

## 6. Úkol na předmět Mnohorozměrná analýza

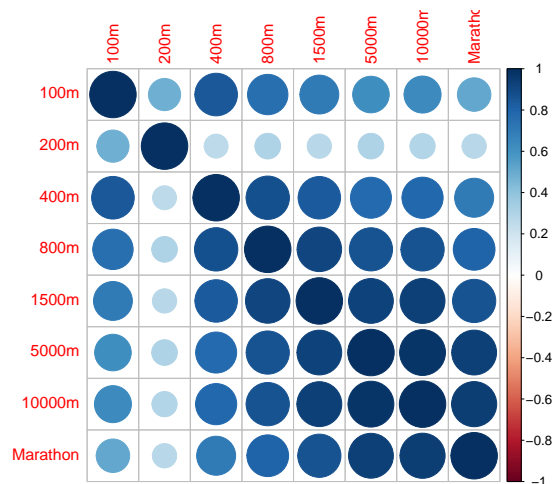
Jaromír Macoun

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, 10000 metrů a maratonský běh.

Na obrázku č. 1 jsou znázorněny po dvou korelace těchto disciplín. Jak je vidět všechny disciplíny spolu velmi silně pozitivně korelují. Ovšem co je zvláštní, tak z tohoto faktu vybočuje běh na 200 metrů, tato disciplína sice stále koreluje se všemi ostatními běhy, nicméně o poznání menší silou. Na základě těchto výsledků se zdá, že již první hlavní komponenta vyjádří velké množství variability dat.

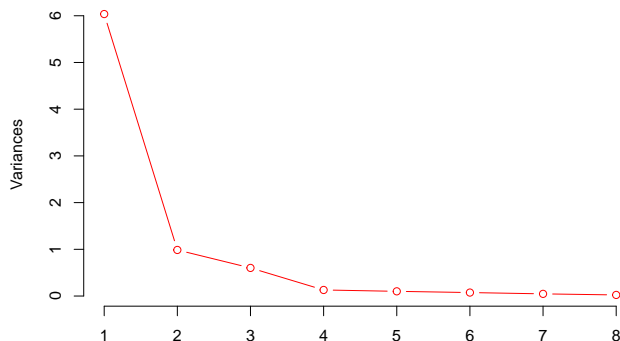


Obrázek 1: Matice korelací

	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

## 2 Metoda hlavních komponent

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, jak je vidět z obrázku č.2 a z tabulky.



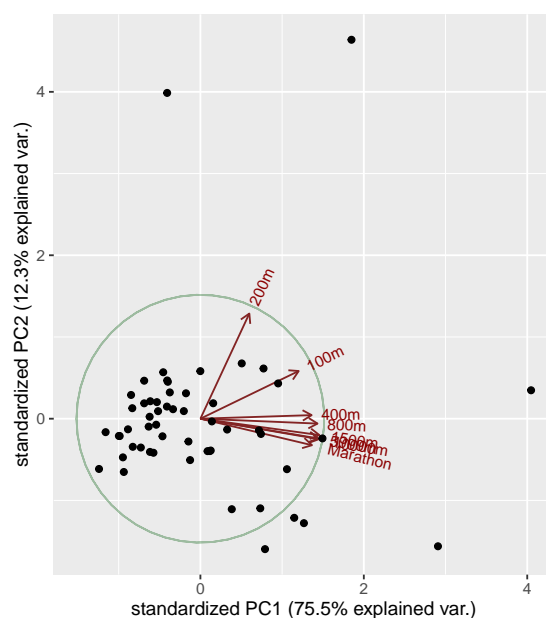
Obrázek 2

Z grafu č. 3 vidíme, že první komponenta dobře vysvětluje všechny běhy kromě již zmiňovaného běhu na 200 metrů, který slabě koreloval s ostatními. Proto je třeba ještě druhá komponenta k získání informace 88% variability. Kromě 200 metrového běhu i sprint na 100 metrů není moc dobře vysvětlen první komponentou.

Z původních 8 proměnných se nám podařilo zredukovat dimenzi na 2 při zachování zhruba 88% "informace".

Dále je zajímaví si všimnout z obrázku č.4, že pozorování Nizozemí a Mauricius nabývají vysoké hodnoty druhé komponenty, naproti tomu Cookovy ostrovy nabývají vysokých hodnot první komponenty.

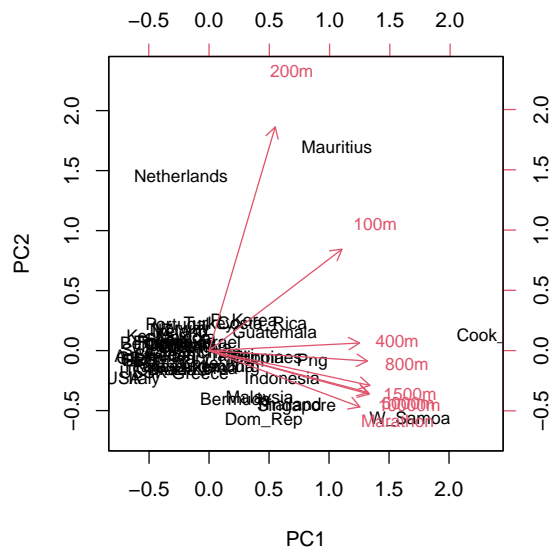
Jak je vidět z obrázku i z následující tabulky, tak Mauricius má relativně vysoké hodnoty (časy) v disciplíně na 200m a relativně (vzhledem k měřítku i na 100m. Hlavní komponenty byly udělané po přeškálování neboť různé disciplíny mají různé jednotky. Podobně Nizozemsko nemá dobrý rekord v disciplíně na 200m.



Obrázek 3

Oproti tomu Cookovy ostrovy mají horší čas než průměr snad ve všech disciplínách. V jediné kde nezaostávají tolik je běh na 200m. (lze vidět že nabývá relativně nízkých hodnot druhé komponenty zato vysokých hodnot první)

	100m	200m	400m	800m	1500m	5000m	10000m	Marathon
průměr	10.47	21.30	46.43	1.79	3.69	13.84	28.99	136.62
Netherlands	10.52	29.95	45.1	1.74	3.62	13.36	27.61	129.02
Mauritius	11.19	33.45	47.7	1.88	3.83	15.06	31.77	152.23
Cook Islands	12.18	23.2	52.94	2.02	4.24	16.7	35.38	164.7



Obrázek 4