Youngjun Woo
UMID: yjwoo
February 24th, 2023

Association between demographic, body measurement variables and blood pressure

We aim to analyze the association between demographic, body measurement variables and blood pressure using NHANES data from the United States. NHANES data consists of 86 variables related to people's demographic, body measures, and blood pressure. Among 86 variables, we selected only the following 13 variables and proceeded with the analysis. After excluding missing values, there are a total of 4380 observations.

Demographic	RIAGENDR	Gender
	RIDAGEYR	Age in years at screening
	DMDHHSIZ	Total number of people in the Household
	INDHHIN2	Annual household income
Body Measures	BMXWT	Weight (kg)
	BMXHT	Standing Height (cm)
	BMXBMI	Body Mass Index (kg/m**2)
	BMXLEG	Upper Leg Length (cm)
	BMXARML	Upper Arm Length (cm)
	BMXARMC	Arm Circumference (cm)
	BMXWAIST	Waist Circumference (cm)
	BMXHIP	Hip Circumference (cm)
Blood Pressure	BPXSY1	Systolic: Blood pres (1st rdg) mm Hg

We used SIR, dimension reduction regression analysis, in order to identify the most important variables that have an impact on blood pressure, while also accounting for potential collinearity between variables. Figure 1 shows that there might be a collinearity issue in our data. By reducing the number of variables in the model, we can better understand the relationships between the remaining variables and blood pressure, and potentially develop a more accurate prediction model.

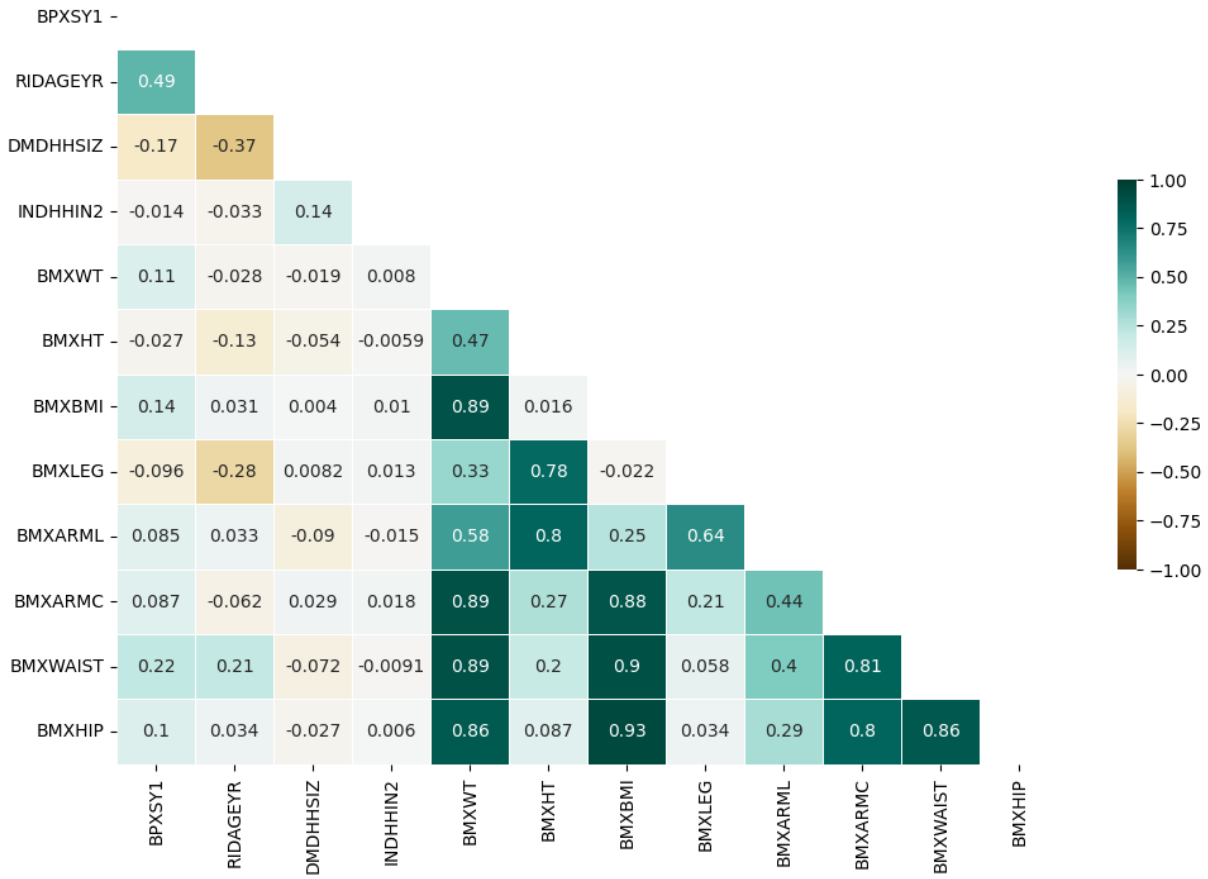

Figure 1: Correlation between variables

How many score variables to choose was determined by comparing models that used different numbers of score variables. Table 1 below shows the score test results between different models. As a result of the score test, after including 2 Score variables, there is no statistical difference between the models even when additional Score variables are included.

	Stat	df	p.value
0D vs >= 1D	1728.49948	110	0.0000000
1D vs >= 2D	347.59491	90	0.0000000
2D vs >= 3D	74.93448	72	0.3833475
3D vs >= 4D	32.23458	56	0.9954854

Table 1: Result of score test comparing nested models with Score variables

First, the meaning of Score1 and Score2 variables can be interpreted through the relationship between Score1 and Score2 and other variables. As shown in Figure 2, the variable

showing the clearest relationship with Score1 is Age. It may have information on various other variables, but the most distinct information represented by Score1 is age.

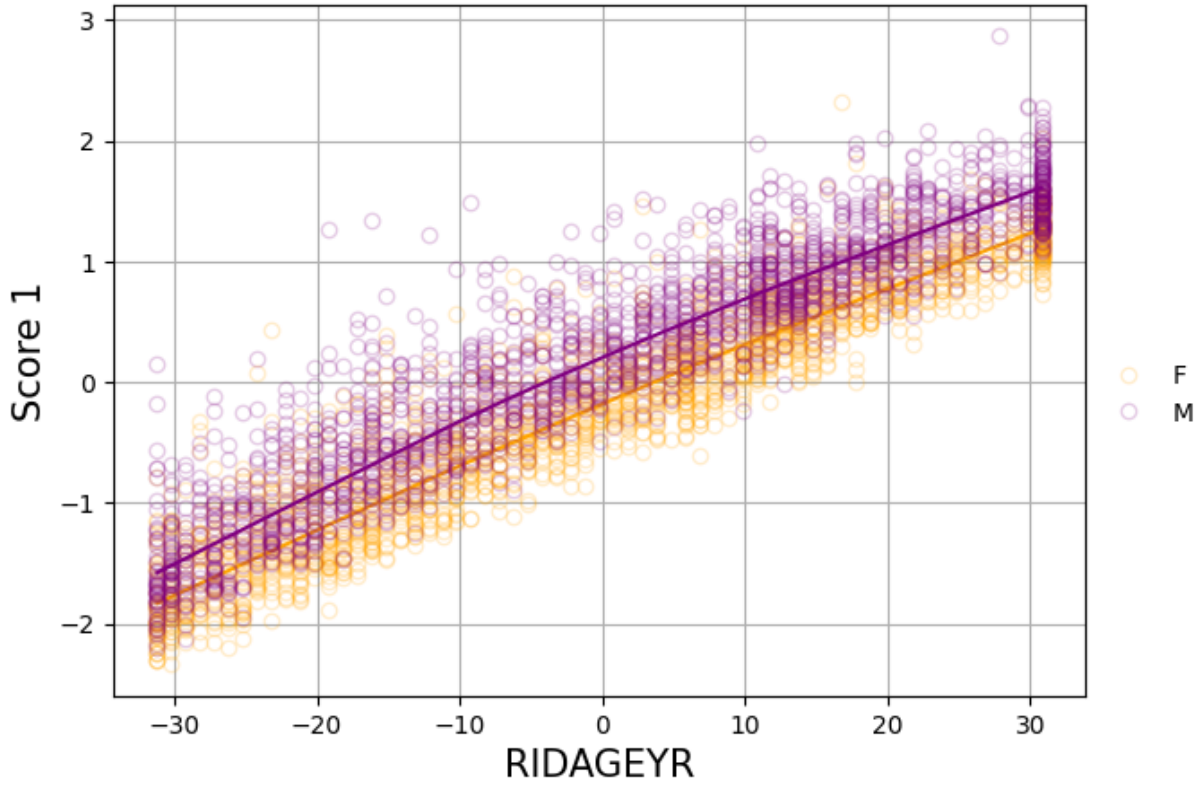

Figure 2: Relationship between Score1 and RIDAGEYR.

Figure 3 shows the variables showing the clearest relationship with Score2. The variables showing the most obvious relationship with Score2 are Weight (kg), Body Mass Index, Upper Arm Length, Arm Circumference, Waist Circumference, and Hip Circumference, all of which are related to body measurements. Since the values of these body measures usually tend to be higher as the degree of obesity increases, Score 2 can be interpreted as representing a person's degree of obesity.

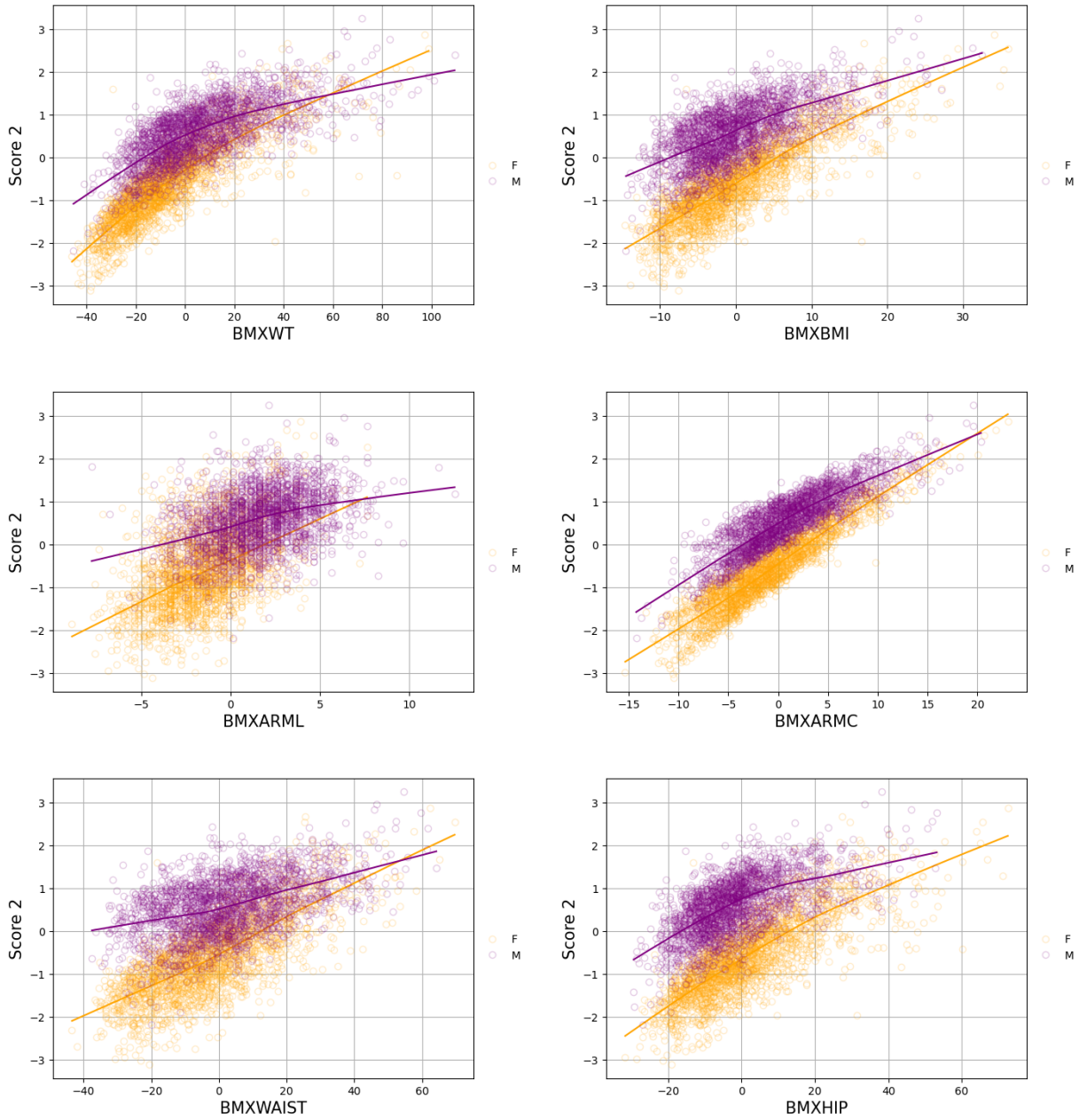

Figure 3: Relationships between Score2 and body measurement related variables.

In addition to age, gender, and body measurement variables, DMDHHSIZ and INDHHIN2, which can affect a person's lifestyle, are also included, but as shown in Figure 4, there is no clear relationship with Score1 and Score2. Therefore, these variables have a smaller effect on blood pressure than other variables.

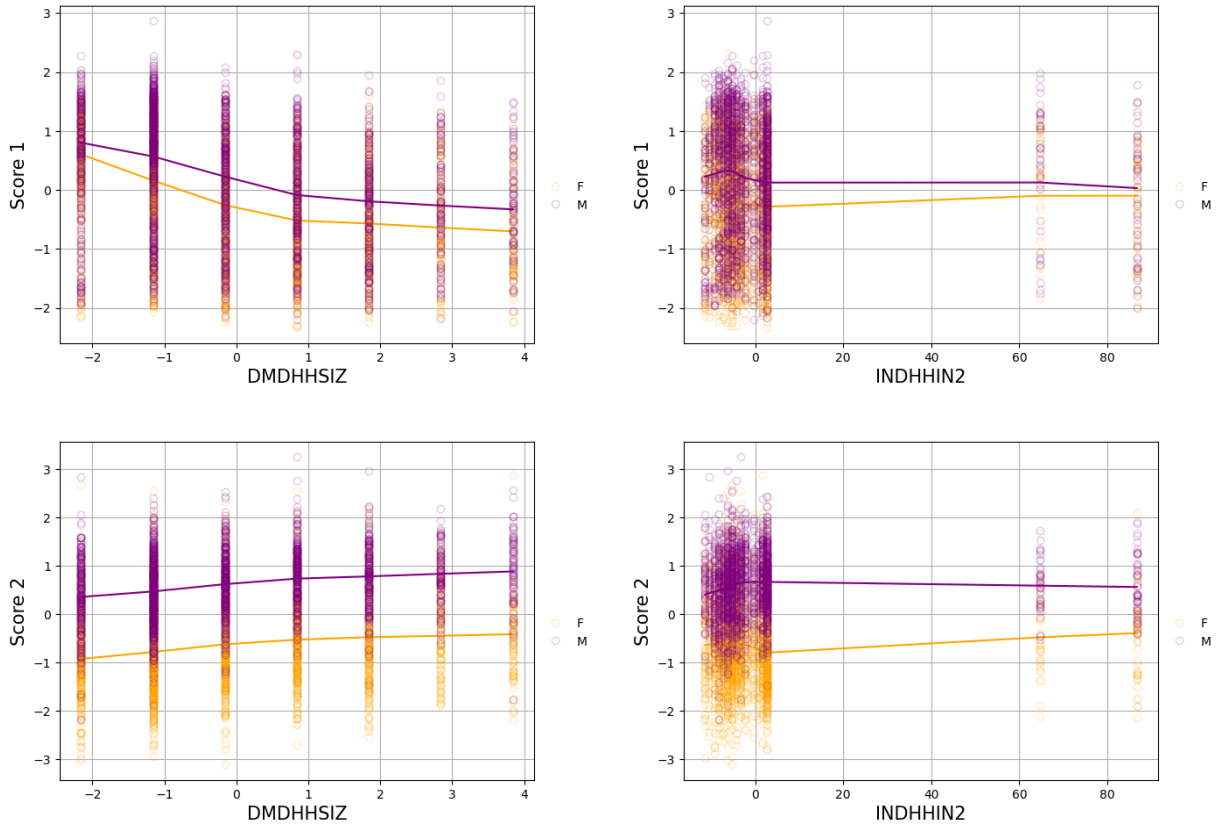

Figure 4: Relationship between Score1, score2 and DMDHHSIZ, INDHHIN2 variables.

Next, we checked the relationship between Score1, Score2 and blood pressure. Figure 5 shows the relationship between Score1 and blood pressure for each of the five groups divided according to Score2. Score1 and blood pressure show a positive relationship in all score2 groups. Since Score1 is a variable related to age, we can interpret that blood pressure increases as age increases. In the case of the blue and orange groups, which have low values of Score2, the increase rate in blood pressure with age is greater than other groups. In other words, it can be interpreted that the blood pressure of the group with high obesity increased faster than the other groups as age increased.

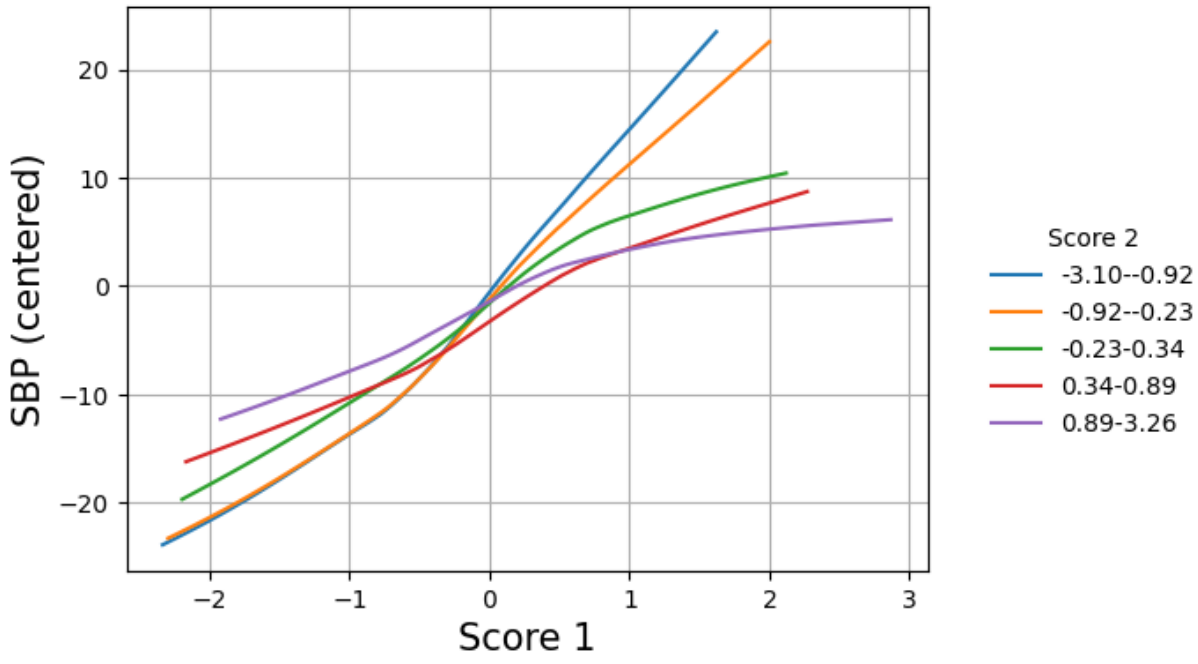

Figure 5: Relationship between Score1 and blood pressure

Figure 6 shows the relationship between Score2 and blood pressure for each of the five groups divided according to Score1. The relationship between Score2 and blood pressure changes depending on the Score1 group. In other words, the information on Score1 must also be checked together to accurately analyze the relationship between Score2 and blood pressure. In the blue and orange groups with low Score1, blood pressure increases as Score2 increases. That is, in the case of the younger group, blood pressure increases as obesity increases. In contrast, in the other three groups with high Score1, blood pressure decreases and then increases as Score2 increases. In other words, in the older group, as obesity increases, blood pressure decreases and then increases. It can be seen that the purple group, which is the oldest, has the largest decrease in blood pressure according to the increase in obesity.

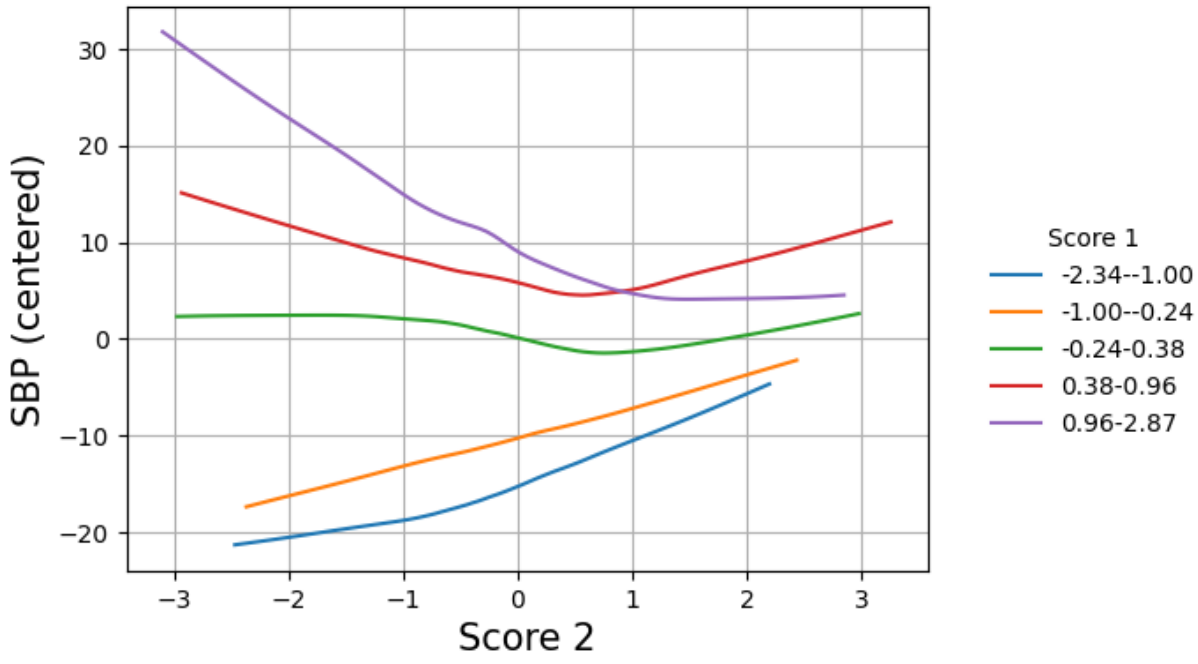

Figure 6: Relationship between Score 2 and blood pressure