

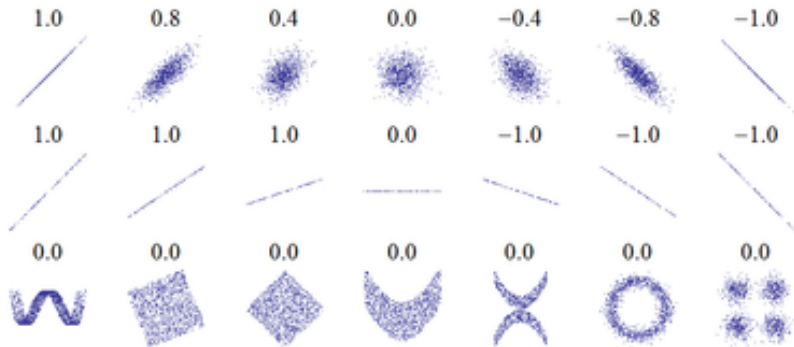
Faktorová analýza a korelační koeficienty

Hana Voňková

Univerzita Karlova v Praze, Pedagogická fakulta

- Korelační koeficient udává **míru lineárního vztahu mezi dvěma proměnnými**.
- Jeho hodnoty se pohybují mezi -1 a 1.
- Podle **znaménka** korelace ("+" či "-") můžeme usoudit, zda je vztah mezi proměnnými **kladný či záporný**. Negativní hodnota korelačního koeficientu naznačuje, že zvětšíme-li hodnotu jedné proměnné, zmenší se hodnoty druhé proměnné. Pozitivní hodnota korelačního koeficientu naznačuje, že zvětšíme-li hodnotu jedné proměnné, zvětší se hodnota i druhé proměnné.
- **Vzdálenost od nuly** indikuje **těsnost** lineárního vztahu mezi dvěma proměnnými
- Je-li hodnota korelačního koeficientu nízká až nulová, neznamená to, že mezi proměnnými nemůže být žádný vztah. Znamená to pouze, že mezi veličinami je lineární vztah zanedbatelný.
- Vysoká hodnota korelačního koeficientu nemusí znamenat, že je mezi proměnnými kauzální vztah. Znamená pouze predikční vztah.

Korelační koeficienty II



- “Více korelačních koeficientů najednou” uspořádaných do obdélníkového schématu

Příklad: **Korelační matice prospěchu žáků** (Anděl, 1978, s. 318)

	Č	R	A	D	Z	M	F	Ch	B
Č	1.00	0.45	0.41	0.40	0.37	0.27	0.16	0.12	0.29
R		1.00	0.32	0.28	0.27	0.31	0.20	0.15	0.28
A			1.00	0.28	0.26	0.20	0.12	0.09	0.21
D				1.00	0.28	0.04	0.00	0.00	0.12
Z					1.00	0.09	0.04	0.03	0.14
M						1.00	0.44	0.33	0.40
F							1.00	0.24	0.28
Ch								1.00	0.21
B									1.00

- vznikla jako matematický model pro hodnocení některých výsledků psychologických testů
- v pedagogice a dalších příbuzných vědách **není často možné měřit koncepty přímo - nejsou pozorovatelné**
- je však možné tyto koncepty, latentní proměnné, **zkoumat pomocí jiných proměnných, které přímo měřit můžeme (jsou pozorovatelné)** a o kterých lze realisticky přepokládat, že jsou indikátory latentní proměnné
- například obecné nadání žáka přímo měřitelné není, ale lze přepokládat, že rozumným indikátorem obecného nadání žáka jsou jeho (pozorovatelné) výsledky v testech z různých předmětů
- faktorová analýza pomáhá nalézt latentní proměnné
- obvykle je jich menší počet než původních proměnných

Intuice při hledání faktorů

- Jaká je struktura korelační matice? Lze najít vysoké a nízké korelační koeficienty?

Příklad: První výstup faktorové analýzy

výsledek v předmětu = $a \cdot \text{Faktor 1} + b \cdot \text{Faktor 2} + \text{chyba}$

	Č	R	A	D	Z	M	F	Ch	B
Faktor 1	0.7	0.6	0.5	0.4	0.6	0.4	0.4	0.3	0.5
Faktor 2	-0.3	-0.1	-0.2	-0.4	-0.3	0.5	0.4	0.3	0.2
Komunalita	0.58	0.37	0.29	0.32	0.25	0.61	0.32	0.18	0.29

- **Faktor 1** - lze pojmenovat jako **faktor obecného nadání** (korelační koeficienty v korelační matici jsou nezáporné, dobré známky dostávají studenti napříč všemi předměty)
- **Faktor 2** - lze pojmenovat jako **faktor bipolární** (odlišuje humanitní a přírodovědné předměty)
- **Komunalita** - jaká část variability je u dané proměnné vysvětlena modelem s faktory 1 a 2
 - **unicita** = 1-komunalita (jaká část variability není u dané

- v příkladu není podstatné, ve které skupině jsou znaménka kladná a ve které jsou záporná
- obecně platí, že **neexistuje právě jedno řešení při hledání faktorové matice**
 - matematicky:

$$y = F * z + e$$

y je výsledek v testu

F faktorová matice - tu je obvykle cílem nalézt

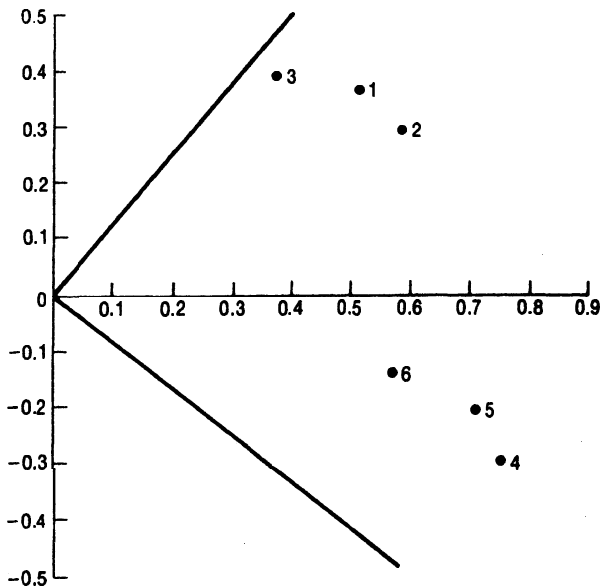
z faktory (faktorové zátěže)

e chyba (nevysvětlená variance)

- $var(y) = F * var(z) * F' + var(e)$, pokud $var(z)$ je jednotková matice, pak $var(y) = F * F' + var(e)$
- správným řešením je nejen F , ale též $F * R$, kde $R * R' = I$, neboť

$$var(y) = F * F' + var(e) = (F * R) * (R' * F') + var(e)$$

Intuitivní chápání rotace graficky - příklad



Ortogonální rotace (ortogonal rotation)

- **rotované faktory jsou nekorelované (“nezávislé”)** \Rightarrow faktorová matice reprezentuje korelační koeficienty mezi proměnnou a daným faktorem
- faktorová matice je tedy snadno interpretovatelná
- obecný faktor má však tendenci “zmizet”
- užití např. při zobecňování výsledků
 - **varimax** rotace - má tendenci dávat faktory, které mají vysoké korelace s jednou množinou proměnných a velmi nízké s dalšími množinami proměnných, nečastěji používaná
 - **quartimax** rotace - má tendenci přinutit proměnnou velmi korelovat s jedním faktorem a velmi málo s ostatními faktory

Šikmá rotace (oblique rotation)

- **rotované faktory mohou být korelované** \Rightarrow faktorová matice nereprezentuje korelační koeficienty mezi proměnnou a daným faktorem
- užití při hledání výsledku, který nejlépe odpovídá datům (“best fit”)
 - **oblmin rotace** - pomocí parametru, který kontroluje velikost korelace mezi faktory, ovlivňuje tvar finální faktorové matice
 - **promax rotace** - vychází z varimax, poskytuje strukturu s využitím co nejnižších koeficientů ve faktorové matici a korelací mezi faktory

- **rotace není magie**, distribuce bodů zůstává zachována, pouze povoluje zvolit nové osy (kolmé či šikmé), tak aby pozice bodů šly lépe, jednodušeji popsat
- rotace jsou také někdy aplikovány na výsledky metody hlavních komponent (“faktory”, které jsou na sebe kolmé a snaží se postupně vysvětlovat co největší část celkové variability dat), ale tím se ztrácí jejich definující vlastnost - přestávají obsahovat maximální část celkové variability dat

- v příkladu jsme zvolili dva faktory
- často však může **řešení s odlišným počtem faktorů** “**vyprodukovat**” **odlišné faktory** (pokud není zvolena metoda hlavním komponent)
- **Jak tedy volit počet faktorů?**
 - někdy lze volit **subjektivně** podle nejvíce přesvědčivé interpretace
 - někdy využít **scree diagram** - ale ten se týká hlavních komponent nikoli faktorů
 - zachycující vlastní čísla jednotlivých faktorů, které reprezentují variabilitu hlavních komponent
 - doporučení: počet faktorů = počet vlastních čísel větších než 1; nebo tam, kde v scree plotu pozorujeme nejvyšší pokles vlastních čísel mezi dvěma faktory
 - **formální statistický test** - lze využít v případě odhadu modelu pomocí metody maximální věrohodnosti, která se využívá nejčastěji (mimo MML existuje mnoho metod na odhad faktorů...)

Je čas předat slovo Martinovi - nutné začít s konkrétní hrou s daty