

Cvičení č. 5.: 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

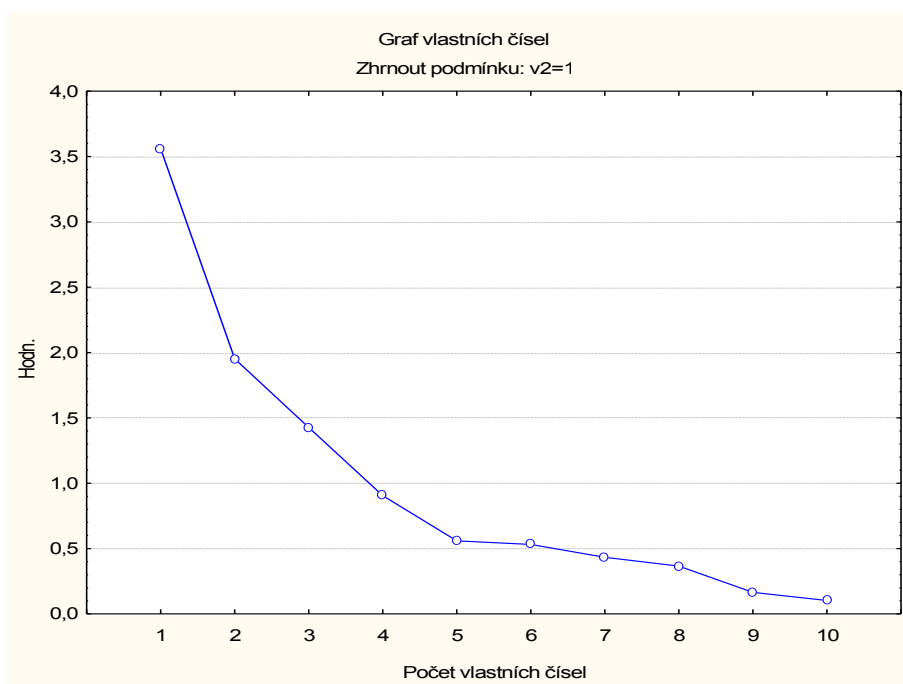
		Korelace (Desetiboj.sta)									
		ChD vynechána případově N=28 Zhmout podmínku: v2=1									
Proměnné		100 skok	skok c kou	výš	400	překá	disk	tyč	oštěp	1500	
Body 100		1,0	0,7	0,3	0,3	0,6	0,4	0,2	0,2	0,0	0,0
Body skok		0,7	1,0	0,2	0,3	0,6	0,4	0,2	0,2	0,1	0,1
Body koule		0,3	0,2	1,0	0,6	0,2	0,2	0,6	0,0	0,3	-0,1
Body výška		0,3	0,2	0,6	1,0	0,1	0,3	0,4	-0,1	0,2	0,0
Body 400		0,6	0,6	0,2	0,1	1,0	0,4	0,1	0,1	0,0	0,4
Body překážka		0,4	0,4	0,2	0,3	0,4	1,0	0,2	0,1	0,0	0,1
Body disk		0,2	0,2	0,6	0,4	0,1	0,2	1,0	-0,1	0,2	-0,1
Body tyčka		0,2	0,2	0,0	-0,1	0,1	0,1	-0,1	1,0	-0,1	-0,1
Body oštěp		0,0	0,1	0,3	0,2	0,0	0,0	0,2	-0,1	1,0	0,2
Body 1500		0,0	0,1	-0,1	0,0	0,4	0,1	-0,1	-0,1	0,2	1,0

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 zvolím Vlastní čísla a poté Sutinový graf.

VI. čísla (Destiboj.sta)				
Extrakce: Hlavní komponenty				
Zhrnout podmínku: v2=1				
Hodn.	vl. číslo	% celk. rozptylu	Kumulativ. vlast. číslo	Kumulativ. %
1	3,55921	35,5921	3,5592	35,592
2	1,95291	19,5291	5,5121	55,121
3	1,42658	14,2658	6,9387	69,387
4	0,90534	9,0534	7,8440	78,440
5	0,55875	5,5875	8,4028	84,028
6	0,53156	5,3156	8,9343	89,343
7	0,43280	4,3280	9,3671	93,671
8	0,36574	3,6574	9,7329	97,329
9	0,16463	1,6463	9,8975	98,975
10	0,10244	1,0244	10,0000	100,000



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	Komunality (Destiboj.sta) Extrakce: Hlavní komponenty Rotace: Varimax pr. Zhrnout podmínku: $\sqrt{2}=1$				
	Z 1 faktorů	Z 2 faktorů	Z 3 faktorů	Z 4 faktorů	Více R ²
Body 100 m	0,6071	0,6883	0,7545	0,7654	0,6529
Body skok dál	0,6572	0,7015	0,7624	0,7624	0,7190
Body koule	0,0187	0,7216	0,7287	0,8010	0,7047
Body výška	0,0692	0,6100	0,6179	0,6287	0,5379
Body 400 m	0,8205	0,8208	0,8293	0,8481	0,7926
Body překážk	0,5158	0,5654	0,5704	0,5706	0,3914
Body disk	0,0106	0,7548	0,7766	0,7772	0,6020
Body tyčka	0,0302	0,0454	0,8982	0,8985	0,2980
Body oštěp	0,0010	0,0727	0,0744	0,9008	0,3969
Body 1500 m	0,2266	0,3989	0,6072	0,8907	0,6771

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 pr.) (Destiboj.sta) Extrakce: Hlavní komponenty (Označené zatěže jsou >,700000) Zhrnout podmínku: $\sqrt{2}=1$			
	Faktor 1	Faktor 2	Faktor 3	Faktor 4
Body 100 m	0,779215	0,284888	0,257295	-0,104642
Body skok dálka	0,810735	0,210319	0,246768	0,005051
Body koule	0,137047	0,838389	0,084332	0,268859
Body výška	0,263195	0,735405	-0,088424	0,104022
Body 400 m	0,905847	-0,016782	-0,092373	0,137160
Body překážky	0,718253	0,222654	0,070625	-0,015438
Body disk	0,103118	0,862698	-0,147690	0,024477
Body tyčka	0,173922	-0,123269	0,923456	0,019420
Body oštěp	-0,031883	0,267834	0,041442	0,909056
Body 1500 m	0,476054	-0,415136	-0,456436	0,532356
Výkl.roz	2,957493	2,422484	1,240182	1,223894
Prp.celk	0,295749	0,242248	0,124018	0,122389

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 (Destiboj.sta) Extrakce: Hlavní komponenty Zhmout podmínku: $v^2=1$	
Proměnné	100 skok c kou výš 400 překá dis tyč oštěp 1500
Body 100	0,7 0,7 0,3 0,3 0,6 0,6 0,2 0,3 -0,1 0,0
Body skok	0,7 0,7 0,3 0,3 0,7 0,6 0,2 0,3 0,0 0,7
Body koule	0,3 0,3 0,8 0,6 0,7 0,2 0,7 0,0 0,4 -0,3
Body výška	0,3 0,3 0,6 0,6 0,2 0,3 0,6 -0,1 0,2 -0,1
Body 400m	0,6 0,7 0,7 0,2 0,8 0,6 0,7 0,0 0,0 0,4
Body překážka	0,6 0,6 0,2 0,3 0,6 0,4 0,2 0,7 0,0 0,2
Body diskus	0,2 0,2 0,7 0,6 0,7 0,2 0,7 -0,1 0,2 -0,3
Body tyčka	0,3 0,3 0,0 -0,1 0,0 0,7 -0,1 0,9 0,0 -0,3
Body oštěp	-0,1 0,0 0,4 0,2 0,0 0,0 0,2 0,0 0,9 0,3
Body 1500m	0,0 0,7 -0,1 -0,1 0,4 0,2 -0,1 -0,1 0,3 0,8

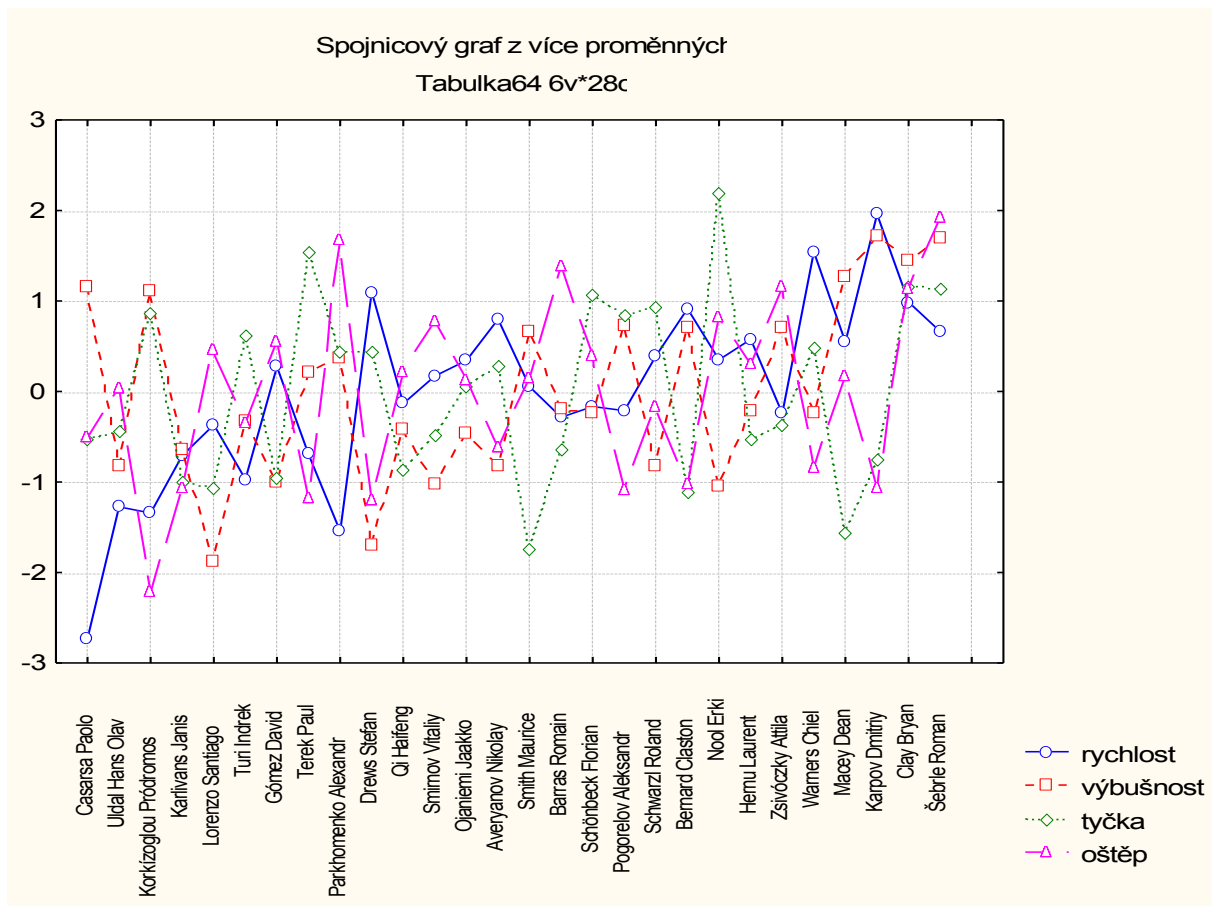
Reziduální korelace (Destiboj.sta) Extrakce: Hlavní komponenty (Označená rezidua jsou > ,100000) Zhmout podmínku: $v^2=1$	
Proměnné	100 skok c kou výš 400 překá dis tyč oštěp 1500
Body 100	0,2 -0,1 0,0 -0,1 -0,1 -0,1 -0,1 -0,1 -0,1 0,0 -0,1
Body skok	-0,1 0,2 -0,1 0,0 -0,1 -0,1 0,0 -0,1 0,0 -0,1
Body koule	0,0 -0,1 0,2 -0,1 0,0 -0,1 -0,1 0,0 -0,1 0,0
Body výška	-0,1 0,0 -0,1 0,3 -0,1 -0,1 -0,1 0,0 -0,1 0,0
Body 400m	-0,1 -0,1 0,0 -0,1 0,7 -0,1 0,0 0,0 -0,1 -0,1
Body překážka	-0,1 -0,1 -0,1 -0,1 -0,1 0,4 -0,1 -0,1 0,0 -0,1
Body diskus	-0,1 0,0 -0,1 -0,1 0,0 -0,1 0,2 0,0 0,0 0,0
Body tyčka	-0,1 -0,1 0,0 0,0 0,0 -0,1 0,0 0,7 -0,1 0,0
Body oštěp	0,0 0,0 -0,1 -0,1 -0,1 0,0 0,0 -0,1 0,7 -0,1
Body 1500m	-0,1 -0,1 0,0 0,0 -0,1 -0,1 0,0 0,0 -0,1 0,7

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 Stát	2 Celkem body	3 rychlost	4 výbušnos	5 tyčka	6 oštěp
Casarsa Paolo	ITA	740	-2,7291	1,1576	-0,5322	-0,5084
Uldal Hans Olav	NOR	749	-1,2738	-0,8229	-0,4444	0,0431
Korkizoglou Pródromos	GRE	757	-1,3422	1,1115	0,8685	-2,2203
Karlivans Janis	LAT	758	-0,7089	-0,6499	-0,9899	-1,0581
Lorenzo Santiago	ARG	759	-0,3682	-1,8676	-1,0575	0,4711
Turi Indrek	EST	770	-0,9708	-0,3308	0,6141	-0,3488
Gómez David	ESP	786	0,2730	-1,0095	-0,9507	0,5593
Terek Paul	USA	789	-0,6932	0,2150	1,5486	-1,1818
Parkhomenko Alexandr	BLR	791	-1,5463	0,3736	0,4467	1,6690
Drews Stefan	GER	792	1,0981	-1,6925	0,4301	-1,2087
Qi Haifeng	CHN	793	-0,1330	-0,4250	-0,8597	0,2141
Smirnov Vitaliy	UZB	799	0,1700	-1,0264	-0,4930	0,7793
Ojaniemi Jaakko	FIN	800	0,3440	-0,4522	0,0505	0,1342
Averyanov Nikolay	RUS	802	0,7988	-0,8223	0,2756	-0,6249
Smith Maurice	JAM	802	0,0526	0,6712	-1,7458	0,1525
Barras Romain	FRA	806	-0,2853	-0,1966	-0,6433	1,3897
Schönbeck Florian	GER	807	-0,1663	-0,2273	1,0600	0,4036
Pogorelov Aleksandr	RUS	808	-0,2113	0,7204	0,8543	-1,0797
Schwarzl Roland	AUT	810	0,3933	-0,8186	0,9273	-0,1611
Bernard Claston	JAM	822	0,9152	0,7157	-1,1186	-1,0235
Nool Erki	EST	823	0,3506	-1,0453	2,1964	0,8258
Hemu Laurent	FRA	823	0,5764	-0,2028	-0,5182	0,2989
Zsivóczky Attila	HUN	828	-0,2417	0,7095	-0,3743	1,1495
Wamers Chiel	NEZ	834	1,5431	-0,2378	0,4871	-0,8402
Macey Dean	GBR	841	0,5461	1,2715	-1,5603	0,1760
Karpov Dmitriy	KAZ	872	1,9578	1,7241	-0,7633	-1,0724
Clay Bryan	USA	882	0,9880	1,4549	1,1474	1,1411
Šebrle Roman	CZE	889	0,6630	1,7027	1,1447	1,9207

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 změním metodu extrakce faktorů a metodu rotace. Na záložce Detaily zvolíme Centroidovou metodu a na záložce Základní výsledky vybereme Varimax normalizovaný.

Vlastní čísla a procento vysvětleného rozptylu:

VI. čísla (Destiboj.sta)				
Extrakce: Hlavní faktory (C				
Zhrnout podmínku: v2=1				
Hoc	vl. čís	% ce rozpt	Kumula vlast. č	Kumula %
1	3,18€	31,8€	3,18€	31,8€
2	1,64€	16,4€	4,832	48,32
3	1,261	12,61	6,093	60,93
4	0,412	4,12	6,50€	65,0€

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

Faktorové zátěže:

Faktor. zátěže (Varimax normaliz.) (Destiboj.st Extrakce: Hlavní faktory (Centroid) (Označené zatěže jsou >,700000) Zhrnout podmínku: v2=1				
Proměnná	Faktor 1	Faktor 2	Faktor 3	Faktor 4
Body 100 m	0,801932	0,169116	0,081079	0,189382
Body skok dálka	0,816233	0,159121	-0,039807	0,129518
Body koule	0,145896	0,944283	0,026342	0,095964
Body výška	0,322114	0,590216	0,107112	-0,103546
Body 400 m	0,807052	0,037248	-0,356082	0,021918
Body překážky	0,636637	0,175575	-0,049495	0,030171
Body disk	0,193464	0,704057	0,192067	-0,262831
Body tyčka	0,218152	-0,096102	0,149190	0,664741
Body oštěp	-0,008345	0,418982	-0,238997	0,011212
Body 1500 m	0,201180	-0,076151	-0,993031	-0,201506
Výkl.roz	2,616575	2,012435	1,251947	0,625656
Prp.celk	0,261658	0,201244	0,125199	0,062566

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.

