Modeling BMI with the Multiple Linear Regression [232 words]

Di Zhen (1717719)

4/29/2020

Introduction

This report summarizes the results of regression modeling on BMI using data from HINTS conducted in 2014. [1]

Method

The data were imported and cleaned. A multiple linear regression of BMI on sedentary leisure time, respondent's sex, general health status and weight was performed. Then the diagnostic tests were conducted on the assumptions about form, errors, predictors and observations. The data analysis is conducted in R (Version 3.6.1).

Table 1: Model Statistics

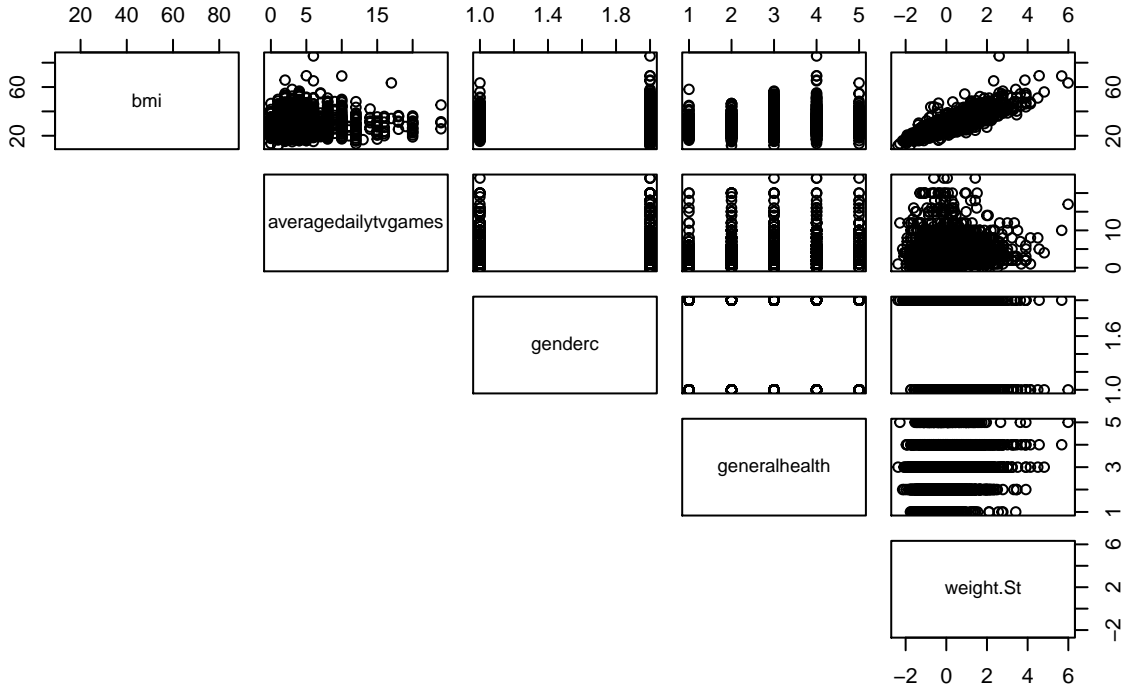|	coefficients	95%_CI	p_value
(Intercept)	24.71	24.39-25.038	<2.00e-16
averagedailytvgames	0.04	0.012-0.076	0.007
genderc2	4.25	4.049-4.456	<2.00e-16
generalhealth2	0.38	0.055-0.714	0.022
generalhealth3	0.84	0.51-1.171	6.64e-07
generalhealth4	1.65	1.252-2.048	6.29e-16
generalhealth5	1.53	0.954-2.101	1.88e-07
weight.St	6.12	6.016-6.22	<2.00e-16

Results

Figure 1. Scatterplot Matrix

Figure 1 shows the scatterplot matrix of all variables. The multiple linear regression equation is:

$$BMI = 24.71 + 0.04 \times averagedailytvgames + 4.25 \times genderc2 + 0.38 \times generalhealth2$$
$$+ 0.84 \times generalhealth3 + 1.65 \times generalhealth4 + 1.53 \times generalhealth5 + 6.12 \times weight$$

For every increased hour the respondents take for TV games, the BMI increased by 0.04 units, holding other variables constant. The effect of covariates is positive and significant (Table 1). The model fits the data well since overall p-value is less than 2.20×10^{-16} and the model accounted for 82.25% variability of the data.

The assumptions of linearity, normality and equal variance of residuals are slightly violated(Figure 2-4). Evidence about independence, nonrandom independent variables and correct measurement are available from the design of data. The VIF values indicated the absence of collienarity (Table 2). And the Cook's

distance indicated outliers in the independent or dependent variables (Figure 5).
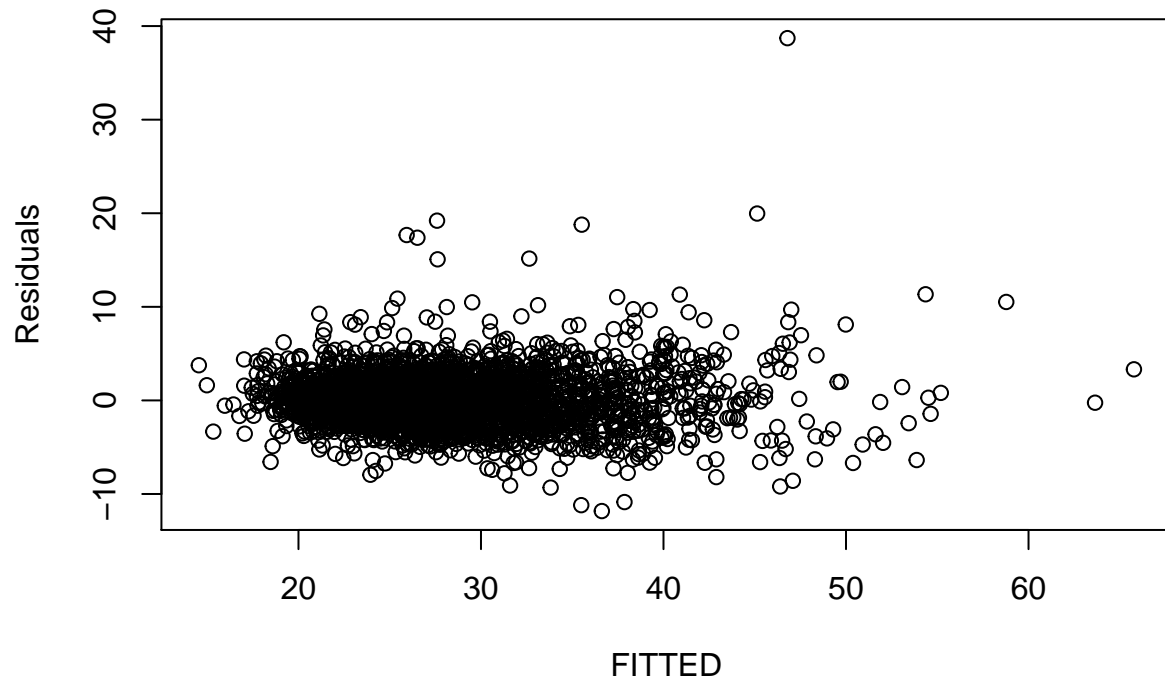
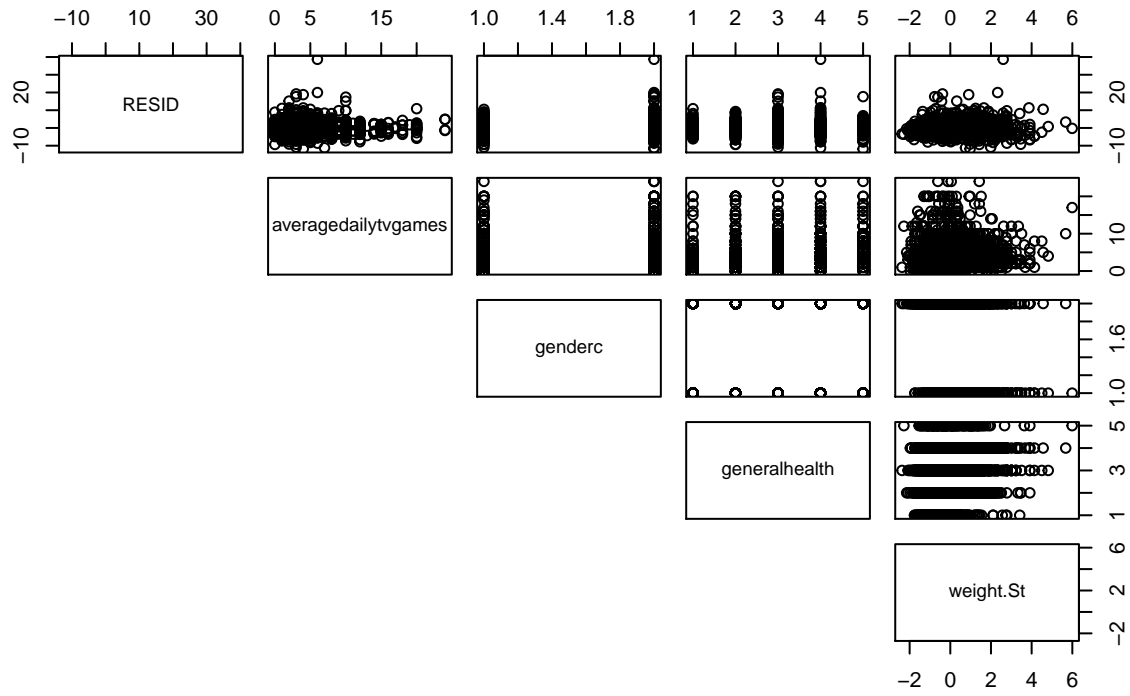## Figure 2. RvF Plot

Figure 3. Residuals against Each Predictors

Table 2: VIF Statistic

	GVIF	Df	GVIF^(1/(2*Df))
averagedailytvgames	1.07	1	1.03
genderc	1.14	1	1.07
generalhealth	1.13	4	1.01
weight.St	1.19	1	1.09

Figure 4. Normal Quantile–Quantile Plot of Ordianry Residuals
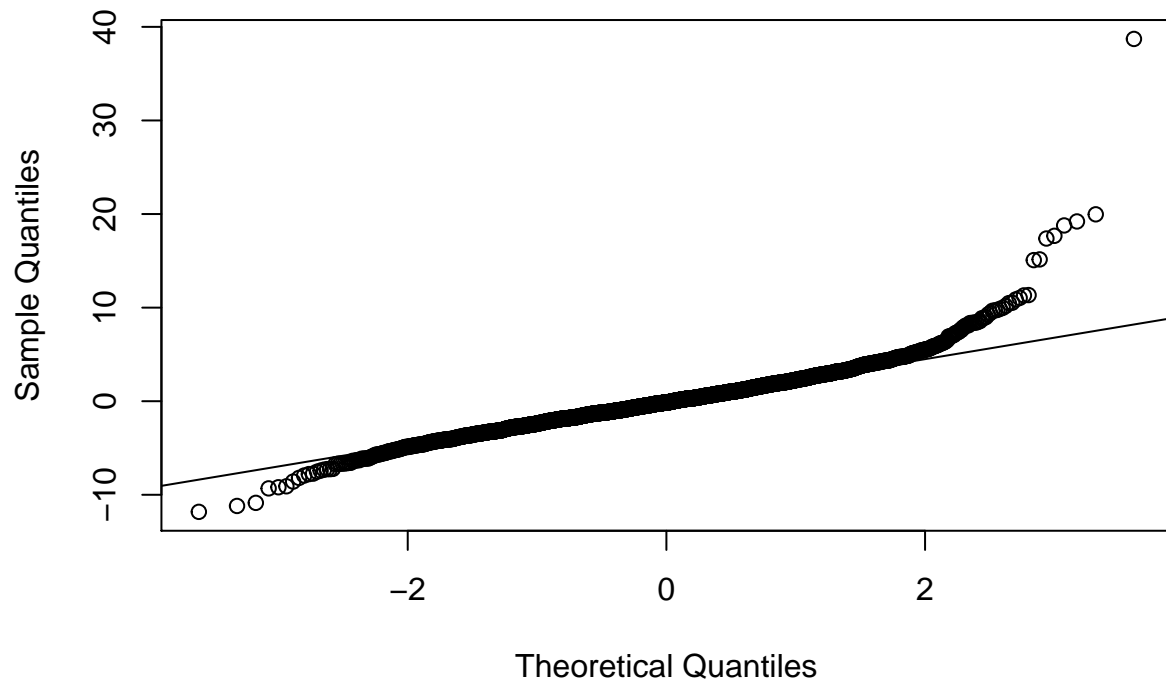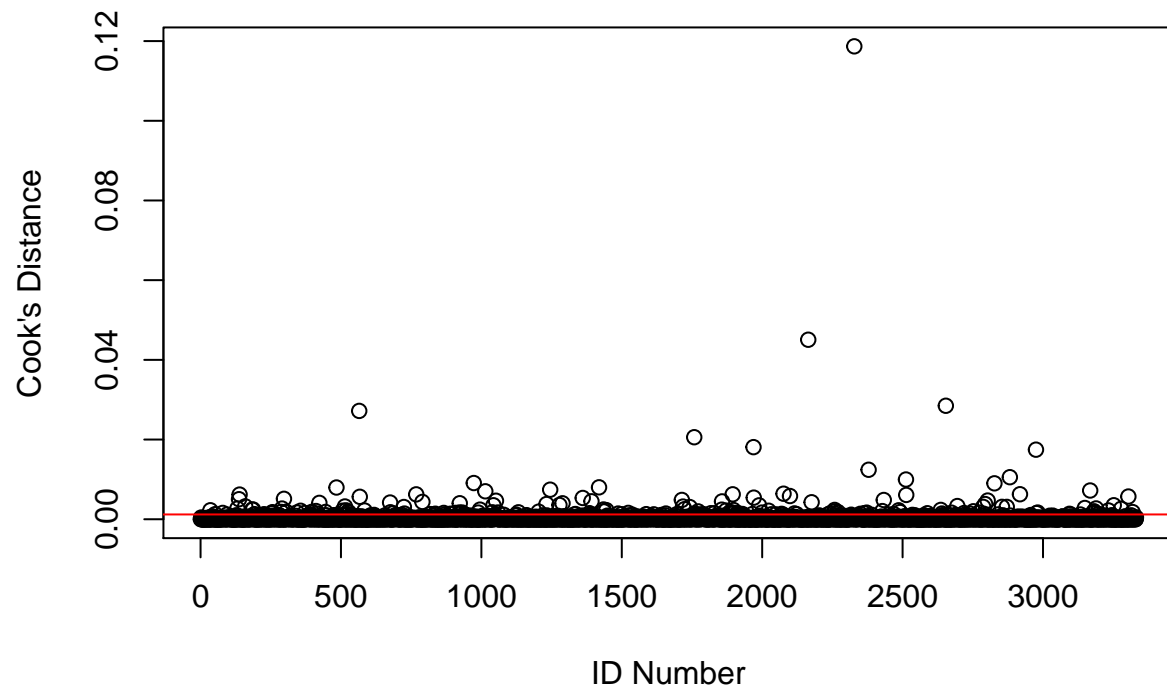
Figure 5. Cook's Distrance by Index Plot

Conclusion

In conclusion, the less time for TV games, the less BMI people will have.

References

[1] Nelson DE, Kreps GL, Hesse BW, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun. 2004;9(5):443-84. doi:10.1080/10810730490504233