

# Dietary Patterns and Mental Distress

by Yunhui Liu

April 2020

- Project Background
- Previewing the Data
- Data Cleaning and Processing
- Visualization
- Statistical Analysis
- Results
- Discussion

# Project Background

- A questionnaire mainly about dietary pattern and mental distress.
- 2636 participants, 27 questions.
- 21 normal questions, including the information of dietary pattern, gender, age, region, education and exercise.
- 6 questions about mental status, asking how often do people have negative mentality.

# Project Background

- "During the past month, about how often did you feel nervous/hopeless/depressed/worthless/...?"
- "All the time" ~ "None of the time"

# Previewing the data

1	1. What is your gender?	2. What is your age?	3. In which region you are living now?	4. What is the highest level of education you have completed?	5. What is your dietary pattern?
2	Male	18-29	North America/Central America	Professional Degree (MD, JD, PharmD, ...)	Mediterranean Diet
3	Male	18-29	Africa	2 or 4 Years of College Degree (AA, BA, BS)	Western Diet
4	Male	18-29	North America/Central America	Professional Degree (MD, JD, PharmD, ...)	Mediterranean Diet
5	Female	18-29	North America/Central America	2 or 4 Years of College Degree (AA, BA, BS)	Mediterranean Diet
6	Female	18-29	North America/Central America	2 or 4 Years of College Degree (AA, BA, BS)	Western Diet
7	Male	18-29	Middle East/North Africa	High School	Mediterranean Diet
8	Male	30-39	North America/Central America	Master's Degree	Mediterranean Diet
9	Male	18-29	North America/Central America	2 or 4 Years of College Degree (AA, BA, BS)	Mediterranean Diet
10	Male	30-39	Asia	Master's Degree	Asian Diet

Figure 1: Basic information data of first 9 participants

# Previewing the data

1	18. On an average week, how many times do you eat fast foods or pre-made food?
2	1 time
3	None
4	None
5	None
6	4 times
7	1 time
8	None
9	3 times
10	None

Figure 2: One dietary pattern of first 9 participants

# Previewing the data

1	21. During the past month, about how often did you feel NERVOUS?
2	Most of the time
3	Some of the time
4	Some of the time
5	Some of the time
6	A little of the time
7	Most of the time
8	Some of the time
9	A little of the time
10	Most of the time

Figure 3: One mental status of first 9 participants

# Data Cleaning and Processing

- N/A and invalid values.
- Levels or variables with too few observations.
- All frequency variables are converted into numeric.
- Create a new response called mental health score and it is defined as the average of 6 responses.



# Data Cleaning and Processing

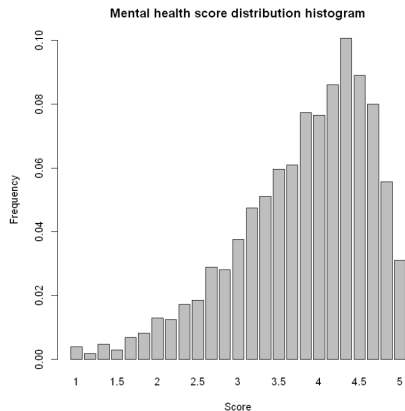


Figure 4: The distribution of mental health score

# Data Cleaning and Processing

1	4. What is the highest level of education you have completed?	COUNTA
2		0
3	2 or 4 Years of College Degree (AA, BA, BS)	926
4	Doctoral Degree	51
5	Graduate	4
6	High School	480
7	Less than High School	14
8	Master's Degree	336
9	Professional	4
10	Professional Degree (MD, JD, PharmD, ...)	75
11	<b>Grand Total</b>	<b>1890</b>

Figure 5: Example of a variable that needs to be collapsed

# Data Cleaning and Processing

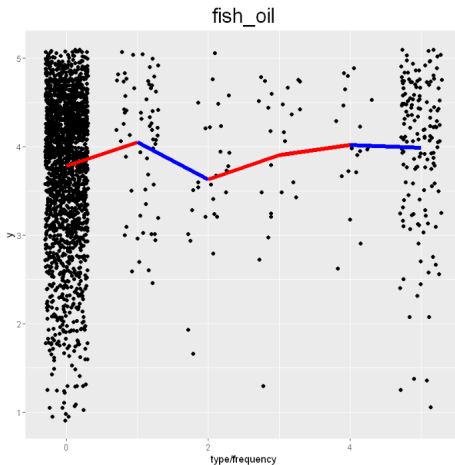


Figure 6: Example of a variable that needs to be removed

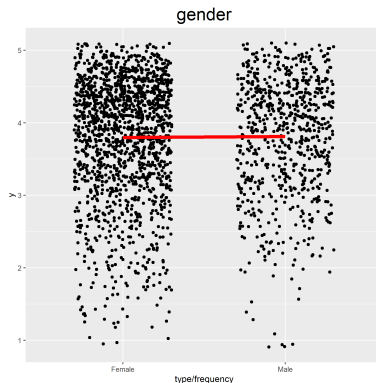


Figure 7: Mental Health Score between Females and Males

- The mean values are connected by lines, indicating increasing by red lines and decreasing by blue lines.
- If the variable is categorical, the color only implies the difference among levels.

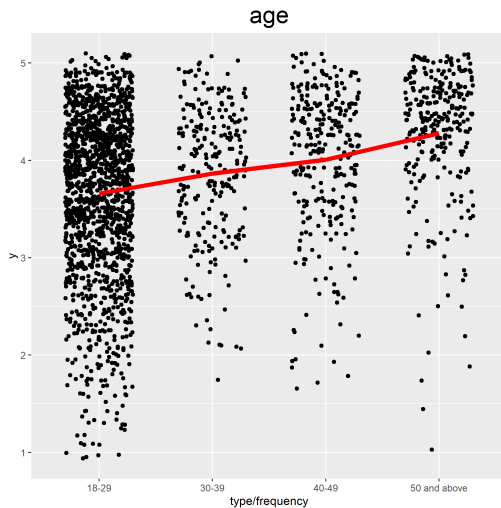


Figure 8: Mental Health Score among different ages

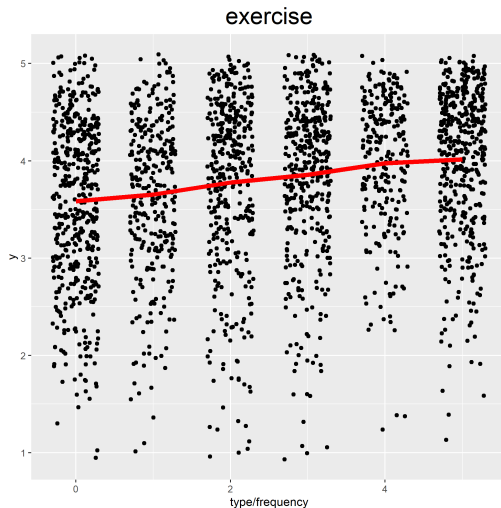


Figure 9: Mental Health Score and the frequency of exercising weekly

# Visualization

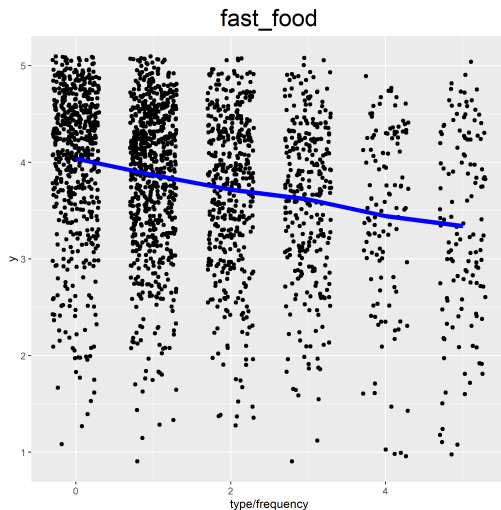


Figure 10: Mental Health Score and the frequency of eating fast food weekly

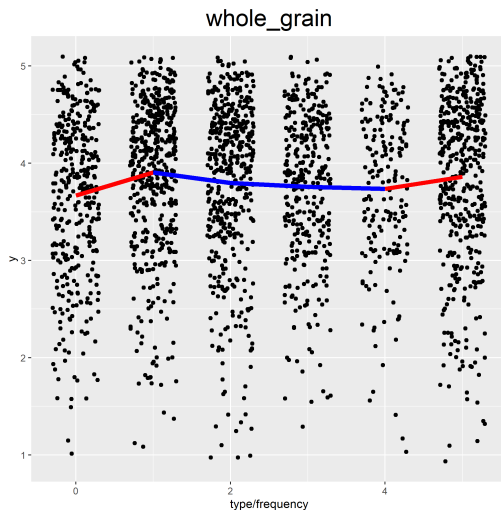


Figure 11: Mental Health Score and the frequency of eating whole grains weekly



- Split randomly into 50% training dataset and 50% test dataset.
- Select variables using the training dataset and test the significance of the selected variables using the test dataset.
- The analysis is conducted on 3 data subsets:
  - The whole data.
  - The data subsets for different age groups.
  - The data subsets for different gender groups.

## Linear Regression

A linear approach to model the relationship between a response and explanatory variables. The model has the following form:

$$y_i = \sum_{j=0}^n \beta_j x_{ij} + \epsilon_i, \quad i = 1 \dots n$$

$\beta_j$  is the coefficient of predictors,  $\epsilon_i$  is the error term,  $i$  is the index of an observation,  $j$  is the index of a predictor,  $p$  is the number of predictors and  $n$  is the number of observations. The  $\beta_j$  is estimated by ordinary least squares.

## Backward AIC

- $AIC = -2\log(L) + 2k$ .  $L$  is the maximum value of the likelihood function,  $k$  is the number of estimated parameters. Choose model with smaller  $AIC$ .
- "Backward" means starting with all variables, deleting the variable whose loss gives the least deterioration of the model fit, and repeating until no further variables can be deleted without a significant loss of fit.

## Cross-Validation LASSO

- Cross validation is an approach to test the effectiveness of models.
- LASSO is a regression analysis method that performs both variable selection and regularization.

## Segmented regression

- The variable is partitioned into intervals and a separate line segment is fit to each interval.

# Statistical Analysis Methods and Terminology

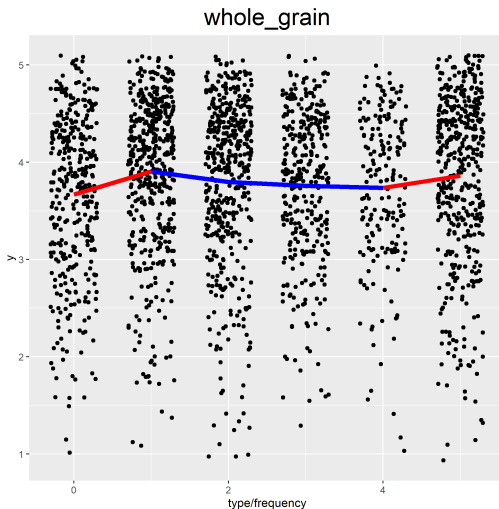


Figure 12: Mental Health Score and the time of eating whole grains weekly

# Statistical Analysis Methods and Terminology

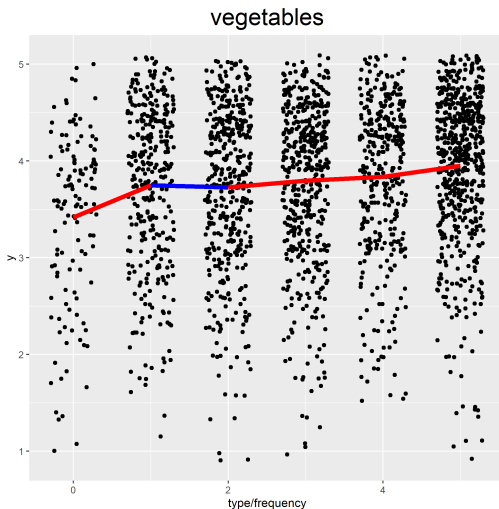


Figure 13: Mental Health Score and the time of eating vegetable weekly

## Significance

- A variable is significant indicating the rejection of the null hypothesis and a linear relationship between the response and the predictor.
- In the analysis, null hypothesis  $H_0$  is that the coefficient of a variable  $\beta_j = 0$ .

## P-value

- The  $p$ -value approach to hypothesis testing uses the calculated probability to determine whether there is evidence to reject the null hypothesis.
- In the analysis, if the variable's  $p$ -value  $\leq 0.05$ , the variable is considered to be significant and we reject  $H_0$ .



# Results of the Analysis on the Whole Data

- Linear regression, Backward AIC and Cross-validation LASSO have agreement on significance of the following variables or levels:
  - Positive: Male, Age over 50, exercise, breakfast, whole grain ( $\leq 1$ )
  - Negative: Coffee, fast food

# Results of the Analysis on the Whole Data

- Matured subjects and male subjects are likely to have higher mental health score.
- People with good mentality usually have stable exercise and breakfast. Eating whole grain could also be a sign of good mental status.
- More consumption of coffee and fast food relates with low mental health score.

# Results of the Analysis on the Data for Different Age Groups

- For **young** (18 ~ 29 years old) group:
  - Positive: **Male**, exercise, **breakfast**, **whole grain** ( $\leq 1$ ), **vegetables** ( $\geq 1$ )
  - Negative: Coffee, **beans**, fast food
- For **matured** group:
  - Positive: **Region North America**, **education**, exercise, **fruit** ( $< 4$ ), **fruit** ( $\geq 4$ )
  - Negative: Coffee, fast food

# Results of the Analysis on the Data for Different Age Groups

- For young group only:
  - Different genders are more likely to have different mental health status.
  - Breakfast and dietary patterns including eating whole grain, vegetables and beans could also be the indicators of the score for young people.

# Results of the Analysis on the Data for Different Age Groups

- For matured group only:
  - People who live in North America usually have higher mental health score than those who live in Africa.
  - People with different education have very different mental status.
  - Fruits.

# Results of the Analysis on the Data for Different Gender Groups

- For **female**:
  - Positive: Age, **region North America**, **region Europe**, exercise, **breakfast**, whole grain ( $\leq 1$ )
  - Negative: Coffee, fast food
- For **male**:
  - Positive: Age, **education**, exercise, whole grain ( $\leq 1$ ), **vegetables ( $\leq 1$ )**, **vegetables ( $> 1$ )**
  - Negative: Coffee, **beans**, fast food

# Results of the Analysis on the Data for Different Gender Groups

- For female group only:
  - Females in North America and Europe have higher mental health score than females in Africa.
  - How often females eating breakfast could also be the indicators of the mental health score.

# Results of the Analysis on the Data for Different Gender Groups

- For male group only:
  - Education is significant among males.
  - Males with different mental status are more likely to have difference in dietary patterns including vegetables and beans.



- Females and males have similar average mental health score in the plot. However, the analysis gives different result.

Name of Variables	Linear Regression		Backward AIC		Cross-Validation LASSO	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Gender_Female	Base line					
Gender_Male	0.1382	0.0089	0.1357	0.0109	0.1417	0.0079

Figure 14: Part of the result of analysis on the whole data

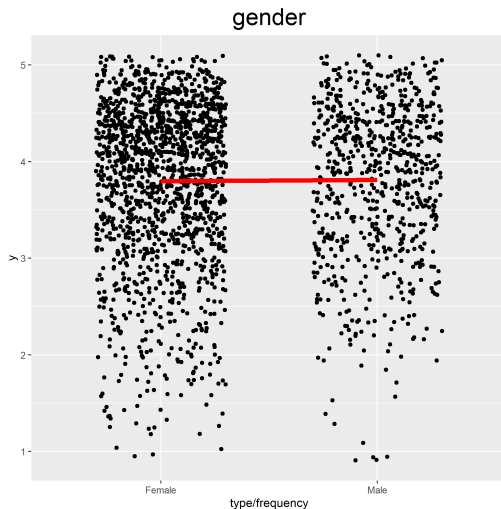


Figure 15: Mental Health Score between Females and Males

- Eating beans sometimes indicates poor mental status?

- Education is only significant when comparing young and matured or females and males, but not when using the whole dataset.

# The End