

# Faktorová analýza

---

- cíl faktorové analýzy
  - základní pojmy, postup
  - interpretace faktorů
  - příklad
-

# Cíle faktorové analýzy

---

- **cílem** faktorové analýzy (exploratorní) je
    - 1) **redukce dat** – zmenšení počtu proměnných odstraněním nadbytečných proměnných (tj. těsně korelujících s ostatními proměnnými)
    - 2) **identifikace struktury dat** – prozkoumat vztahy mezi proměnnými
-

# Výsledek faktorové analýzy

---

- **výsledkem** faktorové analýzy (exploratorní) je vytvoření několika hypotetických proměnných – **faktorů**
    - někdy bývají nazývány **latentní** proměnné
  - faktory jsou **lineárními kombinacemi** původních proměnných (váženými součty)
  - vysvětlují vztahy mezi původními proměnnými
-

# Faktorová analýza

---

- korelace většího množství proměnných se analyzuje tak, že se hledají shluky proměnných, které spolu navzájem korelují silně a s ostatními proměnnými naopak slabě nebo vůbec
  - **faktory se interpretují** podle toho, které proměnné obsahuje daný shluk
-

# Faktorová analýza

---

- cílem je najít **malé množství faktorů**, které vysvětlí **velké množství variability dat**
  - **faktory** jsou navzájem **nekorelované** (=každý měří jinou dimenzi dat)
-

# Korelační matice

---

	matematika	fyzika	angličtina	čeština
matematika				
fyzika	0,893			
angličtina	0,215	0,196		
čeština	0,308	0,262	0,820	
dějepis	0,117	0,065	0,590	0,685

---

# Korelační matice

---

	matematika	fyzika	angličtina	čeština
matematika				
fyzika	0,893			
angličtina	0,215	0,196		
čeština	0,308	0,262	0,820	
dějepis	0,117	0,065	0,590	0,685

---

# Extrakce faktorů

---

- **extrakce** faktorů – na základě matice vztahů mezi proměnnými (např. korelační matice)
    - pokud např. všechny původní proměnné vzájemně silně korelují, pak můžeme celkový rozptyl zachytit pouze jedním faktorem
    - naopak pokud jsou všechny vzájemné korelace slabé, pak potřebujeme k vysvětlení celkového rozptylu tolik faktorů, kolik je původních proměnných
  - **počet** extrahovaných faktorů – do určité míry závisí na rozhodnutí výzkumníka
-



# Korelační matice

---

	matematika	fyzika	angličtina	čeština
matematika				
fyzika	0,893			
angličtina	0,215	0,196		
čeština	0,308	0,262	0,820	
dějepis	0,117	0,065	0,590	0,685

---

# Korelační matice

---

	matematika	fyzika	angličtina	čeština
matematika				
fyzika	0,893			
angličtina	0,215	0,196		
čeština	0,308	0,262	0,820	
dějepis	0,117	0,065	0,590	0,685

---

# Vlastní hodnota

---

- podíl variability původních proměnných, vysvětlených faktorem, se označuje jako **vlastní hodnota** (eigenvalue)
  - používá se i označení vlastní čísla nebo charakteristické kořeny
  - součet vlastních hodnot všech faktorů odpovídá počtu původních proměnných
-

# Faktorová analýza

---

- ❑ cílem je vysvětlit **co největší množství rozptylu** (80-90%) původních proměnných **co nejmenším počtem faktorů**
  - ❑ při rozhodování o počtu extrahovaných faktorů se používá tzv. **sutinový graf** (scree plot), který ukazuje závislost vysvětlené variability na počtu faktorů – znázorňuje pro každý faktor hodnoty **vlastních čísel**
  - ❑ faktory jsou uspořádány podle toho, kolik variability původního souboru proměnných vysvětlují
-

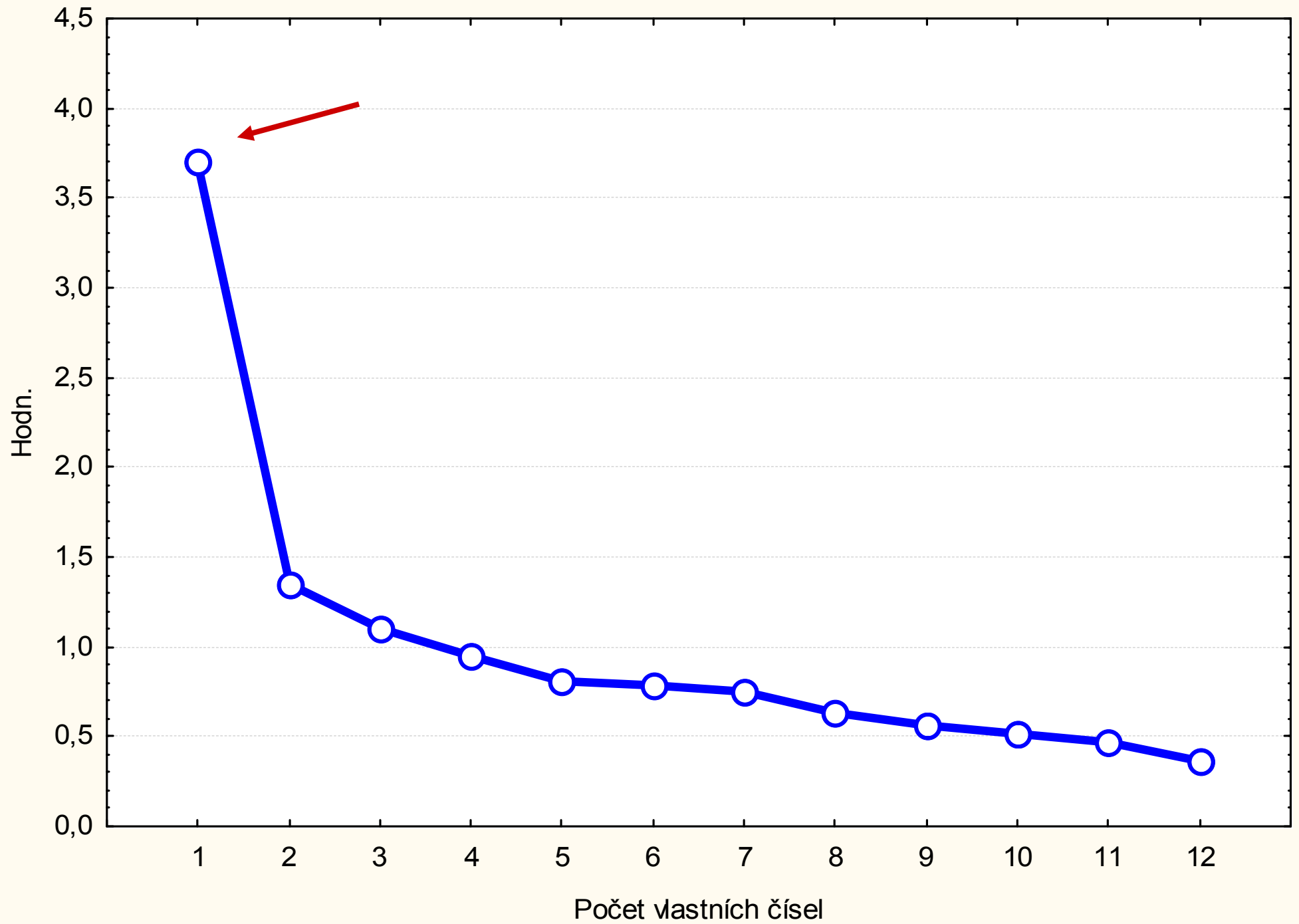
# Extrakce faktorů

---

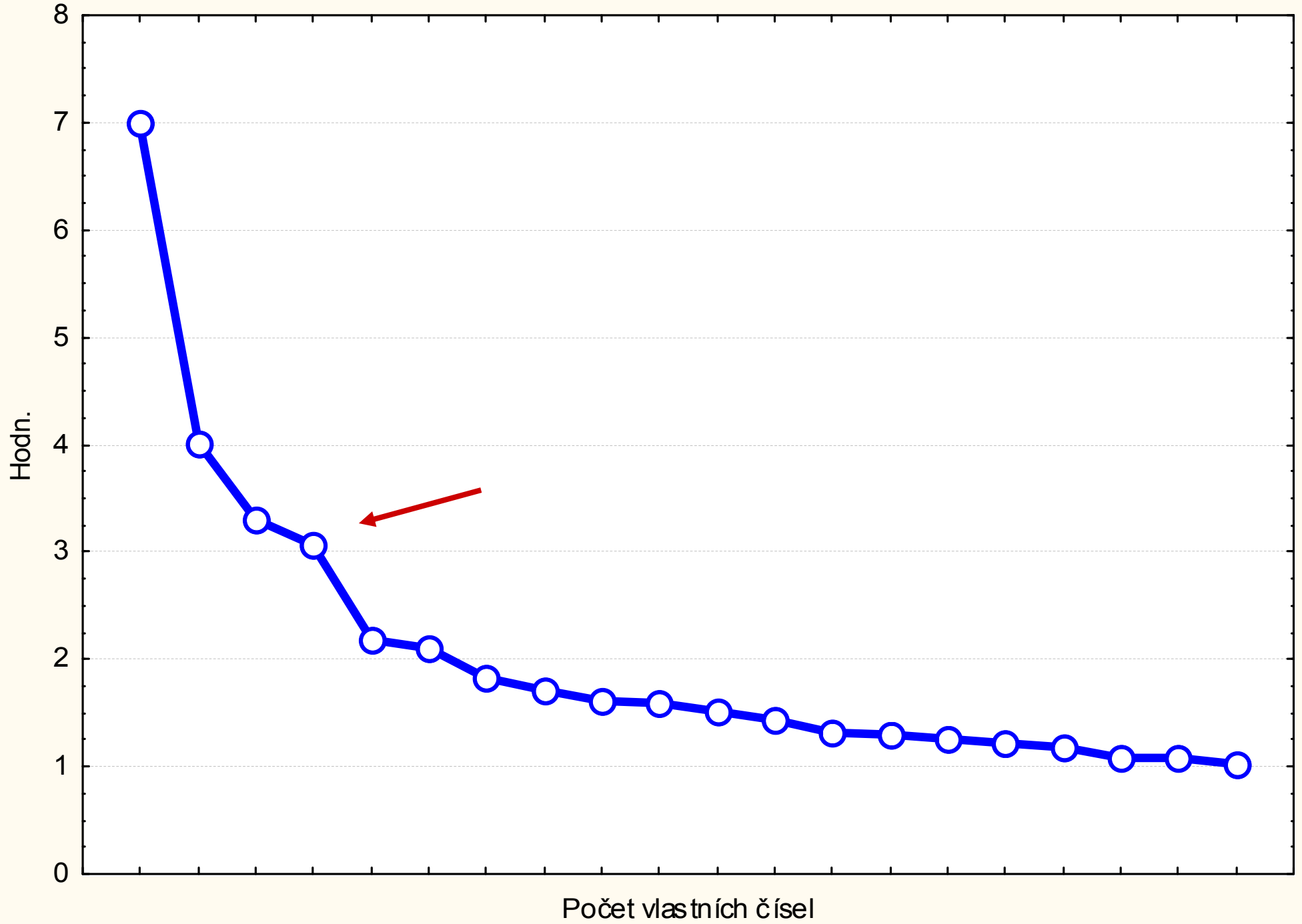
## □ 3 možná kritéria

- vybereme prvních  $x$  faktorů, které vysvětlují **dostatečné množství rozptylu**
  - vybereme faktory s vlastní hodnotou větší než 1 (**Kaiserovo kritérium**)
  - v **sutinovém grafu** najdeme náhlý přechod z prudkého klesání do mírného
-

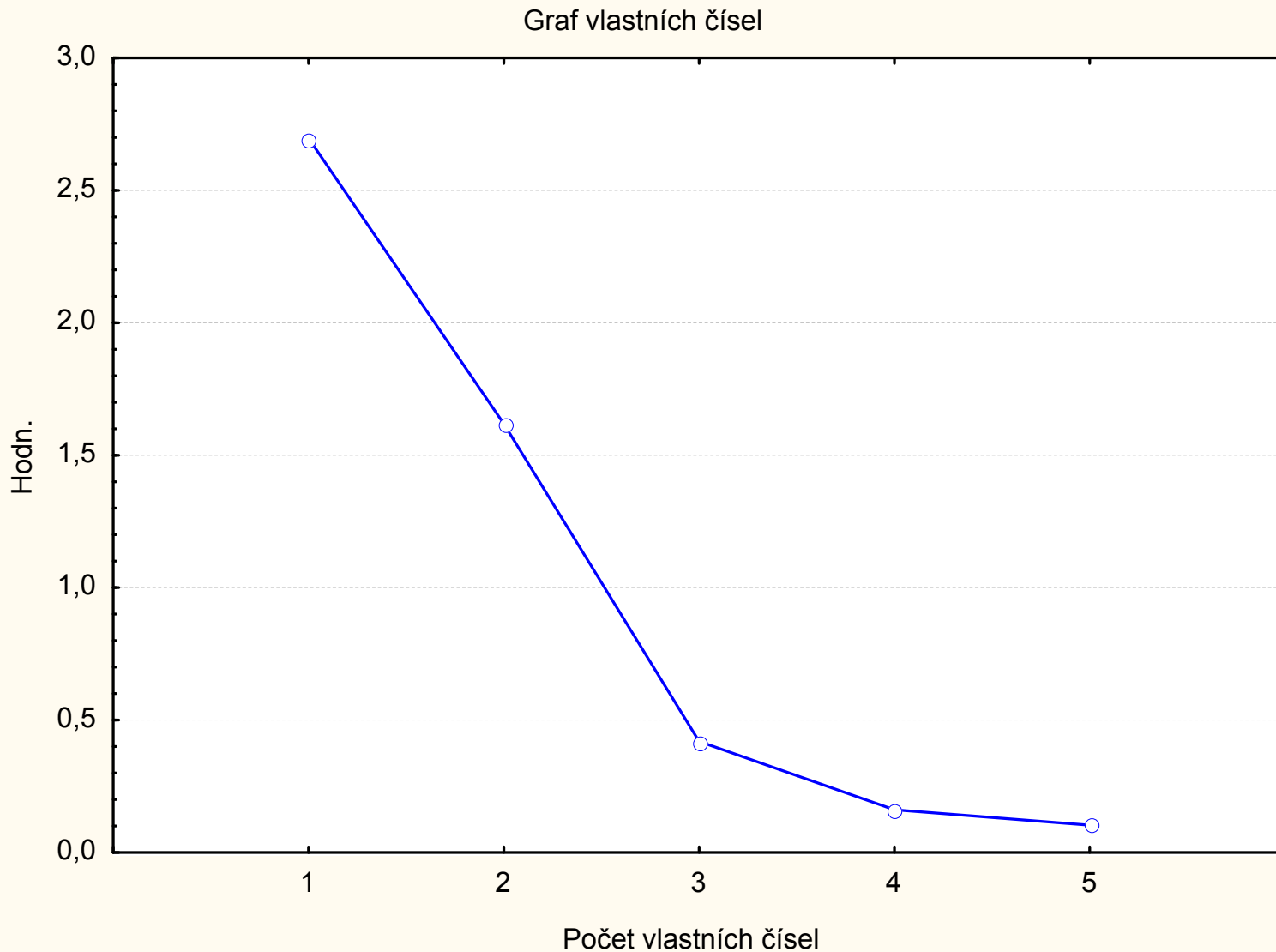
Graf vlastních čísel



Graf vlastních čísel



# Sutinový graf pro korelační matici předmětových testů





# Tabulka vlastních čísel

---

Hodn.	Vl. čísla (matice) Extrakce: Hlavní komponenty			
	vl. číslo	% celk. rozptylu	Kumulativ. vlast. číslo	Kumulativ. %
1	2,696995	53,93990	2,696995	53,93990
2	1,618669	32,37339	4,315664	86,31329
3	0,419285	8,38570	4,734949	94,69900
4	0,161469	3,22937	4,896418	97,92837
5	0,103582	2,07164	5,000000	100,00000

---

# Extrakce faktorů

---

- podle Kaiserova kritéria – 2 faktory
  - podle scutinového grafu – 2 faktory
  - podle tabulky vlastních čísel –
    - 2 faktory vysvětlují 86% rozptylu
    - 3 faktory 95%
-

# Interpretace faktorů

---

- **interpretace** faktorů – faktorová analýza sama o sobě nenabídne označení faktorů (to je opět na výzkumníkovi)
  - faktor bývá označen na základě proměnných, které k němu mají nejtěsnější vztah (nejvyšší tzv. faktorové **náboje/zátěže** – korelace mezi faktorem a položkou)
-

# Faktorové zátěže

---

Proměnná	Faktor. zátěže (Bez rot. ) (matice) Extrakce: Hlavní komponenty (Označené zatěže jsou >,700000)	
	Faktor 1	Faktor 2
matematika	-0,622737	0,745243
fyzika	-0,586717	0,775842
angličtina	-0,821802	-0,367317
čeština	-0,886962	-0,315024
dějepis	-0,709153	-0,476644
Výkl.roz	2,696995	1,618669
Prp.celk	0,539399	0,323734

---

# Komunalita

---

- část variability proměnné, která je vysvětlená extrahovanými faktory, se nazývá **komunalita**
  - zbylá část (1-komunalita) se nazývá **jedinečnost**
  - komunality všech proměnných v modelu by měly být vysoké – ty s nízkými komunalitami je vhodné vyřadit
-

# Tabulka komunalit

---

Proměnná	Komunalita (matice) Extrakce: Hlavní komponenty Rotace: Bez rot.					
	Z 1 faktoru	Z 2 faktorů	Z 3 faktorů	Z 4 faktorů	Z 5 faktorů	Více R <sup>2</sup>
matematika	0,387801	0,943188	0,947648	0,948502	1,000000	0,804809
fyzika	0,344237	0,946168	0,946707	0,951125	1,000000	0,800195
angličtina	0,675358	0,810280	0,945074	0,998675	1,000000	0,676422
čeština	0,786701	0,885942	0,903230	0,998136	1,000000	0,753587
dějepis	0,502898	0,730088	0,992290	0,999979	1,000000	0,485648

---

# Rotace faktorového řešení

---

- **rotace** faktorového řešení – usnadní **interpretaci** faktorů
  - cílem rotace je dosáhnout, aby původní proměnné měly vysoké zátěže s co nejmenším počtem faktorů – a s ostatními faktory měly naopak zátěže nízké (celkový podíl vysvětleného rozptylu ale zůstává stejný)
  - rotace může být **ortogonální** (tj. předpokládá, že faktory jsou nezávislé; např. Varimax) nebo **šikmá** (předpoklad korelace mezi faktory; např. Oblimin)
-

# Faktorové zátěže po rotaci Varimax

---

Proměnná	Faktor. zátěže (Varimax normaliz. ) (matice) Extrakce: Hlavní komponenty (Označené zatěže jsou >,700000)	
	Faktor 1	Faktor 2
matematika	0,127008	0,962838
fyzika	0,080204	0,969399
angličtina	0,890761	0,129712
čeština	0,917801	0,208764
dějepis	0,854144	-0,022936
Výkl.roz	2,387939	1,927725
Prp.celk	0,477588	0,385545

---



# Faktorové skóry

---

- výsledky faktorové analýzy lze uložit v podobě nových proměnných – faktorových skóru, a s nimi pak dále pracovat
  - pro náš příklad bychom proto mohli uložit dvě nové proměnné – výsledek z didaktického testu přírodovědných a z humanitních předmětů (místo původních 5)
-

# Analýza hlavních komponent

---

- analýza hlavních komponent (PCA = principal component analysis)
  - metoda extrakce faktorů – na rozdíl od analýzy hlavních faktorů (PFA) využívá veškerou variabilitu proměnné, zatímco PFA pouze společnou s ostatními proměnnými
  - většinou mají podobné výsledky
  - PCA se používá více (především jako metoda redukce dat), PFA spíše pro objasnění struktury dat
-

# Faktorová analýza - příklad

---

- příklad aplikace FA:
- Osecká, L., Řehulková, O., Macek, P. (1998).

Zdravotní stesky adolescentů:  
struktura a rozdíly mezi pohlavím.

Sborník konference Sociální procesy a osobnost, MU Brno.

---

# Faktorová analýza - příklad

---

- cílem studie bylo mj. vytvořit typologii adolescentů na základě jejich zdravotních obtíží
  - adolescenti v dotazníku označili, jak často trpí každou z 18 nabídnutých zdravotních obtíží
-

# Faktorová analýza - příklad

---

- bolesti hlavy
  - dýchací potíže
  - žaludeční potíže
  - závratě
  - nechutenství
  - nervozita, neklid
  - nespavost
  - noční můry
  - nesoustředěnost
  - nevolnosti
  - silný tlukot srdce
  - třesení rukou
  - náhlé zpotení
  - průjem, zácpa
  - bolesti v zádech
  - krční bolesti
  - bolesti na prsou
  - bolesti v pánvi
-

# Faktorová analýza - příklad

---

- typologie na základě 18 proměnných by byla příliš složitá – je třeba tento počet snížit
  - autoři spočítali faktorovou analýzu (metodou analýzy hlavních komponent) a extrahovali 3 faktory (vysvětlovaly celkem 48% společného rozptylu)
-

# Faktorová analýza - příklad

	F1	F2	F3
nevolnosti	<b>71</b>	17	22
nechutenství	<b>65</b>	23	10
závratě	<b>62</b>	14	30
žaludeční potíže	<b>60</b>	-15	<b>50</b>
bolesti hlavy	<b>58</b>	27	4
nervozita, neklid	<b>56</b>	<b>41</b>	12
třesení rukou	17	<b>69</b>	19
nespavost	38	<b>63</b>	-3
náhlé zpotení	-2	<b>61</b>	35
silný tlukot srdce	16	<b>60</b>	27
nesoustředěnost	37	<b>54</b>	4
noční můry	32	<b>49</b>	20
bolesti v pánvi	4	28	<b>69</b>
průjem, zácpa	21	-9	<b>65</b>
bolesti na prsou	16	36	<b>61</b>
krční bolesti	16	33	<b>52</b>
<del>bolesti v zádech</del>	<del>15</del>	<del>36</del>	<del>42</del>
dýchací potíže	<b>32</b>	21	<b>36</b>
<i>procento rozptylu</i>	<b>17</b>	<b>17</b>	<b>14</b>

# Faktorová analýza - příklad

---

- první faktor nazvali **nevolnosti** –  
sytily ho především tyto potíže:
    - nevolnosti
    - nechutenství
    - závratě
    - žaludeční potíže
    - bolesti hlavy
    - nervozita, neklid
-



# Faktorová analýza - příklad

---

- druhý faktor označili **vegetativní obtíže** – sytily ho především položky:
    - třesení rukou
    - nespavost
    - náhlé zpocení
    - silný tlukot srdce
    - nesoustředěnost
    - noční můry
-

# Faktorová analýza - příklad

---

- třetí faktor označili **bolesti** – slyšely ho především tyto potíže:
    - bolesti v pánvi
    - průjem, zácpa
    - bolesti na prsou
    - krční bolesti
    - bolesti v zádech
-

# Faktorová analýza - příklad

---

- **místo původních 18** proměnných indikujících frekvenci zdravotních potíží měli **nyní 3 proměnné** (lineární kombinace původní proměnných) – nevolnosti, vegetativní potíže a bolesti
  - s nimi pak pracovali při typologii (viz další přednášky)
-

# Literatura

---

- Hendl: kapitoly 13.7 a 13.8
  - článek Osecká, L., Řehulková, O., Macek, P. (1998). Zdravotní stesky adolescentů. In M. Blatný (Ed.): Sociální procesy a osobnost. Brno 1998, str. 135-144.
-