# Section 6. Sociology

## UNHEALTHY DIET AMONG HIGH SCHOOL STUDENTS

*Yucheng Jie* [1]

[1] RDF International School, Shenzhen, China

**Abstract**

***Objective:*** This study aims to: 1) examine the predictors of an Unhealthy diet; 2) build a predictive model for an Unhealthy diet using an artificial neural network and compare its performance to logistic regression model.

***Methods:*** Youth Risk Behavior Surveillance System (YRBSS) 2017 data were used for this study. All the participants who were eligible were randomly assigned into 2 groups: training sample and testing sample. A logistic regression model was built using training sample. Receiver operating characteristic (ROC) was calculated.

***Results:*** About 7.17% of 7221 high school students had unhealthy diet, about 5.9% among the female and 8.8% among the male.

According to the logistic regression, Q1 (how old are you?), Q2 (What is your sex?), race, Q6 (How tall are you without your shoes on), Q31(How old were you when you first tried cigarette smoking, even one or two puffs?), Q60 (How old were you when you had sexual intercourse for the first time?), Q61 (During your life, with how many people have you had sexual intercourse?), Q64 (The last time you had sexual intercourse, did you or your partner use a condom?) were significantly associated with the Unhealthy diet in the high school students.

The area under curve was 0.66. The optional cutoff time is 0.49. The mis-classification error was 0.07. The sensitivity rate is about 0% and the specificity is 99.9%.

***Conclusions:*** In this study, we identified important of predictors of unhealth diet among high school students, for example age, gender and sexual behaviors.

**Keywords:** *Unhealthy diet, gender, behavior, predictors, high school students, model*

### Instruction

Healthy eating in childhood and adolescence is important for proper growth and development and to prevent various health conditions. The 2015–2020 Dietary Guidelines for Americansexternal icon recommend that peo-

ple aged 2 years or older follow a healthy eating pattern that includes the following (US Department of Health and Human Services and US Department of Agriculture, 2015):

1) A variety of fruits and vegetables.
2) Whole grains.

3) Fat-free and low-fat dairy products.

4) A variety of protein foods.

5) Oils.

Childhood obesity is a progressive phenomenon and is a known risk factor for the metabolic syndrome and type-II DM (Adair, L.S., Gordon-Larsen, P., Du S.F., Zhang, B., Popkin, B.M. 2013).

In this study, we aim to: 1) examine the predictors of the having unhealthy diet at high school; 2) build a predictive model for having unhealthy diet using logistic regression model.

## Data and Methods:

### Data:

Youth Risk Behavior Surveillance System (YRBSS) 2017 data were used for this study.

The YRBSS was developed in 1990 to monitor priority health risk behaviors that contribute markedly to the leading causes of death, disability, and social problems among youth and adults in the United States.

### Models:

We also used logistic regression models to calculate the predicted risk. Logistic regression is a part of a category of statistical models called generalized linear models, and it allows one to predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a combination of these. Typically, the dependent variable is dichotomous and the independent variables are either categorical or continuous.

The logistic regression model can be expressed with the formula:

$$\ln(P/1-p) = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \ldots + \beta_n \times X_n$$

### Model evaluation:

The discriminatory ability — the capacity of the model to separate cases from non-cases, with 1.0 and 0.5 meaning perfect and random discrimination, respectively– was determined using receiver operating characteristic (ROC) curve analysis. ROC curves are commonly used to summarize the diagnostic accuracy of risk models and to assess the improvements made to such models that are gained from adding other risk factors. Sensitivity, specificity, and accuracy will be also calculated and compared. For all these measures, there exist statistical tests to determine whether one model exceeds another in discrimination ability.

Optimal Cutoff for Binary Classification maximizes the accuracy.

Mis-Classification Error is the proportion of all events that were incorrectly classified, for a given probability cutoff score.

Sensitivity: probability that a test result will be positive when the disease is present (true positive rate.

Specificity: probability that a test result will be negative when the disease is not present (true negative rate, expressed as a percentage).

### Variables:

The outcome variable is percentage of students who had Unhealthy diet (Percentage of students who drank a can, bottle, or glass of soda or pop three or more times per day (such as Coke, Pepsi, or Sprite, not counting diet soda or diet pop, during the 7 days before the survey)) based on QNSODA3.

**Table 1.** *Variables used in this study*

Q1. How old are you?

    A. 12 years old or younger

    B. 13 years old

    C. 14 years old

    D. 15 years old

    E. 16 years old

    F. 17 years old

    G. 18 years old or older

Q2. What is your sex?

    A. Female

    B. Male

Q3. In what grade are you?
 A. 9th grade
 B. 10th grade
 C. 11th grade
 D. 12th grade
 E. Ungraded or other grade

Q4. Are you Hispanic or Latino?
 A. Yes
 B. No

Q5. What is your race? (Select one or more responses.)
 A. American Indian or Alaska Native
 B. Asian
 C. Black or African American
 D. Native Hawaiian or Other Pacific Islander
 E. White

Q6. How tall are you without your shoes on?

Q7. How much do you weigh without your shoes on?

Q31. Have you ever tried cigarette smoking, even one or two puffs?
 A. Yes
 B. No

Q41. During your life, on how many days have you had at least one drink of alcohol?
 A. 0 days
 B. 1 or 2 days
 C. 3 to 9 days
 D. 10 to 19 days
 E. 20 to 39 days
 F. 40 to 99 days
 G. 100 or more days

Q47. During your life, how many times have you used marijuana?
 A. 0 times
 B. 1 or 2 times
 C. 3 to 9 times
 D. 10 to 19 times
 E. 20 to 39 times
 F. 40 to 99 times
 G. 100 or more times

Q60. Have you ever had sexual intercourse?
 A. Yes
 B. No

Q68. Which of the following best describes you?
 A. Heterosexual (straight)
 B. Gay or lesbian
 C. Bisexual
 D. Not sure

## Results

About 7.17% of 7221 high school students had unhealthy diet, about 5.9% among the female and 8.8% among the male.

Basically, a corrgram is a graphical representation of the cells of a matrix of correlations. The idea is to display the pattern of correlations in terms of their signs and magnitudes using visual thinning and correlation-based variable ordering. Moreover, the cells of the matrix can be shaded or colored to show the correlation value. The positive correlations are shown in blue, while the negative correlations are shown in red; the darker the hue, the greater the magnitude of the correlation.

**Figure 1.** *Matrix of correlations between variables*
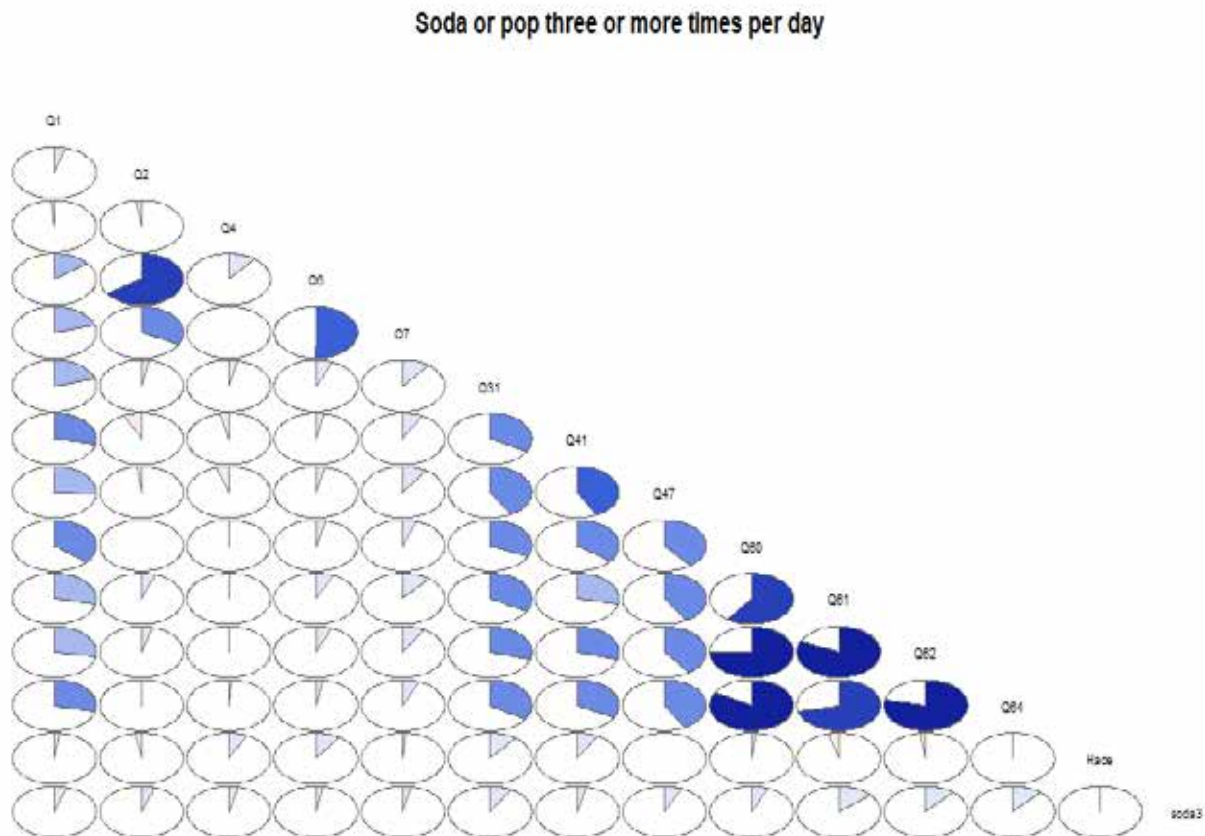


Soda or pop three or more times per day

According to the logistic regression, Q1 (how old are you?), Q2 (What is your sex?), race, Q6 (How tall are you without your shoes on), Q31(How old were you when you first tried cigarette smoking, even one or two puffs?), Q60 (How old were you when you had sexual intercourse for the first time?), Q61 (During your life, with how many people have you had sexual intercourse?), Q64 (The last time you had sexual intercourse, did you or your partner use a condom?) were significantly associated with the Unhealthy diet in the high school students.

**Table 2.** *Logistic Regression*

|  | Estimate | Std. Error | z value | Pr(>\|z\|) |  |
| --- | --- | --- | --- | --- | --- |
| (Intercept) | −1.37 | 1.07 | −1.28 | 0.20 |  |
| Q1 | 0.10 | 0.05 | 2.22 | 0.03 | * |
| factor(Q2)2 | 0.56 | 0.13 | 4.24 | 0.00 | *** |
| factor(Q4)2 | 0.52 | 0.15 | 3.36 | 0.00 | *** |
| factor(Race)2 | −1.56 | 0.46 | −3.36 | 0.00 | *** |
| factor(Race)3 | 0.14 | 0.28 | 0.51 | 0.61 |  |

| | Estimate | Std. Error | z value | Pr(>|z|) | |
|---|---|---|---|---|---|
| factor(Race)4 | −0.63 | 0.49 | −1.28 | 0.20 | |
| factor(Race)5 | −0.19 | 0.26 | −0.72 | 0.47 | |
| Q6 | −1.65 | 0.69 | −2.38 | 0.02 | * |
| Q7 | 0.00 | 0.00 | 0.00 | 1.00 | |
| factor(Q31)2 | 0.79 | 0.24 | 3.25 | 0.00 | ** |
| factor(Q31)3 | 0.76 | 0.25 | 3.01 | 0.00 | ** |
| factor(Q31)4 | 0.66 | 0.20 | 3.25 | 0.00 | ** |
| factor(Q31)5 | 0.50 | 0.17 | 2.98 | 0.00 | ** |
| factor(Q31)6 | 0.43 | 0.17 | 2.57 | 0.01 | * |
| factor(Q31)7 | 0.21 | 0.32 | 0.66 | 0.51 | |
| factor(Q41)2 | 0.36 | 0.22 | 1.65 | 0.10 | . |
| factor(Q41)3 | 0.24 | 0.23 | 1.06 | 0.29 | |
| factor(Q41)4 | 0.25 | 0.19 | 1.29 | 0.20 | |
| factor(Q41)5 | 0.01 | 0.16 | 0.07 | 0.95 | |
| factor(Q41)6 | 0.17 | 0.15 | 1.13 | 0.26 | |
| factor(Q41)7 | −0.14 | 0.28 | −0.52 | 0.60 | |
| factor(Q47)2 | 0.52 | 0.38 | 1.37 | 0.17 | |
| factor(Q47)3 | 0.37 | 0.33 | 1.12 | 0.26 | |
| factor(Q47)4 | 0.10 | 0.22 | 0.43 | 0.67 | |
| factor(Q47)5 | −0.07 | 0.16 | −0.41 | 0.68 | |
| factor(Q47)6 | 0.01 | 0.15 | 0.04 | 0.97 | |
| factor(Q47)7 | 0.16 | 0.27 | 0.60 | 0.55 | |
| factor(Q60)2 | 1.44 | 0.45 | 3.23 | 0.00 | ** |
| factor(Q60)3 | 1.17 | 0.50 | 2.36 | 0.02 | * |
| factor(Q60)4 | 1.05 | 0.47 | 2.21 | 0.03 | * |
| factor(Q60)5 | 0.93 | 0.46 | 2.00 | 0.05 | * |
| factor(Q60)6 | 0.80 | 0.46 | 1.74 | 0.08 | . |
| factor(Q60)7 | 0.59 | 0.48 | 1.24 | 0.22 | |
| factor(Q60)8 | 0.40 | 0.51 | 0.78 | 0.44 | |
| factor(Q61)2 | −0.14 | 0.25 | −0.56 | 0.58 | |
| factor(Q61)3 | −0.66 | 0.27 | −2.49 | 0.01 | * |
| factor(Q61)4 | −0.10 | 0.25 | −0.38 | 0.71 | |
| factor(Q61)5 | 0.05 | 0.27 | 0.17 | 0.86 | |
| factor(Q61)6 | −0.53 | 0.35 | −1.53 | 0.13 | |
| factor(Q62)2 | −0.08 | 0.48 | −0.17 | 0.86 | |
| factor(Q62)3 | −0.04 | 0.47 | −0.08 | 0.94 | |
| factor(Q62)4 | −0.13 | 0.48 | −0.26 | 0.79 | |
| factor(Q62)5 | 0.01 | 0.51 | 0.01 | 0.99 | |
| factor(Q62)6 | 0.06 | 0.61 | 0.10 | 0.92 | |
| factor(Q62)7 | −0.31 | 0.90 | −0.34 | 0.73 | |
| factor(Q64)2 | −0.26 | 0.13 | −1.99 | 0.05 | * |
| Signif. codes: | 0 '***' | 0.001 '**' | 0.01 '*' | 0.05 '.' | 0.1 ' ' |

**Figure.** *ROC in testing sample for Logistic Regression*
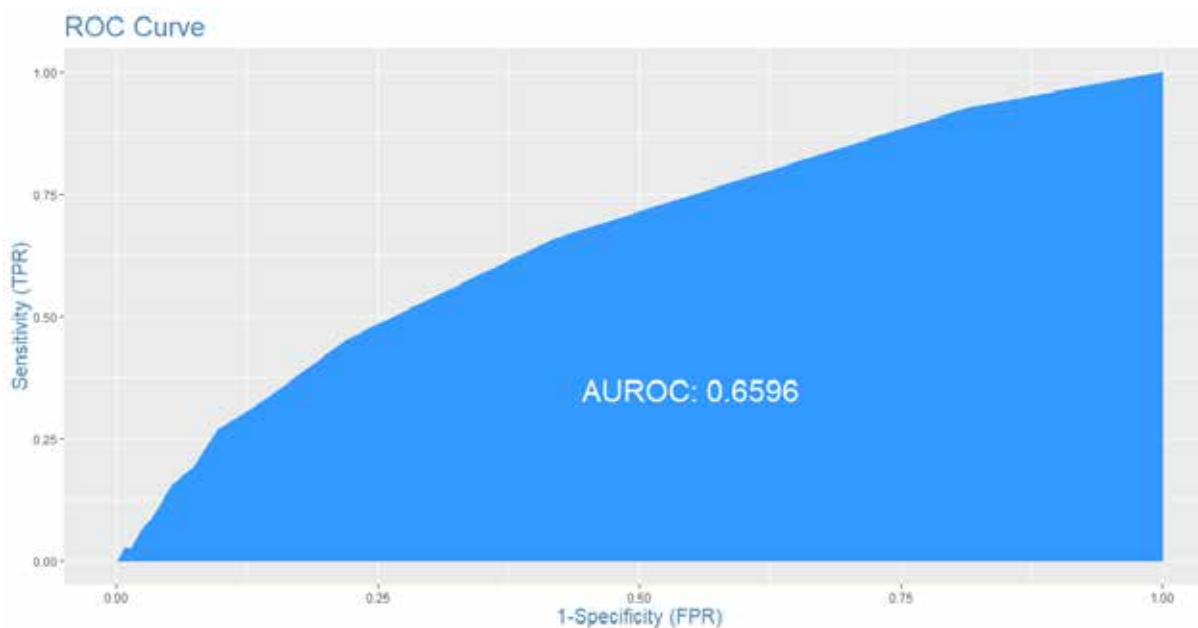


The area under curve was 0.66. The optional cutoff time is 0.49. The mis-classification error was 0.07. The sensitivity rate is about 0% and the specificity is 99.9%.

**Table**

| Cut-off | sensitivity | specificity |
|---------|-------------|-------------|
| 0.3 | 2.3% | 99.2% |
| 0.5 | 0% | 99.97% |
| 0.7 | 0% | 100% |

### Discussions

During 2007–2015, daily soda consumption among U.S. high school students decreased significantly from 33.8% to 20.4%. Although the significant downward trends in daily soda consumption suggest that interventions encouraging reduced consumption of soda are working, overall prevalence of daily soda consumption remains high (Trends in Beverage Consumption Among High School Students — United States, 2007–2015).

About 7.17% of 7221 high school students had unhealthy diet, about 5.9% among the female and 8.8% among the male. According to the logistic regression, Q1 (how old are you?), Q2 (What is your sex?), race, Q6 (How tall are you without your shoes on), Q31(How old were you when you first tried cigarette smoking, even one or two puffs?), Q60 (How old were you when you had sexual intercourse for the first time?), Q61 (During your life, with how many people have you had sexual intercourse?), Q64 (The last time you had sexual intercourse, did you or your partner use a condom?) were significantly associated with the Unhealthy diet in the high school students.

The area under curve was 0.66. The optional cutoff time is 0.49. The mis-classification error was 0.07. The sensitivity rate is about 0% and the specificity is 99.9%.

In this study, we identified important of predictors of unhealth diet among high school students, for example age, gender and sexual behaviors.

# References

US Department of Health and Human Services and US Department of Agriculture. 2015–2020. Dietary Guidelines for Americans. 8th Edition. December 2015. Available at URL: http://health.gov/dietaryguidelines/2015/guidelines

Adair, L. S., Gordon-Larsen, P., Du S. F., Zhang, B., Popkin, B. M. The emergence of cardiometabolic disease risk in Chinese children and adults: consequences of changes in diet, physical activity and obesity. Obes Rev.– 15(Suppl 1): 2013; 49–59.

Trends in Beverage Consumption Among High School Students – United States, 2007–2015. URL: https://www.cdc.gov/mmwr/volumes/66/wr/mm6604a5.htm

Peng, C. J., Lee, K. L., Ingersoll, G. M. An Introduction to Logistic Regression Analysis and Reporting. The Journal of Educational Research,– 96(1).– P. 3–14.

Tabachnick, B., and Fidell, L. Using Multivariate Statistics (4th Ed.). Needham Heights,– MA: Allyn & Bacon, 2001.

Stat Soft. Electronic Statistics Textbook, URL: http://www.statsoft.com/textbook/stathome.html

Stokes, M., Davis, C. S. Categorical Data Analysis Using the SAS System, SAS Institute Inc., 1995.

Contact: jasonjie06@gmail.com