Preselect pipeline

Overview

The preselection pipeline was designed to identify non-athlete volunteers whose food and nutrient intakes align with those of cyclist groups. Dietary intake data was collected using a Short Form Food Frequency Questionnaire (FFQ), previously validated by Cleghorn CL, Harrison RA, Ransley JK, Wilkinson S, Thomas J, Cade JE. **Can a dietary quality score derived from a short-form FFQ assess dietary quality in UK adult population surveys?** *Public Health Nutrition*. 2016. doi:10.1017/S1368980016001099. Additionally, the weekly physical activity levels of potential participants were recorded. The goal was to recruit 10 participants below the WHO's physical activity recommendations and 10 above these recommendations (600 MET-min/week).

The preselection process involved three main steps:

1) Latent Space Projection: Volunteers were mapped into a latent space (via Principal Coordinates Analysis) based on the absolute food intake profiles of cyclist groups.

2) Nutrient Intake Matching: The absolute nutrient intakes of the first 13 cyclists in the cohort were estimated. Non-athlete participants were included if their macronutrient intakes fell within the mean ± 2 standard deviations of the cyclists' macronutrient intakes.

3) Activity Profiling: Volunteers' physical activity levels were analyzed to compare with the WHO's physical activity recommendations.

1) Latent Space Projection

First of all, we import data of one volunteer's questionnaire and the food intakes of the cyclists groups. Their are available in https://github.com/DJrMartin/Gut-Microbiota-and-Physical-Activity/tree/main/data/data_preselect

```
survey <- read.csv(file="~/Desktop/Preselect_pipeline/data_preselect/Volunteers_example_1.
cyclists <- read.csv(file="~/Desktop/Preselect_pipeline/data_preselect/Food_intakes_cyclis
```

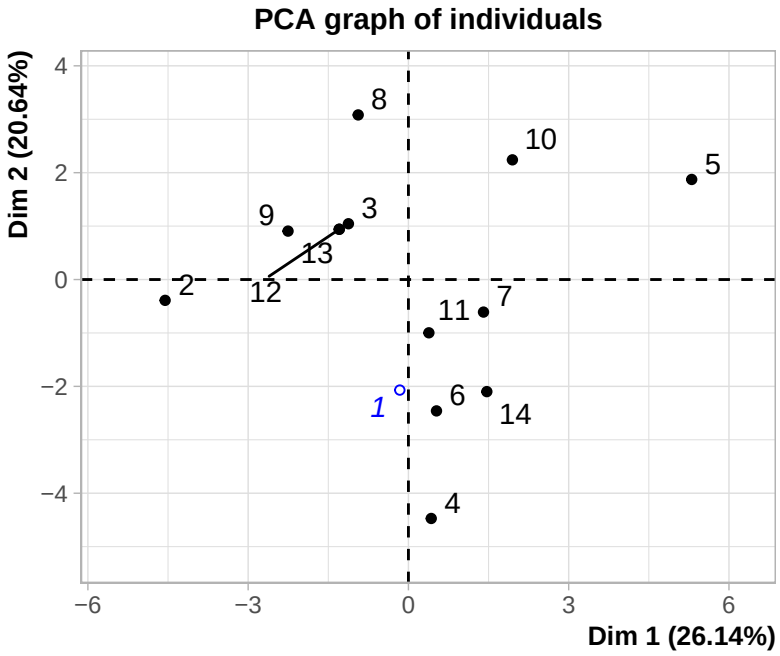
We confirme that the body mass index is above 18 kg/m^2 and under 25 kg/m^2.

```
print(paste('IMC =', round(survey$weight/(survey$height/100)^2, 2)))
```

```
[1] "IMC = 24.49"
```

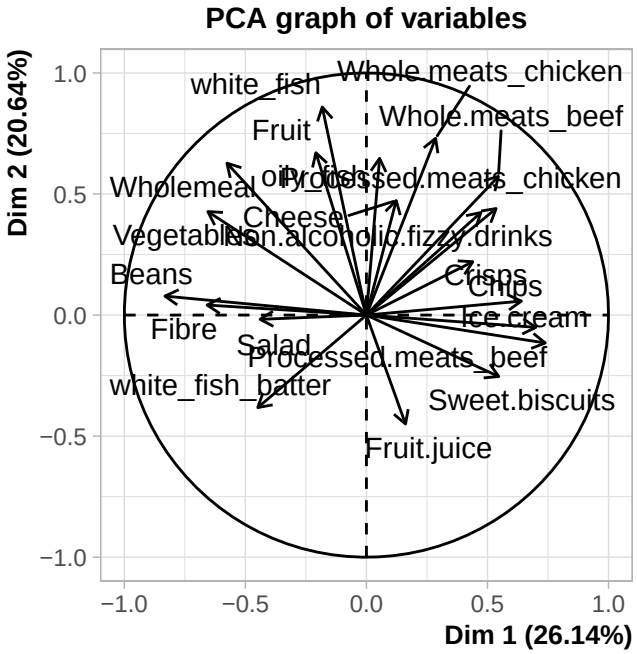
Then we fit a Principal Coordinates Analysis on the food intakes of the cyclist groups, and project the potential volunteer as a supplementary sample. The potential volunteer represents the blue point with the number 1. The cyclist volunteers are represented by the dark points.

```
input.pca <- rbind(survey[,-(1:5)], cyclists)
res.pca <- PCA(input.pca, ind.sup = 1, graph=F)
plot(res.pca, choix='ind')
```



Here, the correlation circle of the variables to interpret the two first axis.

```
# Plot of the variables
plot(res.pca, choix="var")
```



The coordinates of the supplemental individual, corresponding to the potential volunteer seem in the range of the cyclist group.

2) Nutrient Intake Matching

The second verification before including volunteer is to estimate the macro-nutriments (Vegetables, Fat and CHO) and the Dietary Quality Score of the cyclist and the potential volunteer. The corresponding tables to convert food in gramme of macro-nutrients are provided by Cleghorn CL, and al. (2016) in their online supplementary materials. We save their conversion tables in a list called NUTRIENTS_CONVERSION. Here, an example with the first cyclist. The output give the equivalent in grammes of Fruits, Fat, Vegetables, CHO and Oil Fish.

```
nutrient.conversion(cyclists[1,1:13],  # For the vegetables
                    cyclists[1,14:20], # For the meats
                    conversion = nutrients_conversion) # the conversion table
```

```
  Fruits      Fat Vegetables      CHO Oil.Fish DQS
1    120 84.90039        600 61.06378      4.5  12
```

Then, we calculate for the entire cyclists and the potential volunteer…

```
cyclist.nutrients=NULL
for(i in 1:nrow(cyclists)){
  cyclist.nutrients <- rbind(cyclist.nutrients, nutrient.conversion(cyclists[i,1:13],  # F
                     cyclists[i,14:20], # For the meats
                     conversion = nutrients_conversion)) # the conversion table
}
Potential_volunteers <- nutrient.conversion(survey[,6:18],  # For the vegetables
                     survey[,19:25], # For the meats
                     conversion = nutrients_conversion)
```

...and we provide a table visualisation cheking if the potential volunteer respects the inclusion criteria that we choose.

```
mean_tot=apply(cyclist.nutrients, 2, mean)
sd_tot=apply(cyclist.nutrients, 2, sd)

### Moyenne + 2 x sd #######
max.c=mean_tot+2*sd_tot
min.c=mean_tot-2*sd_tot

inclusion=NULL
for(each in c(1:length(Potential_volunteers))){
  if (Potential_volunteers[each]>min.c[each] & Potential_volunteers[each]<max.c[each]){
    inclusion=c(inclusion, "Yes")
  }else{
    inclusion=c(inclusion, "No")
  }
}
df <- data.frame('individu'=as.numeric(Potential_volunteers),
           'cyclistes_moy'=as.numeric(mean_tot),
           "cyclistes_sd"=2*as.numeric(sd_tot), "Inclusion"=inclusion)
colnames(df)=c('individu','cyclistes_moy', "cyclistes_sd", "Inclusion")
rownames(df)=colnames(Potential_volunteers)
kableExtra::kable(df)
```

	individu	cyclistes_moy	cyclistes_sd	Inclusion
Fruits	56.80000	150.00385	207.339795	Yes
Fat	49.10861	65.58080	47.925256	Yes
Vegetables	148.80000	189.23077	341.769273	Yes
CHO	67.13062	65.48581	70.960348	Yes
Oil.Fish	4.50000	11.56154	20.468437	Yes

	individu	cyclistes_moy	cyclistes_sd	Inclusion
DQS	11.00000	11.69231	2.218801	Yes

3) Activity Profiling

For this last section, we explore the level of physical activity of the potential volunteers. In the questionnaire, we include 3 questions to evalute the time spent in minutes per week where the potential volunteer do a low, moderate or intensive physical exercice. We aimed to recruit 10 participants under the WHO recommendations for physical activity levels, and 10 just above these recommendations (i.e. 600 METS/min/week).

1 minute of a low and moderate physical exercice correspond to 4 METS (basal metabolism) while 1 minute of an intensive physical exercice correspond to 8 METS. It has been proposed in the Global Physical Activity Questionnaire, previously developed by Bull FC, Maslin TS, Armstrong T. **Global physical activity questionnaire (GPAQ): nine country reliability and validity study.** *J Phys Act Health.* 2009. doi: 10.1123/jpah.6.6.790.

```
print(paste("Potential volunteer exibit a METS-min/week = ",(survey$low_PA+survey$moderate
```

```
[1] "Potential volunteer exibit a METS-min/week =  600"
```