

7. Úkol na předmět Mnohorozměrná analýza

Jaromír Macoun

Duben 2022

1 Úvod

Zabýváme se v tomto úkolu daty *athletic*, které obsahují záznamy o rekordních časech v atletice. Přesněji řečeno data obsahují národní rekordy 55 zemí v 8 běžeckých disciplínách. Disciplíny to jsou následující: běh na 100, 200, 400, 800, 1500, 5000, 1000 metrů a maratonský běh.

2 Metoda hlavních komponent (viz. minulé cvičení)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	2.4570	0.9931	0.7759	0.3604	0.3173	0.2726	0.2156	0.1535
Proportion of Variance	0.7546	0.1233	0.0752	0.0162	0.0126	0.0093	0.0058	0.0029
Cumulative Proportion	0.7546	0.8779	0.9531	0.9694	0.9819	0.9912	0.9971	1.0000

Jak je vidět z tabulky, již první hlavní komponenta vyjádří 75% variability dat. Společně s druhou vysvětluje 87%, což nám již může stačit. Standardní chyba druhé komponenty je značně menší než první. Proto se zdá, že pro faktorovou analýzu by mohly stačit dvě komponenty.

3 Analýza hlavních komponent

Na základě analýzy hlavních komponent se zdá, že budou stačit dva faktory. Pro jistotu zkusíme ještě třetí faktor a porovnáme. Následující tabulka ukazuje kumulativní vysvětlenou variabilitu, a P-hodnotu zdali je počet faktorů dostačující.

Z tabulky č. 1 je vidět, že stejně jako na základě analýzy hlavních komponent model s dvěma faktory vysvětluje více než 80% variability, avšak P-hodnota pro dva faktory je nízká. Proto na základě tohoto budeme uvažovat model s třemi faktory.

Jak je vidět z tabulky č.2, tak s prvním faktorem jsou velmi silně korelované běhy na 42, 10, 5 a 1.5 kilometrů. Stále silně korelované s prvním faktorem je i běh na 800 metrů. Tedy první faktor by mohl nahrazovat dlouhé běhy. Nicméně

Factors	Cumulative	p val
1	0.713	<0.001
2	0.815	0.023
3	0.865	0.363

Tabulka 1: Počet faktorů

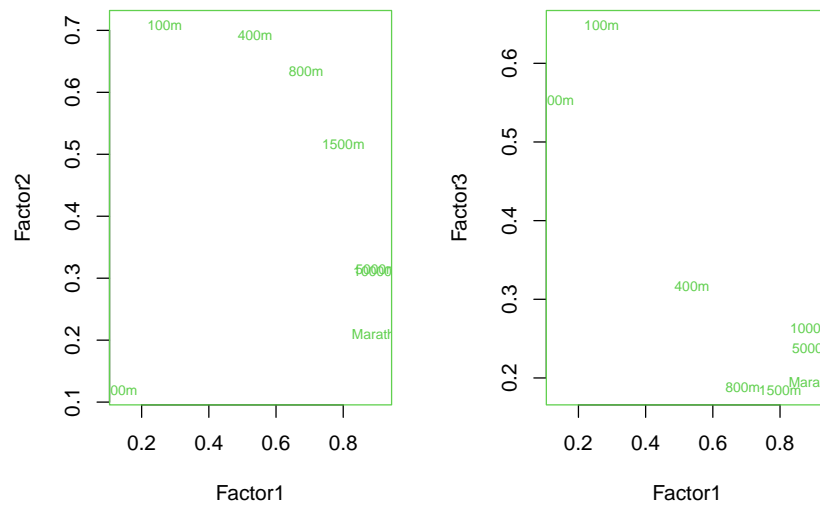
	Factor1	Factor2	Factor3
100m		0.71	0.65
200m			
400m		0.69	
800m	0.69	0.63	
1500m	0.80		
5000m	0.90		
10000m	0.90		
Marathon	0.91		

Tabulka 2: Výstup faktorové analýza

800 metrů běh je také silně korelovaný s druhým faktorem, který je kromě toho korelovaný s během na 400a100 metrů. To, že osmistovka je korelovaná s oběma faktory může být tím, že je to běh, který je na pomezí rychlých a vytrvalostních disciplín, (anaerobních a aerobních běhů). Dále je vidět, že faktor 3 koreluje se sprintem na 100 metrů.

I na základě obrázku č. 1 je vidět, že se běh na 200 metrů odlišuje od všech ostatních. Dále je vidět postupně se zmenšující se korelace se zkracovanou vzdáleností běhu na první faktor.

Tedy kdybychom potřebovali využít v nějakém modelu, tak by asi stačily pouze dva faktory dlouhé a krátké běhy s tím že by se 200 metrů vzalo jako samostatná disciplína. Kdybychom trvali na použití všech třech faktorů, tak třetí by nejlépe vystihovala běh na 100 a 200 metrů. Korelace mezi třetím faktorem a a během na 200 metrů je 0.55.



Obrázek 1