

Kvantitativní výzkum

SZc033, SZ6031

Petr Květon

Úkol - kvantitativní metodologie

- Úkol je zadán ve dvou variantách – zpracujete pouze jednu variantu, která vám lépe vyhovuje
- Zpracovaný úkol vložíte do odevzdávárny v IS MU
- Pro práci v SPSS využijte stručný úvod na konci této prezentace a také tzv. cheat_sheet, který je ke stažení ve studijních materiálech
- Seznámení se SPSS na youtube:
 - <https://www.youtube.com/watch?v=ArVgSihGGXM>
 - https://www.youtube.com/watch?v=BMrQQ_jUH4E
- SPSS je pro studenty MU k dispozici zdarma ke stažení na adrese inet.muni.cz (nabídka software), případně je free alternativa PSPP na adrese <http://pspp.awardspace.info/>

Úkol - kvantitativní metodologie

- Varianta 1

- Využít data z pozorování jedné třídy
- Sestavit kvantitativní pozorovací protokol na vybrané jevy u žáků (minimálně 3)
- Provést deskriptivní (frekvenční) analýzu

- OBSAH PRÁCE

- Titulní strana
- Popis zkoumaného vzorku
- Popis metody sběru dat
- Prezentace výsledků
 - Tabulovaně
 - Graficky
 - Verbálně
- Závěr

Úkol - kvantitativní metodologie

- Varianta 2

- Využít data z matice videogames.sav (ke stažení ve studijních materiálech)
- Kompletní informace k jednotlivým proměnným jsou uloženy přímo v souboru
- Ve studijních materiálech je také soubor se stručným návodem k SPSS
- Analyzovat v SPSS
- Zvolit nejméně 4 proměnné, které vás zajímají a analyzovat (popsat deskriptivní statistikou)
- Pokuste se aspoň 1 proměnnou analyzovat dle skupin (např. dle pohlaví, ročníku, nebo typu školy)

- OBSAH PRÁCE

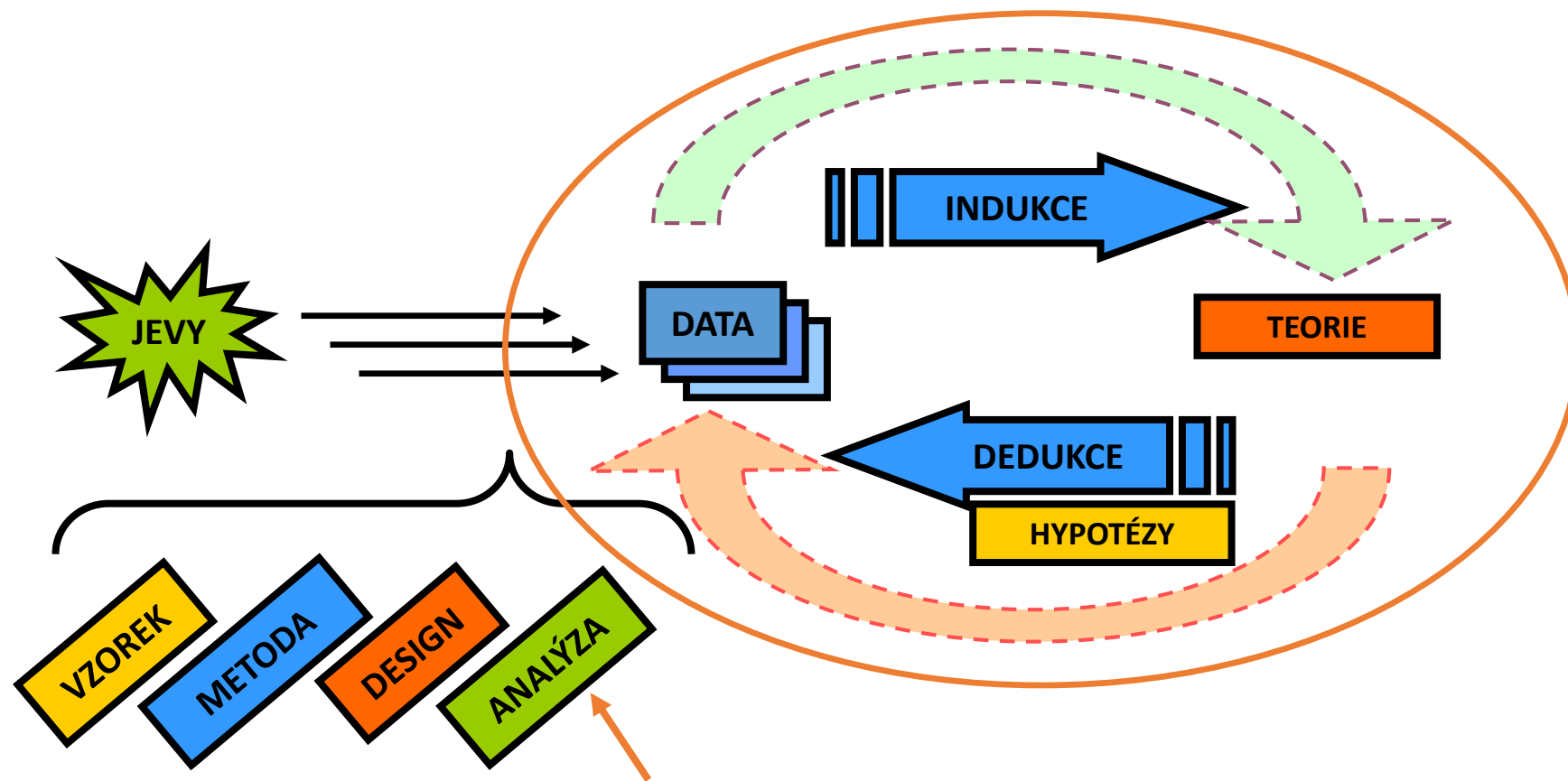
- Titulní strana
- Popis zkoumaného vzorku
- Prezentace výsledků
 - Tabulovaně
 - Graficky
 - Verbálně
- Závěr

Tabulka P1 Deskriptivní statistiky výzkumného souboru (uvedena jsou validní procenta)

	kategorie	Frequency
Pohlaví	žena	52,9%
	muž	47,1%
Typ školy	základní škola	65,3%
	střední odborné učiliště	9,1%
	střední odborná škola s maturitou	15,0%
	gymnázium	10,5%
Ročník	7. třída	33,0%
	9. třída (nebo její ekvivalent)	35,0%
	2. ročník SŠ	34,7%
Lokalita	Brno	36,9%
	Ostrava	28,8%
	Praha	34,3%
Rodinná situace	kompletní rodina	65,5%
	nekompletní rodina	34,5%
Nejvyšší vzdělání v rodině	základní nebo středoškolské bez maturity	17,0%
	středoškolské s maturitou	44,7%
	vysokoškolské	35,0%

Poznámka. Kompletní rodina - oba biologičtí rodiče žijí společně s dítětem; nejvyšší vzdělání v rodině - nejvyšší úroveň vzdělání v rodičovské dyádě.

Když se řekne... věda



AJ: phenomena, (empirical) data, facts, induction, theory, deduction, hypotheses, sample, method, design, analysis, exploratory and confirmatory research, statistics

Fáze výzkumu

- Následuje nejčastěji po všech přípravných fázích
 - Vymezení tématu
 - studium literatury (rešerše)
 - formulování výzkumné otázky a hypotéz
 - nastavení výzkumného designu
 - Výběr či tvorba výzkumného nástroje (operacionalizace)
 - volba výzkumného vzorku
 - ověřování konstrukce a vlastností výzkumného nástroje aj. v předvýzkumu
 - sběr dat

- Následuje fáze zpracování dat

Strategie a taktiky

- Strategie – výzkumný design
- Taktika – výzkumný nástroj

Kvantitativní výzkumné designy (strategie)

- **Experiment**

- Systematická výzkumná strategie manipulace s určitým jevem/jevy – NP
- kontrola jiných jevů (udržení na konstantní úrovni) - IP
- pozorování/měření výsledků – ZP

- explanační funkce experimentu umožňuje (za předpokladu, že jsme výzkum dobře naplánovali a provedli) určit příčinný vztah mezi jevy

Stanfordský vězeňský experiment

Kvantitativní výzkumné designy

- **Korelační studie**

- zabývá se zkoumáním vztahů mezi přirozeně se objevujícími proměnnými a studiem individuálních rozdílů; založena na měření proměnných a zjišťování vztahů mezi nimi
- úvaha o příčinných vztazích je vždy problematická – problém direkcionality zjištěn vztah mezi jevy A a B, je A příčinou B, nebo je tomu naopak?
- specifickým problémem je tzv. problém třetí proměnné jev C je společnou příčinou jevu A a B
- přes uvedené problémy mohou výsledky korelačních studií sloužit k predikci jevů

Specifické výzkumné strategie

- **Longitudinální studie**

- Sledování stejné skupiny osob po delší dobu (v extrémních případech po celou dobu života) – technické obtíže
- Umožňuje vysoudit kauzalitu

Výzkumné nástroje (taktiky)

Pozorování

Rozhovor

Analýza produktů činnosti

Dotazník

Test

Nástroje produkují data...

Hromadný sběr dat

- Dotazníky (ev. jiné nástroje generující množství dat)
 - Tužka –papír:
 - Sebereme a očíslovíme (typicky levý horní roh – např. 001 – 999)
 - Vytvoříme datovou matici („přepis dotazníku do PC“)
 - Google form či jiný / podobný elektronický sběr dat (předchozí krok odpadá)
 - Možno v Excelu (či LibreOffice Calculator) ev. rovnou ve statistickém programu
 - Kde vzít Excel - <https://it.muni.cz/sluzby/microsoft-office-365>
 - Kde vzít statistický software – viz návod doc. Vlčkové... nebo zkuste <https://www.gnu.org/software/pspp/>
 - První sloupec ID (číslo dotazníku/ů)
 - Další postupně jednotlivé odpovědi (např. „ot_1“ až „ot_99“)
 - Jednotlivé typy odpovědí kódujeme
 - Chybějící údaje kódujeme vysokým číslem, které nemůže být v intervalu použitého pro kódování (např. „99“ v případě likertovské škály míry souhlasu s výrokiem v dotazníku „1-5“)

Typy dat(proměnných)

Kategoriální - Diskrétní - Kvalitativní

- Binární / Nominální (kategoriální)
 - Binární: Pohlaví (m/ž), stav (ženatý/svobodný), kuřák/nekuřák
 - Nominální – obsahují více kategorií, které nelze vzájemně seřadit: krevní skupina, stav (ženatý/svobodný/rozvedený/ovdovělý), národnost (čech/slovák/vietnamec/rom), typ školy
- Ordinální (pořadová data)
 - Obsahují více kategorií, ale lze říci, která kategorie je vyšší, která nižší – např. dosažené vzdělání, pořadí na matematické olympiádě, **školní známky**, Likertovy škály
 - Možné statistické postupy: medián, kvantily; můžeme zjišťovat těsnost vztahů mezi proměnnými (používá se postupů adaptovaných pro tuto úroveň měření, např. Spearmanův pořadový korelační koeficient)

Typy dat(proměnných)

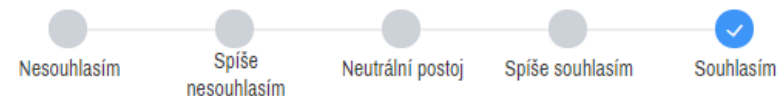
Spojité – Kvantitativní proměnné – Kardinální

- Intervalové a poměrové (kardinální)
 - Intervalové - známe rozdíly mezi sousedními body, jsou konstantní; nemají nulu (např. teplota na Celsiové nebo Fahrenheit stupnici)
 - Poměrové – stejné jako intervalové, ale mají stanovenou nulu (výška, váha, stupně Kelvina)
- Možné statistické postupy:
 - Nejvyšší úroveň měření – je možné počítat průměr, SD, a navazující postupy (parametrické porovnávání skupin, aj.)

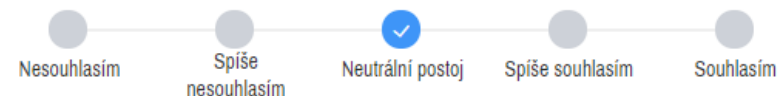
Zvláštnosti v klasifikaci proměnných

- Často jsou v praxi ordinální data ve výpočtech používána jako by byla spojitá, např. Likertovy škály (většinou od 7 bodů)
- Snížení úrovně měření, pokud je to užitečné a opodstatněné (většinou se nedoporučuje)
 - kardinální proměnná (intelligence)
 - -> ordinální (kategorie - slabomyslnost, podprůměr, průměr, nadprůměr, genialita)
 - -> dichotomická (geniální vs. zbytek)

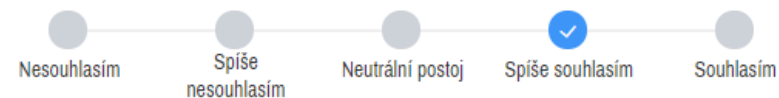
B4. Každý/každá, s kým v tomto podniku pracuji, má pocit, že provozní bezpečnost je jeho/její osobní odpovědností.



B5. Mému/mé nadřízenému/nadřízené na provozní bezpečnosti velmi záleží.



B6. Pracovníci a pracovnice mají vysoký stupeň důvěry ve vedení, pokud jde o provozní bezpečnost.



Jaký typ proměnné vznikne z každé položky dotazníku?

➤ Q1: Jaký je Tvůj oblíbený předmět?

Matematika Angličtina **Tělocvik** Fyzika Biologie

• Q2: Pohlaví:

Muž Žena

• Q3: Počet hodin strávených přípravou na zkoušku

Vypsat číslem

• Q4: Myslím si, že jsem dobrý v matematice:

Zcela souhlasím Souhlasím **Ani jedno ani druhé** Nesouhlasím Zcela nesouhlasím

Jaký typ proměnné vznikne z každé položky dotazníku?

➤ Q1: Jaký je Tvůj oblíbený předmět?

Matematika Angličtina Tělocvik Fyzika Biologie

Nominalní

• Q2: Pohlaví:

Muž Žena

Nominalní
(dichotomická)

• Q3: Počet hodin strávených přípravou na zkoušku

Vypsat číslem

poměrová

• Q4: Myslím si, že jsem dobrý v matematice:

Zcela souhlasím

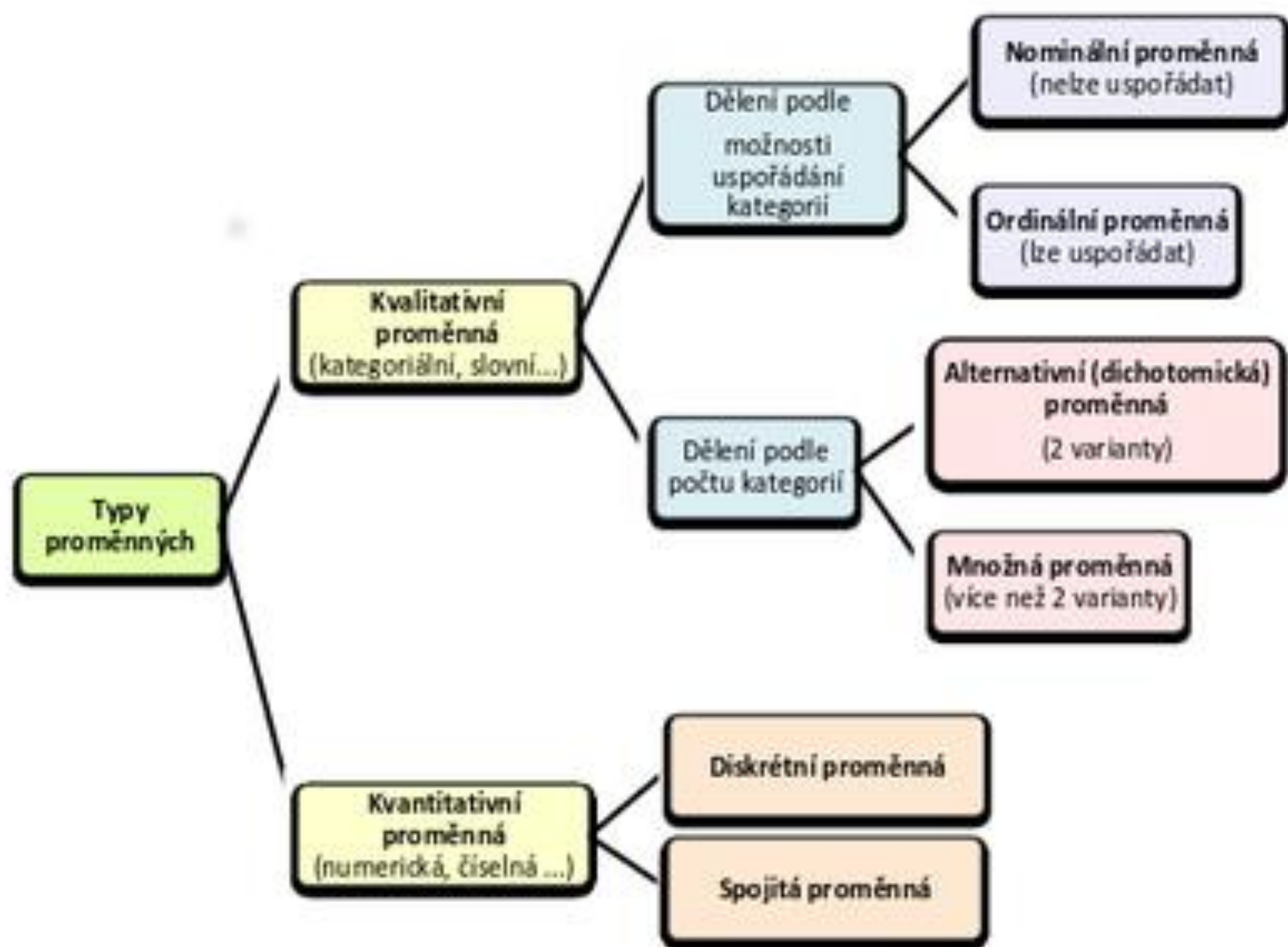
Souhlasím

Ani jedno
ani druhé

Nesouhlasím

Zcela nesouhlasím

Ordinalní



Popisná statistika

- rozdělení hodnot
- míry centrální tendence
- míry variability
- grafické zobrazení

Popisná statistika - Úvod

- užívá se k popisu základních vlastností dat
- poskytuje jednoduché shrnutí hodnot proměnných ve výběrovém souboru
- předchází indukční statistiku (která odvozuje zjištění ze vzorku na populaci)
- techniky deskriptivní statistiky pomáhají redukovat větší množství dat do zvládnutelné podoby
- touto redukcí např. údajů o rychlosti čtení u 200 žáků na jeden ukazatel, např. na hodnotu průměru, samozřejmě část informací ztratíme

Popisná statistika

- pro každou proměnnou obvykle popisujeme 3 charakteristiky
- rozdělení hodnot (i graficky), středovou hodnotu a míru rozptýlení hodnot kolem tohoto středu

Rozdělení hodnot

- rozdělení (distribuce) hodnot - souhrn četností jednotlivých kategorií nebo intervalů hodnot proměnné
- kromě grafů je základní možností, jak zobrazit rozložení hodnot proměnné **tabulka četností** – seznam kategorií proměnné a u nich počet osob, které do každé kategorie spadají

Rozdělení hodnot – tabulka četností

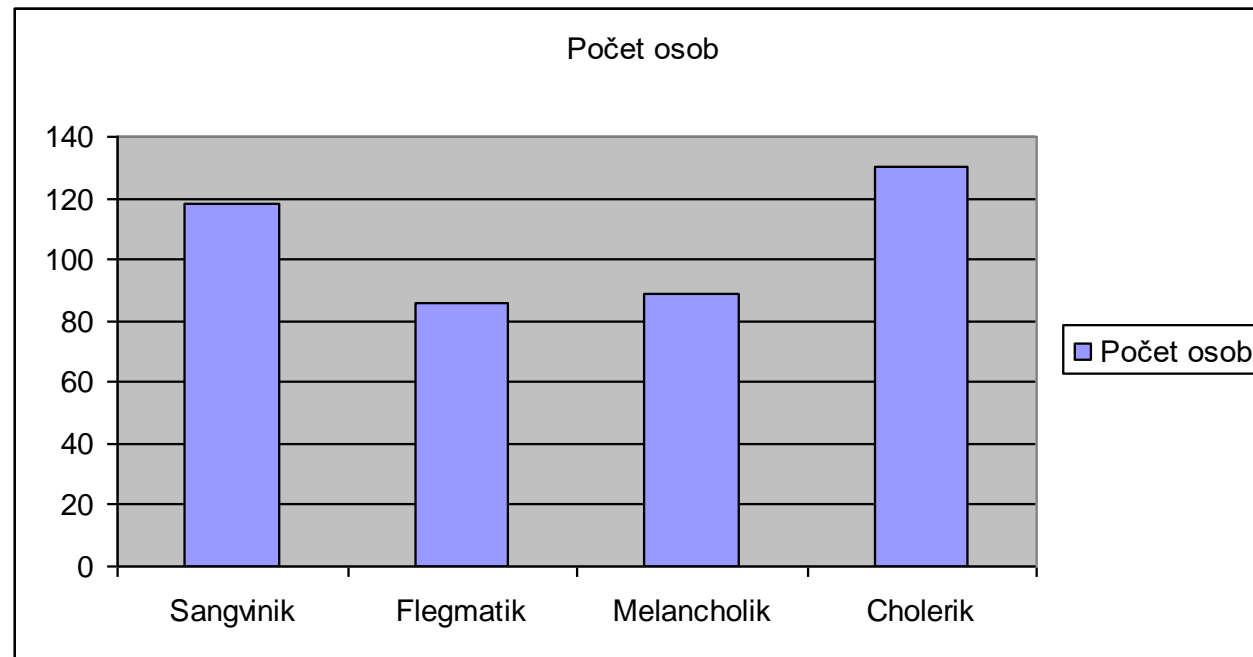
	počet osob	%
Sangvinik	118	28
Flegmatik	86	20
Melancholik	89	21
Cholerik	130	31
<i>celkem</i>	423	100

Rozdělení hodnot

- vždy je třeba uvést celkový počet osob (N)
- relativní četnosti mohou být uvedeny buď jako procenta (8%) nebo podíly (0.08)
- může jít rovněž o poměr (*ratio*) dvou kategorií (např. poměr dívek a chlapců s ADHD 1:4 (nebo 0,25))

Rozdělení hodnot

- stejná data je možno zobrazit i **graficky** (v příkladu sloupcový diagram – barchart)



Rozdělení hodnot

- pokud proměnná nabývá mnoha hodnot, je vhodnější je **sloučit do kategorií (intervalů)**
- počet intervalů by měl být přiměřený počtu hodnot
- někdy se používá tzv. Sturgesovo pravidlo $k = 1 + 3,3 \log_{10}(n)$
- podle něj by pro 200 hodnot byl vhodný počet intervalů 9

Rozdělení hodnot

IQ	počet	%	<i>kumul.%</i>
méně než 86	11	10	10
86 – 100	36	34	44
101 – 115	34	32	76
116 – 130	20	19	95
131 a více	5	5	100
<i>celkem</i>	<i>106</i>	<i>100</i>	

Míry centrální tendence

- míry centrální tendence (středu, polohy) jsou výsledkem snahy najít typickou hodnotu pro daný znak
- nejčastěji používané modus, medián, aritmetický průměr, méně často harmonický a geometrický průměr

Míry centrální tendence

- **modus** – nejčastěji se vyskytující hodnota (např. u příkladu s temperamentem to byl *choleric*)
- jediná použitelná charakteristika polohy pro nominální data; u pořadových a kardinálních (intervalových nebo poměrových) jsou většinou více typickými charakteristikami medián nebo průměr

Míry centrální tendence

- **medián** - prostřední hodnota v řadě hodnot uspořádaných podle velikosti (50. percentil)
- je jen pro data, která je možno podle velikosti uspořádat, tj. pořadová a kardinální
- dělí soubor na dvě poloviny (pro sudý počet hodnot je medián průměrem dvou prostředních pozorování)

Míry centrální tendence

Medián

- používá se především, pokud chceme eliminovat vliv extrémních hodnot
- příklad – průměrný plat 20 tisíc může u 10 osob znamenat, že 9 z nich má 10 tisíc a jedna 110 tisíc; použijeme-li medián – 10 tisíc, získáme více typickou hodnotu
- můžeme ho vyčíst z tabulky četností, pokud jsou uvedeny kumulativní četnosti
- <https://www.czso.cz/csu/czso/cri/prumerne-mzdy-2-ctvrtleti-2019>

Míry centrální tendence

- **aritmetický průměr** – součet všech hodnot znaku dělený jejich počtem
- lze použít u kardinálních proměnných
- vzorec: $\mu = \Sigma X/N$ (pro populaci)
- nebo $m = \Sigma x/n$ (pro výběr)

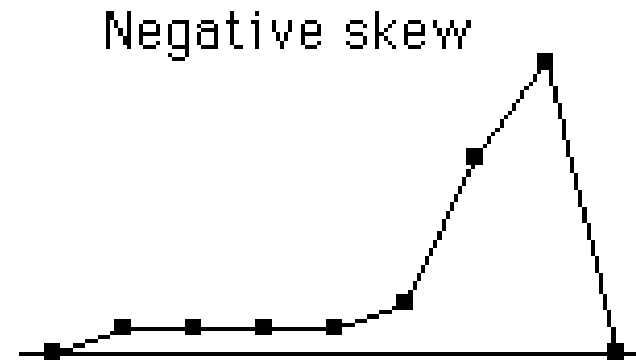
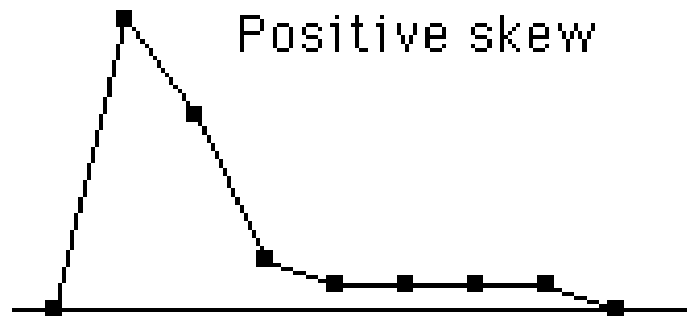
Míry centrální tendence

- průměr zahrnuje každou hodnotu znaku – což je jak výhoda, tak nevýhoda (citlivý na extrémní hodnoty)
- to je možno vyřešit použitím tzv. useknutého průměru (*trimmed mean*), který se počítá tak, že se vynechá určité % hodnot z obou stran rozdělení, např. 5% nejnižších a 5% nejvyšších (běžná praxe v hodnocení na ZŠ – škrtne se nejhorší známka).
- průměr špatně reprezentuje nehomogenní skupiny
- příklad – 30 osob v parku, průměrný věk 12.5 roku, průměrná výška 130 cm: nemusí jít o školní děti, ale o 15 matek se 4-letými dětmi

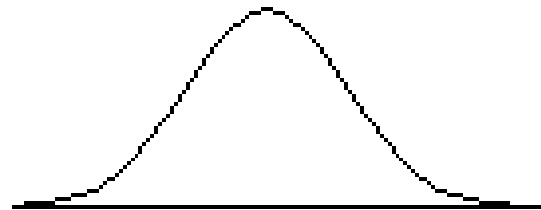
Míry centrální tendence

- porovnáním hodnoty průměru a mediánu získáme představu o šikmosti rozdělení hodnot
- pokud je průměr větší než medián – kladně (doprava) zešikmeno
- průměr menší než medián – záporně (doleva) zešikmeno
- průměr = medián – symetrické rozdělení

Míry centrální tendence



Symmetric distribution
(No skew)



Míry centrální tendence

- pro znaky s normálním rozdělením hodnot je průměr **nejúčinnější** charakteristikou (tj. nejvíce stabilní pro různé výběrové soubory) – dá se nejlépe použít pro odhad parametru populace z charakteristik výběru
- je nejčastěji užívanou mírou centrální tendence

Míry centrální tendence

- kterou statistiku uvádět v případě, že se můžete rozhodnout?
- **průměr** – pokud může být spočítán a pokud není rozdělení příliš šikmé
- **modus** – pokud je rozdělení multimodální (neexistuje jediná typická hodnota)
- **medián** – pokud je rozdělení šikmé a unimodální

Míry centrální tendence

- **příklad** – určete modus, medián

18 5 128 2 14 87 50 87 70

Míry centrální tendence

- modus = 87 (2x)

- medián =

2 5 14 18 **50** 70 87 87 128

- průměr = $461/9 = 51,22$

Míry variability

- míry variability popisují kolísání
v rozdělení hodnot
- užívá se rozpětí, mezikvartilové rozpětí, rozptyl,
směrodatná odchylka

Míry variability

- **rozpětí** (variační šíře, variační rozpětí) – rozdíl mezi nejvyšší a nejnižší hodnotou
- značně ovlivněno extrémními hodnotami, není dobrým odhadem parametru populace

Míry variability

- **mezikvartilové rozpětí** (interkvartilová odchylka) – rozdíl mezi hodnotou horního kvartilu a dolního kvartilu
- **kvartily** – dělí soubor na 4 stejné části; horní kvartil odděluje 25% nejvyšších hodnot (75. percentil), dolní 25% nejnižších (25. percentil)
- mezikvartilové rozpětí udává rozpětí pro středních 50% hodnot (=délka obdélníku v krabicovém diagramu)
- není (podobně jako medián) citlivé na extrémní hodnoty

Míry variability

- **rozptyl** (střední kvadratická odchylka průměru) - ukazuje, jak jsou hodnoty rozptýleny kolem průměru

- v populaci

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$$

- ve výběru

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2$$

Míry variability

- více než rozptyl se používá jeho odmocnina – **směrodatná odchylka průměru**
- oba ukazatele slouží jako vhodné doplnění průměru – získáme představu o jeho věrohodnosti, tj. jak dobře reprezentuje všechny hodnoty

Míry variability

- příklad – porovnejte variabilitu u těchto dvou rozložení hodnot (jde o počet správně vyřešených úloh v didaktickém testu u výběru osob ze dvou tříd ZŠ)

a)

4 5 4 3 5 5 3 4 3

b)

8 2 12 1 4 3 5 0 1

Míry variability

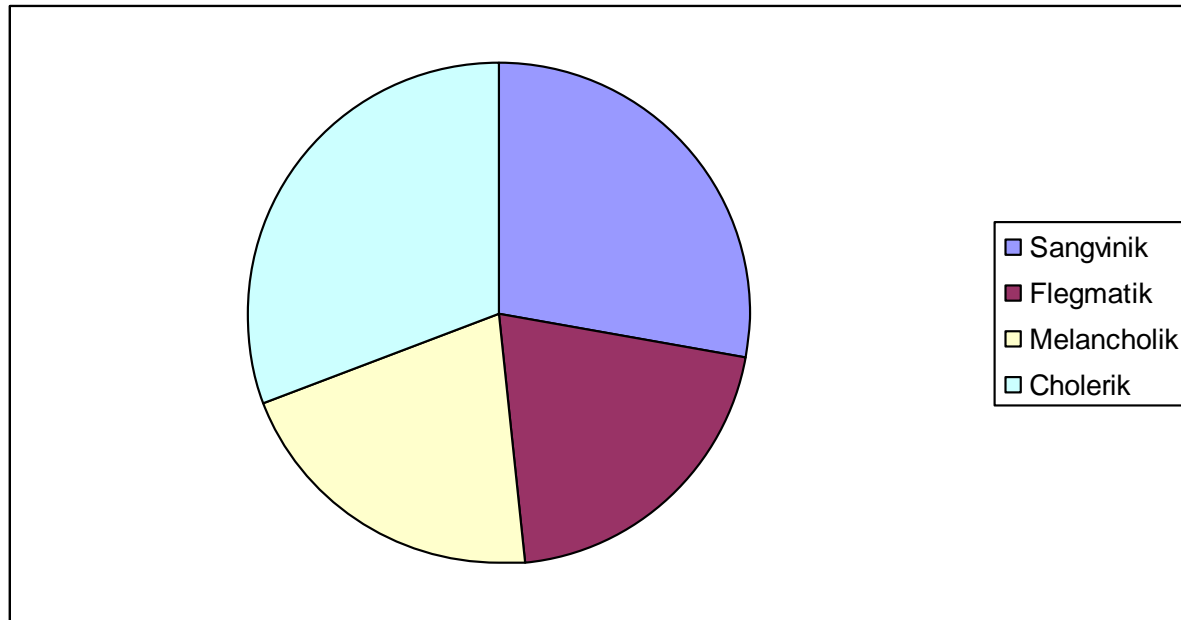
- řešení příkladu
- $m_a = 4, s_a = 0.87$
- $m_b = 4, s_b = 3.87$
- u prvního rozdělení je průměr lepší reprezentací hodnot; u druhého jsou hodnoty kolem průměru hodně rozptýleny

Grafy

- pouze základní typy
- pro kategoriální data - sloupcový diagram, výsečový graf
- pro intervalová data – histogram, krabicový diagram
- grafy je možno znázornit v kategorizované formě – pro jednotlivé kategorie další proměnné (např. pro muže a ženy)
- grafy pro vztah dvou a více proměnných budou probrány později

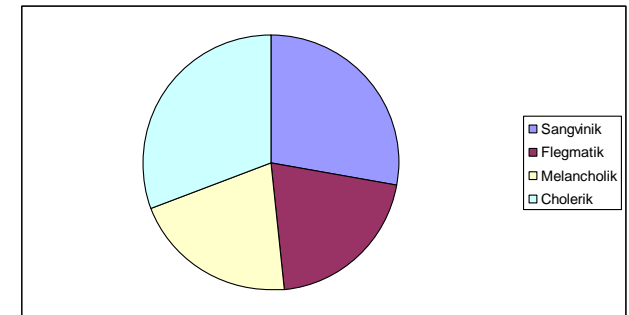
Výsečový graf

- koláčový diagram, pie chart – užívá se více v populárních publikacích než v odborných



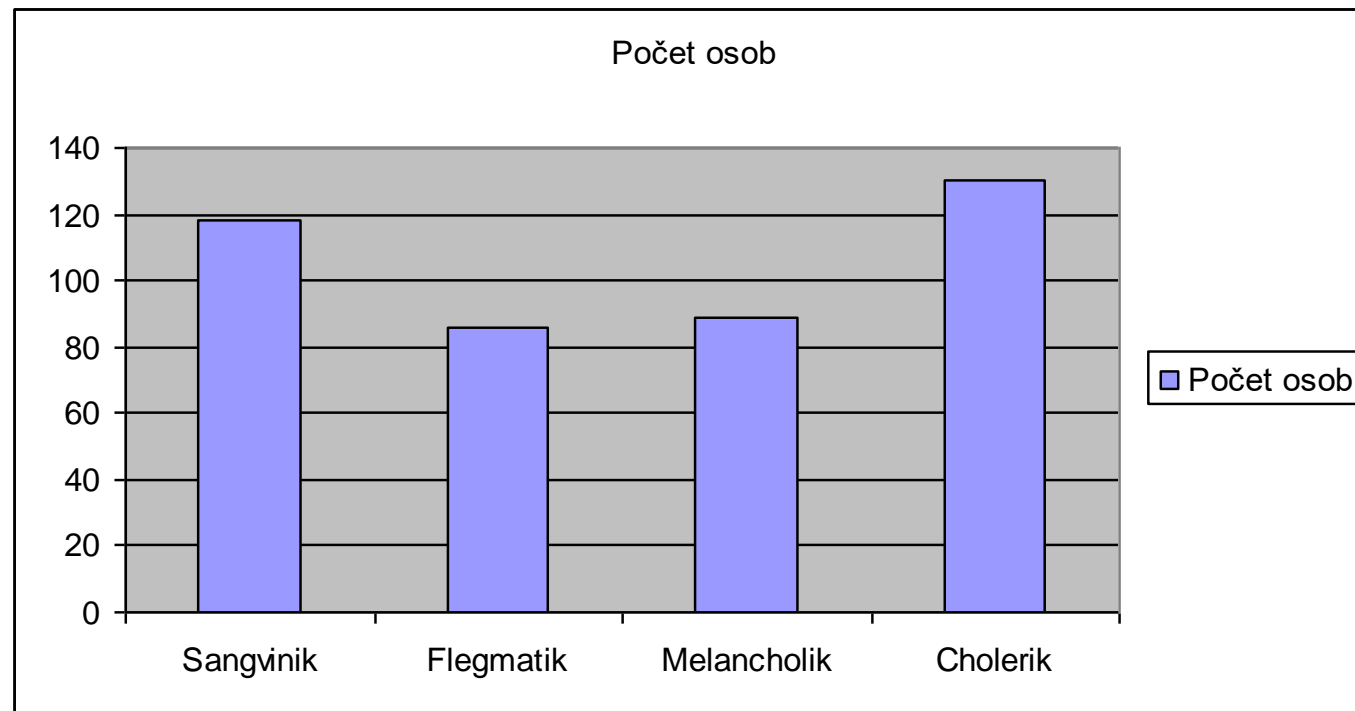
Výsečový graf

- každá výseč by měla být označena % a uveden celkový počet případů
- ideální pro 3-7 kategorií
- **výhody:** srozumitelný
- **nevýhody:** jen pro kategoriální data; neukazuje přesné údaje (pokud nejsou vyznačeny); srovnání více skupin osob problematické



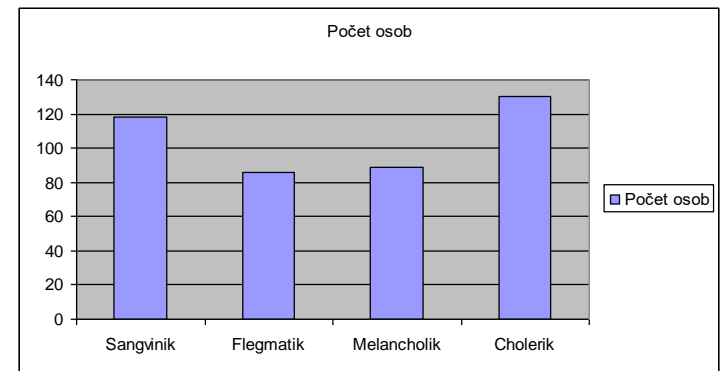
Sloupcový diagram

- bar chart



Sloupcový diagram

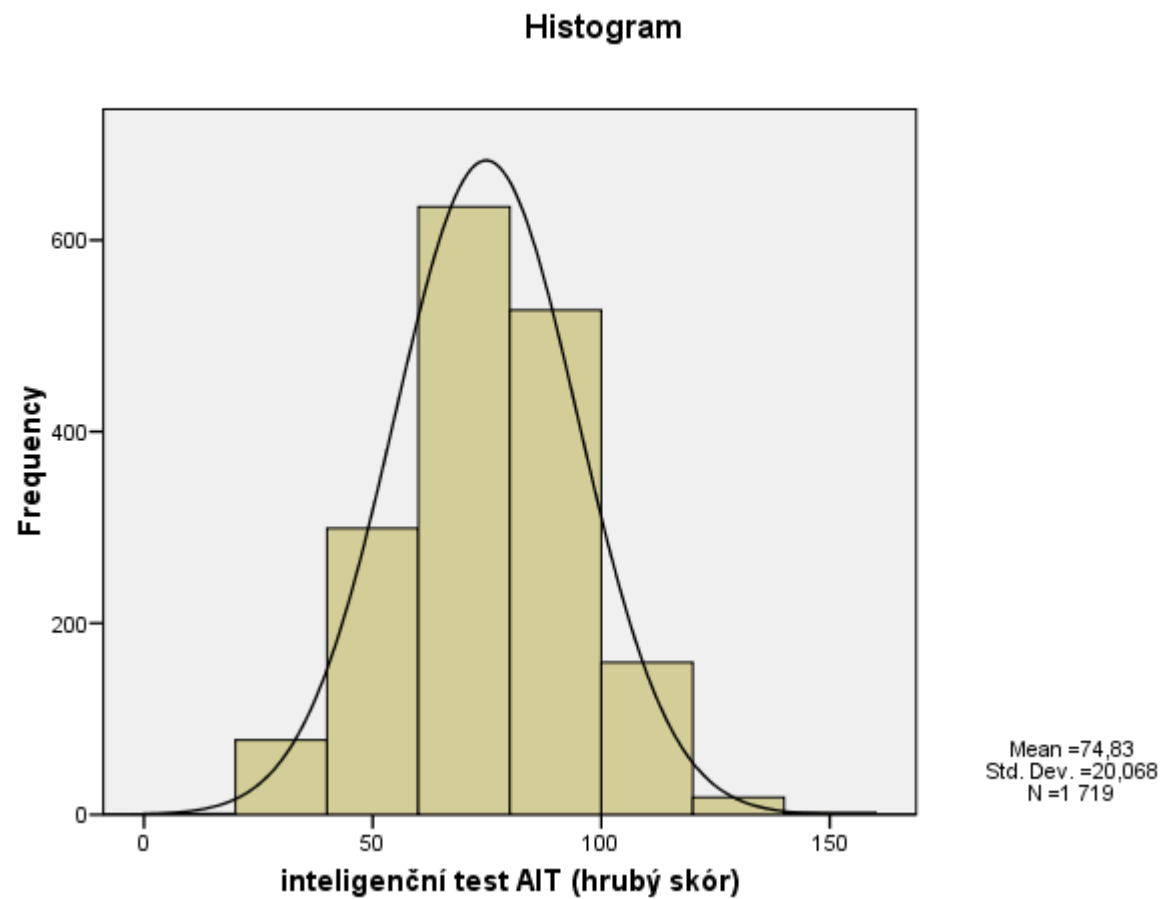
- pro kategoriální data, může být orientován horizontálně či vertikálně
- jednotlivé sloupce odděleny mezerou
- **výhody:** srozumitelný, je možno v jednom grafu porovnat četnosti pro více skupin osob



Histogram

- často užívaný
- podobný sloupcovému diagramu, ale je pro intervalová data
- jednotlivé sloupce reprezentují nikoliv jednotlivé kategorie, ale intervaly hodnot (sloupce jsou bez mezer)
- tvar histogramu závisí také na šířce intervalů
- **výhody**: umožňuje detekovat odlehlá pozorování, srovnání s normálním rozdělením
- **nevýhody**: nezjistíte přesné hodnoty jednotlivých případů, obvykle se nezobrazují data pro více skupin případů

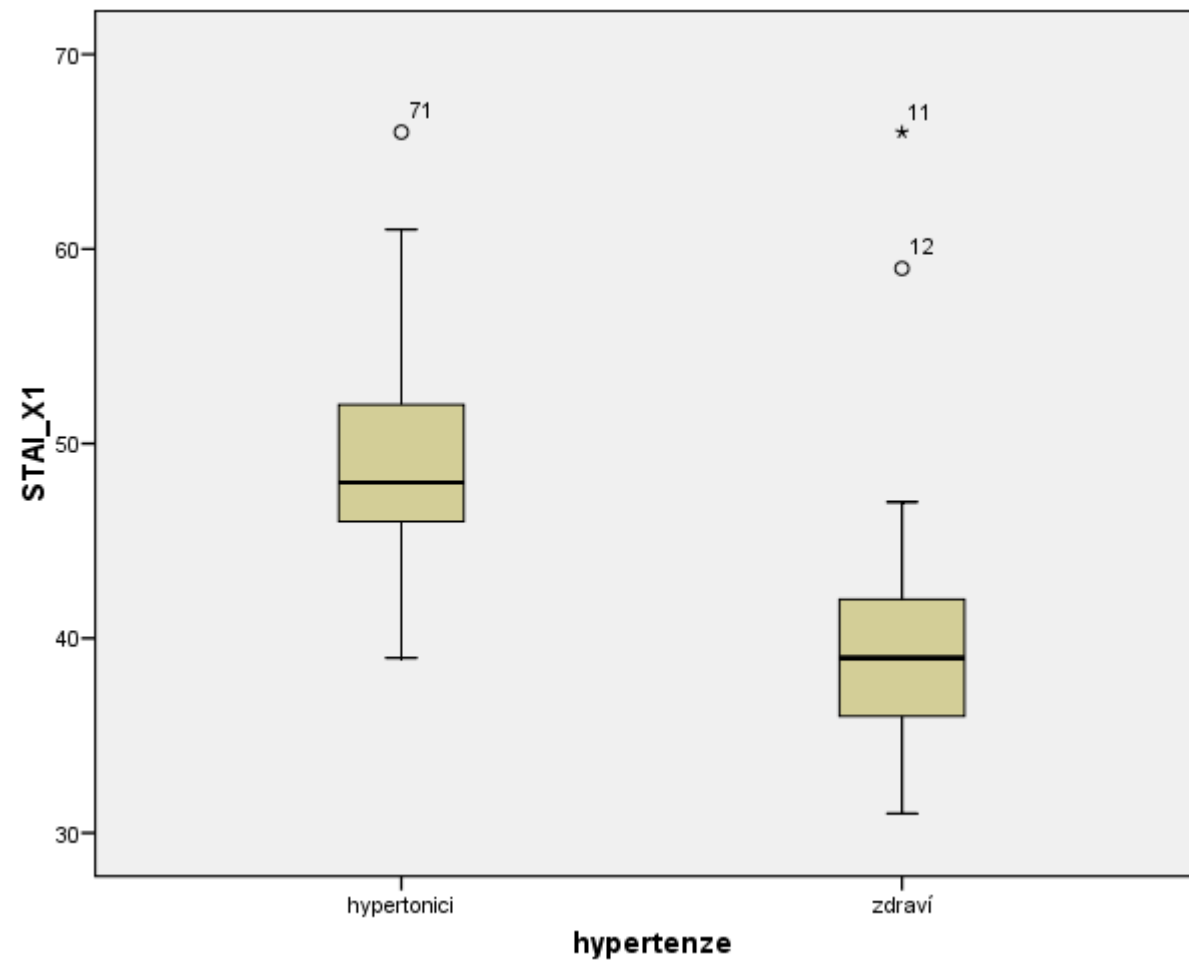
Histogram



Krabicový diagram

- boxplot, vousatá krabička
- poskytuje bohaté zobrazení důležitých aspektů rozdělení hodnot
- délka krabice odpovídá interkvartilové odchylce; uvnitř krabice je vyznačen medián
- v některých variantách grafu jde např. o směrodatnou odchylku a průměr
- „vousy“ je ohraničeno rozmezí hodnot

Krabicový diagram



Induktivní statistika

- Deskriptivní statistika se týká výhradně výběrového souboru (vzorku)
- Induktivní statistika se snaží zobecnit zjištěné údaje na celou uvažovanou populaci.
- Základním nástrojem je tzv. testování hypotéz
 - Většinou vychází z předpokladu, že neexistuje vztah mezi jevy nebo není rozdíl mezi skupinami
 - Statistická významnost – umožňuje rozhodnout, zda lze přijmout postulovanou hypotézu

Testování hypotéz

- Test chí-kvadrát: doplněk kontingenční tabulky – vztah dvou kategoriálních proměnných. Existuje rozdíl mezi chlapci a dívkami v oblíbenosti učebních předmětů?
- Porovnání průměrů: T-test či analýza rozptylu (experimentální vs. Kontrolní skupina)
- Korelační koeficient: Existence vztahu mezi jevy
 - Rozsah -1 až 1 (0 = žádný vztah)
- A jiné...

Pearsonův korelační koeficient

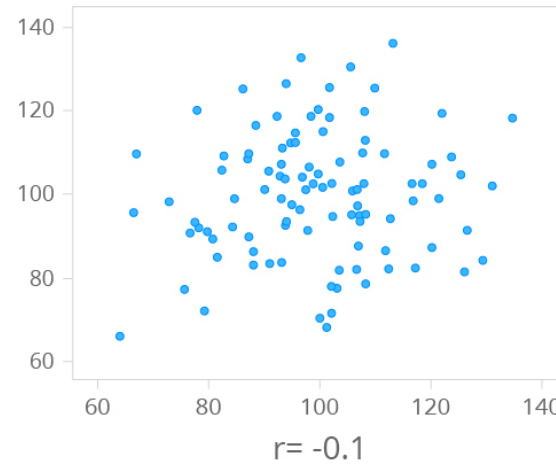
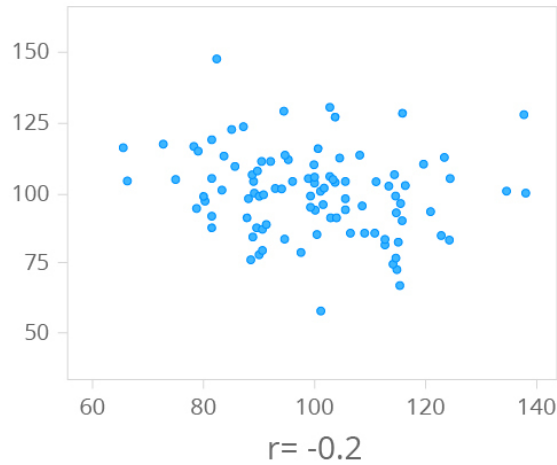
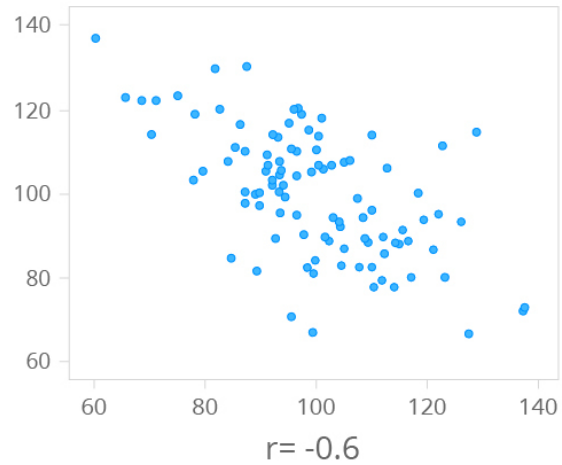
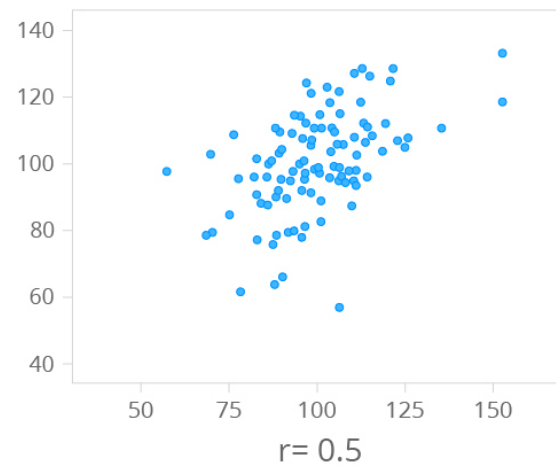
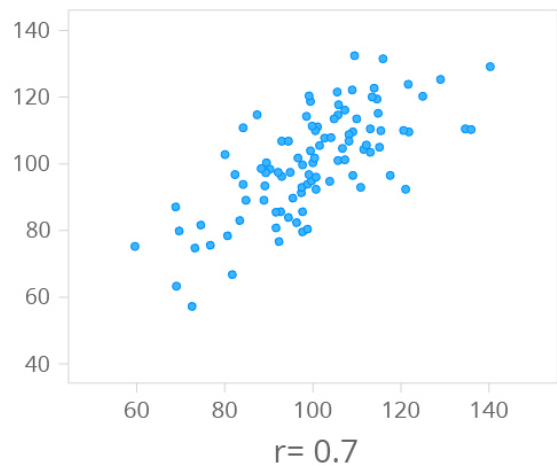
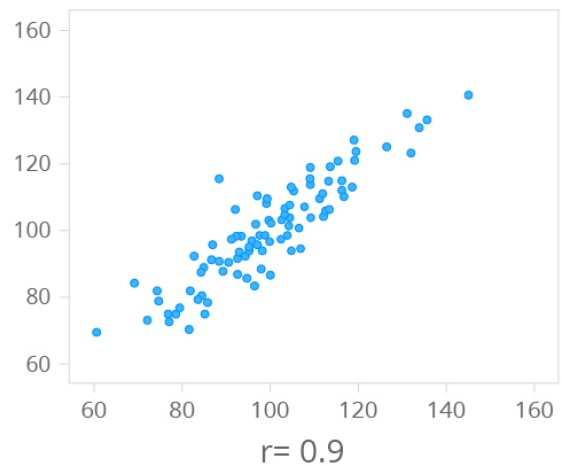
- u intervalových a poměrových dat můžeme jako míru asociace - vztahu mezi proměnnými - použít **Pearsonův korelační koeficient**
- **korelace**
 - *ko* = s, spolu, vzájemně
 - *relace* = vztah
 - *korelace* = vzájemný vztah proměnných

Pearsonův korelační koeficient

- absolutní hodnota koeficientu vyjadřuje **sílu (těsnost) vztahu**
- znaménko (+ nebo -) **směr vztahu**
- **rozsah -1 až +1**
- označuje se **r**

Pearsonův korelační koeficient

- sám o sobě je deskriptivní statistikou, ale podobně jako u ostatních měř asociace je možno spočítat **statistickou významnost (=zda se se významně liší od nuly**, tj. zda nějaký vztah mezi proměnnými vůbec existuje)
- závisí na velikosti výběru – čím vyšší, tím nižší koeficient vychází průkazný
- je mírou **pouze pro lineární vztahy**
- před výpočtem je vhodné zobrazit vztah mezi proměnnými graficky – tzv. **scatter** (dvourozměrný bodový diagram)

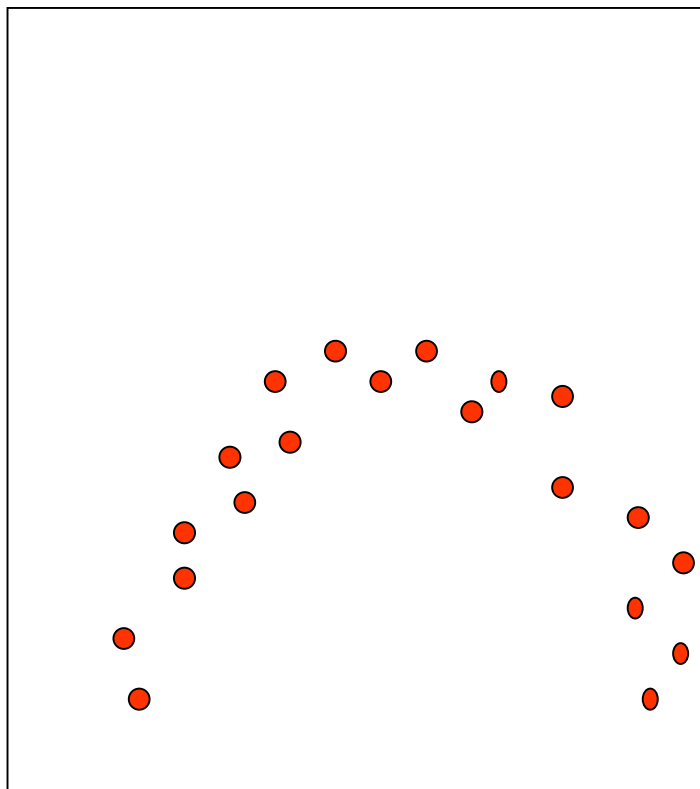


Interpretace r

- není shoda v tom, jaká hodnota r je považována za těsný vztah
- interpretace navržená Guilfordem:
 - <0.20 zanedbatelný vztah
 - $0.20-0.40$ nepříliš těsný vztah
 - $0.40-0.70$ středně těsný vztah
 - $0.70-0.90$ velmi těsný vztah
 - >0.90 extrémně těsný vztah

Scatter

- **nelineární vztah**
- $r = 0$



Excel

- Není statistický program
- Mnohé jde zpracovat i v Excelu
- Problém, jak pracovat s chybějícími daty
 - ručně

Microsoft Excel - database_disertacka

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

Nápověda - zadejte dotaz

C616

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Cislo	škola	trida	hlavi	doxa-Aj	znamka-Aj	znalost-Aj	mluveni-Aj	cteni-Aj	psani Aj	poslech-Aj	doxa-Nj	znamka-Nj	znalost-Nj	mluveni NJ	cteni-Nj	psani-Nj	poslec
2	1	Taborska	3.A	z	3	2	3	2	1	2	3	8	3	4	4	3	4	
3	2	Taborska	3.A	m	11	3	3	2	2	2	3	3	4	4	4	4	4	
4	3	Taborska	3.A	m	10	2	3	2	2	2	2	3	2	4	3	3	3	
5	4	Taborska	3.A	m	10	4	4	4	4	3	4	3	4	4	4	4	4	
6	5	Taborska	3.A	z	10	3	4	4	4	3	4	3	3	4	4	4	4	
7	6	Taborska	3.A	z	3	2	4	3	3	3	4	10	2	3	2	3	2	
8	7			m	3	3	3	3	3	3	3	7	3	3	2	3	2	
9	8	Taborska	3.A	m	7	3	2	2	1	3	1	3	2	3	2	3	3	
10	9	Taborska	3.A	m	8	3	3	2	2	2	2	3	2	4	3	3	3	
11	10	Taborska	3.A	m	10	2	2	2	2	1	2	6	2	4	2	3	1	
12	11	Taborska	3.A	z	8	3	4	4	3	4	2	6	1	3	2	2	3	
13	12	Taborska	3.A	m	3	2	3					9	3	2	2	2	4	
14	13	Taborska	3.A	m	3	3	4	3	2	3	3	9	2	3	2	1	3	
15	14	Taborska	3.A	z	8	3	3	3	3	4	4	3	2	4	4	4	4	
16	15	Taborska	3.A	z	3	3	4	4	3	3	3	11	2	3	3	2	1	
17	16	Taborska	3.A	m	3	3	4	4	3	4	3	10	2	3	3	3	2	
18	17	Taborska	3.B	z	10	2	3	3	2	2	3	3	2	4	4	3	3	
19	18	Taborska	3.B	z	10	4	3	3	2	2	3	3	3	4	4	4	4	
20	19	Taborska	3.B	z	10	3	3	3	2	3	4	3	2	4	4	3	3	
21	20	Taborska	3.B	z	8	3	3	3	2	4	3	3	3	4	4	4	4	
22	21	Taborska	3.B	m	10	2	2	3	1	2	3	3	3	3	4	3	3	
23	22	Taborska	3.B	m	12	1	2	2	1	1	1	7	2	3	3	1	3	
24	23	Taborska	3.B	z	10	1		3	2	3	3	3	1	1	1	1	1	
25	24	Taborska	3.B	m	8	2	2	2	1	1	2	3	2	3	3	3	3	
26	25	Taborska	3.B	z	10	2	4	4	2	4	2	5	2				4	
27	26	Taborska	3.B	z	10	2	2	3	3	3	3	3	1				4	
28	27	Taborska	3.B	z	9	1	3	3	2	3	2	4	2				4	
29	28	Taborska	3.B	z	10	3	4	3	3	3	4	3	2				3	
30	29	Taborska	3.B	m	6	3	2	3	3	3	2	3	3				4	
31	30	Taborska	3.B	m	11	2	2	2	2	3	1	3	2	4	4	4	4	
32	31	Taborska	3.B	z	10	2	3	3	2	3	3	3	3	4	4	3	3	

database List2 List3

Kreslení Automatické tvary

Připraven

**datová
matice**

Microsoft Excel - database_disertacka

Nápověda – zadejte dotaz

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

Seřadit...
Filtr
Formulář...
Souhrny...
Ověření...
Tabulka...
Text do sloupců...
Sloučit...
Skupina a přehled
Kontingenční tabulka a graf...
Importovat externí data
Aktualizovat data

	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE
1	otiv-kultura	otiv-zajem	partn.skola	otiv-rodice	otiv-DA																							
2	0	0	0	0																								
3	0	1	0	1																								
4	0	0	0	0																								
5	0	0	0	1																								
6	0	0	0	1																								
7	0	0	0	0																								
8	0	0	0	0																								
9	0	0	0	0																								
10	0	0	0	0																								
11	0	0	0	0																								
12	0	0	0	0																								
13	0	0	0	0																								
14	1	1	0	0	0																							
15	0	0	0	1	0																							
16	0	1	0	0	0																							
17	0	0	0	0	0																							
18	0	0	0	0	0																							
19	0	1	0	0	0																							
20	1	1	0	0	0																							
21	0	1	0	0	1	cteni knize																						
22	0	0	0	0	1	potrebuji to																						
23	1	0	0	1	0																							
24	1	1	0	0	0																							
25	0	0	0	1	0																							
26	0	0	0	1	0																							
27	0	0	0	0	1	pribuzni v d																						
28	1	1	0	0	0																							
29	0	0	0	0	0																							
30	0	0	0	0	0																							
31	1	0	0	0	0																							
32	0	0	0	0	0																							

Data lze filtrovat – vybrat si pro analýzy jen např. dívky

database / List2 / List3

Kreslení Automatické tvary

Připraven

The image shows a screenshot of the Microsoft Excel application window. The title bar reads "Microsoft Excel - databaze_disertacka". The menu bar includes "Soubor", "Úpravy", "Zobrazit", "Vložit", "Formát", "Nástroje", "Data", "Okno", and "Nápověda". The "Vložit" menu is open, displaying options such as "Buňky...", "Řádek", "Sloupec", "List", "Graf...", "Symbol...", "Konec stránky", "Funkce..." (highlighted), "Název", "Komentář", "Obrázek", "Diagram", "Objekt...", and "Hypertextový odkaz... Ctrl+K". The spreadsheet grid shows columns A through R and rows 600 through 631. Column C is highlighted in yellow. A text box with a brown border on the right side of the spreadsheet contains the text: "Pomocí Excelu lze počítat mnohé statistické funkce". The taskbar at the bottom shows the Start button and several open applications: "stare database", "zaklady_statistiky -...", "stat_zprac_dat - Mi...", "Microsoft PowerPoi...", and "Microsoft Excel - ...". The system tray shows the time as 9:32.

Vyhledat funkci:

Zadejte stručný popis požadované činnosti a potom klepněte na tlačítko Přejít.

Přejít

Vybrat kategorii: Statistické

Vybrat funkci:

MAX
MAXA
MEDIAN
MIN
MINA
MODE
NEGBINOMDIST

Použití funkcí v
Excelu

AVERAGEA(hodnota1;hodnota2;...)

Vrátí průměrnou hodnotu (aritmetický průměr) argumentů. Text a logická hodnota NEPRAVDA mají hodnotu 0, logická hodnota PRAVDA má hodnotu 1. Argumenty mohou být čísla, názvy, matice nebo odkazy.

[Nápověda k této funkci](#)

OK

Storno

Microsoft Excel - Kopie - PISA strategie

Nápověda - zadejte dotaz

S3

Table 4.5
Index of control strategies and performance on the combined reading literacy scale, by national quarters of the index
Results based on students' self-reports

Tabulka výsledků

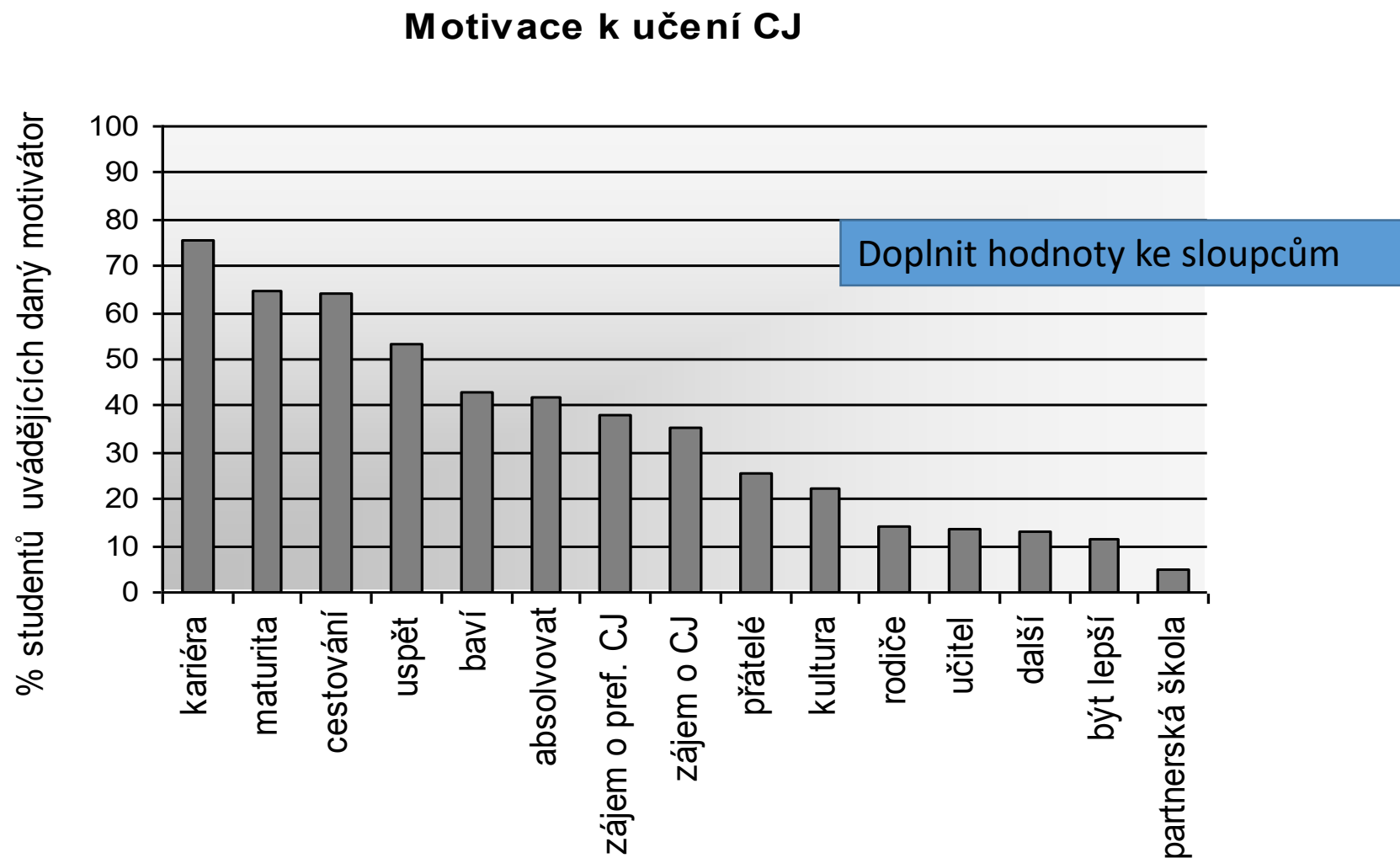
Country	Index of control strategies ¹													
	All students		Males		Females		Bottom quarter		Second quarter		Third quarter		Top quarter	
	Mean index	S.E.	Mean index	S.E.	Mean index	S.E.	Mean index	S.E.	Mean index	S.E.	Mean index	S.E.	Mean index	S.E.
OECD Countries														
Australia	0,02	(0,02)	-0,05	(0,02)	0,10	(0,03)	-1,20	(0,02)	-0,26	(0,01)	0,30	(0,01)	1,24	(0,02)
Austria	0,40	(0,02)	0,31	(0,03)	0,48	(0,02)	-0,72	(0,02)	0,12	(0,01)	0,64	(0,01)	1,54	(0,02)
Belgium (F)	0,14	(0,02)	0,07	(0,02)	0,21	(0,03)	-0,96	(0,02)	-0,11	(0,01)	0,37	(0,01)	1,25	(0,02)
Czech Rep	0,27	(0,02)	0,11	(0,03)	0,42	(0,02)	-0,84	(0,01)	-0,02	(0,00)	0,50	(0,01)	1,45	(0,02)
Denmark	-0,23	(0,01)	-0,24	(0,02)	-0,22	(0,02)	-1,25	(0,02)	-0,48	(0,01)	0,00	(0,01)	0,81	(0,02)
Finland	-0,47	(0,02)	-0,52	(0,02)	-0,42	(0,02)	-1,54	(0,02)	-0,71	(0,01)	-0,21	(0,01)	0,58	(0,02)
Germany	0,24	(0,02)	0,14	(0,04)	0,33	(0,02)	-0,94	(0,02)	-0,05	(0,01)	0,48	(0,00)	1,45	(0,03)
Hungary	0,21	(0,02)	0,09	(0,03)	0,33	(0,03)	-0,91	(0,02)	-0,05	(0,01)	0,45	(0,01)	1,34	(0,02)
Iceland	-0,35	(0,02)	-0,36	(0,03)	-0,34	(0,02)	-1,53	(0,02)	-0,60	(0,01)	-0,08	(0,01)	0,79	(0,02)
Ireland	0,07	(0,02)	-0,10	(0,04)	0,23	(0,03)	-1,28	(0,03)	-0,22	(0,01)	0,37	(0,01)	1,39	(0,02)
Italy	0,23	(0,02)	0,05	(0,04)	0,41	(0,02)	-0,94	(0,03)	-0,04	(0,01)	0,48	(0,00)	1,41	(0,01)
Korea	-0,44	(0,02)	-0,47	(0,03)	-0,41	(0,03)	-1,74	(0,02)	-0,70	(0,01)	-0,12	(0,01)	0,79	(0,02)
Luxembou	0,05	(0,02)	-0,10	(0,03)	0,19	(0,03)	-1,26	(0,03)	-0,23	(0,01)	0,33	(0,01)	1,34	(0,03)
Mexico	0,16	(0,02)	0,06	(0,03)	0,25	(0,02)	-0,98	(0,01)	-0,18	(0,01)	0,39	(0,01)	1,41	(0,02)
New Zeala	0,07	(0,03)	-0,03	(0,03)	0,17	(0,03)	-1,12	(0,02)	-0,23	(0,01)	0,30	(0,01)	1,32	(0,02)
Norway	-0,58	(0,02)	-0,50	(0,03)	-0,66	(0,02)	-1,76	(0,02)	-0,81	(0,01)	-0,28	(0,01)	0,54	(0,02)
Portugal	0,19	(0,02)	0,03	(0,02)	0,34	(0,02)	-0,90	(0,02)	-0,13	(0,01)	0,39	(0,01)	1,41	(0,02)
Sweden	0,03	(0,02)	0,04	(0,03)	0,02	(0,02)	-1,09	(0,02)	-0,22	(0,01)	0,29	(0,01)	1,17	(0,02)
Switzerlan	0,11	(0,02)	0,00	(0,03)	0,22	(0,03)	-1,00	(0,02)	-0,15	(0,00)	0,35	(0,01)	1,26	(0,03)
United Sta	-0,08	(0,03)	-0,26	(0,04)	0,09	(0,04)	-1,44	(0,03)	-0,40	(0,01)	0,24	(0,01)	1,30	(0,03)
OECD	0,01	(0,02)	-0,12	(0,02)	0,14	(0,02)	-1,24	(0,02)	-0,29	(0,01)	0,29	(0,00)	1,28	(0,01)
OECD	0,00	(0,01)	-0,09	(0,01)	0,09	(0,01)	-1,17	(0,01)	-0,28	(0,00)	0,26	(0,00)	1,19	(0,01)
Non-OECD countries														

List1 / List2 / List3 /

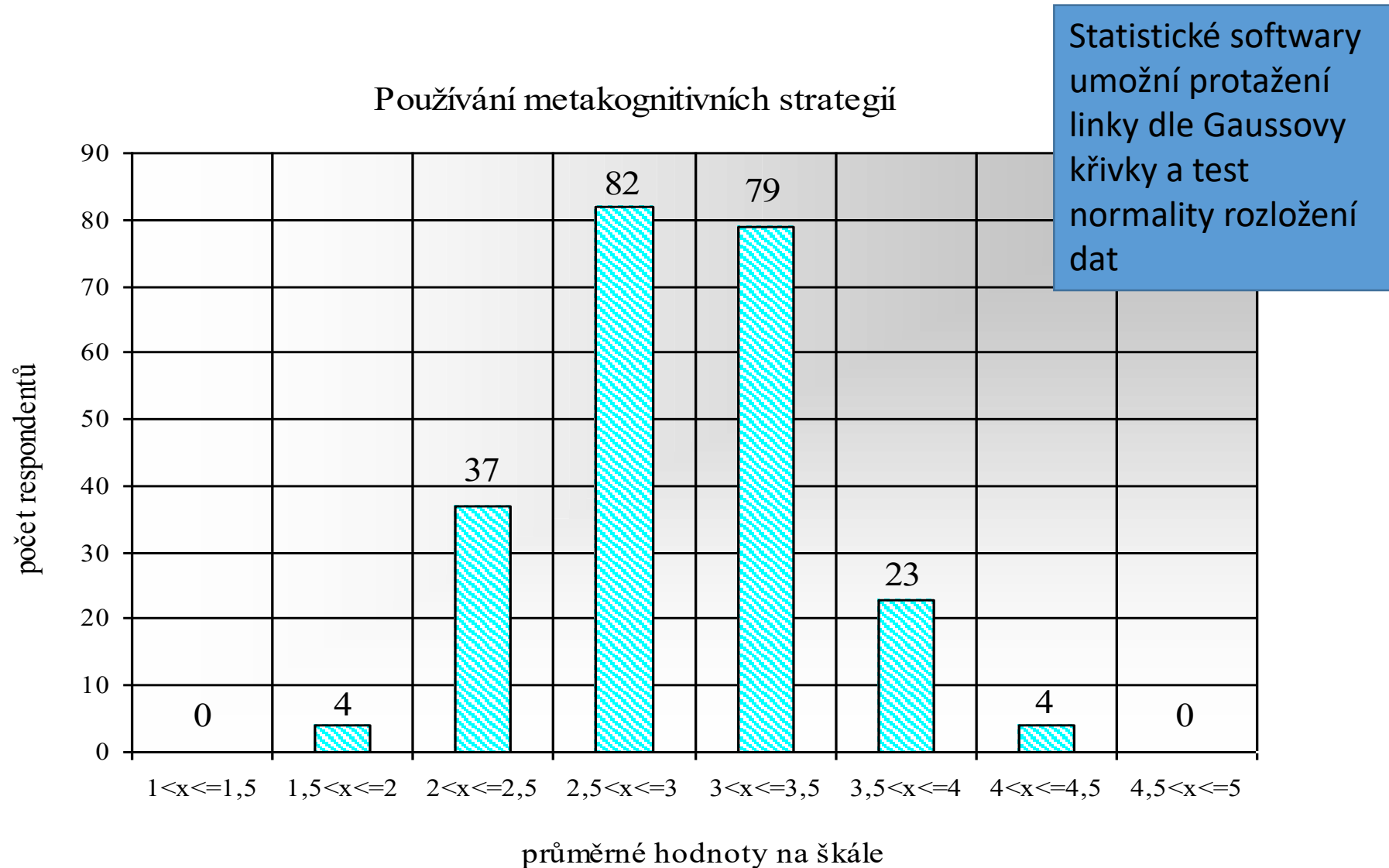
Kreslení Automatické tvary

Připraven

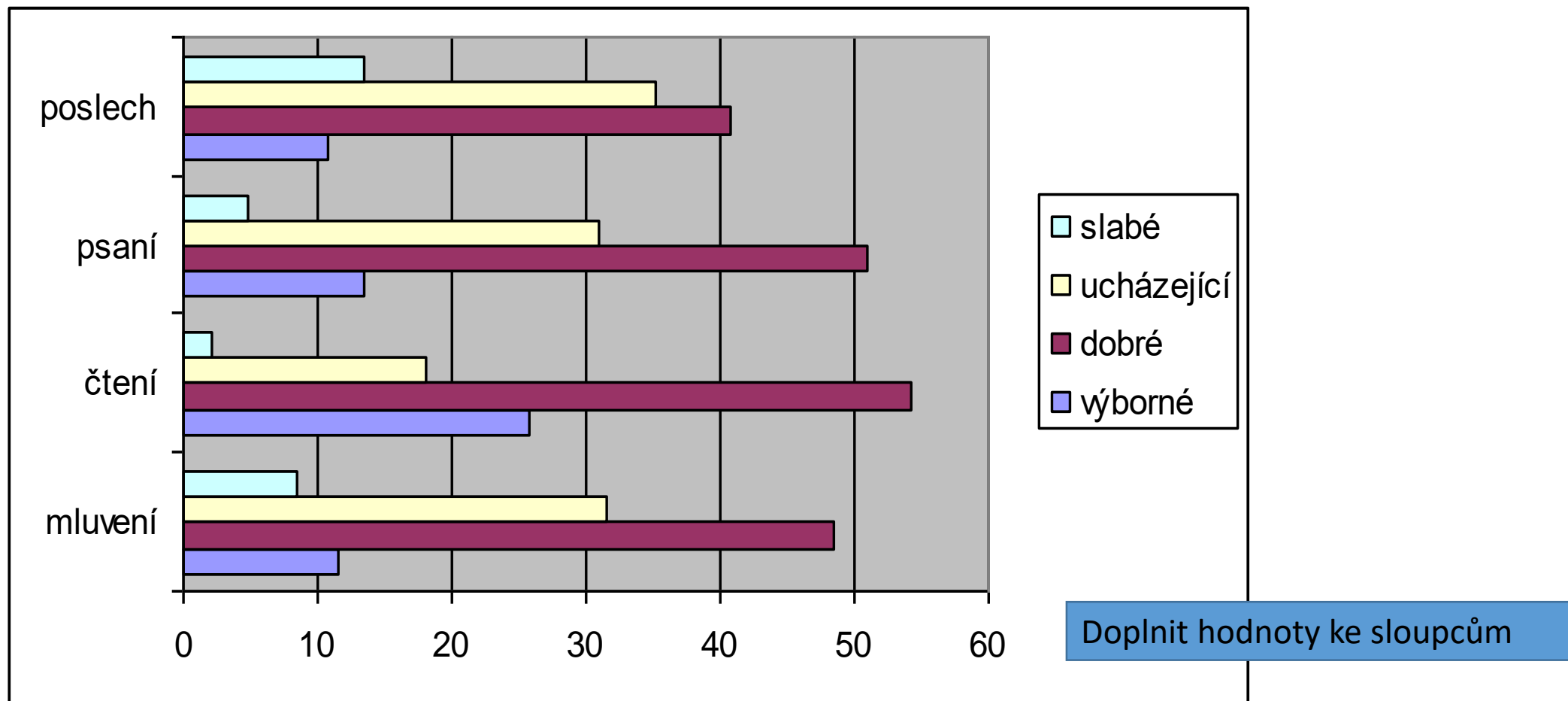
Ukázka grafického znázornění dat (Excel, sloupcový graf)



Ukázka grafického znázornění dat (Excel, histogram)

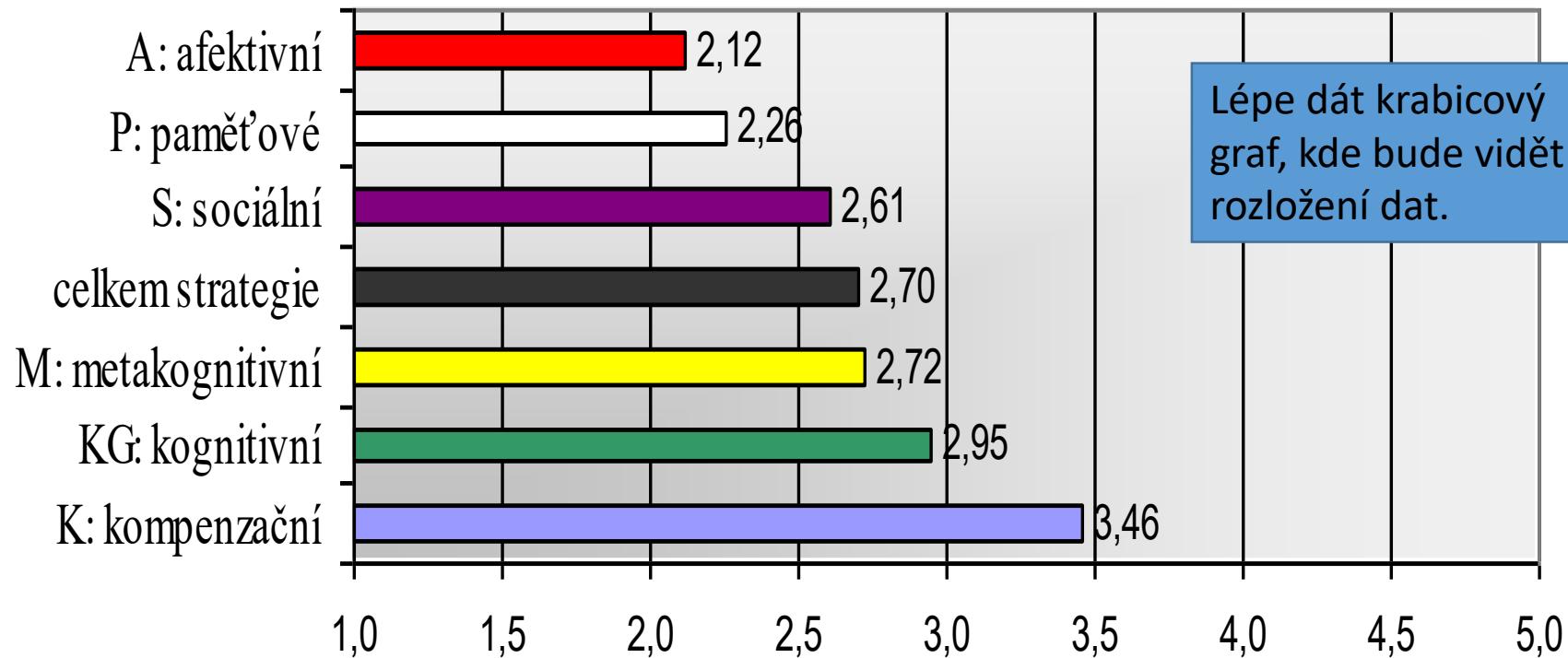


Ukázka grafického znázornění dat (Excel, pruhový graf)

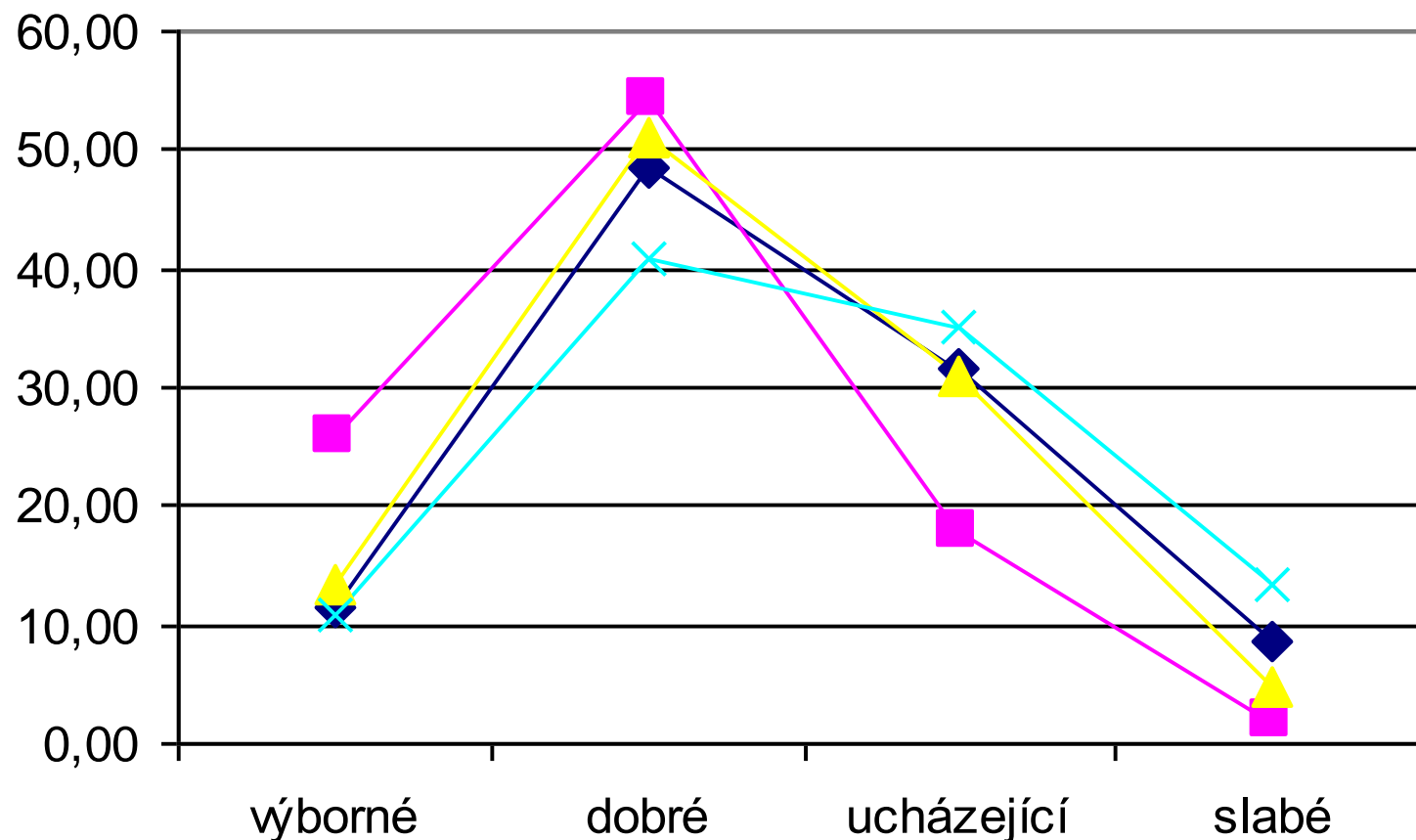


Ukázka grafického znázornění dat (Excel, pruhový graf)

Používání skupin strategií

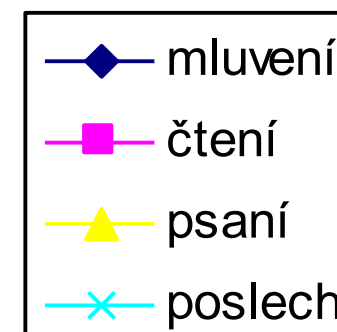


Ukázka grafického znázornění dat (Excel)



Chyba: znázornění
spojnic, nejde o
trend, nejde o
souvislá data!

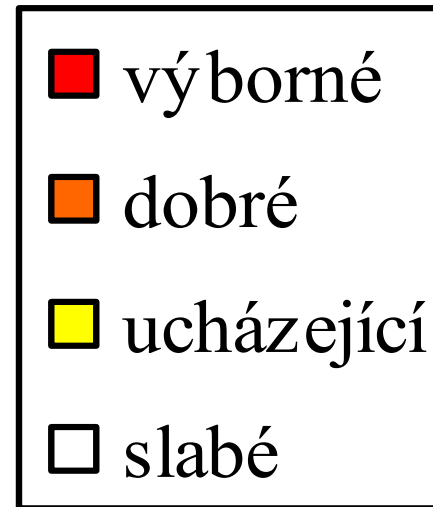
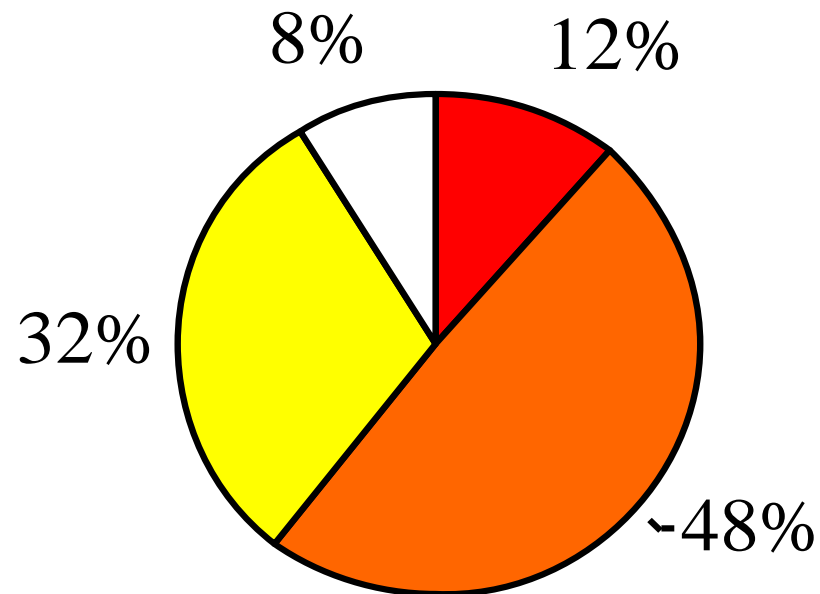
Viz také klima
učitelského sboru aj.



Ukázka grafického znázornění dat (Excel, koláčový graf)

mluvení

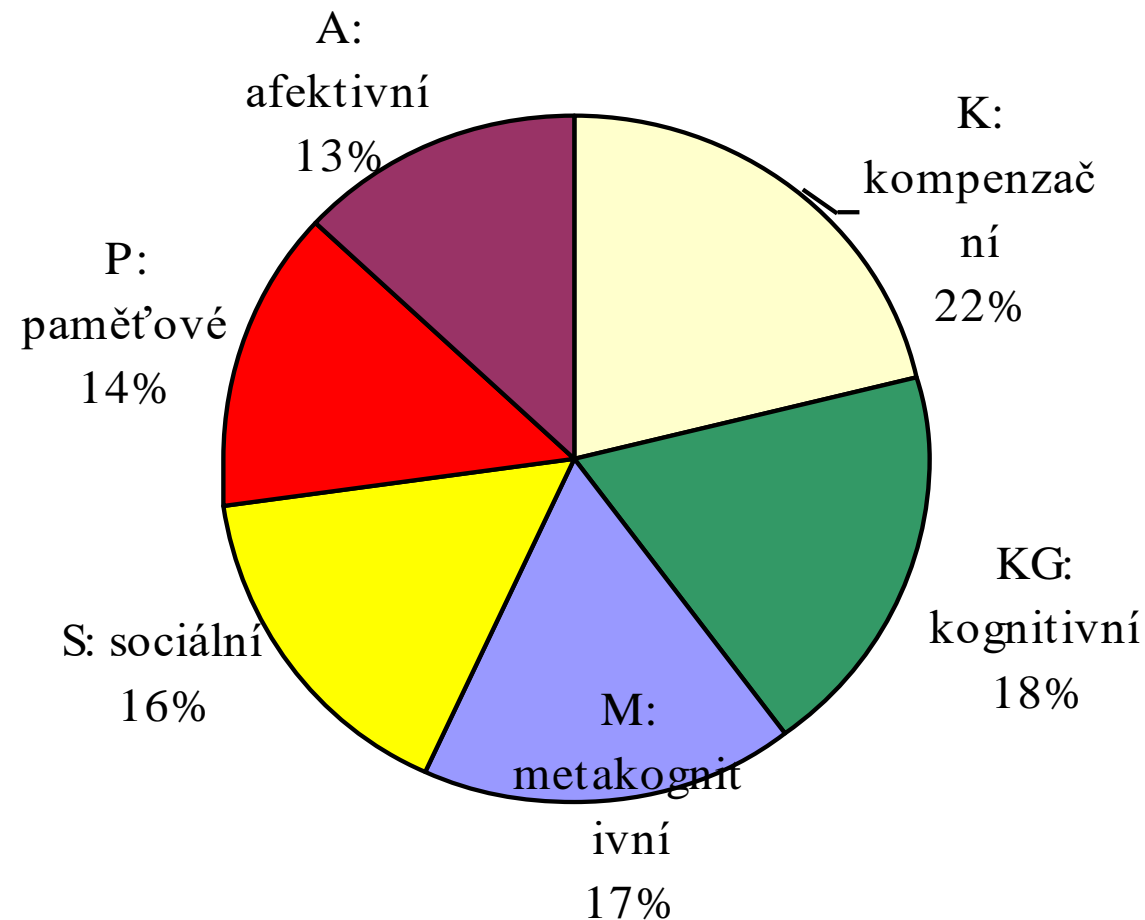
Pozor na zkreslení koláči,
jde dát i poměrový
pruhový graf



Ukázka grafického znázornění dat (Excel, koláčový graf)

Nepostrádá tento graf smysl?

Poměr strategií v procesu učení



Zpracování dat ve statistickém softwaru

Co by měl umět:

- vytvoření kódovacího rámce,
- zakódování dat,
- transformace dat,
- statistická analýza dat.

Hromadný sběr dat

- Dotazníky (ev. jiné nástroje generující množství dat)
 - Tužka –papír:
 - Sebereme a očíslujeme (typicky levý horní roh – např. 001 – 999)
 - Vytvoříme datovou matici („přepis dotazníku do PC“)
 - Google form či jiný / podobný elektronický sběr dat (předchozí krok odpadá)
 - Možno v Excelu (či LibreOffice Calculator) ev. rovnou ve statistickém programu
 - Kde vzít Excel - <https://it.muni.cz/sluzby/microsoft-office-365>
 - Kde vzít statistický software – viz návod doc. Vlčkové... nebo zkuste <https://www.gnu.org/software/pspp/>
 - První sloupec ID (číslo dotazníku/ů)
 - Další postupně jednotlivé odpovědi (např. „ot_1“ až „ot_99“)
 - Jednotlivé typy odpovědí kódujeme
 - Chybějící údaje kódujeme vysokým číslem, které nemůže být v intervalu použitého pro kódování (např. „99“ v případě likertovské škály míry souhlasu s výrokem v dotazníku „1-5“)

SPSS / PSPP

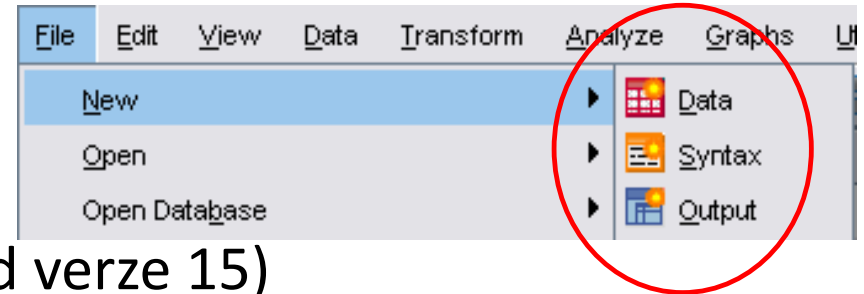
Stat. software

Tři typy oken v SPSS / PSPP

- **Data editor** (záložky **Variable View** a **Data View**)
 - správa datového souboru
 - soubory s koncovkou ***.sav**

- **Output** → výstupy analýz

→ soubory s koncovkou ***.spv** (od verze 15)



- **Syntax** → zadávání pomocí příkazového řádku
 - soubory s koncovkou ***.sps**

Jak to vypadá

The screenshot displays the SPSS interface with two windows open:

- Output - PSPPIRE Output Viewer:** This window shows the results of a descriptive analysis. It includes a menu bar (File, Edit, Windows, Help) and a list of variables: GET, SORT CASES, SPLIT FILE, DESCRPTIVES, EXAMINE, and DESCRPTIVES. The main area shows a table with columns for 'Variable', 'N', 'Mean', and 'Std. Deviation'. The variables listed are 'year', 'LOWER OR UPPER DIVISION', 'ethnicit', 'quiz1', and 'final'. The 'year' variable has 19 cases with a mean of 1.00 and a standard deviation of 0.00. The other variables have 3 cases each.
- *students.sav [DataSet1] - PSPPIRE Data Editor:** This window shows the data editor with a menu bar (File, Edit, View, Data, Transform, Analyze, Utilities, Windows, Help) and a toolbar. The 'Descriptives' dialog box is open, showing the following settings:
 - Variables:** quiz1, quiz2
 - Statistics:** Mean, Standard deviation, Minimum, Maximum, Variance (checked), Range, Sum, Standard error, Kurtosis.
 - Options:** Exclude entire case if any selected variable is missing (unchecked), Include user-missing data in analysis (checked), Save Z-scores of selected variables as new variables (unchecked).

The 'Data Editor' window shows a table with columns for 'Case', 'id', 'lastname', 'firstnam', and several quiz scores. The 'Descriptives' dialog box is highlighted with an orange circle, and the 'Data View' tab is also highlighted with an orange circle.

Proměnné - úpravy

The screenshot displays the SPSS Data Editor interface for a dataset named '*students.sav [DataSet1]'. The main window shows a list of variables with their properties. The variable 'ethnicit' is selected, and a dialog box titled 'psppire' is open, allowing for the modification of its value labels. The dialog shows a list of value labels: 1 = 'AMERICAN INDIAN', 2 = 'ASIAN', 3 = 'AFRO-AMERICAN', 4 = 'CAUCASIAN', and 5 = 'HISPANIC'. The 'Value' field is set to 2, and the 'Value Label' field is set to 'ASIAN'. The 'Variable View' tab is selected at the bottom of the window.

Variab	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Column	Align	Measure
1	id	Numeric	6	0		None	None	8	Right	Scale
2	lastname	String	10			None	None	10	Left	Nominal
3	firstnam	String	10			None	None	10	Left	Nominal
4	gender	Numeric	1	0		{1, FEMALE}...	None	8	Right	Ordinal
5	ethnicit	Numeric	1	0						Ordinal
6	year	Numeric	1	0	YEAR IN SCHOOL					Ordinal
7	lowup	Numeric	1	0	LOWER OR UPPER DIVIS					Ordinal
8	section	Numeric	1	0						Ordinal
9	hsgpa	Numeric	4	2	High School GPA					Scale
10	colgpa	Numeric	8	2	College GPA					Scale
11	extrcred	Numeric	1	0	DID EXTRA CREDIT PRO					Ordinal
12	review	Numeric	1	0	ATTENDED REVIEW SES					Ordinal
13	quiz1	Numeric	2	0						Ordinal
14	quiz2	Numeric	2	0						Ordinal
15	quiz3	Numeric	2	0						Ordinal
16	quiz4	Numeric	2	0		None	None	8	Right	Ordinal
17	quiz5	Numeric	2	0		None	None	8	Right	Ordinal
18	final	Numeric	2	0		None	None	8	Right	Scale
19	total	Numeric	3	0		None	None	8	Right	Scale
20	percent	Numeric	4	1		None	None	8	Right	Scale
21	grade	String	8			None	None	8	Left	Nominal
22	passfail	String	1			None	None	1	Left	Nominal

Tři základní informace o proměnných:

- **jméno proměnné** **VARIABLE NAME**
max. 8 znaků bez diakritiky, mezer; musí začínat písmenem
(např. *pohlavi*) → **bezpodmínečně v datech musí být**

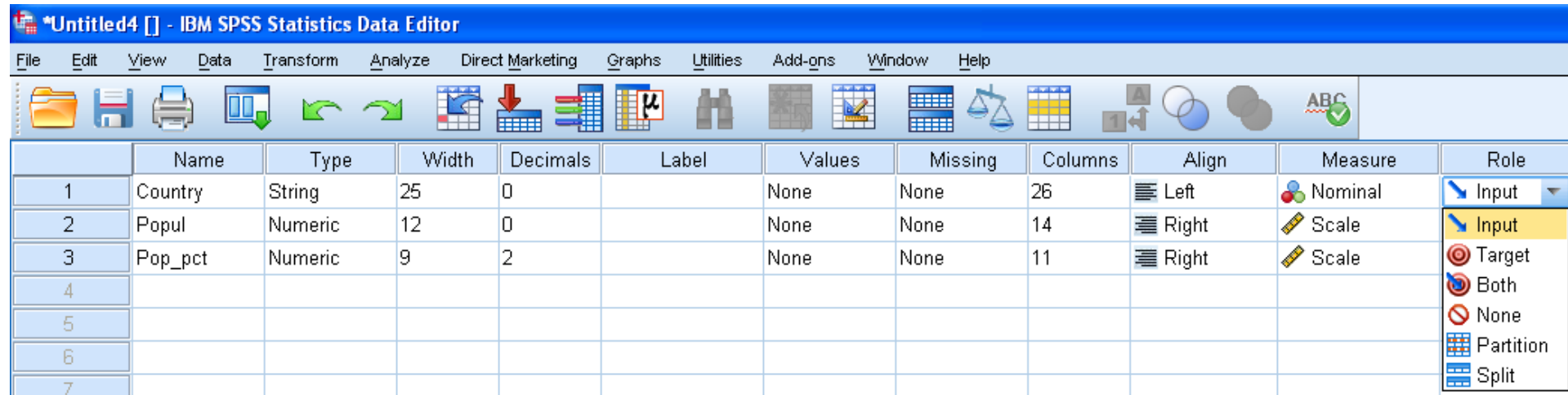
pro uživatele může být dále užitečná nadstavba:

- **popiska proměnné** **VARIABLE LABEL**
Delší popis názvu znaku, lze s diakritikou, např. (zkrácené)
znění otázky v dotazníku, který se bude objevovat ve
výstupech (např. *Pohlaví respondenta*)
- **popisky kategorií (hodnot)** **proměnné VALUE LABELS**
Popis kategorií znaku, které se budou objevovat ve výstupech
(např. *1 = muž, 2 = žena*)

The screenshot shows the 'Variable View' tab in PASW Statistics Data Editor. The table below represents the data shown in the screenshot, with red boxes highlighting the 'Name', 'Label', and 'Values' columns.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	pohlavi	Numeric	8	0	Pohlaví respondenta	{1, muž}...	None	8	Right	Nominal
2	vek	Numeric	8	0	Věk	None	None	8	Right	Scale
3	q1	Numeric	8	0	Spokojenost s prací	{1, nespokojen...}	None	8	Right	Ordinal
4	q2	Numeric	8	0	Spokojenost s životem	{1, nespokojen...}	None	8	Right	Ordinal

Ve Variable View je toho ale více



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Country	String	25	0		None	None	26	Left	Nominal	Input
2	Popul	Numeric	12	0		None	None	14	Right	Scale	Input
3	Pop_pct	Numeric	9	2		None	None	11	Right	Scale	Target
4											Both
5											None
6											Partition
7											Split

Nejdůležitější jsou:

- **Type:** druh záznamu dat (v zásadě buď jako číslo = Numeric nebo jako slova = String). Preferujeme Numeric
- **Decimals:** počet desetinných míst, co se bude zobrazovat ve výstupech (samotný záznam dat ale zachovává více desetinných míst)
- **Measure:** typ proměnné (dříve to nevadilo, novější verze vyžadují pro určitý typ analýz)
- **Missings:** uživatelsky definované hodnoty (o tom více jinde)
- v novějších verzích (SPSS 22) ještě přibyla **role proměnné v analýze** (závislá, nezávislá, třídící atd.)

Překódování a transformace

The screenshot shows the SPSS Data Editor interface with the 'Compute Variable' dialog box open. The dialog is configured to create a new variable named 'log_colgpa' using the expression 'LN(colgpa)'. The 'Target Variable' field contains 'log_colgpa' and the 'Numeric Expressions' field contains 'LN(colgpa)'. The 'Functions' list includes 'LN(number)'. The 'Data View' tab is selected at the bottom, which is circled in orange.

Case	id	lastname	firstnam	gender	ethnicit	year	lowup	section	hsgpa	colgpa	extrcred	review	quiz1	quiz2	qi
1	302400	JONES													6
2	106484	VILLARRUZ													6
3	664653	KHAN													3
4	595177	WILLIAMS													5
5	506467	SCARBROUG													6
6	681855	GRISWOLD													4
7	721311	SONG													2
8	237983	LEE													4
9	725987	BATILLER													4
10	615115	VASENIUS													4
11	979028	NEUHARTH													5
12	140219	GUADIZ													3
13	908754	MARQUEZ													3
14	417003	EVANGELIST													3
15	818528	CARRINGTON													1
16	938666	SUAREZ-TAN													3
17	354601	CARPIO	MARY	1	2	2	1	1	2.03	2.40	1	2	10	1	1
18	307894	TORRENCE	GWEN	1	3	2	1	2	2.09	2.21	2	2	6	6	6
19	983522	SLOAT	AARON	2	3	3	2	3	2.11	2.45	1	1	4	6	6
20	108642	VALAZQUEZ	SCOTT	2	4	3	2	2	2.19	3.50	2	1	10	1	1
21	287617	CUMMINGS	DAVENA	1	5	3	2	3	2.21	3.82	1	2	9	1	1

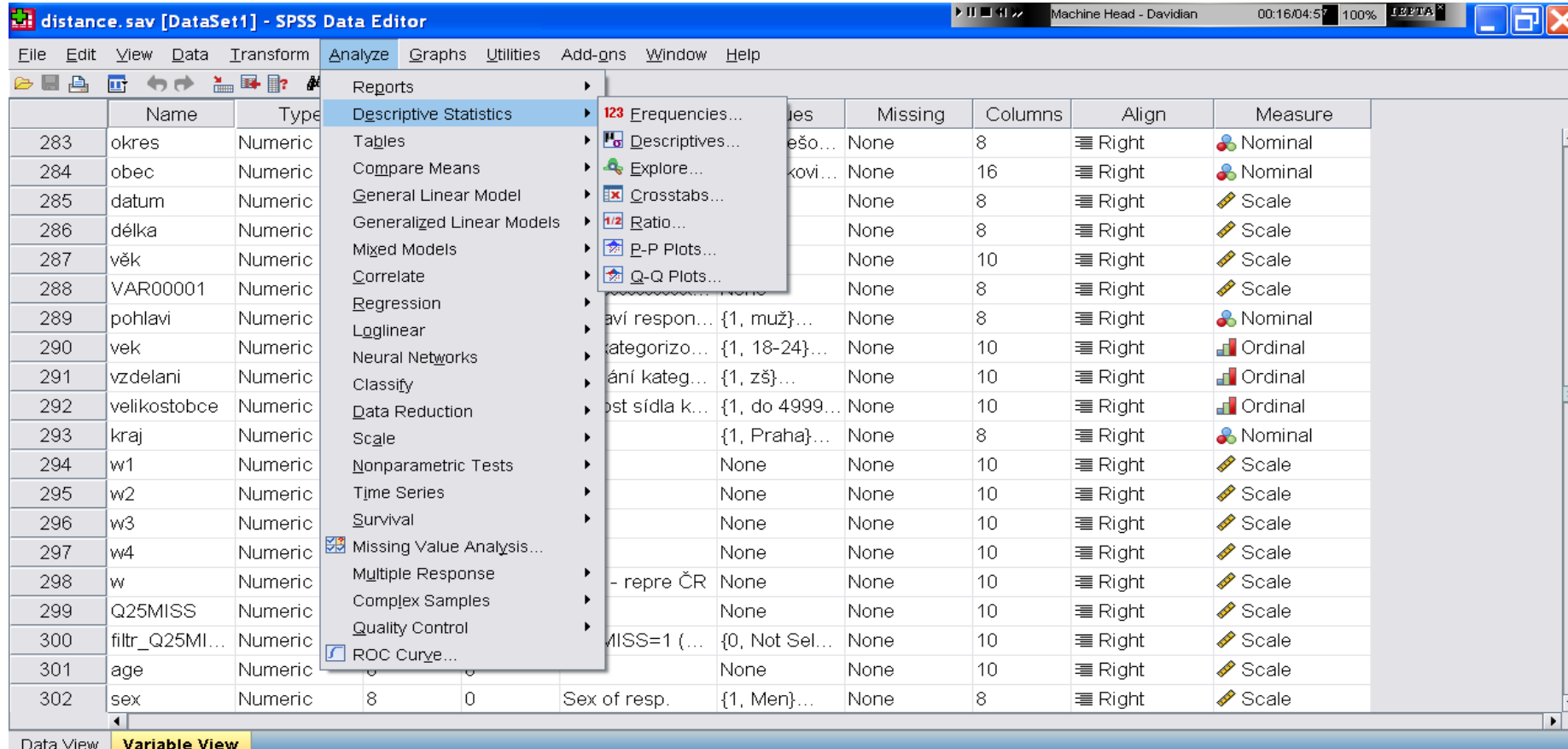
Transformace dat → *Transform*

- Výpočet/vytváření nových (syntetických) znaků
→ **COMPUTE** (nebo pro „načítání“ **COUNT**)
- Rekódování
→ **RECODE** (do stejné nebo nové proměnné)

Úpravy dat a výstupů → *Data*

- Uspořádání případů → **SORT CASES**
- Rozdělení na podsoubory (při analýze dle skupin) → **SPLIT FILE**
- Výběr případů (filtrování) → **SELECT CASES**

Analýzy



The screenshot shows the SPSS Data Editor interface with the 'Analyze' menu open. The 'Descriptive Statistics' submenu is expanded, showing options like 'Frequencies...', 'Descriptives...', 'Explore...', 'Crosstabs...', 'Ratio...', 'P-P Plots...', and 'Q-Q Plots...'. The background table lists variables such as 'okres', 'obec', 'datum', 'délka', 'věk', 'VAR00001', 'pohlavi', 'vek', 'vzdelani', 'velikostobce', 'kraj', 'w1', 'w2', 'w3', 'w4', 'w', 'Q25MISS', 'filtr_Q25MI...', 'age', and 'sex'.

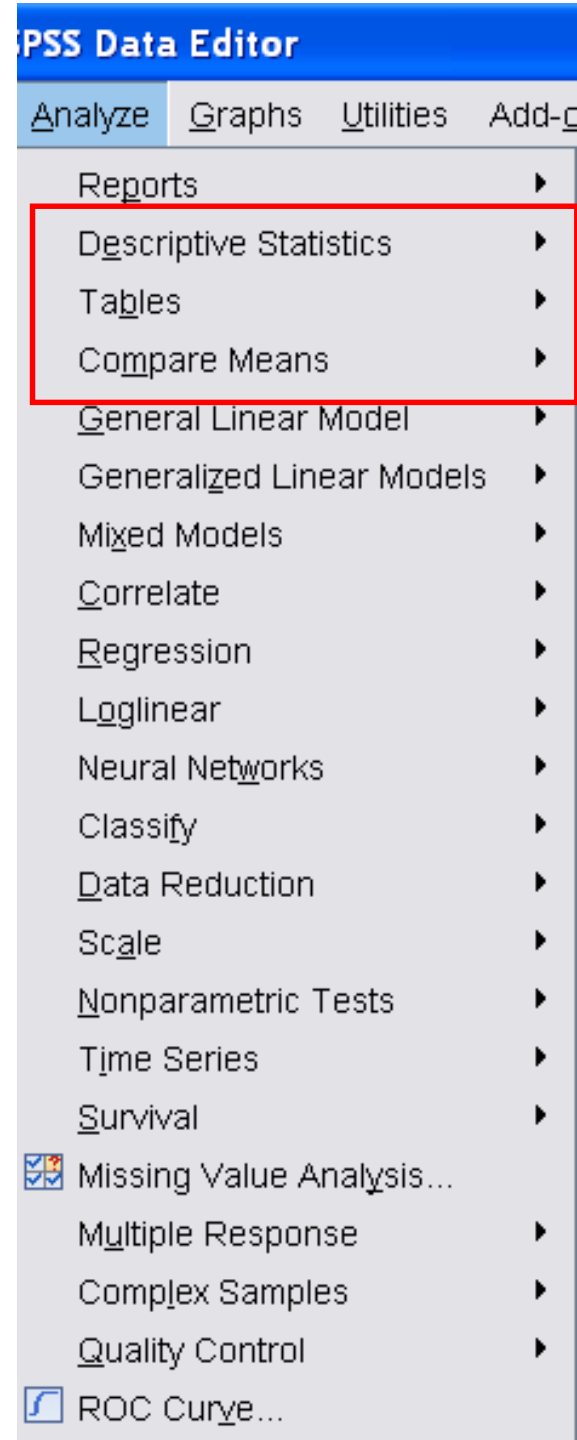
Name	Type	Missing	Columns	Align	Measure
283 okres	Numeric		8	Right	Nominal
284 obec	Numeric		16	Right	Nominal
285 datum	Numeric		8	Right	Scale
286 délka	Numeric		8	Right	Scale
287 věk	Numeric		10	Right	Scale
288 VAR00001	Numeric		8	Right	Scale
289 pohlavi	Numeric		8	Right	Nominal
290 vek	Numeric		10	Right	Ordinal
291 vzdelani	Numeric		10	Right	Ordinal
292 velikostobce	Numeric		10	Right	Ordinal
293 kraj	Numeric		8	Right	Nominal
294 w1	Numeric		10	Right	Scale
295 w2	Numeric		10	Right	Scale
296 w3	Numeric		10	Right	Scale
297 w4	Numeric		10	Right	Scale
298 w	Numeric		10	Right	Scale
299 Q25MISS	Numeric		10	Right	Scale
300 filtr_Q25MI...	Numeric		10	Right	Scale
301 age	Numeric		10	Right	Scale
302 sex	Numeric	8	0	Sex of resp. {1, Men}...	Scale

→ příkazy v sekci hlavního menu **Analyze**

Popisné statistiky: **Descriptive statistics** → **Frequencies / Descriptives / Explore**

Analýzy → *Analyze*

- **Descriptive statistics**
- **Tables**
- **Compare means**
- **Correlate**
- Data Reduction
- **Nonparametric Tests**
- Missing Value Analysis



Grafy → Graphs

The screenshot shows the PASW Statistics Viewer interface. The main window displays a bar chart titled "Graph" with the y-axis labeled "Mean" ranging from 0 to 4. The x-axis categories are "Práce", "Rodina", "Přátelé a známí", "Volný čas", "Politika", and "Náboženství". The bars represent the mean values for each category. A red circle highlights the "Graphs" menu, which is open, showing various chart types. The "Legacy Dialogs" option is selected, and the "Interactive" submenu is also visible.

Output
Log
Graph
Title
Notes
Bar of mean(q30_a, q30_b, ...)

GRAPH
/BAR(SIMPLE)=I

Graph

Mean

Práce Rodina Přátelé a známí Volný čas Politika Náboženství

MEAN(q30_e) MEAN(q30_f)

Graphs menu options:
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
Interactive
Bar...
Dot...
Line...
Ribbon...
Drop-Line...
Area...
Pie...
Boxplot...
Error Bar...
Histogram...
Scatterplot...