

# MAT5317 Categorical Assignment 1

Teng Li(7373086)  
Zhize Lu(300075114)  
Chutong Zhang(300311325)

## Introduction

We were given the data set of The National Health and Nutrition Examination Survey (NHANES). The survey program has been conducted as a series of surveys designed to assess the health and nutritional status of adults and children in the United States since the 1960s, according to CDC (2023). It combines in-person face-to-face interviews and physical examinations of participants for data collection.

The survey data wasn't a simple random sample, however. According to CDC's National Health and Nutrition Examination Survey: Plan and Operations, 1999–2010 (G et al. 2013), the sampling strategy consists of several stages: 1. Selection of counties as primary sampling units (PSU). 2. selection of segments within PSUs that constitute blocks of households. 3. Selection of specific households within segments. 4. Selection of individuals within a household.

We aim to study the relationship between the weight variable and the other health related variables of the data.

## Method

We began our study by doing an exploratory analysis among the variables through various tables and charts. We then performed several hypothesis tests on some of the variables. Lastly we did a linear regression model fit to the response variable “weight” with other variables and confounders.

## Conclusion

## References

2023. [https://www.cdc.gov/nchs/nhanes/about\\_nhanes.htm](https://www.cdc.gov/nchs/nhanes/about_nhanes.htm).  
G, Zipf, Chiappa M, Porter KS, et al. 2013. “National Health and Nutrition Examination Survey: Plan and Operations, 1999–2010.” *National Center for Health Statistics* 1 (56).

Table 1: Data Variable Definition

Variables	Size	Type	Example	Number.Unique	Number.Missing	Comment
id	6482	integer	1, 2, 3	6482	0	Identification Code (1 - 6482)
gender	6482	factor	Male, Female	2	0	Gender (1: Male, 2: Female)
age	6482	integer	34, 16, 60	65	0	Age (Years)
marstat	6482	factor	Married, NA, Widowed	6	629	Marital Status (1: Married, 2: Widowed, 3: Divorced, 4: Separated, 5: Never Married, 6: Living Together)
samplewt	6482	numeric	80100.544, 13953.078, 20090.339	2499	0	Statistical Weight (4084.478 - 153810.3)
psu	6482	integer	1, 2	2	0	Pseudo-PSU (1, 2)
strata	6482	integer	9, 10, 1	15	0	Pseudo-Stratum (1 - 15)
tchol	6482	integer	135, 192, 202	251	395	Total Cholesterol (mg/dL)
hdl	6482	integer	50, 60, 45	112	395	HDL-Cholesterol (mg/dL)
sysbp	6482	integer	114, 112, 154	61	553	Systolic Blood Pressure (mm Hg)
dbp	6482	integer	88, 62, 70	40	594	Diastolic Blood Pressure (mm Hg)
wt	6482	numeric	87.400002, 72.300003, 116.8	957	37	Weight (kg)
ht	6482	numeric	164.7, 181.3, 166	527	37	Standing Height (cm)
bmi	6482	numeric	32.22, 22, 42.39	2276	37	Body mass Index (Kg/m <sup>2</sup> )
vigwrk	6482	factor	No, Yes, NA	2	1	Vigorous Work Activity (1: Yes, 2: No)
modwrk	6482	factor	No, Yes, NA	2	1	Moderate Work Activity (1: Yes, 2: No)
wlkbik	6482	factor	No, Yes, NA	2	1	Walk or Bicycle (1: Yes, 2: No)
vigrecrex	6482	factor	No, Yes, NA	2	1	Vigorous Recreational Activities (1: Yes, 2: No)
modrecrex	6482	factor	No, Yes, NA	2	2	Moderate Recreational Activities (1: Yes, 2: No)
sedmin	6482	integer	480, 240, 720	37	79	Minutes of Sedentary Activity per Week (0 - 840)
obese	6482	factor	No, Yes, NA	2	37	BMI>35 (1: No, 2: Yes)