
Z	-1,280
Asymp. Sig. (2-tailed)	,200

a. Grouping Variable: Pohlaví

Záver: Nie je rozdiel v počte vyfajčených cigariet medzi chlapcami a dievčatami.

ANOVA

- Jednoduchá analýza rozptylu (analysis of variance)
- Umožňuje testovať rozdiely u 3 a viac populácií naraz
- Post-Hoc testy – párové porovnania
- Neparametrická alternatíva:
 - Kruskal-Wallis test
 - Neobsahuje Post-Hoc testy
 - Párové porovnanie je potrebné urobiť U-testom pre každú dvojicu

Ako na to v SPSS:

Analyze -> Compare Means -> One-Way ANOVA: Dependent List, Factor- OK

Analyze -> Nonparametric tests -> Legacy dialogs -> K independent samples ✓
Kruskal-Wallis H - OK



Analýza vzťahov medzi premennými

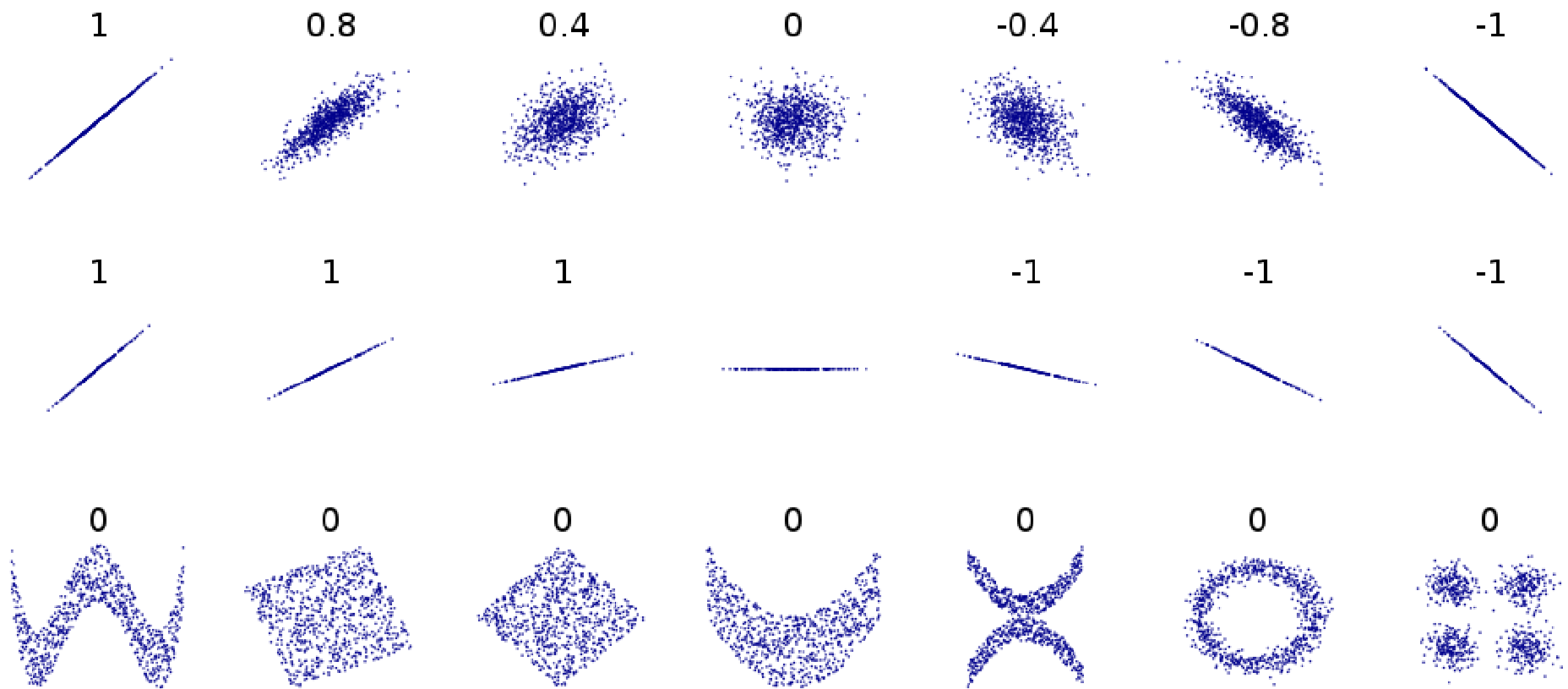
Korelácie

- Pearsonova korelácia (koeficient r)
 - r nadobúda hodnoty od -1 po +1
 - -1 negatívna asociácia (nepriama úmera)
 - +1 pozitívna asociácia (priama úmera)
 - kardinálna vs. kardinálna premenná (napr. hrubé skóre)
- Interpretácia koeficientu
 - Úrovne sú stanovené arbitrárne, nemali by byť chápané veľmi striktné
 - Interpretácia závisí od kontextu a zámeru (vo fyzike je mála aj $r=0.9$)
 - Nemožno hovoriť o kauzalite

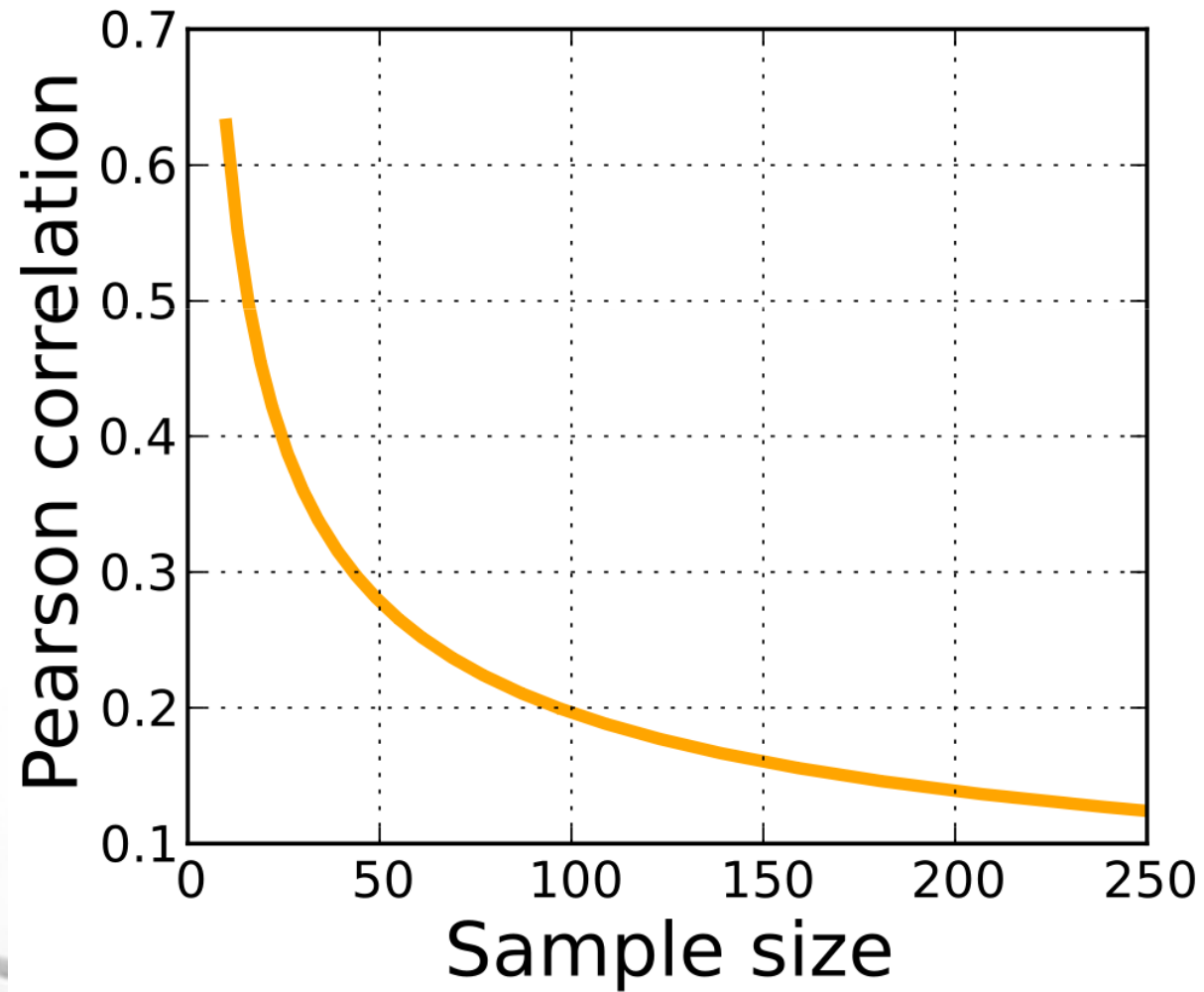
Sila vzťahu/korelácie	Negatívne hodnoty	Pozitívne hodnoty
Žiaden vzťah	-0.09 do 0.0	0.0 do 0.09
Malá	-0.3 do -0.1	0.1 do 0.3
Stredná	-0.5 do -0.3	0.3 do 0.5
silná	-1.0 do -0.5	0.5 do 1.0

Korelácie

- Vzťah premenných v grafickej a číselnej podobe (r koeficient)



Korelácie



Korelácie

- Spearmanove Rho (ρ)
 - poradová korelácia
 - **Rho** nadobúda hodnoty od -1 po +1
 - ordinálna vs. ordinálna premenná
 - ordinálna vs. kardinálna premenná

Ako na to v SPSS:

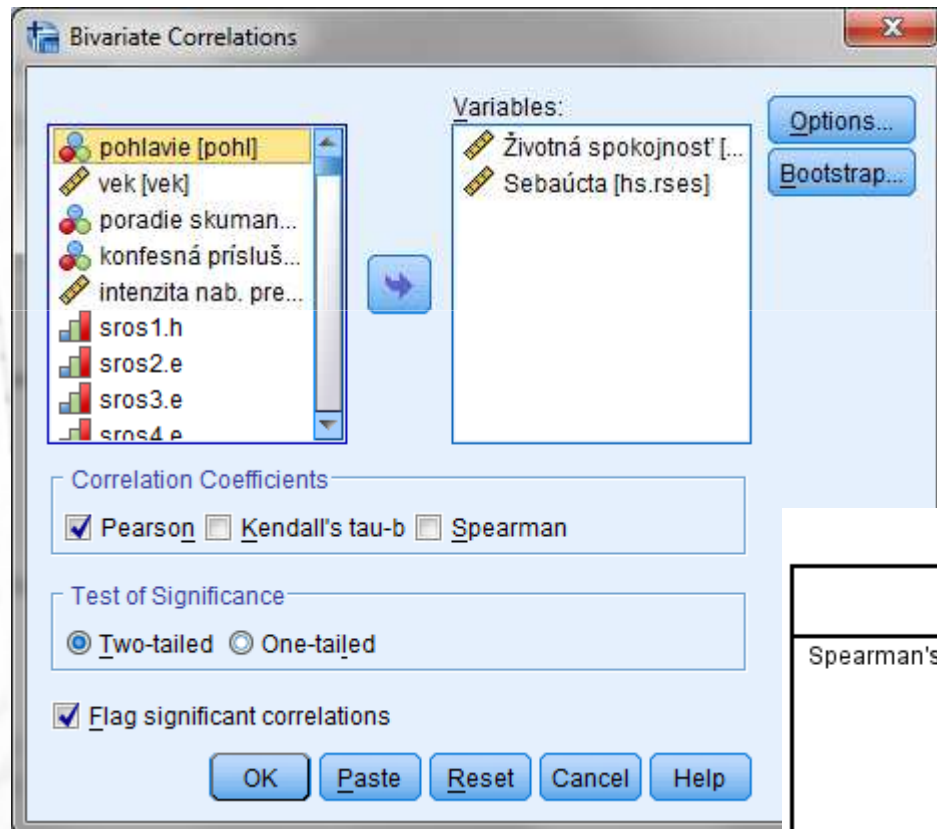
Analyze -> Correlate -> Bivariate: Variables, Pearson/Spearman- OK

- Parciálne korelácie
 - Umožňujú „očistiť“ koreláciu 2 premenných o podiel interferujúcej premennej

Korelácie

Výskumná otázka:

Existuje súvislosť medzi životnou spokojnosťou a sebaúctou respondenta?



Correlations

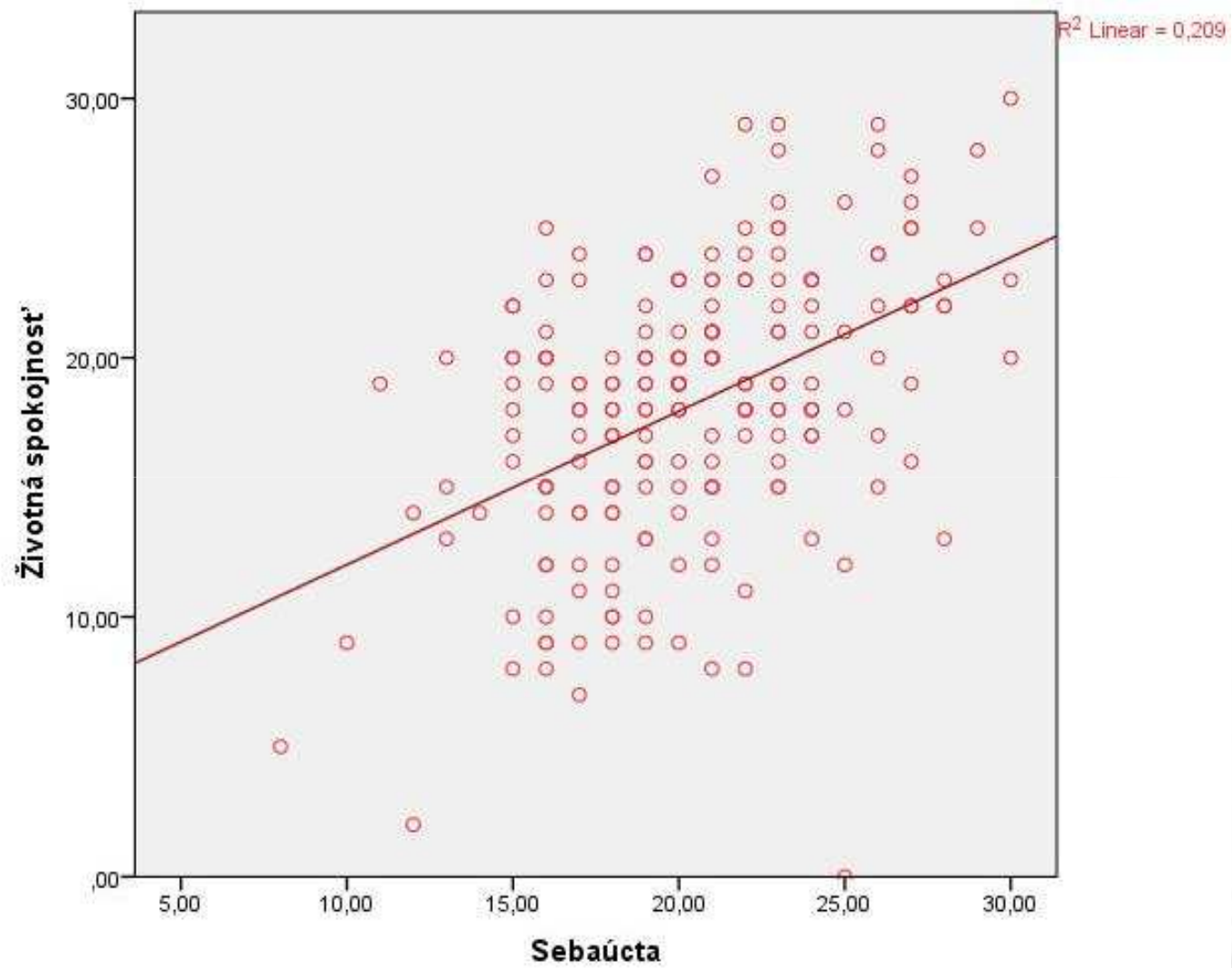
		Životná spokojnosť	Sebaúcta
Životná spokojnosť	Pearson Correlation	1	,458**
	Sig. (2-tailed)		,000
	N	204	202
Sebaúcta	Pearson Correlation	,458**	1
	Sig. (2-tailed)	,000	
	N	202	202

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations

		Životná spokojnosť	Sebaúcta
Spearman's rho	Životná spokojnosť	Correlation Coefficient	1,000
		Sig. (2-tailed)	,000
		N	204
Sebaúcta		Correlation Coefficient	,437**
		Sig. (2-tailed)	,000
		N	202

** . Correlation is significant at the 0.01 level (2-tailed).





„Ja verím len štatistikám, ktoré som
si sám sfalšoval.“

Winston S. Churchill



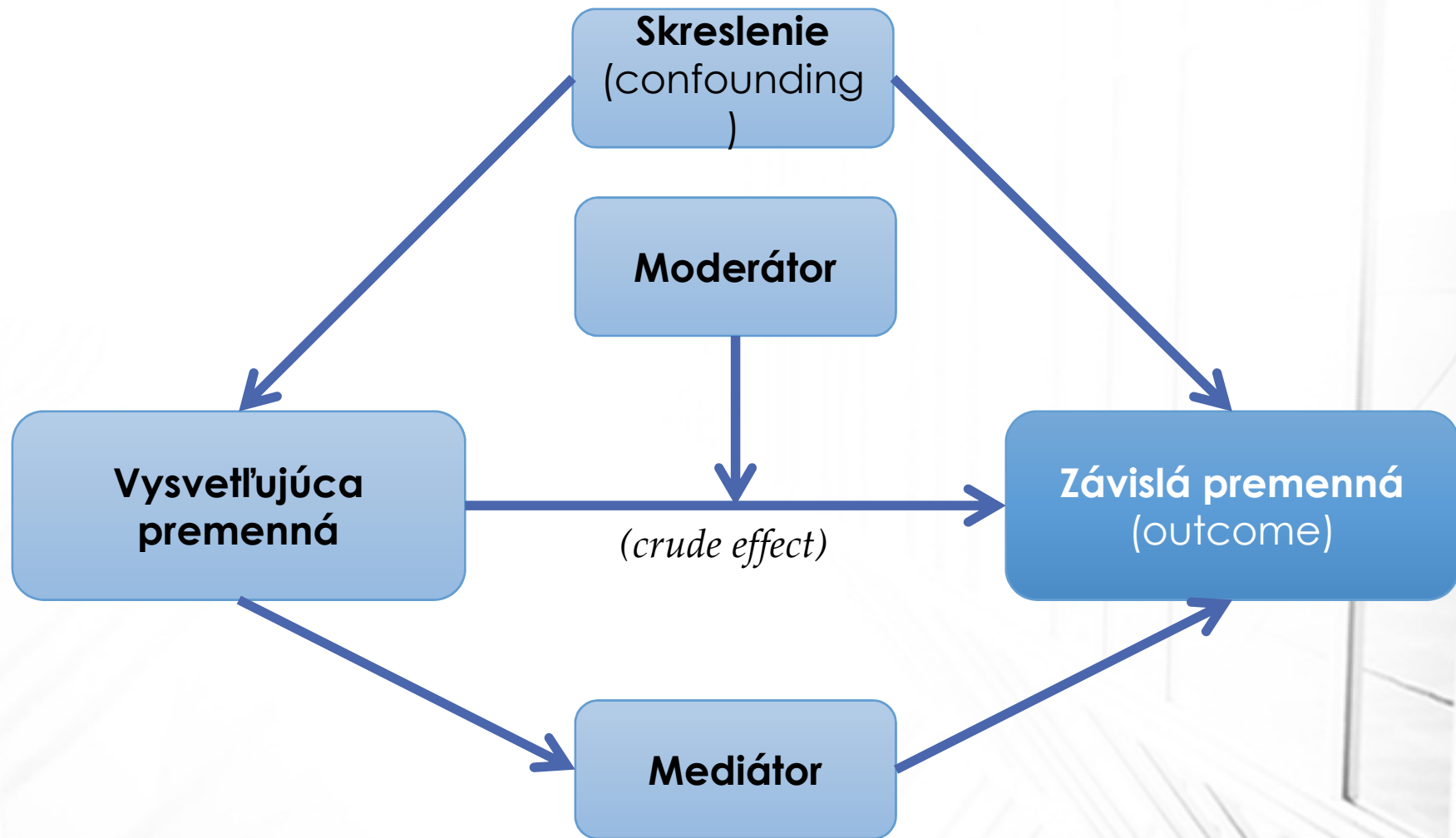
Ďakujem za pozornosť

Regresná analýza

- Umožňuje odhadnúť vzťahy medzi premennými
- Techniky modelovania
- Jedna závislá, jedna alebo viac nezávislých (vysvetľujúcich) premenných
- Umožňuje predikciu zmien závislej premennej v prípade, že dôjde k zmene v nezávislých premenných
- Za špecifických podmienok môžeme usudzovať o kauzálnom vzťahu (inak len experiment)

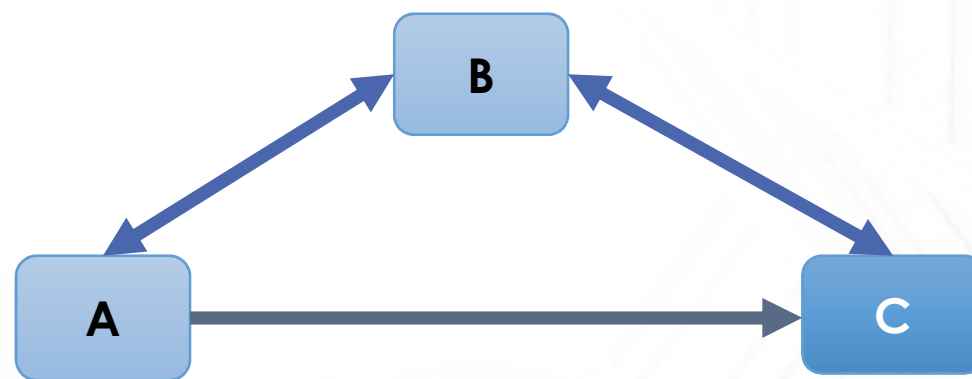
Regresná analýza

- Druhy:
 - **Lineárna**
 - (nezávislá premenná: kardinálna)
 - **Logistická**
 - (nezávislá premenná: dichotomická)
 - **Multinomiálna**
 - (nezávislá premenná: kategorická, 3 a viac kategórií)
- Pojmy:
 - Crude effect
 - Adjusted effect
 - Moderácia (interakcia)
 - Mediácia
 - Skreslenie (confounding)
 - Skreslenie (bias) – závisí od dizajnu štúdie a faktormi súvisiacimi so získavaním údajov



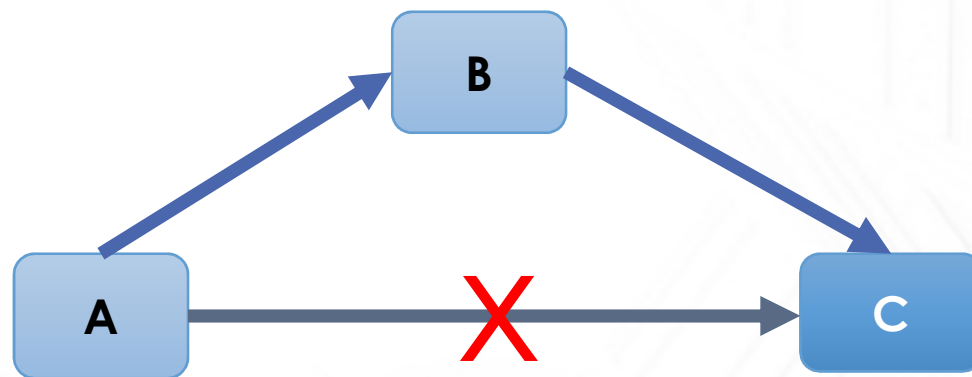
Skreslenie (confounding)

- 1.) prediktor (A) a confounder (B) musia byť prepojený so závislou premennou (C)
- 2.) prediktor a confounder musia byť prepojený (*interkorelácia*)
- 3.) skreslenie nesmie byť kauzálnym následkom prediktora



Mediácia

- Sprostredkujúci efekt
- Matematicky sa nelíši od confoundingu
- 1.) prediktor (A) a mediátor (B) musia byť prepojený so závislou premennou (C)
- 2.) prediktor a mediátor musia byť prepojený (*interkorelácia*)
- 3.) skreslenie **musí** byť kauzálnym následkom prediktora (A zapríčiňuje B)



Mediácia vs. Skreslenie

- Adjustácia na moderátor/confounder
 - zníženie alebo eliminácia efektu A na C (pokles OR, alebo aj strata signifikancie)
- Príklad mediácie:
 - Stres -> fajčenie -> rakovina
 - Obezita -> cholesterol -> vysoký krvný tlak
 - Etnicita (Róm) -> nižšie vzdelanie -> horšie zdravie
- Príklad skreslenia:
 - Pitie kávy -> infarkt myokardu (conf. fajčenie)

Moderácia

- Ak vzťah dvoch premenných závisí od úrovne tretej premennej
- Úroveň moderátora ovplyvňuje mení vzťah 2 premenných
- Častý moderátor:
 - Pohlavie, etnicita, rasa,....

Napr. : Aspirín -> zníženie rizika infarktu (ale len u mužov)

Moderácia

- Zisťuje sa inakciou dvoch prediktorov (napr. etnicita a pohlavie)
- Do analýzy zadáme:
 - Prediktor 1
 - Prediktor 2
 - Prediktor 1 * Prediktor 2
- Ak je interakcia signifikantná je vhodné stratifikovať analýzy (osobitne analyzovať mužov a osobitne ženy)

Binárna logistická regresia

- Závislou premennou môže byť len dichotomická premenná (0/1)
- Diagnóza (napr. má depresiu=1, nemá depresiu=0)
- Pozor na kódovanie! (riziko opačných výsledkov)
- Vysvetľujúce premenné môžu byť kategorickej alebo kontinuálnej povahy (potrebné zadať)
- Pri definovaní kategorickej povahy si môžeme vybrať referenčnú kategóriu (na rozdiel od lineárnej regresie, nemusíme už vytvárať „dummy variables“ – deje sa to automaticky)

Ako na to v SPSS:

Analyze -> Regression -> Binary Logistic....



Štatistika je ako bikiny – to čo odhaľuje je síce uchvacujúce, ale to čo zakrýva je rozhodujúce.

Aaron Levenstein

Table 3

Associations of ethnicity with health outcomes: odds ratios (OR) or beta coefficients (B), and the significance of model change (smc) after inclusion of additional variables.

	SRH bad/fair (N = 918)		Injuries (N = 921)		Healthcare use (N = 922)		Health complaints (N = 790)		Difficulties (N = 837)		Prosocial scale (N = 899)	
	OR (95% CI)	smc	OR (95% CI)	smc	OR (95% CI)	smc	B (95% CI)	smc	OR (95% CI)	smc	OR (95% CI)	smc
Model 1												
Ethnicity (Roma vs. non-Roma)	3.01 (2.25, 4.01)***	***	1.57 (1.14, 2.17)**	**	2.77 (2.09, 3.66)***	***	-1.01 (-1.44, -0.58)***	***	1.04 (0.74, 1.45)	ns	0.19 (0.10, 0.35)***	***
Model 2												
Ethnicity (Roma vs. non-Roma)	3.00 (2.25, 4.01)***	**	1.67 (1.20, 2.32)**	***	2.76 (2.09, 3.65)***	ns	-0.98 (-1.40, -0.55)***	***	1.04 (0.74, 1.45)		0.19 (0.10, 0.35)***	***
Gender (males vs. females)	0.65 (0.49, 0.87)**		2.80 (1.99, 3.93)***		0.91 (0.69, 1.19)		0.85 (0.43, 1.26)**	***	0.79 (0.57, 1.09)		2.40 (1.56, 3.69)***	
Model 3												
Ethnicity (Roma vs. non-Roma)	1.96 (1.25, 3.06)**	ns	1.42 (0.84, 2.41)	ns	1.86 (1.21, 2.87)**	*	-0.76 (-1.42, -0.10)***	ns	0.97 (0.56, 1.68)	ns	0.24 (0.10, 0.57)***	ns
Gender (males vs. females)	0.65 (0.49, 0.87)**		2.81 (2.00, 3.95)***		0.91 (0.69, 1.19)		0.83 (0.42, 1.25)**	***	0.80 (0.58, 1.11)		2.40 (1.56, 3.70)***	
Parents' highest education ^a												
apprenticeship	0.89 (0.57, 1.38)		0.81 (0.49, 1.35)		1.23 (0.79, 1.92)		0.50 (-0.17, 1.16)		0.82 (0.48, 1.38)		1.05 (0.36, 3.09)	
secondary	0.55 (0.32, 0.97)*		0.79 (0.42, 1.50)		0.66 (0.39, 1.13)		0.54 (-0.28, 1.36)		0.71 (0.37, 1.40)		1.38 (0.44, 4.34)	
university	0.53 (0.29, 0.96)*		0.70 (0.35, 1.39)		0.63 (0.35, 1.11)		0.44 (-0.44, 1.31)		0.98 (0.49, 1.96)		1.40 (0.44, 4.50)	
Model 4												
Ethnicity (Roma vs. non-Roma)	2.42 (1.50, 3.90)***	**	1.58 (0.91, 2.75)	ns	1.65 (1.05, 2.61)*	ns	-1.17 (-1.86, -0.48)***	***	1.33 (0.75, 2.39)	***	0.40 (0.16, 0.99)*	***
Gender (males vs. females)	0.64 (0.48, 0.85)**		2.78 (1.97, 3.91)***		0.92 (0.70, 1.20)		0.88 (0.46, 1.29)***	***	0.77 (0.55, 1.07)		2.36 (1.52, 3.65)***	
Parents' highest education ^a												
apprenticeship	0.88 (0.56, 1.37)		0.80 (0.48, 1.34)		1.24 (0.79, 1.94)		0.54 (-0.12, 1.20)		0.79 (0.47, 1.35)		0.94 (0.31, 2.84)	
secondary	0.54 (0.31, 0.95)*		0.79 (0.41, 1.49)		0.67 (0.39, 1.14)		0.59 (-0.22, 1.41)		0.69 (0.35, 1.36)		1.27 (0.39, 4.15)	
university	0.53 (0.29, 0.96)*		0.70 (0.35, 1.39)		0.62 (0.35, 1.10)		0.43 (-0.44, 1.29)		0.98 (0.49, 1.98)		1.32 (0.40, 4.41)	
Social desirability	0.84 (0.74, 0.96)**		0.91 (0.79, 1.05)		1.10 (0.98, 1.24)		0.33 (0.15, 0.50)***		0.76 (0.65, 0.88)***		0.58 (0.46, 0.74)***	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.smc – Significance of model change for the added variable; Improvement of fit of the model due to the addition of the variable concerned (χ^2 -test in the case of logistic regression and the F -test in the case of linear regression).^a Elementary education was set as the reference category.



Ďakujem za pozornosť