

Cvičení č. 9.: Provedení faktorové analýzy

Příklad: Výsledky desetiboje z olympiády v Aténách 2004 (Příklad je převzat z knihy Meloun M., Militký J., Hill, M.: Počítačová analýza vícerozměrných dat v příkladech. Academia Praha 2005)

V datovém souboru Desetiboj.sta jsou uloženy výsledky 39 závodníků - mužů, kteří se v roce 2004 zúčastnili desetiboje na olympiádě v Aténách. Zajímají nás výsledky jednotlivých disciplín, tj. proměnné v14 – v23. Budeme se snažit najít menší počet společných faktorů, které vysvětlují variabilitu výsledků závodníků v desetiboji. Přitom budeme uvažovat jenom závodníky, kteří desetiboj dokončili, tj. v proměnné Dokončil je 1.

Řešení v systému STATISTICA:

Sestavení korelační matice:

Statistiky – Vícerozměrné průzkumné techniky – Faktorová analýza – Select cases – zaškrtneme Zapnout filtr – vybereme ty případy, kdy v2=1 – OK - Proměnné v14 až v23 – OK – OK. Na záložce Popisné statistiky zvolíme Přehled korelací, průměrů, směrodatných odchylek – Korelace

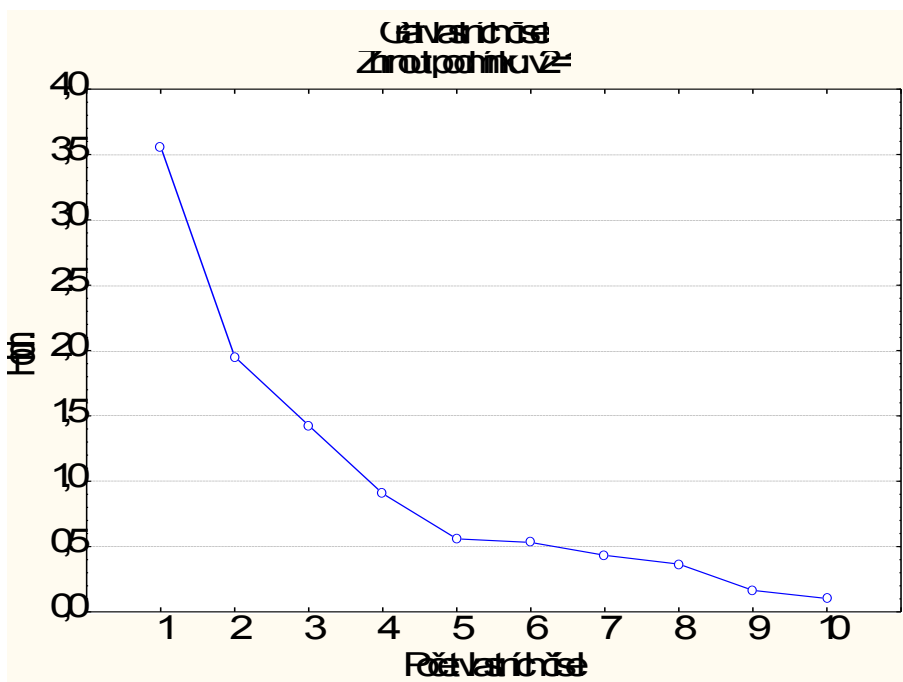
		Korelace (Desetiboj.sta) ChD vynechána případ N=28 Zhrnout podmínku: v2=									
Prom		1	2	3	4	5	6	7	8	9	10
Body	1	1	0	0	0	0	0	0	0	0	0
Body	0	1	0	0	0	0	0	0	0	0	0
Body	0	0	1	0	0	0	0	0	0	0	0
Body	0	0	0	1	0	0	0	0	0	0	0
Body	0	0	0	0	1	0	0	0	0	0	0
Body	0	0	0	0	0	1	0	0	0	0	0
Body	0	0	0	0	0	0	1	0	0	0	0
Body	0	0	0	0	0	0	0	1	0	0	0
Body	0	0	0	0	0	0	0	0	1	0	0
Body	0	0	0	0	0	0	0	0	0	1	0
Body	0	0	0	0	0	0	0	0	0	0	1

Některé korelace mezi proměnnými jsou dostatečně vysoké, zřejmě tedy má smysl provádět faktorovou analýzu.

Vypočteme vlastní čísla výběrové korelační matice, zjistíme procento vysvětleného rozptylu a nakreslíme sutinový graf.

Na záložce Základní nastavení změním Max. počet faktorů na 10 a Min. vlastní číslo na 0 – OK – na záložce Výklad rozptylu zvolíme Vlastní čísla a poté Sutinový graf.

VI. čísla (Destiboj.sta) Extrakce: Hlavní komp Zhrnout podmínku: v2=				
Hoc	vl. čí	% čí rozp.	Kumu vlast.	Kumu %
1	3,55	35,5	3,55	35,5
2	1,95	19,5	5,51	55,1
3	1,42	14,2	6,93	69,3
4	0,90	9,0	7,84	78,4
5	0,55	5,5	8,40	84,0
6	0,53	5,3	8,93	89,3
7	0,43	4,3	9,36	93,6
8	0,36	3,6	9,72	97,2
9	0,16	1,6	9,88	98,8
10	0,10	1,0	10,0	100,0



Zkusíme pracovat se čtyřmi faktory., které vysvětlují asi 78% variability obsažené v datech. Zlom v sutinovém grafu je sice až u 5 faktorů, ale to už je příliš velký počet.

Spočteme komunalitu pro první čtyři faktory. Na záložce Základní nastavení zadáme Max. počet faktorů 4 – OK. Na záložce Zákl. výsledky zvolíme Rotace faktorů Varimax prostý. Na záložce Výklad rozptylu zvolíme Komunalitu.

Proměnná	Komunalita (Děsibo)				
	Extrakce: Hlavní kom	Rotace: Varimax pr.	Zhrnout podmínku: \		
	Λ fakt	Λ fakt	Λ fakt	Λ fakt	Víc R ²
Body 100	0,60	0,68	0,75	0,76	0,65
Body skok	0,65	0,70	0,76	0,76	0,71
Body koule	0,01	0,72	0,72	0,80	0,70
Body výška	0,06	0,67	0,67	0,62	0,53
Body 400	0,82	0,82	0,82	0,82	0,79
Body přehod	0,57	0,56	0,51	0,51	0,39
Body disk	0,01	0,75	0,71	0,71	0,60
Body tyč	0,03	0,02	0,89	0,89	0,29
Body oštěp	0,00	0,01	0,01	0,90	0,39
Body 1500	0,22	0,39	0,60	0,89	0,61

Vidíme, že např. variabilita proměnné Body na 100 m je ze 76,5% vysvětlena prvními čtyřmi faktory.

Nyní získáme odhad matice rotovaných faktorových zátěží: na záložce Zátěže zvolíme Shrnutí: Faktorové zátěže.

Proměnná	Faktor. zátěže (Varimax)			
	Extrakce: Hlavní kom	(Označené zátěže jsou	Zhrnout podmínku: v2=	
	Fakt 1	Fakt 2	Fakt 3	Fakt 4
Body 100	0,77	0,28	0,25	-0,10
Body skok	0,81	0,21	0,24	0,00
Body koule	0,13	0,83	0,08	0,26
Body výška	0,26	0,73	-0,08	0,10
Body 400	0,90	-0,01	-0,09	0,13
Body přehod	0,71	0,22	0,07	-0,01
Body disk	0,10	0,86	-0,14	0,02
Body tyč	0,17	-0,12	0,92	0,01
Body oštěp	-0,03	0,26	0,04	0,90
Body 1500	0,47	-0,41	-0,45	0,53
Vykl.roz	2,95	2,42	1,24	1,22
Prp.celk	0,29	0,24	0,12	0,12

První faktor vysoce koreluje s výsledky krátkých běhů a skoku do dálky. Lze ho označit jako rychlost. Druhý faktor koreluje s výsledky hodů koule, disku a skoku do výšky. Je možné ho interpretovat jako schopnost zkoncentrovat výbušnou energii do jediného okamžiku. Třetí faktor koreluje s výsledkem skoku o tyči. Vzhledem k vysokému korelačnímu koeficientu ho lze ztotožnit s touto proměnnou. To samé platí o čtvrtém faktoru, který vysoce koreluje s výsledkem hodů oštěpem. Proměnné body oštěp a Body tyčka jsou tedy unikátní a bez výraznějšího vztahu ke znakům ostatním proměnným.

Kvalitu získaného faktorového modelu posoudíme též pomocí odhadnuté korelační a reziduální korelační matice. Na záložce Výklad rozptylu vybereme Reprod./rezid. korelace.

		Reprodukované korelace Extrakce: Hlavní komponenty Zhrnout podmínku: $\lambda^2 \geq 0,1$									
Proměnná		1 (skokový)	2 (předtýl)	3 (tyčka)	4 (oštěp)	5 (stát)	6 (celkem)	7 (boby)	8 (boby)	9 (boby)	10 (boby)
Boby		1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000

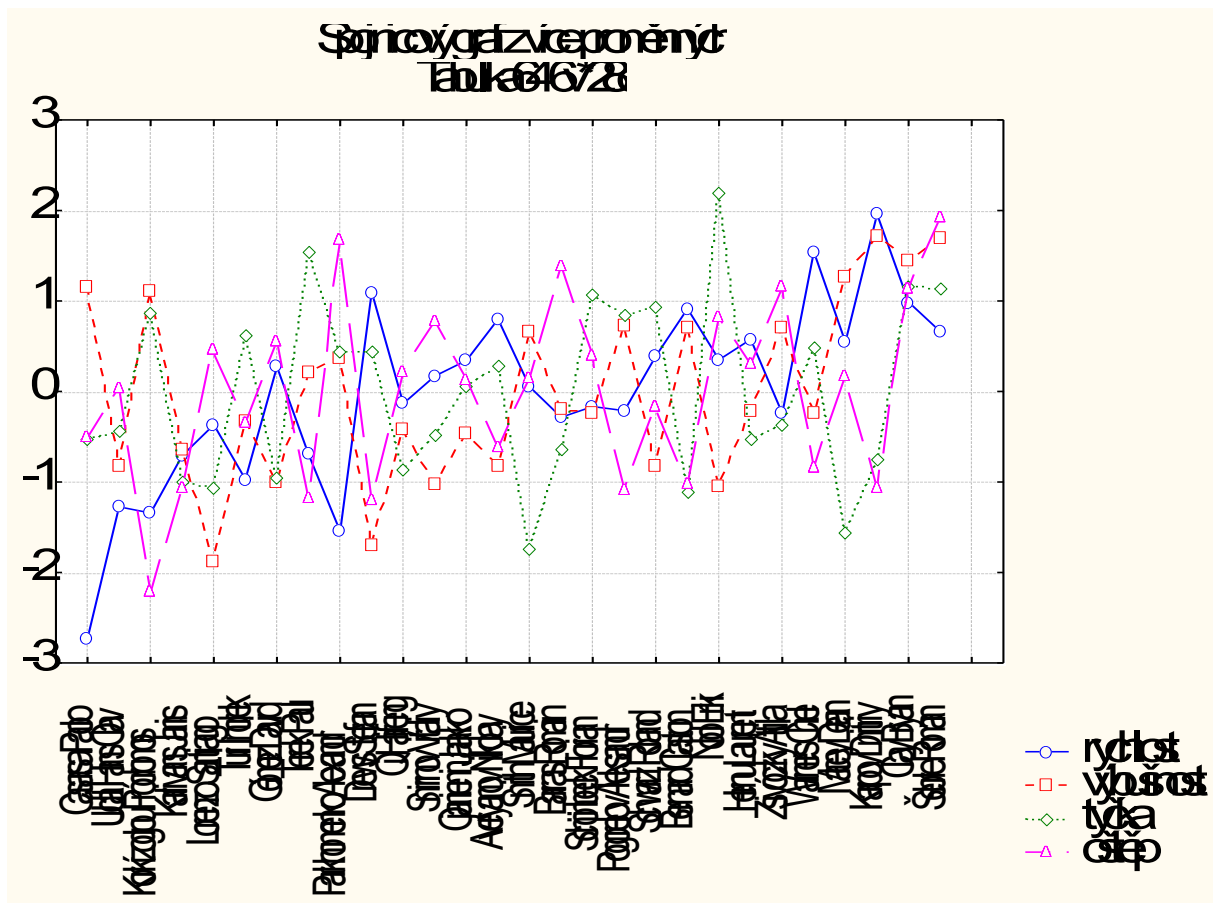
		Reziduální korelace (D) Extrakce: Hlavní komponenty (Označená rezidua jsou) Zhrnout podmínku: $\lambda^2 \geq 0,1$									
Proměnná		1 (skokový)	2 (předtýl)	3 (tyčka)	4 (oštěp)	5 (stát)	6 (celkem)	7 (boby)	8 (boby)	9 (boby)	10 (boby)
Boby		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,000
Boby		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000

Nyní uložíme faktorová skóre. Na záložce Skóre vybereme Uložit faktorová skóre. Uložíme je společně s proměnnými Stát a Celkem body. Faktor 1 pak přejmenujeme na rychlost, faktor 2 na výbušnost, faktor 3 na tyčka a faktor 4 na oštěp. Závodníky ještě seřadíme podle dosaženého počtu bodů.

	Faktor. skóre pro 4					
	1 Stá	2 Celken	3 rychl	4 výbuš	5 tyčk	6 oště
Casarsa Pac	IT	74	-2,71	1,15	-0,51	-0,50
Uldal Hans C	NO	74	-1,21	-0,81	-0,41	0,04
Korkizoglou	GR	75	-1,31	1,11	0,81	-2,21
Karlivans Jar	LT	75	-0,71	-0,61	-0,91	-1,01
Lorenzo Sar	AF	75	-0,31	-1,81	-1,01	0,41
Turi Indrek	ES	77	-0,91	-0,31	0,61	-0,31
Gómez Davi	ES	78	0,21	-1,01	-0,91	0,51
Ierek Paul	US	78	-0,61	0,21	1,51	-1,11
Parkhomenk	BL	79	-1,51	0,31	0,41	1,61
Drews Steta	GE	79	1,01	-1,61	0,41	-1,21
Qi Haiteng	CH	79	-0,11	-0,41	-0,81	0,21
Smirnov Vita	UZ	79	0,11	-1,01	-0,41	0,71
Ojaniemi Ja	FI	80	0,31	-0,41	0,01	0,11
Averyanov N	RU	80	0,71	-0,81	0,21	-0,61
Smith Mauri	JA	80	0,01	0,61	-1,71	0,11
Barras Rom	FR	80	-0,21	-0,11	-0,61	1,31
Schönbeck I	GE	80	-0,11	-0,21	1,01	0,41
Pogorelov Al	RU	80	-0,21	0,71	0,81	-1,01
Schwarzl Ro	AT	81	0,31	-0,81	0,91	-0,11
Bernard Clas	JA	82	0,91	0,71	-1,11	-1,01
Nool Erki	ES	82	0,31	-1,01	2,11	0,81
Hernu Laure	FR	82	0,51	-0,21	-0,51	0,21
Zsivoczky A	HU	82	-0,21	0,71	-0,31	1,11
Warners Chi	NE	83	1,51	-0,21	0,41	-0,81
Macey Dear	GE	84	0,51	1,21	-1,51	0,11
Karpov Dmit	KA	87	1,91	1,71	-0,71	-1,01
Clay Bryan	US	88	0,91	1,41	1,11	1,11
Sebrle Rom	CZ	88	0,61	1,71	1,11	1,91

Nyní sestrojíme spojnicový graf faktorových skóre.

Grafy – 2D Grafy – Spojnicové grafy (Proměnné) – Proměnné rychlost – oštěp – OK,
zapneme Vícenásobný – OK



Na první pohled zde nedominuje žádný z faktorů. Znamená to, že k vítězství je potřeba souhra všech. Co se týká jednotlivých závodníků, vidíme např., že Roman Šebrle má jedny z nejlepších skóre u všech faktorů, proto také vyhrál na těchto OH.

Podívejme se ještě, jak se změni výsledky, když změni metodu extrakce faktorů a metodu rotace. Na záložce Details zvolíme Centroidovou metodu a na záložce Základní výsledky vybereme Varimax normalizovaný.

Vlastní čísla a procento vysvětleného rozptylu:

VI. čísla (Des				
Extrakce: Hja				
Zhrnout podn				
H ₀	VI.	% roz	Kur vlas	Kur %
1	3,1	31	3,1	31
2	1,6	16	4,8	48
3	1,2	12	6,0	60
4	0,4	4	6,5	65

Poněkud pokleslo procento vysvětleného rozptylu, z 78% na 65%.

Faktorové zátěže:

Proměnná	Faktor. zátěže (Varimá Extrakce: Hlavní faktor (Označené zátěže jsou Zhrnout podmínku: $v^2=$			
	Fakt 1	Fakt 2	Fakt 3	Fakt 4
Body 100	0,80	0,16	0,08	0,18
Body skok	0,81	0,15	-0,05	0,12
Body kolár	0,14	0,94	0,02	0,09
Body výška	0,32	0,59	0,10	-0,10
Body 400	0,80	0,03	-0,35	0,02
Body pře	0,63	0,17	-0,04	0,03
Body dis	0,19	0,70	0,19	-0,20
Body tyč	0,21	-0,05	0,14	0,66
Body os	-0,00	0,41	-0,25	0,01
Body 1500	0,20	-0,01	-0,95	-0,20
Vykl.roz	2,61	2,01	1,25	0,62
Prp.celk	0,26	0,20	0,12	0,06

Na rozdíl od metody hlavních komponent koreluje třetí faktor s proměnnou Body 1500 m, lze ho tedy interpretovat jako vytrvalost.

Faktorovou strukturu můžeme též znázornit graficky v prostoru faktorových zátěží. Vytvoří se shluky jednotlivých proměnných, přičemž každý shluk reprezentuje takovou skupinu disciplín, kterou lze vysvětlit působením stejného faktoru.

Na záložce Zátěže zvolíme Graf zátěží, 3D.

