**Phân tích thành tích Decathlon**

Các cuộc thi mười môn phối hợp tôn vinh các vận động viên hoàn thành tất cả các môn thi. Mỗi vận động viên đều có điểm mạnh và điểm yếu. Mục tiêu của bài tập là chính là sự dao động này. Kết quả của 4 cuộc thi mười môn phối hợp: Thế vận hội Olympic Athens năm 2004, Thế vận hội Rio năm 2016, Giải vô địch châu Âu 2018 và cuộc thi tài năng năm 2018 của 20 vận động viên được lưu trong tệp decathlon.csv.

Đối với mỗi vận động viên, kết quả bao gồm thành tích của họ trong mỗi nội dung trong số 10 nội dung thi đấu, số điểm của họ (trong mỗi nội dung thi đấu, vận động viên kiếm được điểm dựa trên thành tích của họ) và thứ hạng cuối cùng của họ.

Các nội dung thi đấu diễn ra theo thứ tự sau: ngày thứ nhất [thi 5 nội dung]: (1) chạy 100 mét, (2) nhảy xa, (3) đẩy tạ, (4) nhảy cao), và (5) chạy 400 mét; ngày thứ hai [thi 5 nội dung nữa]: (6) Chạy 110 mét rào, (7) ném đĩa, (8) nhảy sào, (9) ném lao, và (10) chạy 1.500 mét.

*Một số gợi ý để phân tích:*

* Mô tả dữ liệu (các biến? cá nhân?) và xác định các mục tiêu của PCA được thực hiện trên những dữ liệu này.
* Bạn có thể biết điều gì từ dữ liệu rút gọn quy tâm (sử dụng hàm scale)? Tại sao việc rút gọn quy tâm dữ liệu lại quan trọng trong ví dụ này
* Tính ma trận tương quan (có hàm cor và round để làm tròn kết quả).
* Sau đó, sử dụng hàm Factoshiny của gói Factoshiny để triển khai Phân tích thành phần chính của bảng dữ liệu này.
* Chúng ta nên sử dụng những biến nào để xây dựng PCA? Các biến này được gọi là biến active. Biểu diễn các biến khác như dưới dạng biến minh họa (biến phụ).
* Nhận xét về tỷ lệ phần trăm quán tính được giải thích bởi trục đầu tiên của PCA, theo mặt phẳng chính. Theo quan điểm nào thì 2 trục đầu tiên không giải thích quá nhiều về quán tính tổng sẽ thích hợp hơn?
* 5 vận động viên thể hiện xuất sắc nhất trên mặt phẳng chính là ai?
* Ai là vận động viên đóng góp nhiều nhất vào việc xây dựng trục chính thứ nhất của ACP? Và trong lần thứ 2? Đóng góp đáng kể đó có ý nghĩa gì?
* Từ vòng tròn tương quan, có thể nói gì về các mối tương quan sau:
* 100m vượt rào - 400m
* 100 m - Nhảy xa
* Đĩa - Đẩy tạ
* Đĩa - Nhảy xa
* Nhảy sào - 1500 m
* So sánh hồ sơ của MAYER và LELIEVRE. Sau đó là của Kaul và Duckworth. Bạn có thể nói gì về sự tương tự giữa Karpov và Clay? Giữa Ziemek và Bernard?
* Giải thích các trục chính của PCA (sử dụng biểu đồ cá nhân và biểu đồ các biến chủ động và minh họa).
* Từ 4 biến số được chiếu tốt nhất trên trục 3, hãy so sánh profile của các vận động viên đối lập trên trục này.
* Những người chiến thắng mười môn phối hợp xuất sắc trong những môn nào? Tồn tại trường hợp môn thi nào ít có ảnh hưởng đến thành tích (số điểm) trong mười môn phối hợp không?
* So sánh thành tích đạt được ở 4 cuộc thi. Bạn có thể nói gì về thành tích trung bình trong cuộc thi này?