

SODEZZ
Sociální determinanty zdraví u sociálně a zdravotně znevýhodněných a jiných skupin populace
(CZ.1.07/2.3.00/20.0063)

Mgr. Peter Kolarčík, PhD

„Štatistické spracovanie údajov“



SODEZZ
Sociální determinanty zdraví u sociálně a zdravotně znevýhodněných a jiných skupin populace
(CZ.1.07/2.3.00/20.0063)

Mgr. Peter Kolarčík, PhD

„Štatistické spracovanie údajov“



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

VIPPA2 TVORBA
EXCELENTNÍHO
TÝMU
výzkum a inovace
v podpoře pohybové aktivity

Posílení odborného potenciálu výzkumných týmů v oblasti podpory pohybové aktivity na
Univerzitě Palackého - CZ.1.07/2.3.00/20.0171

Mgr. Peter Kolarčík, PhD

„Práca s SPSS“



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Štatistické spracovanie údajov

alebo

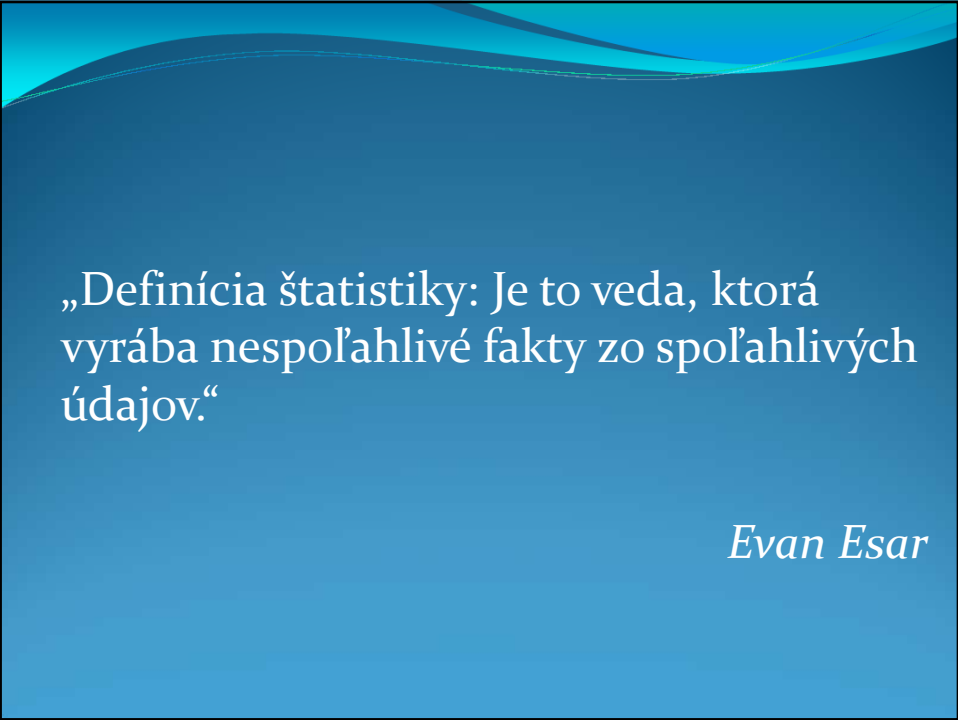
ako sa popasovať s číslami v SPSS

Olomouc 18.1.2013

Kurz: „Metodologie a komunikace
v kvantitativním výzkumu“

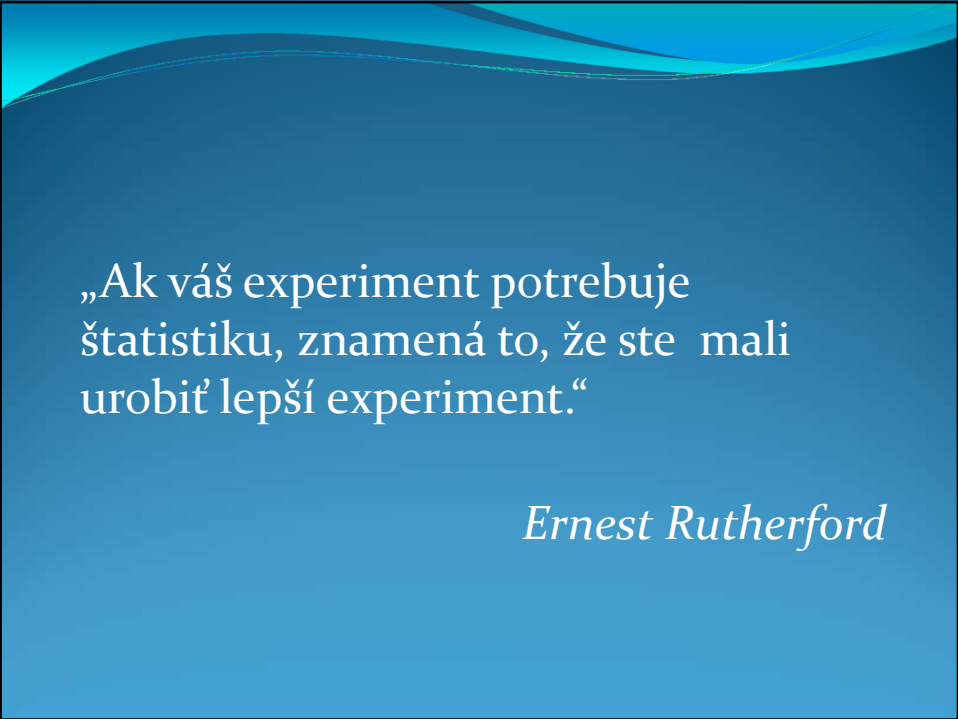
Mgr. Peter Kolarčík, PhD.

Oddelenie psychológie zdravia
Ústav verejného zdravotníctva
Lekárska Fakulta UPJŠ



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

Evan Esar



„Ak váš experiment potrebuje
štatistiku, znamená to, že ste mali
urobiť lepší experiment.“

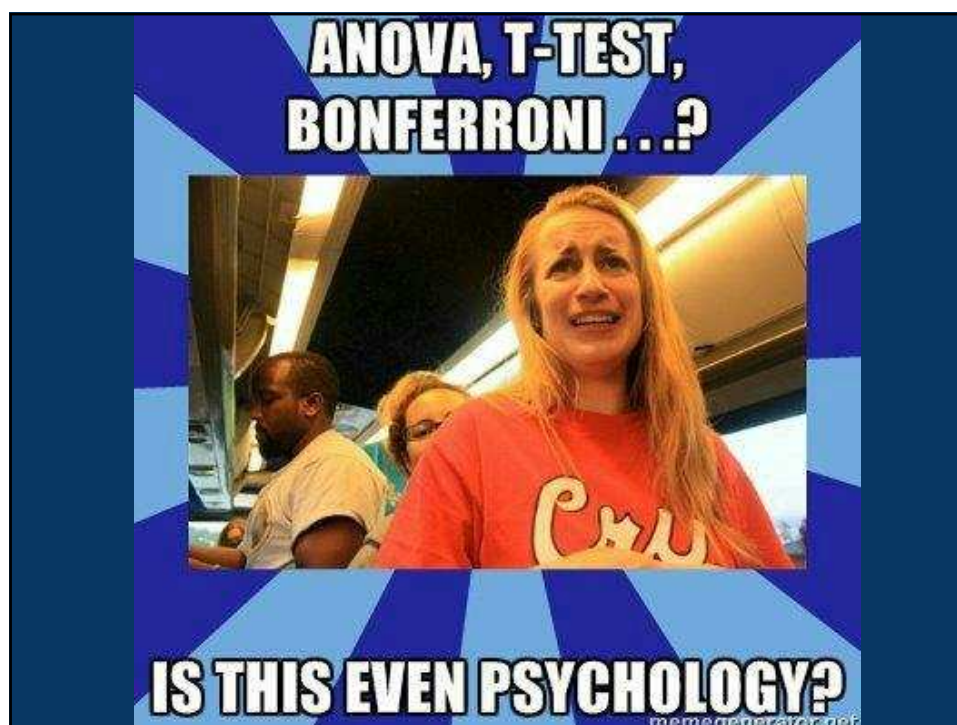
Ernest Rutherford

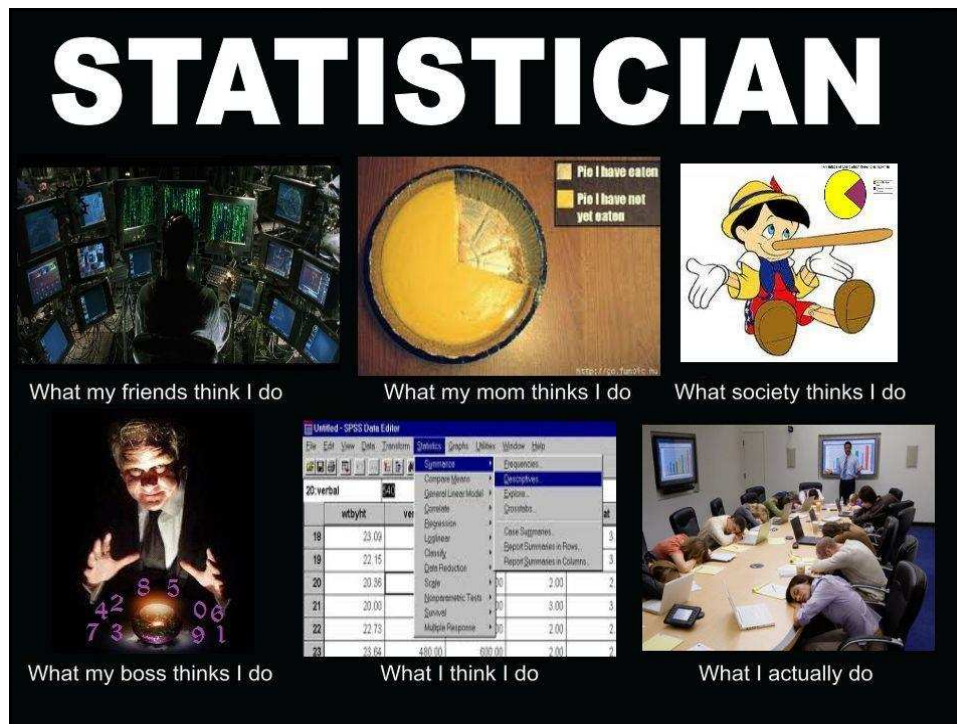
Je dokázané, že oslavovanie narodenín
je **zdravé**.

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

S. den Hartog

Ph D. Thesis University of Groningen





O čom budeme hovoriť?

- Premenné
- Hypotézy
- Štatistická signifikancia
- Frekvencie
- Deskripcia/inferencia
- Rozptyl
- Variancia
- Distribúcia
- Korelácia
- Regresia
- atď.

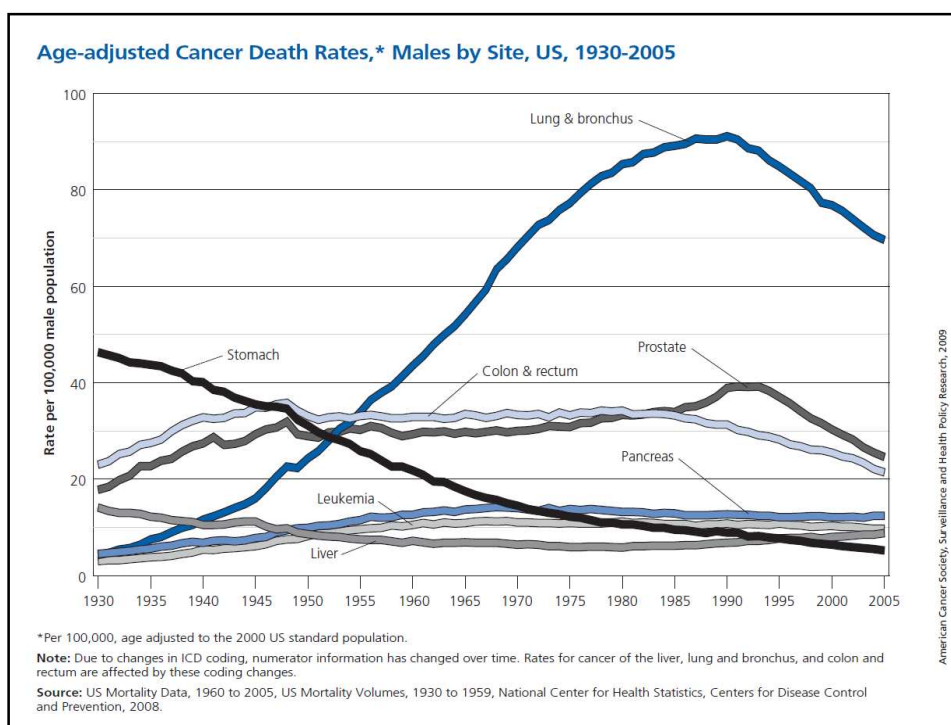
Rýchly úvod do štatistiky

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	Nezoradená množina	modus, Chi-kvadrát	One to One (equality (=))
Ordinálna (poradová)	Zoradená 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, factorová 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)

Stanley Smith Stevens



Deskripcia/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)

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 pravdivosť/nepravdivosť

	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)

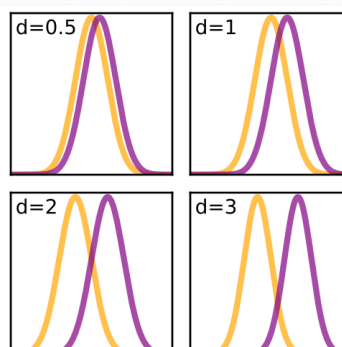
“Štatistika nikdy nemôže nahradiť
zdravý úsudok“

Henry Clay

Signifikancia

- 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



Signifikancia

- **Power analyses** (G*Power – nie je 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 S/N (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

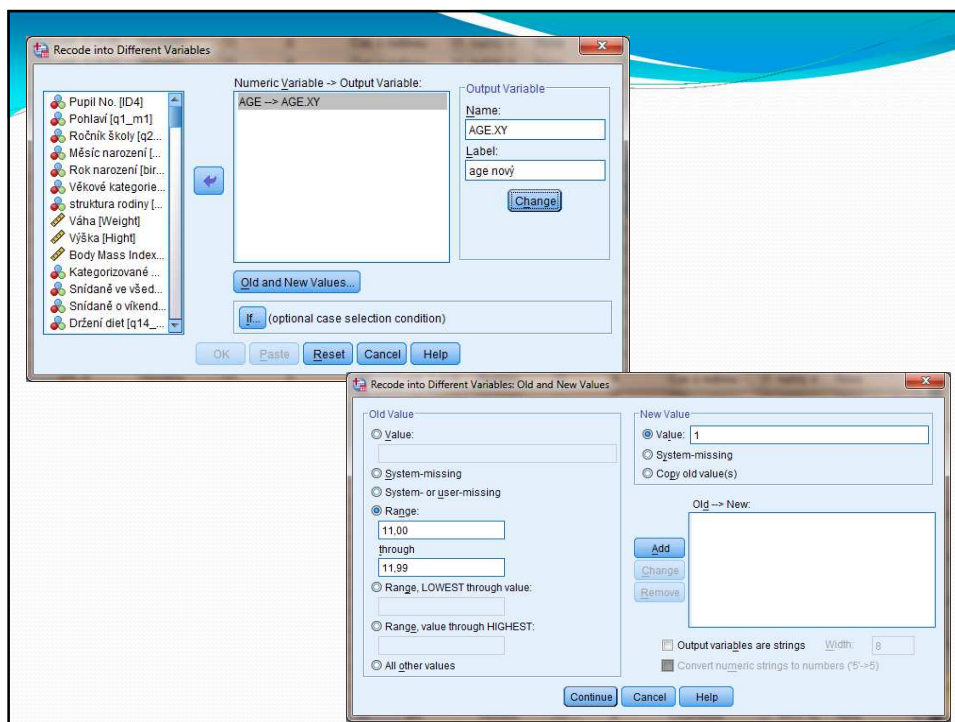
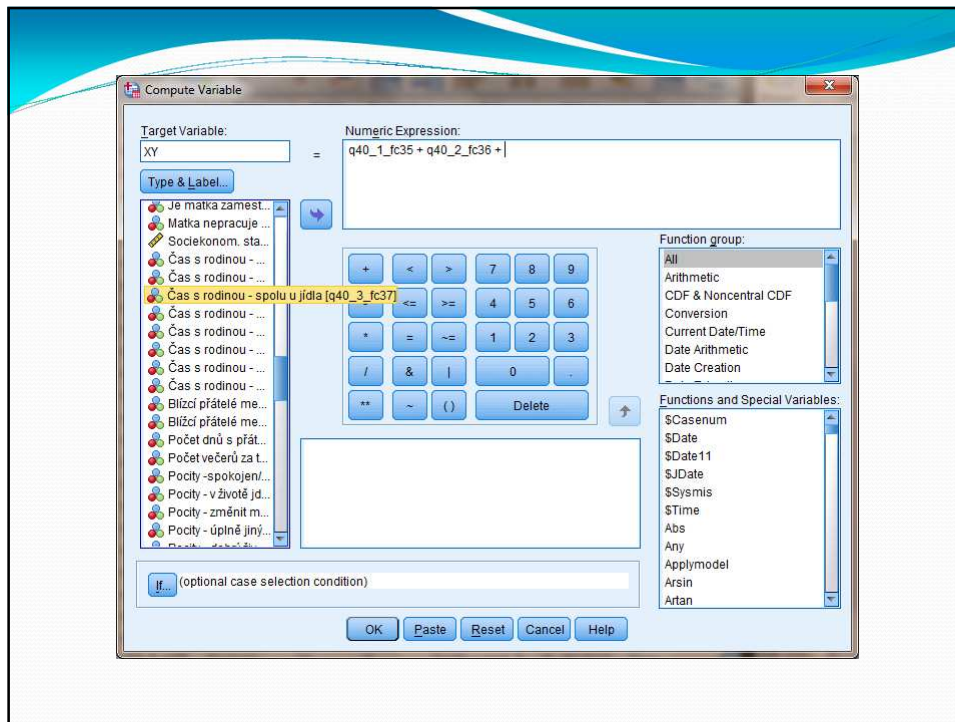
- Nie len prijatie/zamietnutie nulovej hypotézy
- Úroveň signifikancie
- Zhodnotenie reálnych rozdielov/vzťahov
- Zohľadniť charakteristiky testovaného súboru (kontext) – súčasť diskusie

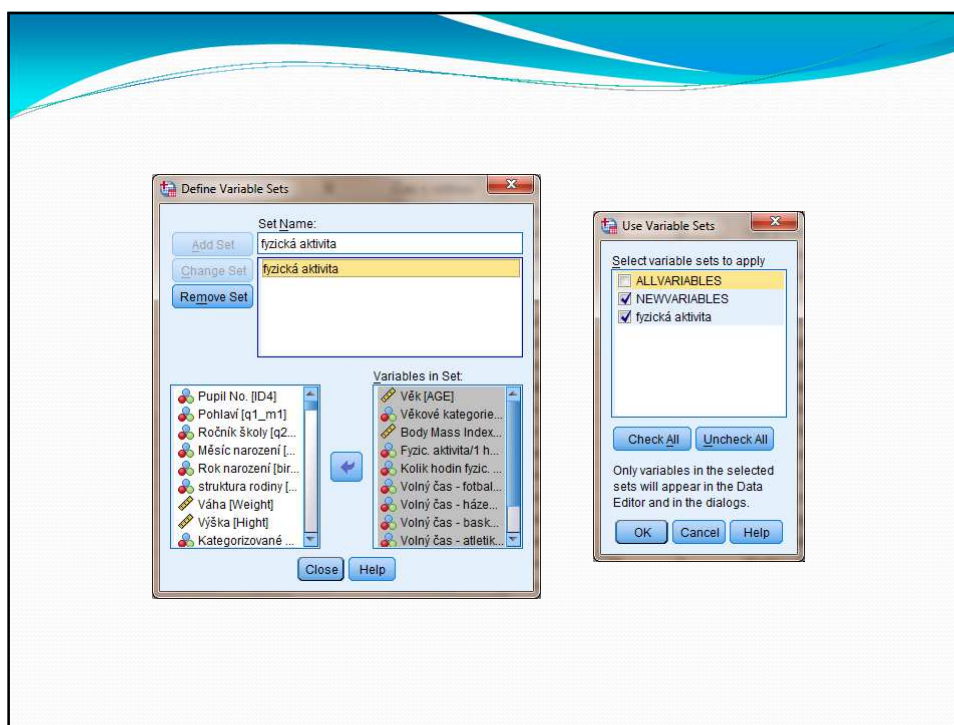
Príprava databázy

- Vytvorenie premenných
 - Pohlavie, vek, ... Položky škál alebo škálové skóre,...
 - Missing values
 - import z Excelu?
- Definovanie vlastností a hodnôt premennej:
 - číselná, dátum, reťazec textu,...
 - Nominálna, ordinálna, kardinálna (škálová)
 - Názov a popis premennej
 - Definovanie hodnôt (čo ktoré číslo znamená)
- Úprava premenných
 - Výpočet skóre škál, transformácia hodnôt, rekódovanie, dichotomizácia
 - Definovanie súborov premenných pre analýzy

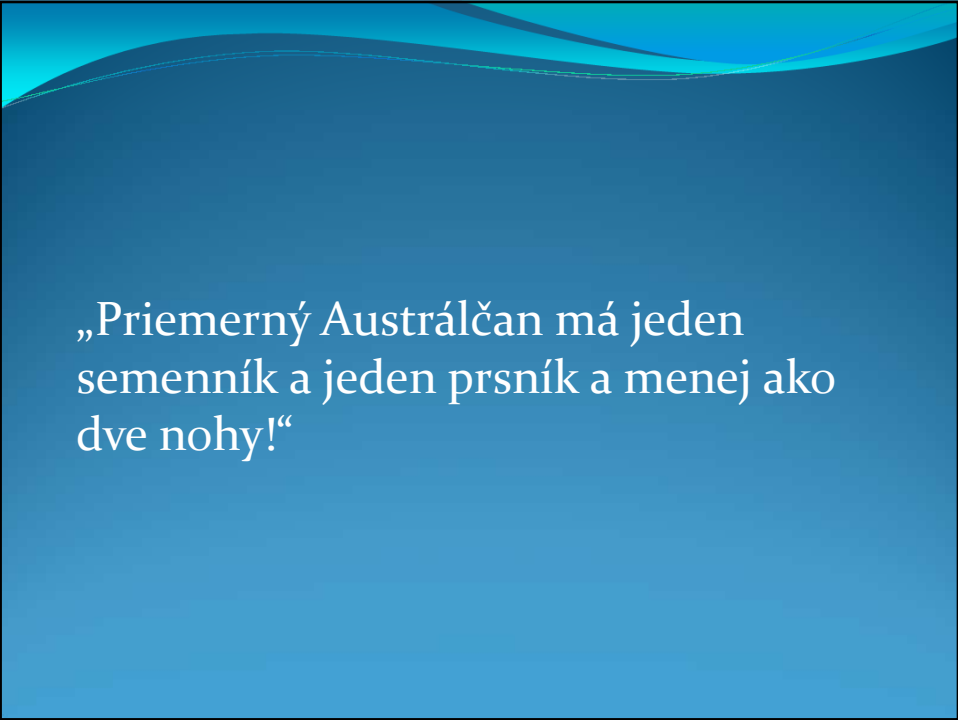
Úprava premenných

- Výpočet skóre škál
Ako na to v SPSS:
 Transform -> Compute variable -> položka1 + položka 2 +....+ položka n - OK
- Transformácia hodnôt – z-skóre
Ako na to v SPSS:
 Analyze -> Descriptive statistics-> Descriptives: ✓ Save standardized value as variables -OK
- Rekódovanie hodnôt, dichotomizácia
Ako na to v SPSS:
 Transform -> Recode into Different variable: položka1 -> Old and new values ...
- Definovanie súborov premenných
Ako na to v SPSS:
 Utilities -> Define Variable sets -> Názov nového setu -> položka1 + položka 2 +....+ položka n – OKUse variable sets





Vedeli ste, že veľká väčšina ľudí má viac ako priemerný počet nôh?



„Priemerný Austrálčan má jeden
semenník a jeden prsník a menej ako
dve nohy!“



Deskriptívna analýza

Popis súboru

• Frekvenčné tabuľky

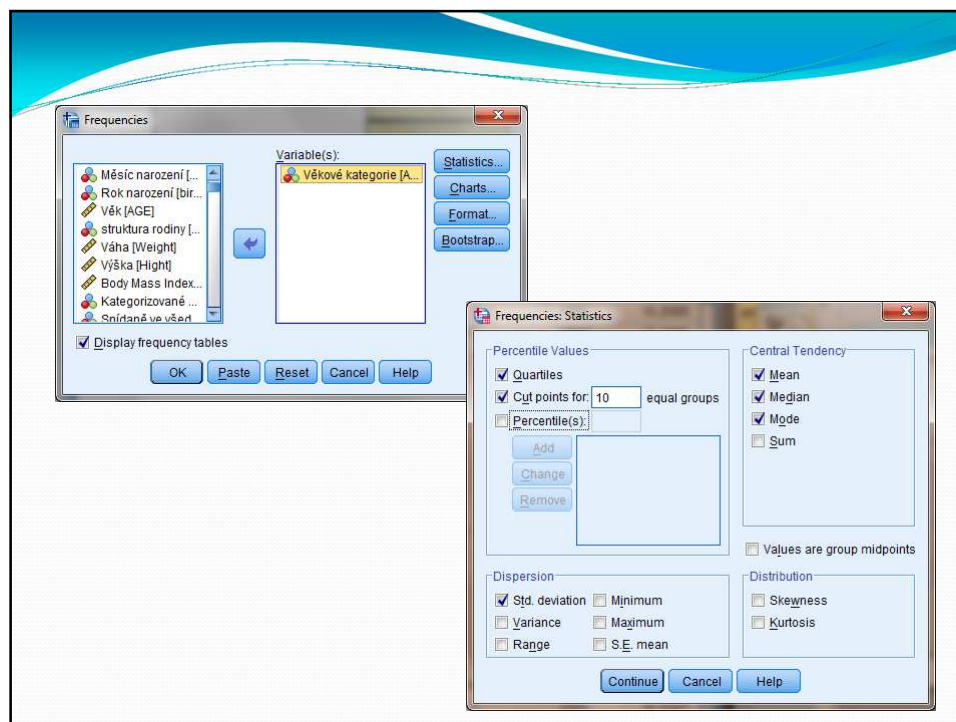
Ako na to v SPSS:

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

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

Statistics		
Vekové kategórie		
N	Valid	2182
	Missing	14
Mean		2,02
Median		2,00
Mode		3
Std. Deviation		,818

Vekové kategórie					
		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		

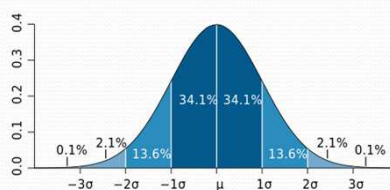
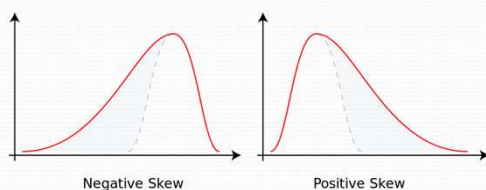


Popis súboru

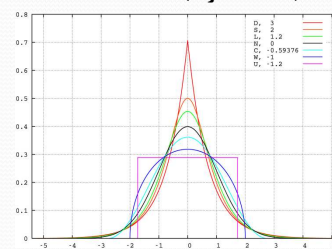
- Miery centrálnej tendencie:
 - **Modus** – najčastejšie sa vyskytujúca hodnota
 - **Medián** – polohový stred, 50-ty percentil
 - **Priemer** – aritmetický priemer
- Nominálne premenné: modus
- Ordinálne premenné: modus, medián
- Kardinálne premenné: modus, medián, priemer

Popis súboru

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



- Strmost' (kurtosis) distribúcie (vydutie)



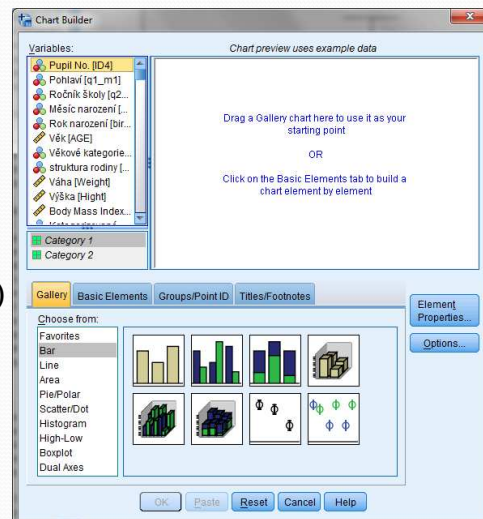
Popis súboru

- Grafy

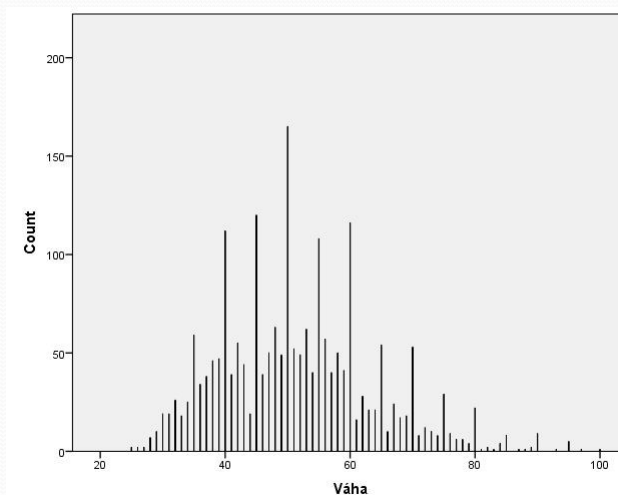
Ako na to v SPSS:

Graphs -> Chart Builder:

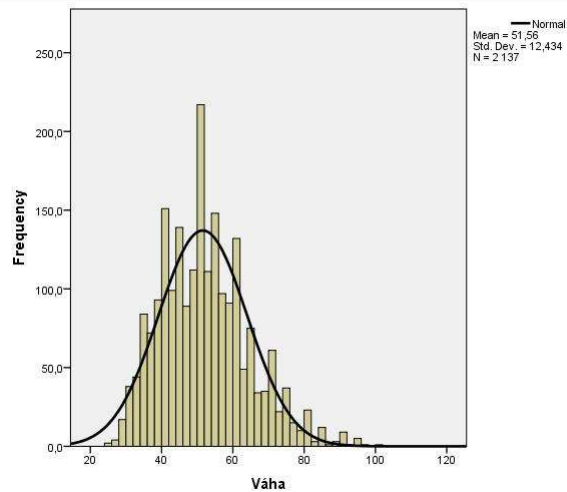
- Stĺpcový graf
- Čiarový graf
- Koláčový graf
- Bodový graf (scatter plot)
- Histogram
- Boxplot



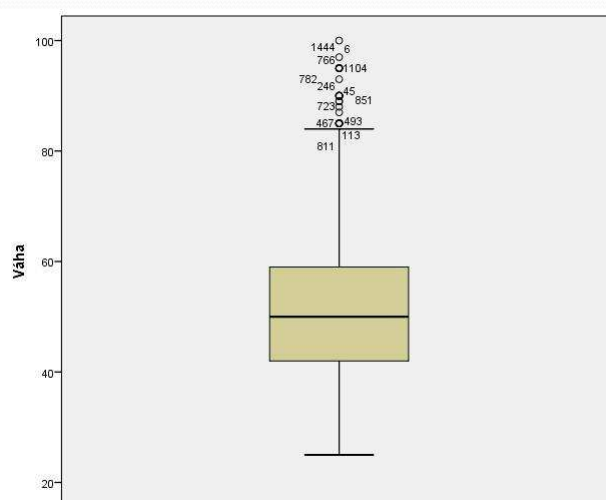
Stĺpcový graf



Histogram



Boxplot



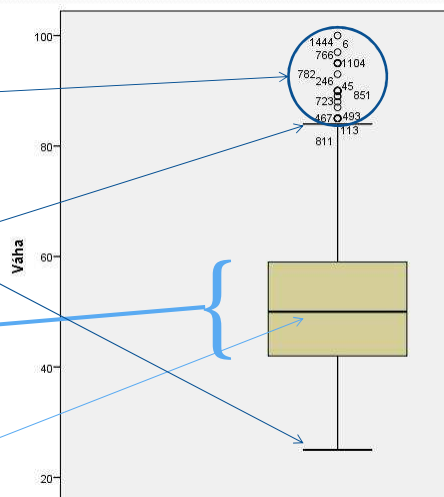
Boxplot - vysvetlenie

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

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

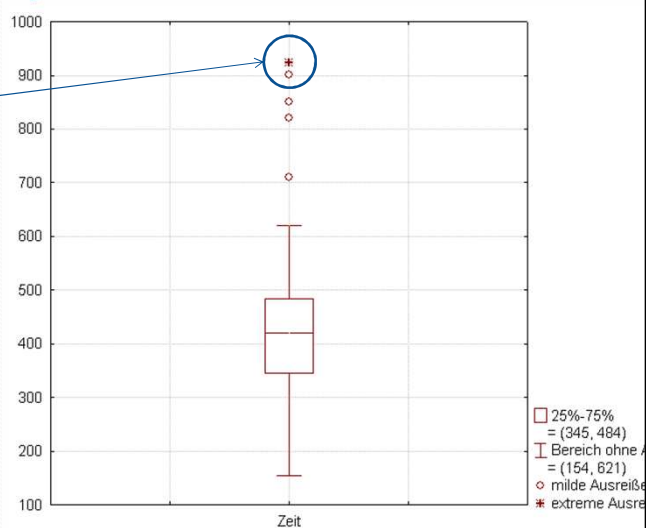
Medzikvartilové
rozpätie (IQR)

Medián
(polohový stred)

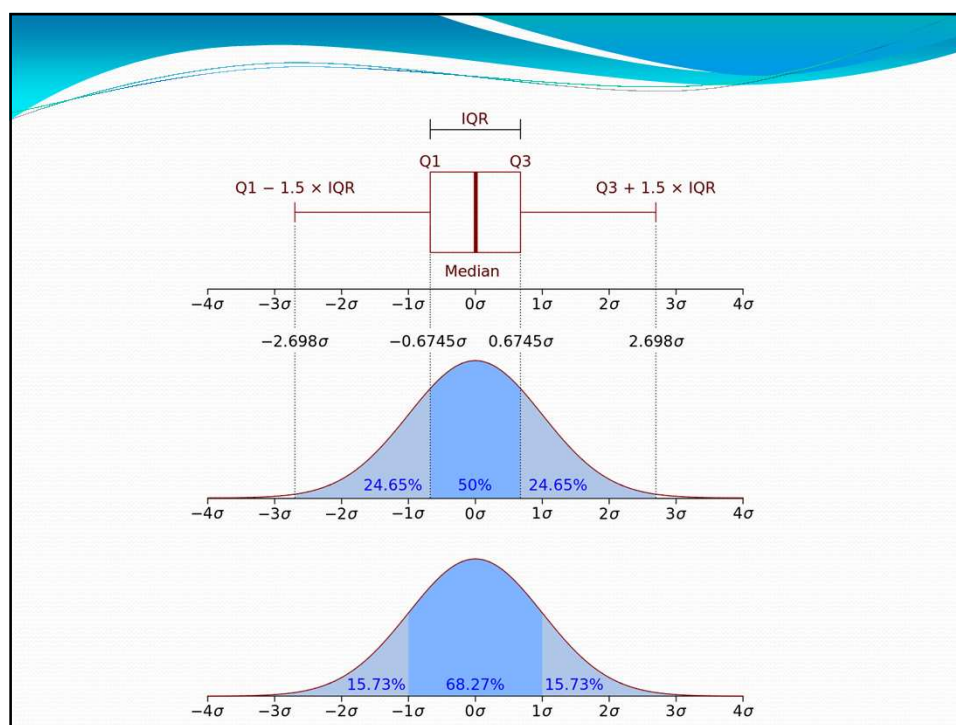
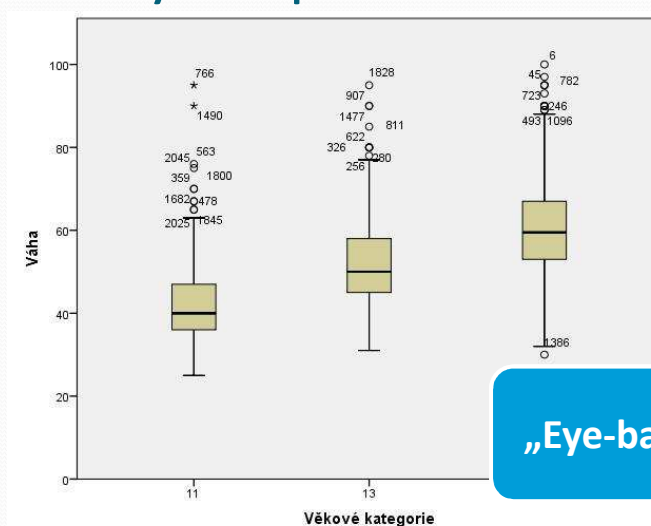


Boxplot - vysvetlenie

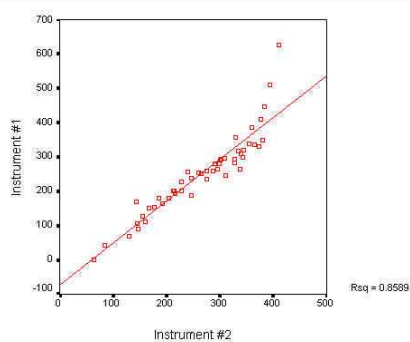
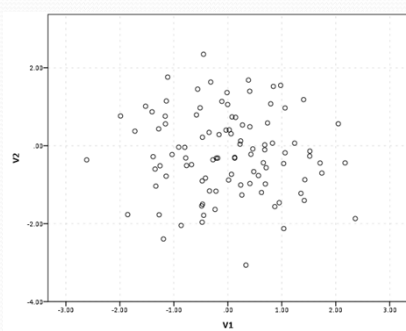
Extrémne
hodnoty
nad ± 3 násobok
IQR



Združený boxplot



Scatter plot

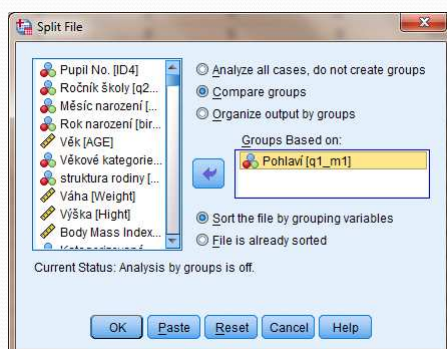


Stratifikácia súboru

- Členenie na podskupiny
- Umožňuje simultánnu analýzu podskupín (muži, ženy, a pod) – *split file*
- Môžeme tiež analyzovať len jednu podskupinu (len muži, len 11-roční) – *select cases*

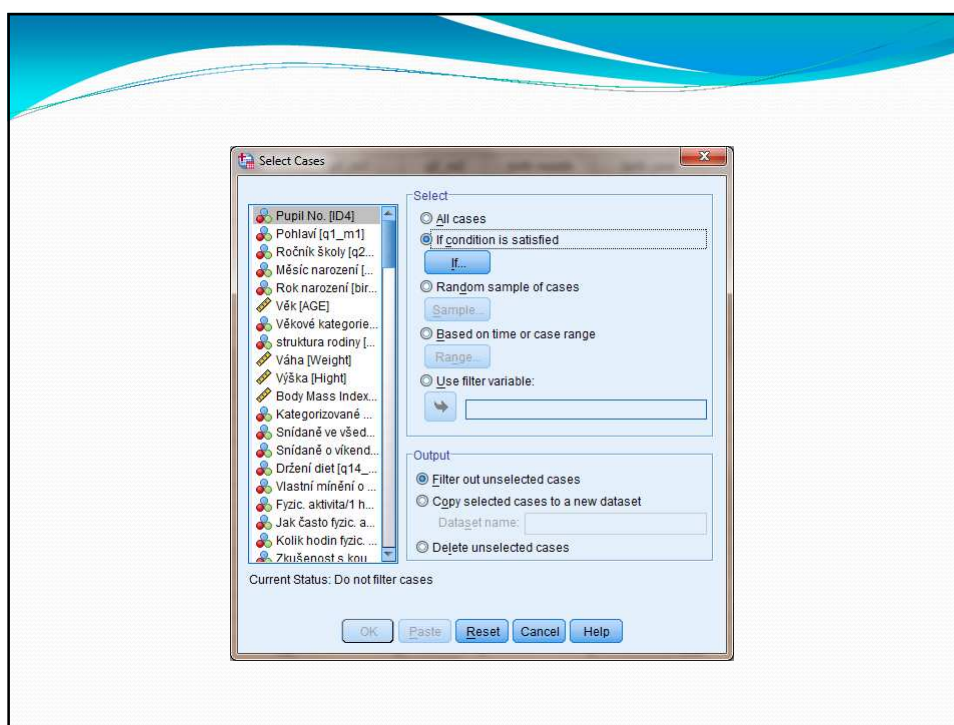
Ako na to v SPSS:

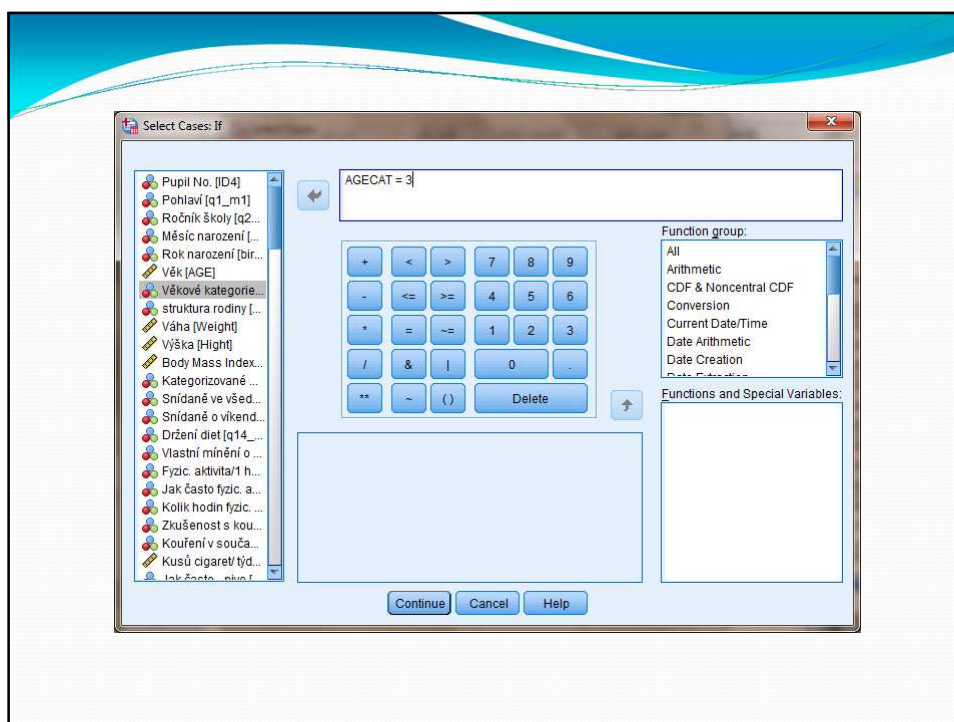
Data -> Split File alebo Select Cases



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	chlapeček	7	50,0	50,0
		divka	7	50,0	100,0
	Total	14	100,0	100,0	
11	Valid	chlapeček	355	50,1	50,1
		divka	354	49,9	100,0
	Total	709	100,0	100,0	
13	Valid	chlapeček	331	46,0	46,0
		divka	389	54,0	100,0
	Total	720	100,0	100,0	
15	Valid	chlapeček	344	45,7	45,7
		divka	409	54,3	100,0
	Total	753	100,0	100,0	



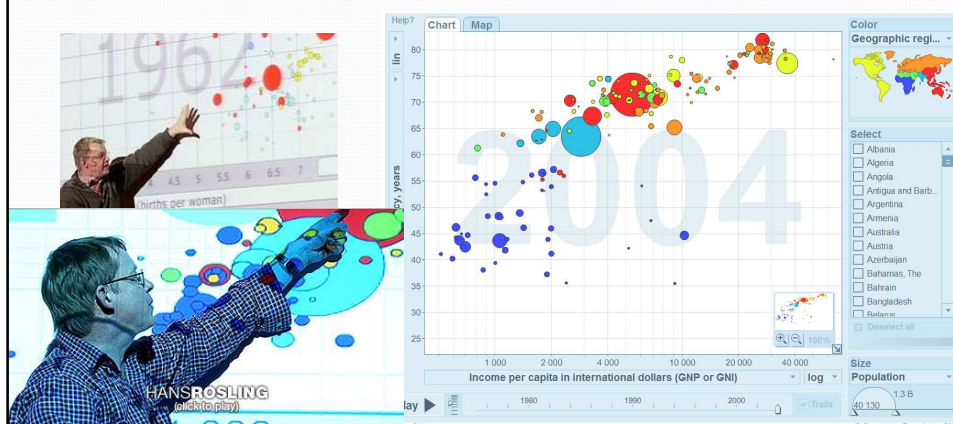


- Proporcia chlapcov a dívek len v skupine 15-ročných

Pohlaví					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	chlapec	344	45,7	45,7	45,7
	dívka	409	54,3	54,3	100,0
	Total	753	100,0	100,0	

Príklad umenia deskripcie

- Hans Rosling (TED talks)

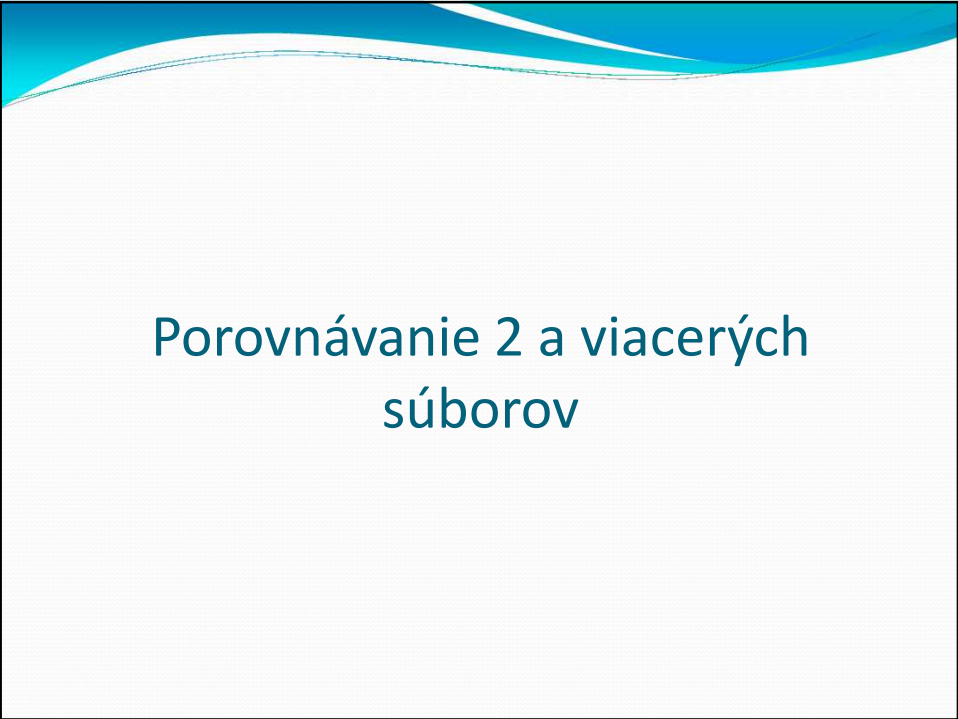


So štatistikou je to ako s príbehom o obchodníkovi, ktorého sa pýtali, ako môže predávať králičie sendviče tak lacno. "No," vysvetľoval, "musím do nich tiež dať aj nejaký ten kúsok konského mäsa. Ale miešam to 50:50. Jeden kôň, jeden králik."

Darrel Huff



Inferenčná štatistická analýza



Porovnávanie 2 a viacerých súborov

Rozdiely (nominálna úroveň)

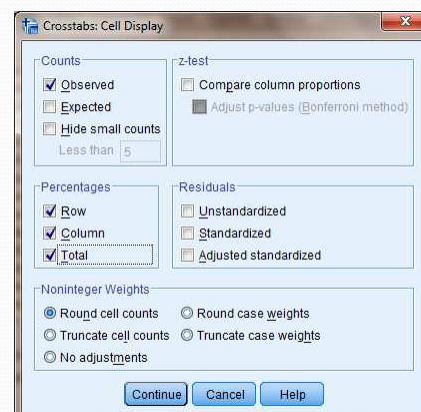
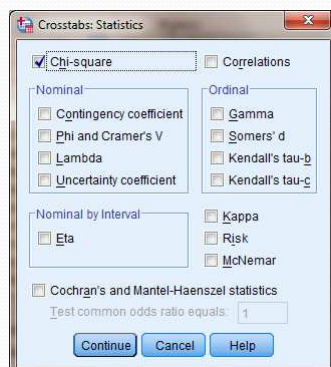
- Pearsonov Chi kvadrát

Ako na to v SPSS:

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

Výskumná otázka:

Majú viac sexuálnych skúseností
dievčatá alebo chlapci?



Rozdiely (nominálna úroveň)

- Pearsonov Chi kvadrát

Pohlaví * Mě/a někdy sex. styk Crosstabulation					
			Mě/a někdy sex. styk		Total
			ano	ne	
Pohlaví	chlapec	Count	54	181	235
		% within Pohlaví	23,0%	77,0%	100,0%
		% within Mě/a někdy sex. styk	40,0%	42,9%	42,2%
		% of Total	9,7%	32,5%	42,2%
	divka	Count	81	241	322
		% within Pohlaví	25,2%	74,8%	100,0%
		% within Mě/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 Mě/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)	Exact
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 13,5.

b. Computed only for a 2x2 table

Záver: Chlapci a dievčatá sa nelíšia v sexuálnych skúsenostiach





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

www.guinness.de

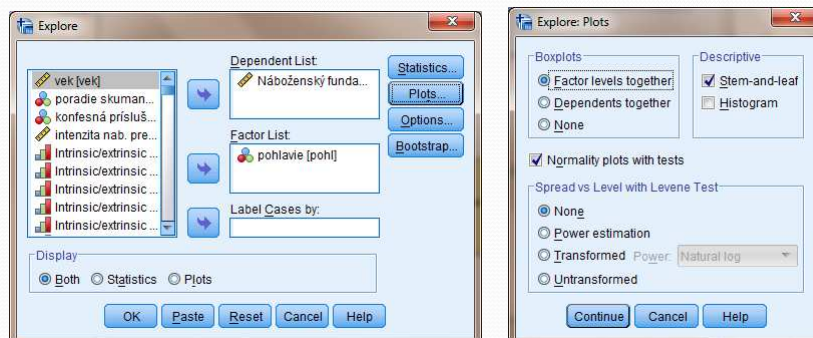
Studentov t-test

- Normalita distribúcie
- 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: V Normality plots with tests - OK

Standard Deviations From The Mean	Cumulative %	Z Scores	T Scores
-4σ	0.1%	-4.0	20
-3σ	0.1%	-3.0	30
-2σ	2.3%	-2.0	40
-1σ	15.9%	-1.0	50
0	50%	0	60
+1σ	84.1%	+1.0	70
+2σ	97.7%	+2.0	80
+3σ	99.9%	+3.0	90
+4σ	99.9%	+4.0	100

Overenie normality distribúcie



Ako na to v SPSS:

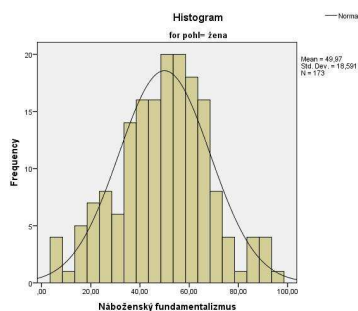
Analyze -> Descriptive stat. -> Explore – Plots: ☒ Normality plots with tests - OK

Overenie normality distribúcie

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

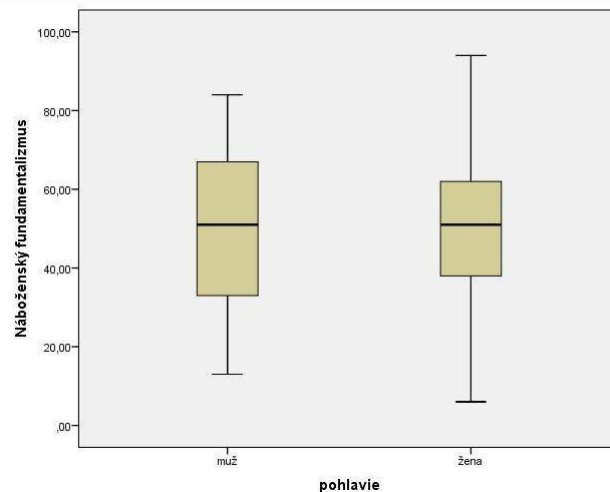
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



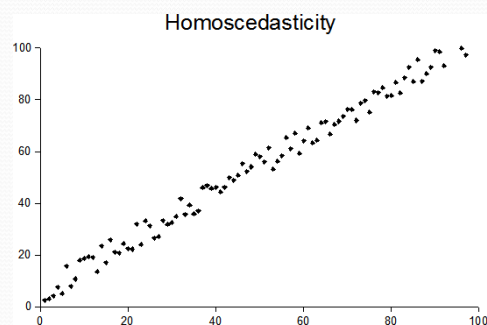
Záver: Prijímame H_0 o normálnej distribúcii hodnôt v populácii (aj u mužov aj u žien)

Overenie normality distribúcie

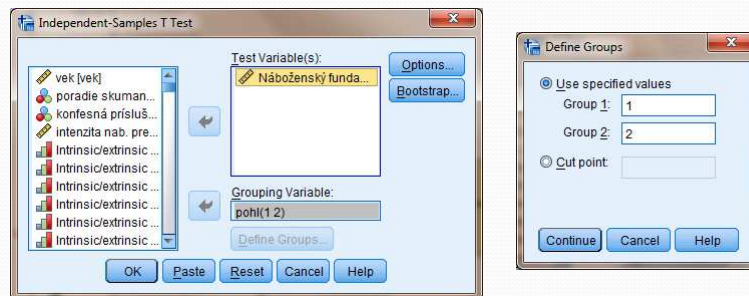


Studentov t-test

- 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	
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 (ordinálna úroveň)

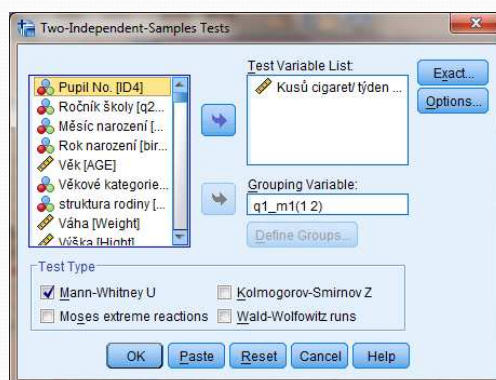
- Mann-Whitneyho U-test

Ako na to v SPSS:

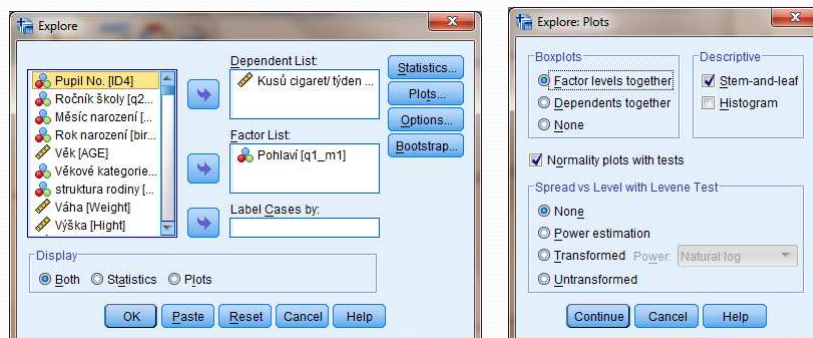
Analyze -> Nonparametric tests -> Legacy dialogs -> 2 independent samples ✓ Mann-Whitney U - OK

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

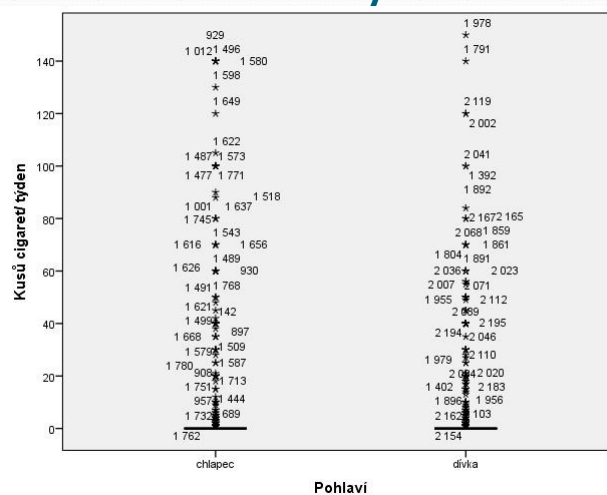
Overenie normality distribúcie

Tests of Normality							
	Pohlaví	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Kusů cigarety 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

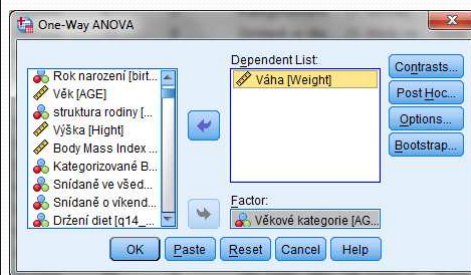
Záver: Nie je rozdiel v počte vyfajčených cigariet medzi chlapcami a dievčatami

ANOVA

- Jednoduchá analýza rozptylu (analysis of variance)
- Umožňuje testovať rozdiely u 3 a viac populácií naraz

Ako na to v SPSS:

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



Výskumná otázka:

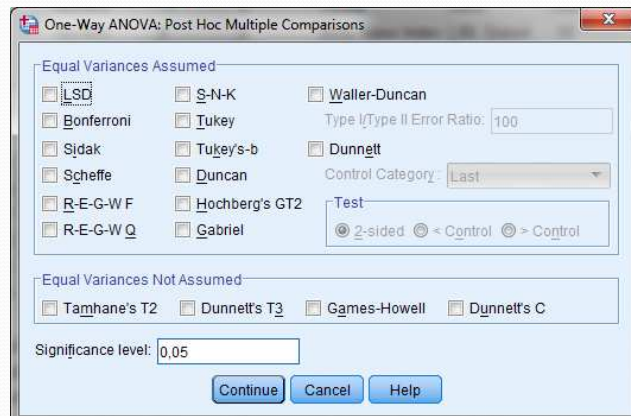
Je rozdiel v telesnej hmotnosti medzi žiakmi v jednotlivých vekových kategóriách?

Váha

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	123701,466	2	61850,733	639,003	,000
Within Groups	205780,976	2126	96,793		
Total	329482,442	2128			

ANOVA

- Post-Hoc testy – párové porovnania



ANOVA

- Post-Hoc testy – párové porovnania

Multiple Comparisons

Dependent Variable: Váha

	(I) Věkové kategorie	(J) Věkové kategorie	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	11	13	-10,008*	,525	,000	-11,24	-8,78
		15	-18,657*	,522	,000	-19,88	-17,43
	13	11	10,008*	,525	,000	8,78	11,24
		15	-8,649*	,520	,000	-9,87	-7,43
	15	11	18,657*	,522	,000	17,43	19,88
		13	8,649*	,520	,000	7,43	9,87
LSD	11	13	-10,008*	,525	,000	-11,04	-8,98
		15	-18,657*	,522	,000	-19,68	-17,63
	13	11	10,008*	,525	,000	8,98	11,04
		15	-8,649*	,520	,000	-9,67	-7,63
	15	11	18,657*	,522	,000	17,63	19,68
		13	8,649*	,520	,000	7,63	9,67

*. The mean difference is significant at the 0.05 level.

ANOVA

- Neparametrická alternatíva:

- Kruskal-Wallis test

Ako na to v SPSS:

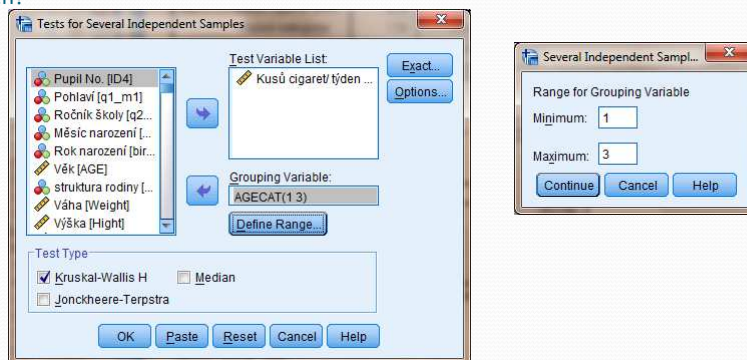
Analyze -> Nonparametric tests -> Legacy dialogs -> K independent samples ✓ Kruskal-Wallis H - OK

- Neobsahuje Post-Hoc testy
- Párové porovnanie je potrebné urobiť U-testom pre každú dvojicu
- Friedman test (nie je v SPSS)

Kruskal-Wallis test

Výskumná otázka:

Je rozdiel medzi žiakmi v jednotlivých vekových kategóriách v počte vyfajčených cigariet za týždeň?



Kruskal-Wallis test

Ranks			
	Věkové kategorie	N	Mean Rank
Kusů cigaret/ týden	11	673	883,75
	13	674	991,48
	15	703	1193,82
	Total	2050	

Test Statistics ^{a,b}	
	Kusů cigaret/ týden
Chi-Square	235,638
df	2
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable:
Věkové kategorie

Analýza vzťahov medzi premennými

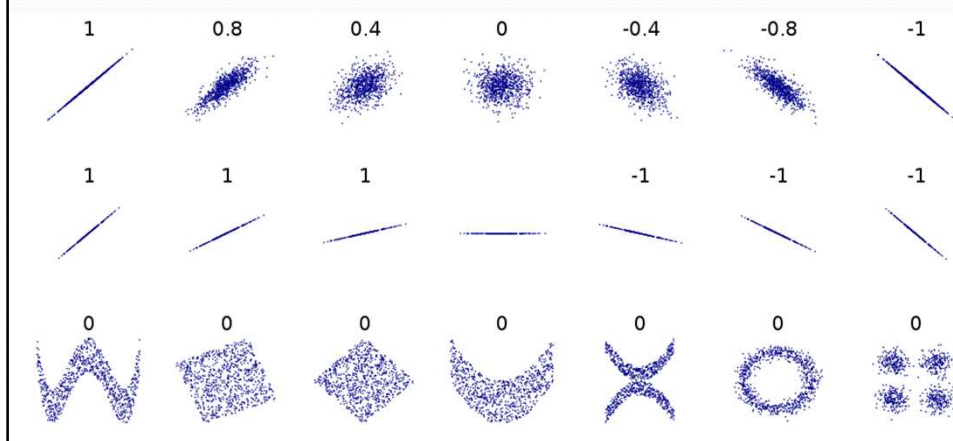
Korelácie

- 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
 - Úrovně 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

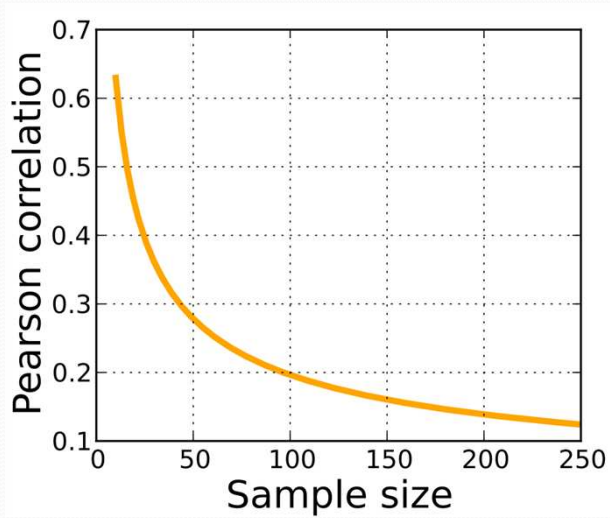
Sila vzťahu/korelácie	Negatívne hodnoty	Pozitívne hodnoty
Žiaden vzťah	-0.09 do 0.0	0.0 do 0.09
Malá	-0.3 do -0.1	0.1 do 0.3
Stredná	-0.5 do -0.3	0.3 do 0.5
silná	-1.0 do -0.5	0.5 do 1.0

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:

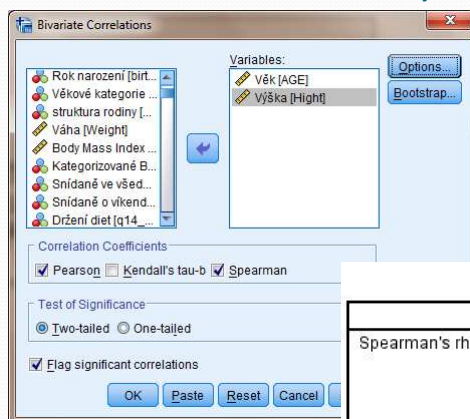
Analyze -> Correlate -> Bivariate: Variables, $\checkmark$ Pearson/Spearman- OK

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

Korelácie

Výskumná otázka:

Existuje súvislosť medzi vekom a výškou respondenta?

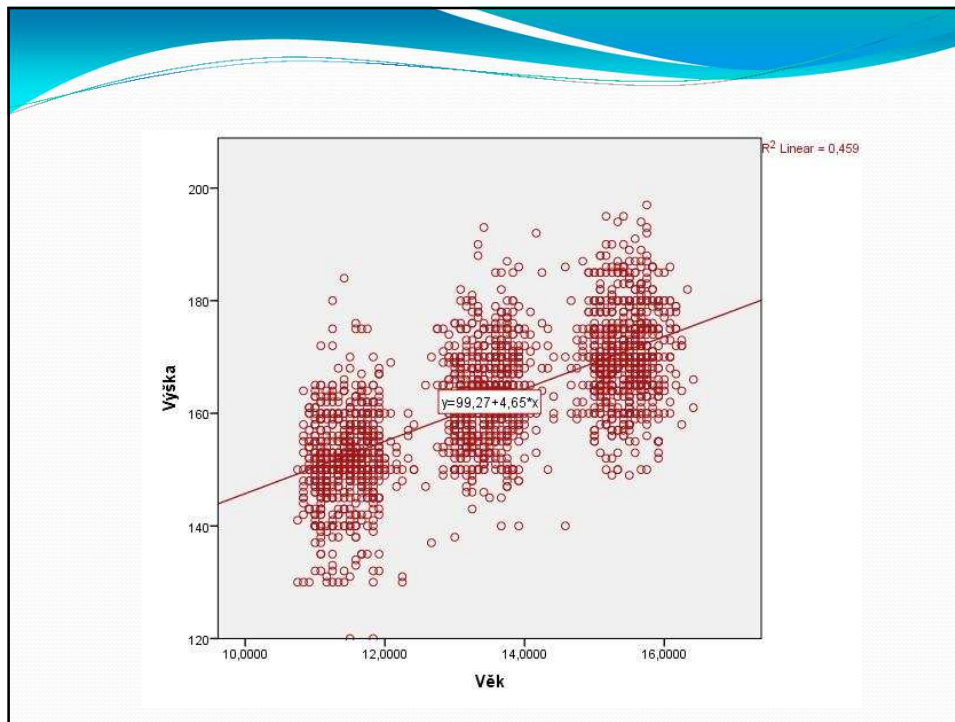


Correlations			Věk	Výška
Věk	Pearson Correlation		1	,678**
	Sig. (2-tailed)			,000
	N		2182	2141
Výška	Pearson Correlation		,678**	1
	Sig. (2-tailed)		,000	
	N		2141	2149

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

Correlations			Věk	Výška
Spearman's rho	Věk	Correlation Coefficient	1,000	,682**
		Sig. (2-tailed)	.	,000
		N	2182	2141
Výška		Correlation Coefficient	,682**	1,000
		Sig. (2-tailed)	,000	.
		N	2141	2149

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

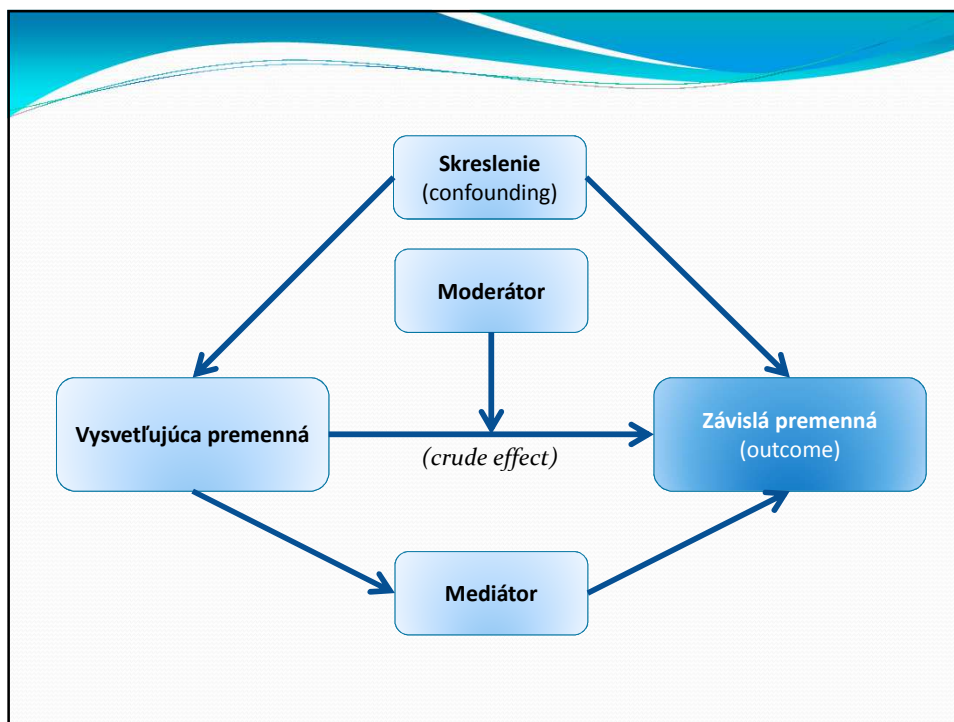


Regresná analýza

- 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)

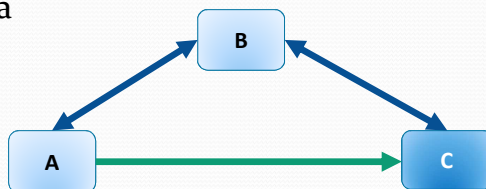
Regresná analýza

- 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)

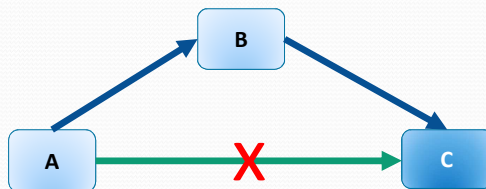
- 1.) prediktor (A) a confounder (B) musia byť prepojený so závislou premennou (C)
- 2.) prediktor a confounder musia byť prepojený (*interkorelácia*)
- 3.) skreslenie nesmie byť kauzálnym následkom prediktora



M.A. Babyak: Understanding confounding and mediation. *Evid Based Mental Health* 2009 12: 68-71

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 a 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)

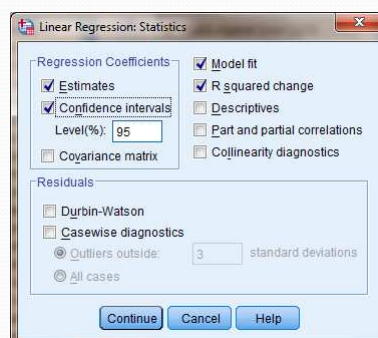
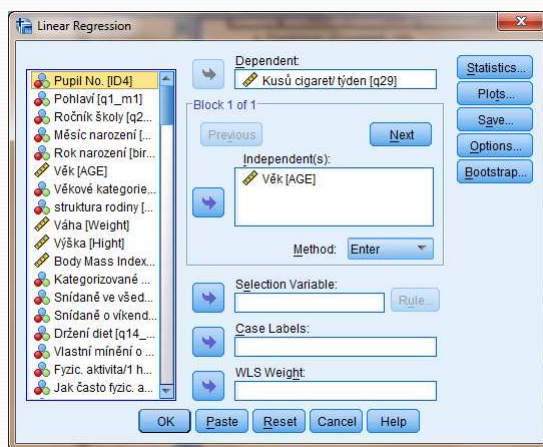
Moderácia

- 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)

Lineárna regresia

Výskumná otázka:

Má vek respondenta vplyv na množstvo týždenne vyfajčených cigariet?



Ako na to v SPSS:

Analyze -> Regression -> Linear...

Lineárna regresia – crude effect

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,223 ^a	,050	,049	15,566	,050	107,055	1	2048	,000

a. Predictors: (Constant), Vek

ANOVA ^a						
Model	Sum of Squares	df	Mean Square	F	Sig.	
1 Regression	25939,447	1	25939,447	107,055	,000 ^b	
Residual	496229,689	2048	242,300			
Total	522169,136	2049				

a. Dependent Variable: Kusů cigareť týden

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-24,505	2,788		-8,791	,000	-29,972	-19,038
	Vek	2,123	,205	,223	10,347	,000	1,721	2,526

a. Dependent Variable: Kusů cigareť týden

Záver: Vek respondenta vysvetľuje množstvo týždenne vyfajčených cigariet. Čím vyšší vek, tým viac vyfajčených cigariet žiaci udávajú.

Lineárna regresia – adjusted effect

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,223 ^a	,050	,049	15,591	,050	106,668	1	2041	,000
2	,234 ^b	,055	,054	15,554	,005	10,870	1	2040	,001

a. Predictors: (Constant)

b. Predictors: (Constant)

ANOVA ^a						
Model	Sum of Squares	df	Mean Square	F	Sig.	
1 Regression	25929,200	1	25929,200	106,668	,000 ^b	
Residual	496133,426	2041	243,084			
Total	522062,627	2042				
2 Regression	28558,774	2	14279,387	59,027	,000 ^c	
Residual	493503,852	2040	241,914			
Total	522062,627	2042				

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-24,572	2,800		-8,775	,000	-30,064	-19,080
	Vek	2,129	,206	,223	10,328	,000	1,725	2,533
2	(Constant)	-21,554	2,940		-7,332	,000	-27,319	-15,789
	Vek	2,082	,206	,218	10,099	,000	1,677	2,486
	Fyzic. aktivita/1 hod/7 dní	-,553	,168	-,071	-3,297	,001	-,882	-,224

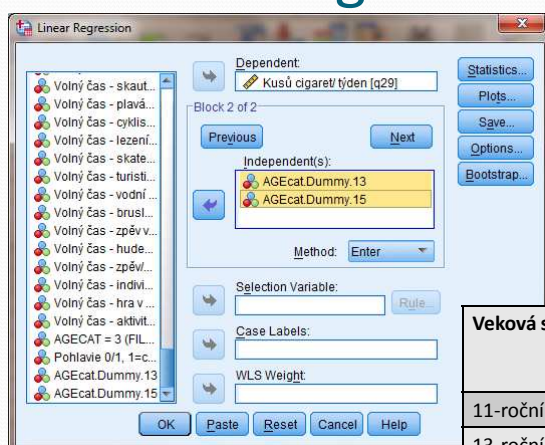
a. Dependent Variable: Kusů cigareť týden

Lineárna regresia – kategorizovaný prediktor

- Môžeme zadať aj kategoriickú premennú, ale len v podobe **0/1** (dichotomická)
- Pohlavie (1 muž, 0 žena) (0 je referenčná kategória)
- Pri multikategoriálnych premenných musíme vytvoriť tzv. „dummy variables“
 - Napr. vzdelanie: ZŠ=1, SŠ=2, VŠ=3
 - Transformácia:
 - ak referenčná kategória ZŠ
 - Rekódujeme premenné:

vzdelanie	Hodnota v databáze (dummy SŠ)	Hodnota v databáze (dummy VŠ)
ZŠ	0	0
SŠ	1	0
VŠ	0	1

Lineárna regresia



Veková skupina	Hodnota v databáze (dummy SŠ)	Hodnota v databáze (dummy VŠ)
11-roční	0	0
13-roční	1	0
15-roční	0	1

Lineárna regresia

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,020 ^a	,000	,000	15,964	,000	,832	1	2048	,362
2	,239 ^b	,057	,056	15,514	,057	61,353	2	2046	,000

a. Predictors: (Con

b. Predictors: (Con

ANOVA ^a					
Model		Sum of Squares	df	Mean Square	Sig.
1	Regression	211,959	1	211,959	,832
	Residual	521957,177	2048	254,862	,362 ^b
	Total	522169,136	2049		
2	Regression	29744,410	3	9914,803	
	Residual	492424,726	2046	240,677	
	Total	522169,136	2049		

a. Dependent Variable: Kusô cigarety týden

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	t
		B	Std. Error	Beta	
1	(Constant)	3,810	,487		7,821
	Pohlavie 0/1, 1=chlapec	,644	,706	,020	,912
2	(Constant)	,053	,689		,076
	Pohlavie 0/1, 1=chlapec	,781	,686	,024	1,138
	AGEcat.Dummy.13	2,027	,846	,060	2,397
	AGEcat.Dummy.15	8,822	,837	,262	10,541

a. Dependent Variable: Kusô cigarety týden

Žiaci v skupine 15-ročných s vyššou pravdepodobnosťou fajčia viac cigariet ako 11-roční žiaci

Binárna logistická regresia

- Závislou premennou môže byť len dichotomická premenná (o/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)

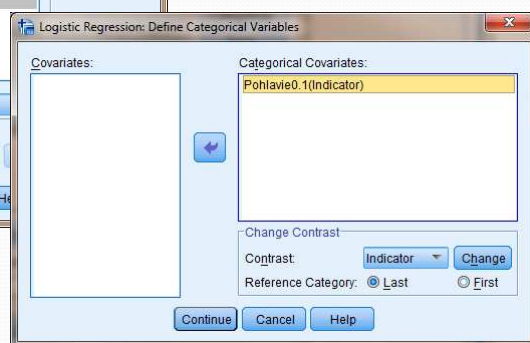
Ako na to v SPSS:

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

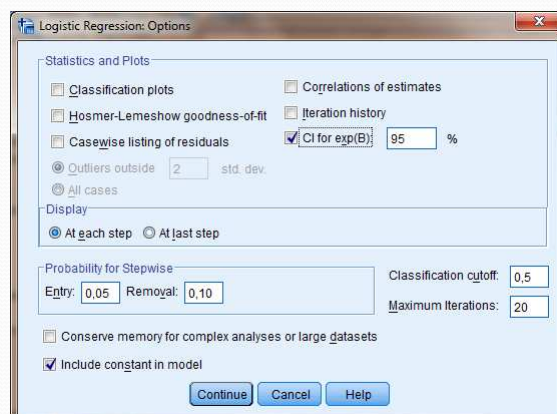
Binárna logistická regresia



Výskumná otázka:
Má pohlavie respondenta
súvislosť so sexuálnou
skúsenosťou?



Binárna logistická regresia



Binárna logistická regresia

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Pohlavie0.1(1)	,115	,202	,325	1	,569	1,122	,755	1,666
	Constant	-1,210	,155	60,845	1	,000	,298		

a. Variable(s) entered on step 1: Pohlavie0.1.

a. Variable(s) entered on step 1: Pohlavie0.1.

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Pohlavie0.1(1)	,105	,205	,260	1	,610	1,110	,742	1,661
	MBMI	,019	,033	,321	1	,571	1,019	,954	1,088
	Constant	-1,589	,728	4,767	1	,029	,204		

a. Variable(s) entered on step 1: Pohlavie0.1, MBMI.

Záver 1: Pohlavie respondenta nesúvisí so sexuálnou skúsenosťou (nie je prediktorom)?

Záver 2: BMI respondenta nesúvisí so sexuálnou skúsenosťou (nie je prediktorom)?

Binárna logistická regresia

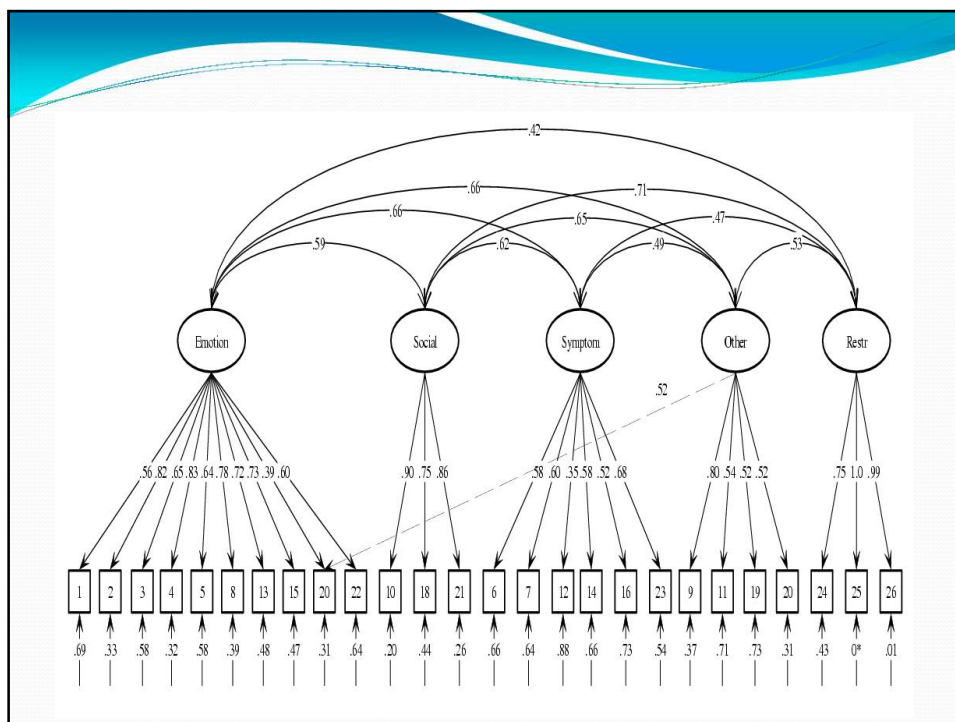
Categorical Variables Codings				
		Frequency	Parameter coding	
			(1)	(2)
Kategorizované BMI	normal	478	1,000	,000
	nadváha	68	,000	1,000
	obezita	8	,000	,000
Pohlavie 0/1, 1=chlapec	,00	319	1,000	
	1,00	235	,000	

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Pohlavie0.1(1)	,096	,204	,219	1	,640	1,100	,737	1,642
	BMI_kat			,702	2	,704			
	BMI_kat(1)	,809	1,075	,565	1	,452	2,245	,273	18,474
	BMI_kat(2)	,683	1,108	,380	1	,538	1,980	,226	17,387
	Constant	-1,983	1,072	3,419	1	,064	,138		

a. Variable(s) entered on step 1: Pohlavie0.1, BMI_kat.

Faktorová analýza

- Na základe analýzy variability pozorovaných premenných abstrahuje latentné – nepozorované premenné
- pozorované premenné odrážajú variáciu menšieho počtu latentných premenných
- Latentné premenné – faktory
- FA vychádza z analýzy hlavných komponentov (Principal component analysis, PCA)
- Psychologická metóda :D
 - Charles Spearman – výskum inteligencie (g faktor)
 - Raymond Cattell – osobnostný dotazník 16PF

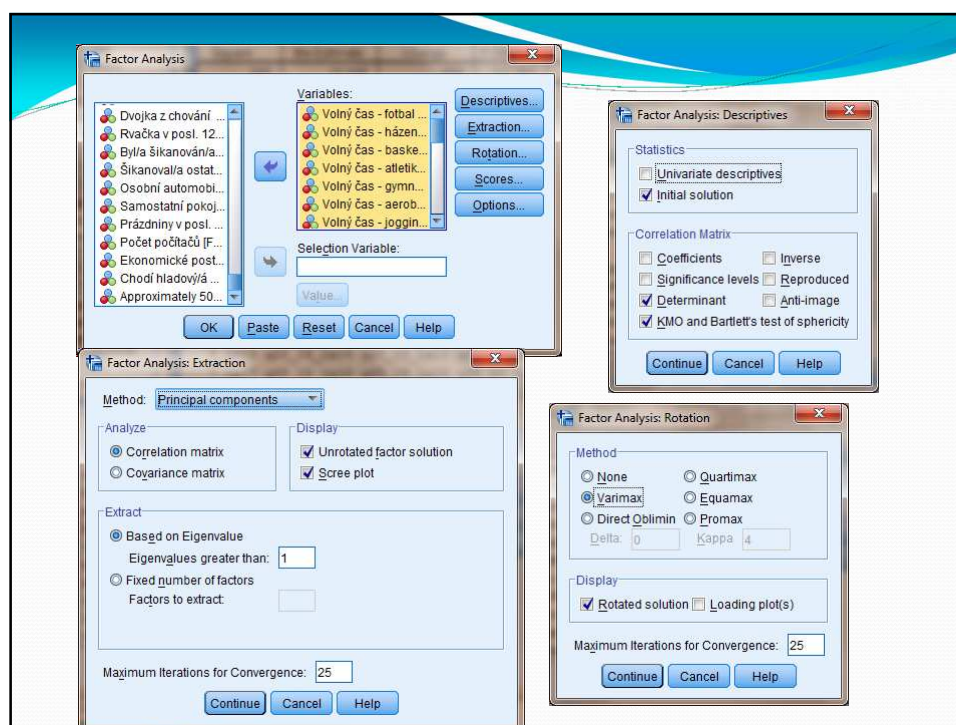


Faktorová analýza

- Vychádza z korelácií (korelačná matica)
- Determinant korelačnej matice (nad 0,00001)
- Keiser-Meier-Olkinova miera adekvátnosti výberu (nad 0,7)
- Bartlettov test sféricity: potrebujeme zamietnuť H_0 (sig. < 0,05)
- Nesplnenie podmienok značí, že dáta nie sú vhodné na faktorovú analýzu

Ako na to v SPSS:

Analyze -> Dimension reduction -> Factor ...



Faktorová analýza

- Počet faktorov závisí:
 - Eigenvalue (vlastná hodnota) nad 1.0
 - Definovaním počtu faktorov
 - Percentom vysvetlenej variancie
 - Posúdením sutinového grafu (scree plot)
- Rotácia – zlepšuje utriedenie položiek a tým interpretáciu faktorov (v porovnaní s nerotovaným riešením)
- Pomenovanie výsledných faktorov je tvorivá činnosť
 - Snaha vystihnúť spoločné znaky premenných vo faktore

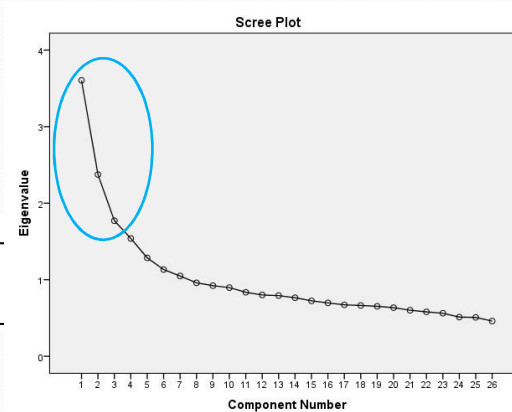
Faktorová analýza

Correlation Matrix^a

a. Determinant = ,025

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy,		,777
Bartlett's Test of Sphericity	Approx. Chi-Square	7128,029
	df	325
	Sig.	,000



Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,605	13,864	13,864	3,605	13,864	13,864	2,363	9,090	9,090
2	2,374	9,132	22,997	2,374	9,132	22,997	2,212	8,510	17,599
3	1,771	6,811	29,808	1,771	6,811	29,808	1,984	7,631	25,231
4	1,540	5,922	35,729	1,540	5,922	35,729	1,733	6,666	31,896
5	1,286	4,946	40,675	1,286	4,946	40,675	1,655	6,365	38,261
6	1,133	4,357	45,032	1,133	4,357	45,032	1,443	5,549	43,810
7	1,048	4,033	49,065	1,048	4,033	49,065	1,366	5,255	49,065
8	,960	3,693	52,757						
9	,922	3,547	56,304						
10	,897	3,449	59,753						
11	,836	3,216	62,969						
12	,801	3,081	66,050						
13	,793	3,050	69,100						
14	,765	2,941	72,041						
15	,724	2,786	74,827						
16	,697	2,682	77,510						
17	,672	2,585	80,095						
18	,663	2,552	82,647						
19	,653	2,510	85,157						
20	,635	2,444	87,601						
21	,602	2,314	89,914						
22	,580	2,232	92,147						
23	,562	2,162	94,309						
24	,512	1,969	96,278						
25	,508	1,953	98,232						
26	,460	1,768	100,000						

Extraction Method: Principal Component Analysis.

Factor Analysis: Extraction

Method: **Principal components**

Analyze:

☒ Correlation matrix

☐ Covariance matrix

Display:

☒ Unrotated factor solution

☒ Scree plot

Extract:

☐ Based on Eigenvalue

Eigenvalues greater than: 1

☒ Fixed number of factors

Factors to extract: 3

Maximum iterations for Convergence: 25

Continue Cancel Help

Component Matrix ^a			
	Component		
	1	2	3
Volný čas - jogging	,497	-,262	
Volný čas - plavání	,486		
Volný čas - lezení, rafting	,474		
Volný čas - atletika	,463		
Volný čas - skateboard, koleč. brusle	,461		-,311
Volný čas - vodní aktivity	,458		-,424
Volný čas - gymnastika	,446		
Volný čas - aerobik	,422		
Volný čas - házená	,416		
Volný čas - basket, jiné míč. hry	,413	-,292	
Volný čas - bruslení, hokej	,398		-,272
Volný čas - tanec	,382		-,352
Volný čas - vzpírání/posilování	,379	-,312	,273
Volný čas - cyklistika	,367		
Volný čas - turistika, rybaření	,312		
Volný čas - aktivity v církev. společ.	,267		
Volný čas - skauting			
Volný čas - zpěv ve sboru	,370	,573	
Volný čas - individ. lekce hudby	,335	,554	
Volný čas - hudeb. nástroj v orchestru	,347	,548	
Volný čas - zpěv/hra v kapele	,324	,548	
Volný čas - fotbal		-,473	,293
Volný čas - hra v divadle	,276	,369	
Volný čas - box/kickbox	,256		,568
Volný čas - karate, judo..			,522
Volný čas - zápas		-,372	,490

Extraction Method: Principal Component Analysis.
a. 3 components extracted.

Component Matrix ^a			
	Component		
	1	2	3
Volný čas - jogging	,497	-,262	
Volný čas - plavání	,486		
Volný čas - lezení, rafting	,474		
Volný čas - atletika	,463		
Volný čas - skateboard, koleč. brusle	,461		-,311
Volný čas - vodní aktivity	,458		-,424
Volný čas - gymnastika	,446		
Volný čas - aerobik	,422		
Volný čas - házená	,416		
Volný čas - basket, jiné míč. hry	,413	-,292	
Volný čas - bruslení, hokej	,398		-,272
Volný čas - tanec	,382		-,352
Volný čas - vzpírání/posilování	,379	-,312	,273
Volný čas - cyklistika	,367		
Volný čas - turistika, rybaření	,312		
Volný čas - aktivity v církev. společ.	,267		
Volný čas - skauting			
Volný čas - zpěv ve sboru	,370	,573	
Volný čas - individ. lekce hudby	,335	,554	
Volný čas - hudeb. nástroj v orchestru	,347	,548	
Volný čas - zpěv/hra v kapele	,324	,548	
Volný čas - fotbal		-,473	,293
Volný čas - hra v divadle	,276	,369	
Volný čas - box/kickbox	,256		,568
Volný čas - karate, judo..			,522
Volný čas - zápas		-,372	,490

Extraction Method: Principal Component Analysis.
a. 3 components extracted.

Rotated Component Matrix ^a			
	Component		
	1	2	3
Volný čas - vodní aktivity	,604		
Volný čas - plavání	,560		
Volný čas - skateboard, koleč. brusle	,547		
Volný čas - bruslení, hokej	,522		
Volný čas - jogging	,493		,271
Volný čas - házená	,457		
Volný čas - basket, jiné míč. hry	,457		
Volný čas - cyklistika	,435		
Volný čas - atletika	,433		,297
Volný čas - tanec	,421	,265	-,252
Volný čas - gymnastika	,368		
Volný čas - aerobik	,353	,319	
Volný čas - turistika, rybaření	,284		
Volný čas - zpěv ve sboru		,682	
Volný čas - individ. lekce hudby		,661	
Volný čas - hudeb. nástroj v orchestru		,659	
Volný čas - zpěv/hra v kapele		,656	
Volný čas - hra v divadle		,474	
Volný čas - aktivity v církev. společ.		,359	
Volný čas - skauting		,278	
Volný čas - zápas			,652
Volný čas - box/kickbox			,571
Volný čas - karate, judo..			,553
Volný čas - fotbal			,515
Volný čas - vzpírání/posilování	,265		,495
Volný čas - lezení, rafting	,304		,364

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Overovanie reliability škál a dotazníkov

Odhad reliability

- Reliabilita = spoľahlivosť merania
- Reliabilita ako interná konzistencia testu (split-half, paralelné formy, Cronbach alfa) alebo stabilita v čase (test-retest)
- Test retest (korelácia)
- Split-half
- Paralelné formy
- Cronbachove α (vhodné nad 0,70)

The screenshot shows the SPSS Reliability Analysis dialog box and its Statistics sub-dialog. In the main dialog, the 'Items' list contains several 'Fundamentalism' items (snf1 to snf9). The 'Model' is set to 'Alpha'. The 'Statistics' sub-dialog is open, showing options for 'Descriptives for' (Item, Scale, Scale if item deleted), 'Summaries' (Means, Variances, Covariances, Correlations), 'Inter-Item' (Correlations, Covariances), and 'ANOVA Table' (None, F test, Friedman chi-square, Cochran chi-square). The 'Model' is set to 'Two-Way Mixed' and 'Type' is 'Consistency'.

Reliability Analysis

Items:

- Fundamentalism [snf1]
- Fundamentalism [snf2.r]
- Fundamentalism [snf3]
- Fundamentalism [snf4.r]
- Fundamentalism [snf5]
- Fundamentalism [snf6]
- Fundamentalism [snf7.r]
- Fundamentalism [snf8]
- Fundamentalism [snf9.r]

Model: Alpha

Scale label:

OK Paste Reset Cancel Help

Reliability Analysis: Statistics

Descriptives for:

- ☐ Item
- ☐ Scale
- ☒ Scale if item deleted

Summaries:

- ☐ Means
- ☐ Variances
- ☐ Covariances
- ☐ Correlations

Inter-Item:

- ☒ Correlations
- ☐ Covariances

ANOVA Table:

- ☒ None
- ☐ F test
- ☐ Friedman chi-square
- ☐ Cochran chi-square

Hotelling's T-square

Intraclass correlation coefficient

Model: Two-Way Mixed

Type: Consistency

Confidence interval: 95 %

Test value: 0

Continue Cancel Help

Ako na to v SPSS:
Analyze -> Scale -> Reliability analysis ...

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.863	.863	12

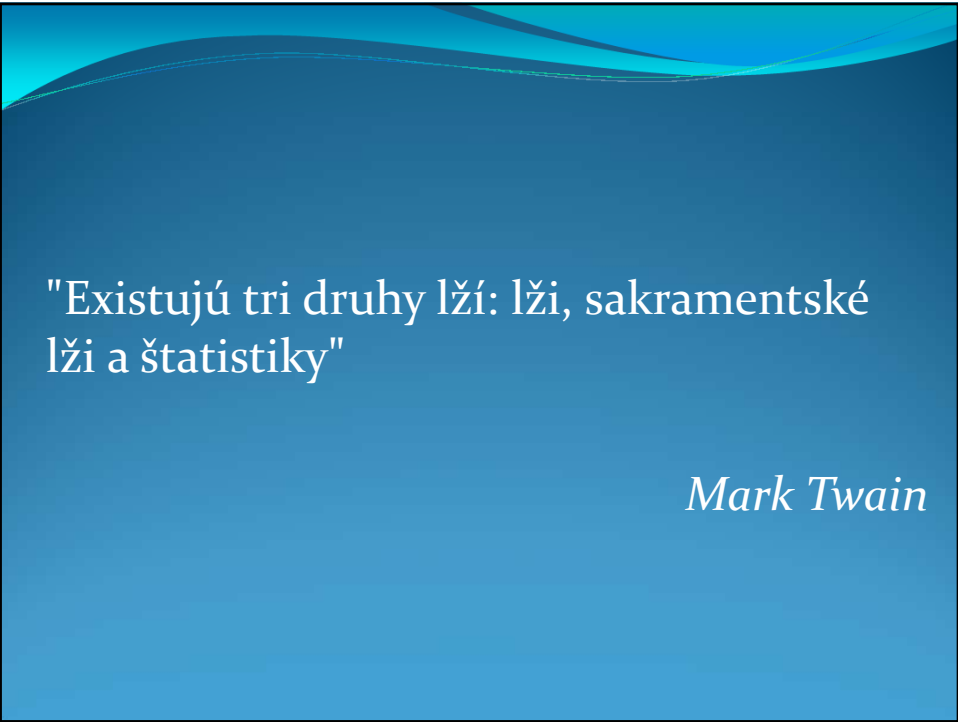
Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Fundamentalism	45,3202	308,773	,572	,436	,850
Fundamentalism	44,9015	314,337	,473	,386	,856
Fundamentalism	45,1478	289,909	,646	,477	,844
Fundamentalism	46,4877	301,370	,564	,391	,851
Fundamentalism	45,5714	314,513	,499	,380	,855
Fundamentalism	46,8571	313,994	,422	,351	,860
Fundamentalism	46,2266	311,008	,493	,409	,855
Fundamentalism	46,7537	310,048	,515	,417	,854
Fundamentalism	45,1182	294,095	,634	,478	,845
Fundamentalism	45,2611	305,550	,648	,525	,846
Fundamentalism	45,7044	317,734	,488	,303	,855
Fundamentalism	46,3251	304,894	,560	,403	,851

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Intrinsic/extrinsic relig	60,4780	160,574	,272	,171	,765
Intrinsic/extrinsic relig	60,1073	160,881	,212	,274	,768
Intrinsic/extrinsic relig	59,9659	160,808	,229	,419	,767
Intrinsic/extrinsic relig	60,8195	172,933	-,131	,452	,787
Intrinsic/extrinsic relig	61,2732	158,356	,379	,495	,760
Intrinsic/extrinsic relig	60,4244	157,559	,253	,259	,767
Intrinsic/extrinsic relig	60,1951	171,864	-,102	,329	,787
nsic relig	60,0439	150,464	,457	,485	,753
nsic relig	60,3610	161,497	,225	,439	,767
nsic relig	60,8878	164,022	,191	,218	,768
nsic relig	60,1902	163,469	,176	,177	,769
nsic relig	60,0049	153,103	,497	,451	,753
nsic relig	60,7122	168,098	,006	,468	,780
Intrinsic/extrinsic relig	60,0195	149,588	,615	,692	,746
Intrinsic/extrinsic relig	59,9268	150,088	,595	,572	,747
Intrinsic/extrinsic relig	61,5902	169,204	-,011	,190	,778
Intrinsic/extrinsic relig	62,0000	173,765	-,178	,293	,783
Intrinsic/extrinsic relig	60,5659	157,129	,377	,408	,759
Intrinsic/extrinsic relig	60,5024	164,241	,201	,339	,768
Intrinsic/extrinsic relig	60,2195	157,623	,380	,367	,759
Intrinsic/extrinsic relig	59,7707	152,913	,480	,461	,753
Intrinsic/extrinsic relig	59,7610	152,800	,534	,486	,751
Intrinsic/extrinsic relig	61,0000	160,500	,264	,245	,765
Intrinsic/extrinsic relig	59,7756	150,832	,553	,558	,749
Intrinsic/extrinsic relig	60,1610	153,881	,455	,636	,755
Intrinsic/extrinsic relig	59,4780	161,663	,255	,321	,766
Intrinsic/extrinsic relig	60,2293	154,276	,463	,706	,755
Intrinsic/extrinsic relig	60,3951	154,142	,459	,725	,755

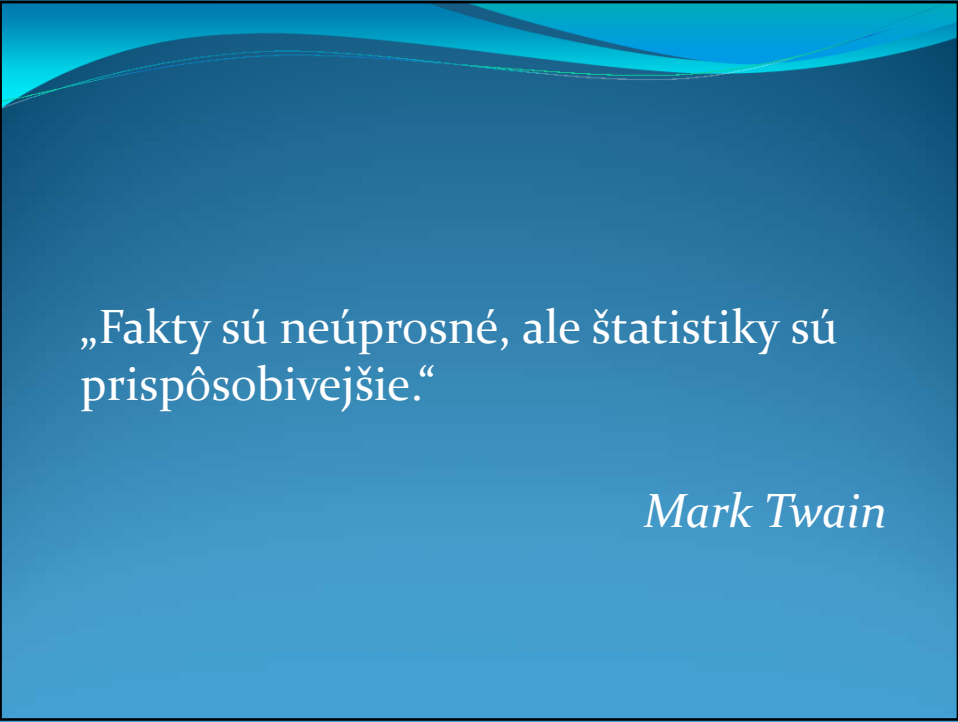
Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,771	,769	28

Zneužívanie štatistiky



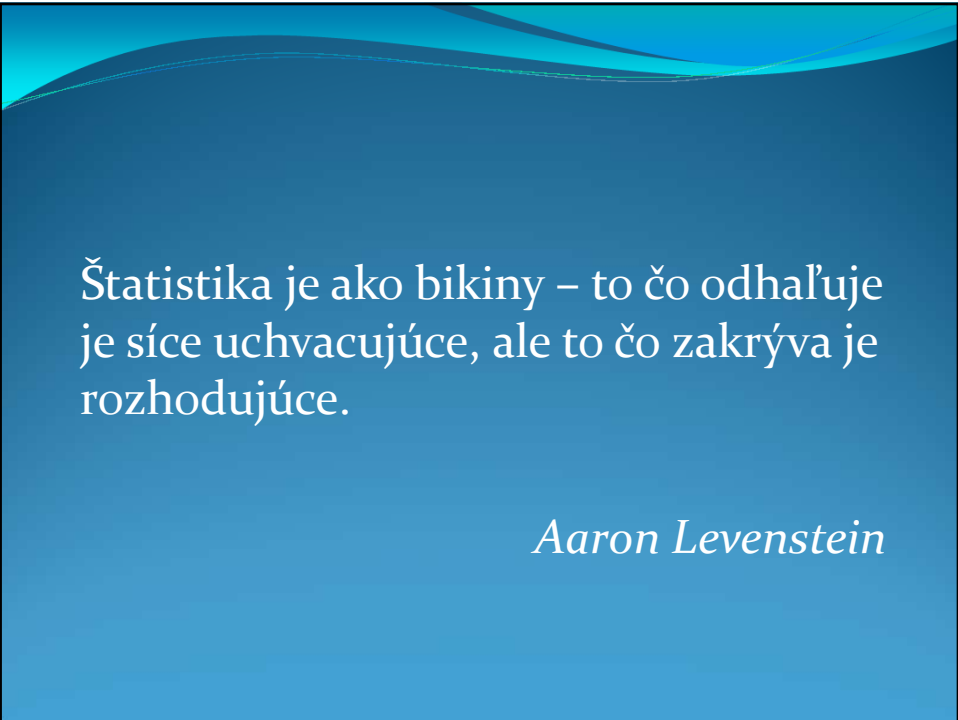
"Existujú tri druhy lži: lži, sakramentské
lži a štatistiky"

Mark Twain



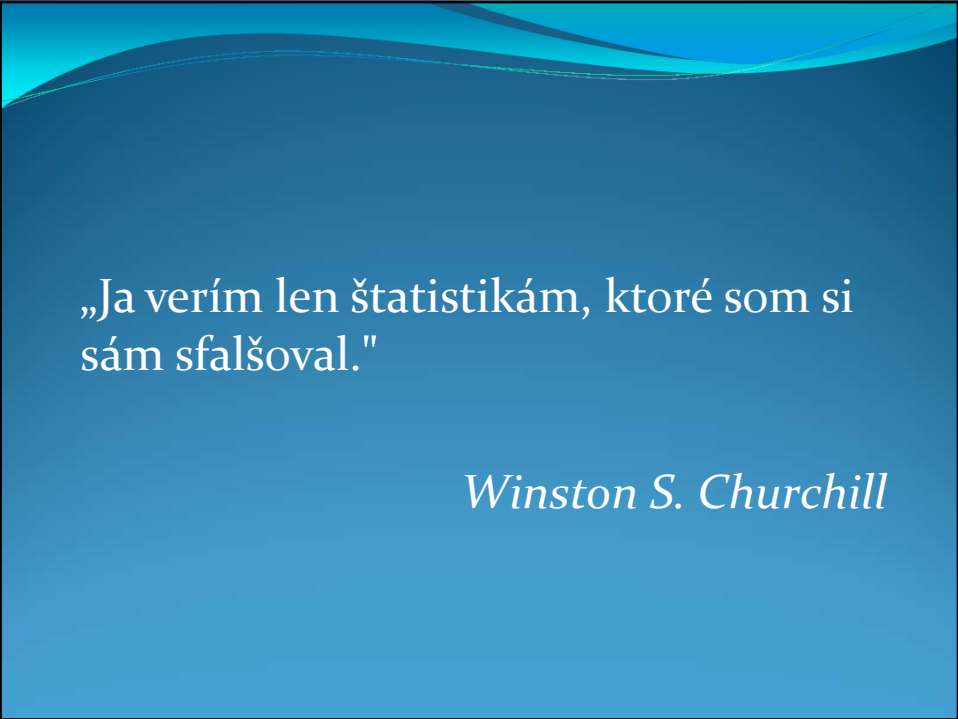
„Fakty sú neúprosné, ale štatistiky sú
prispôsobivejšie.“

Mark Twain



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

Aaron Levenstein



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

Winston S. Churchill

Skreslenia výsledkov

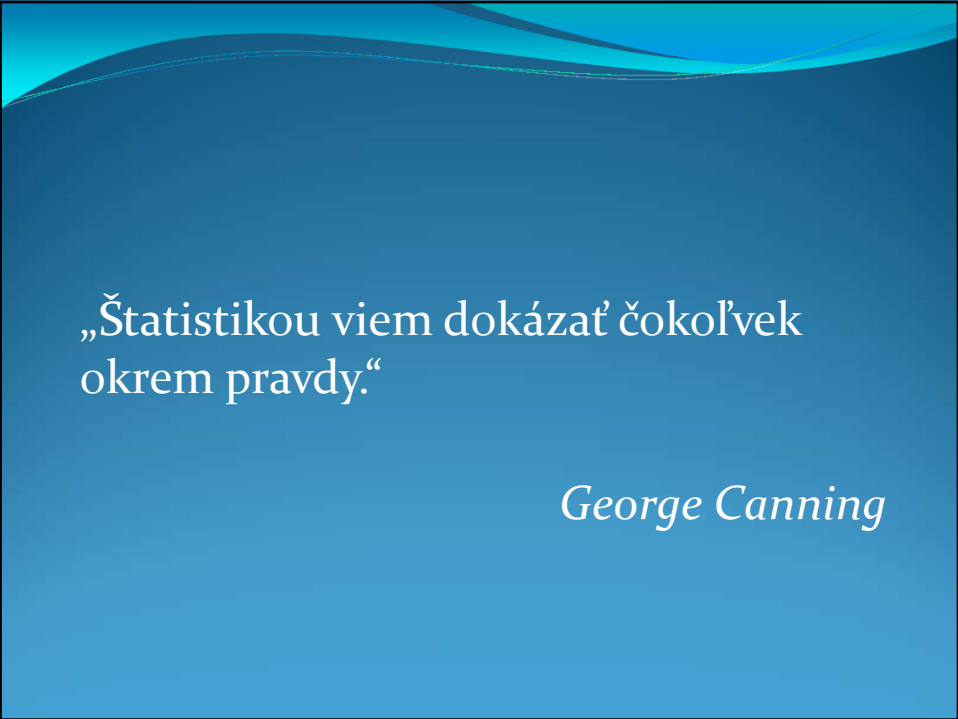
- Zámerné/nezámerné, náhodné
- Hľadanie spôsobov ako interpretovať len údaje, ktoré sú priaznivé pre prezentujúceho
- Falošné štatistiky – poškodzujú hľadanie poznania (náprava môže trvať desaťročia, stať veľa peňazí/životov)

- Preosievanie dát/fishing
 - hľadanie vzťahov a rozdielov, bez vopred stanoveného zámeru, hypotézy
 - 5% šanca nájdenia korelácie medzi 2 náhodnými premennými
- Manipulácia s dátami
 - selektívne reportovanie (uprednostňovanie výsledkov ktoré podporujú preferovanú hypotézu a ignorovanie protirečiacich; napr. experimenty s ESP)
 - vyrábanie falošných dát (exProf. Diederik Stapel – sociálny psychológ, NL, 2011, priznal sa k manipulácii dát po niekoľko rokov)



„Štatistiku často používame ako, keď
opitý človek používa stĺpy kandelábrov
– skôr pre svoju oporu než pre
osvetlenie.“

Andrew Lang



„Štatistikou viem dokázať čokoľvek
okrem pravdy.“

George Canning

Problémové učenie

Problémové učenie

- Úloha 1:
 - a) Vypočítajte vek respondentov z mesiaca a roku narodenia, ku šiestemu mesiacu roku 2010 (príkaz compute)
 - b) Kategorizujte kontinuálnu premennú veku do 3 kategórií 11-, 13- a 15-ročných (príkaz recode into different variable)

Problémové učenie

- Úloha 2:
- Vypočítajte hrubé skóre škály FAS (príkaz compute) (položky FAS.1-4)

Problémové učenie

- Úloha 3:
- Vytvorte premennú, ktorá by rozdeľovala respondentov na tých ktorí pijú alkohol (akýkoľvek) aspoň raz týždenne a tých, ktorý ho pijú menej často alebo vôbec – dichotomizácia (compute/recode)

Problémové učenie

- Úloha 4:
 - a) Popíšte súbor štúdie HBSC
 - Pohlavie (N, %)
 - Vek (priemer, SD)
 - Skóre FAS (priemer, SD, medián, modus, range)
 - Pije/nepije alkohol (N, %)
 - b) Popíšte vekové podskupiny (11, 13, 15-ročných)
 - Pohlavie (N, %)
 - Vek (priemer, SD)
 - Skóre FAS (priemer, SD, medián, modus, range)
 - Pije/nepije alkohol (N, %)

Problémové učenie

- Úloha 5:
- Vytvorte graf:
 - Koláčový pre pohlavie
 - Histogram a boxplot pre vek
 - Histogram a boxplot pre skóre FAS

Problémové učenie

- Úloha 6:
 - a) Existuje rozdiel v úrovni bohatstva (FAS) u tých ktorý pijú akýkoľvek alkohol aspoň raz týždenne a tými, ktorý ho pijú menej často/vôbec?
 - b) Vytvorte boxplot pre skóre FAS stratifikovaný podľa dichotomizovanej premennej pitia alkoholu

Problémové učenie

- Úloha 7:
 - a) Líšia sa respondenti v jednotlivých vekových kategóriách (11, 13, 15-roční) v úrovni rodinného blahobytu vyjadrenom škálou FAS? Ak áno, tak medzi ktorými dvojicami vekových kategórií je rozdiel.

Problémové učenie

- Úloha 8:
 - a) Overte internú konzistenciu škál:
 - SROS, S. fundamentalizmu, S. ortodoxnosti, Levenson LOC, R. salencie, S. zivotnej spokojnosti, God mediated control, Depresivita, Self-esteem
 - b) Vypočítajte ich hrubé skóre, v prípade viacdimeználnych škál – skóre pre subškály
 - c) Overte faktorovú štruktúru škál SROS, Levenson LOC, Self-esteem, S. fundamentalizmu

Problémové učenie

- Úloha 9:

Problémové učenie

- Úloha 10:

Ďakujem za pozornosť