

Supplement to Explore the factors that may be associated with obesity

Xueyan Hu

2024-04-22

This shows some materials that could go into a supplementary file. Often you want/need references here too. You can use the same reference bib file for this and the main text (as done here) or have separate bib files.

For illustrative purposes, I'm doing the supplement as pdf. For this to work, you need a (La)TeX system installed. It's easy. Just follow [these steps](#).

1 Overview

All of the raw data file in csv, processed data in RDS file, descriptive analysis, exploratory analysis, and full data analysis codes and products are included in this project folder.

2 Code and file information

- assets: storing reference pdf files and citation style files
- data:
 - raw-data: codebook created by me and original data file downloaded from website
 - processed-data: cleaned data saved in RDS file
- products: manuscript in qmd file and its render output with same name in doc file
- R: including all codes needed
 - analysis-code: including codes for simple exploratory analysis and machine learning models fitting both in qmd file and R script file

- eda-code: including codes for simple exploratory analysis and machine learning models fitting both in qmd file and R script file
 - processing-code: including codes for data cleaning process both in qmd and R script file
- renv: storage for all of the packages used in this project in R
- results:
 - figures: all of the plots in png file created and saved from eda-code and analysis-code
 - tables: summary tables created and saved from eda-code and 2 linear model tables from analysis-code

3 Additional Method Details

I used linear regression model, LASSO model, random forest model for continuous variable BMI as the outcome; and I used discriminant analysis model and multinomial logistic regression for categorical variables obesity level as the outcome.

4 Additional results

Table 1 shows a summary of the dataset according to each variable.

Table 1: a dataset summary

Gender	Age	History	Water	AlcoholFAF		MTRANS	BMI	Obesity
Min. :1.000	Min. :14.00	Length:2111	Min. :1.000	1: 639	Min. : 1.0	Length:2111	Min. :13.00	Length:2111
1st Qu.:1.000	1st Qu.:19.95	Class :character	1st Qu.:1.585	2:1401	1st Qu.: 113.5	Class :character	1st Qu.:24.33	Class :character
Median :1.000	Median :22.78	Mode :character	Median :2.000	3: 70	Median : 590.0	Mode :character	Median :28.72	Mode :character
Mean :1.494	Mean :24.31	NA	Mean :2.008	4: 1	Mean : 540.0	NA	Mean :29.70	NA
3rd Qu.:2.000	3rd Qu.:26.00	NA	3rd Qu.:2.477	NA	3rd Qu.: 920.5	NA	3rd Qu.:36.02	NA
Max. :2.000	Max. :61.00	NA	Max. :3.000	NA	Max. :1190.0	NA	Max. :50.81	NA

Table 2 shows a more detailed summary of the dataset using skim function.

Table 2: a detailed dataset summary

[illegible]

skinski	type	evan	anis	logr	ch	ent	tra	char	acty	fct	of	bit	dsp	ed	nt	tpc	cic	ms	dim	pfi	mpe	pfi	pfi	pfi	hist
num	EAF	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	540.04726788033	59000	90000	0	0	19000	0	0	0	0	0
num	BMI	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	29.7800511230946325	87000	90000	0	0	15658	1	1	175		

Table 3: linear model fit 1

Table 4 shows linear model fitting result using Alcohol as predictors and BMI as outcome.

	Coefficients	Estimate	Std.Error	t.value	Pr..t..
(Intercept)	(Intercept)	27.0641363	0.3081965	87.8145560	0.0000000
Alcohol2	Alcohol2	3.9793185	0.3718979	10.7000307	0.0000000
Alcohol3	Alcohol3	-0.0829854	0.9808481	-0.0846058	0.9325829
Alcohol4	Alcohol4	-4.5727868	7.7968225	-0.5864936	0.5576067

Table 5: linear model fit 3

5

Table 5: linear model fit 3

	Coefficients	Estimate	Std.Error	t.value	Pr..t..
FAF	FAF	-	0.1747592	-	0.0000011
		0.8542115		4.8879333	
MTRANSBike	MTRANSBike	1.2102005	2.4625079	0.4914504	0.6231593
MTRANSMotorbike	MTRANSMotorbike	2.0417391	1.9741124	1.0342568	0.3011352
MTRANSPublic_Transportation	MTRANSPublic_Transportation	4.7167338	0.4369877	10.7937442	0.0000000
MTRANSWalking	MTRANSWalking	0.4936115	0.9517146	0.5186549	0.6040560

Table 6 shows linear model fitting result using Gender * Age as predictors and BMI as outcome.

Table 6: linear model fit 4

	Coefficients	Estimate	Std.Error	t.value	Pr..t..
(Intercept)	(Intercept)	10.6897420	2.1240354	5.032751	5e-07
Gender	Gender	7.5693791	1.3307676	5.687980	0e+00
Age	Age	0.7162389	0.0839784	8.528847	0e+00
Gender:Age	Gender:Age	-0.2685120	0.0529636	-5.069747	4e-07

5 Discussion

For reproducing this project, raw-data folder will be downloaded and stored in a new folder called data and then included in a larger folder with all the codes together. Processing-code qmd file under R folder will be opened and run so the raw data will go through data cleaning process and the cleaned data file will be saved in a folder called process-data under the data folder. Then eda qmd file can be opened and run using processed data file and the outputs will be saved in a newly created fold called results. Results should include tables and figures two subfolders.

Then, exploratory-analysis qmd file under R folder will be involved for the whole data analysis and result will be seen directly in the file.

After all of the qmd files under R folder are run out, the final manuscript will be generated. There are reference related files under assets folders so it will be made sure download and then manuscript qmd file under products fold will be find and run with a doc file saved beside the qmd file.

Supplementary material file is under supplement subfolder and a pdf file will be created and saved after running this qmd file.

6 References