# Graduate Credit Assignment

Asma Naeem

## ABSTRACT

Individuals with higher levels of education have overall healthier and happier lifestyles. This study is aimed to determine whether university students with a major in health sciences have eating patterns that differ from those of students in other social science majors. The sample of 924 inland Spain students from the University of Castilla-La Mancha in the cities of Albacete, Ciudad Real, Cuenca, Talavera de la Reina, and Toledo were majoring in health science degrees, including nursing, medicine, pharmacy, or social science degrees like business and administration, economics, law, etc. Students recorded their socioeconomic and demographic characteristics and food intake patterns on a survey questionnaire. Participants were asked about their consumption of 141 foods divided into 12 groups. Researchers calculated each food group's mean daily/weekly servings based on the answers. We consolidated these 12 food groups into five per the Food Guide Pyramid and compared the respondent's daily serving intake in the five groups with the recommended servings. Health Science majors reported significantly more positive eating habits than Social Science majors. Both genders show almost equal deviation from the recommended total servings of five food groups. Students' family socioeconomic status, home location (urban, semi-urban, or rural), or whether the student cooks for themselves have no significant effect on their unhealthy eating scores.

## REPORT

### INTRODUCTION

Developing a balanced and healthy diet at a young age can improve a student's performance and set them up for a lifetime of healthy eating.

Eating nutrient-dense foods helps students stay attentive and productive, which is vital for concentrating on lectures and remembering information. A healthy diet can also help prevent illnesses, so students won't need to take sick days off from class.

Food impacts the human brain's function. For example, greasy foods take longer to digest, so eating them before bed can keep students awake when they need to sleep. This habit can negatively impact student success before a test or presentation.

Similarly, trans fats, found in fried food and processed baked goods, can cause decreased memory function. In addition, the foods that cause blood sugar levels to rise quickly and then crash, such as desserts and white bread, can cause a decrease in attention span and mood. Instead of eating simple carbs, it would be better to consume complex carbs like legumes and whole brains that steadily send sugar to the brain and help students focus on their schoolwork. Students perform better when they are well-nourished.

For this reason, we want to study barriers that impede university students from eating healthily and answer the following questions:

- Are the students in health science degrees more conscious of a balanced diet?
- Is there a difference in food intake patterns of male and female students?
- Do the students who cook for themselves eat healthier?
- Does the socioeconomic status of the student's family impact their food choices?
- Does the location of the student's family home make a difference in their food habits?

We used the *exploratory analysis approach* to find relationships in a dataset comprising observations of food consumption from the inland Spain University of Castilla-La Mancha students. We compared the student's daily food intake in the five food pyramid groups with the recommended servings.

## METHODS

A survey was undertaken with a sample of 924 inland Spain university students from the University of Castilla-La Mancha in the cities of Albacete, Ciudad Real, Cuenca, Talavera de la Reina, and Toledo. The survey comprised food intake questions alongside lifestyle and sociodemographic questions. Respondents were asked about their consumption of 141 foods divided into 12 groups. The mean daily/weekly servings for each food group were generated from the answers. We consolidated these 12 food groups' data into five per the Food Guide Pyramid to compare respondents' daily food intake with recommended servings. Family socioeconomic status was based on the information given about parental occupation. Health science degrees included nursing, medicine, pharmacy, etc., and social science degrees included business and administration, economics, law, etc.

The following new variables are constructed from ones existing in the data:

- potatoDaily – Average Daily consumption of potatoes in servings = potato/7
- redmeatDaily – Average Daily consumption of red meat in servings = redmeat/7,
- whitemeatDaily - Average Daily consumption of white meat in servings = whitemeat/7,
- fishDaily - Average Daily consumption of fish in servings =fish/7,
- eggDaily - Average Daily consumption of eggs in servings = egg/7,
- legumesDaily - Average Daily consumption of legumes in servings = legumes/7,
- protein - Daily consumption of proteins in servings = Sum of olivesnuts, redmeatDaily, whitemeatDaily, fishDaily, eggDaily, legumesDaily servings.
- starchyFoods - Daily consumption of starchy Foods in servings = Sum of breadpasta, and potatoDaily servings.
- veggie - Daily consumption of vegetables in servings = Sum of herbsspices, veget servings.
- unHealthyStarchyFoodsServings – Difference in the consumption of starchy foods from recommended daily servings (6-11). Consumption of 6 – 11 servings will be counted as 0 difference from the recommendation.
- unHealthyVeggieServings - Difference in consumption of vegetables from recommended daily servings (3-5). Consumption of 3 - 5 servings will be counted as 0 difference from the recommendation.
- unHealthyFruitServings- Difference in consumption of fruits from recommended daily servings (2-4). Consumption of 2 – 4 servings will be counted as 0 difference from the recommendation.
- unHealthyProteinServings - Difference in consumption of proteins from recommended daily servings (2-4). Consumption of 2 – 4 servings will be counted as 0 difference from the recommendation.
- unHealthyDairyServings - Difference in consumption of dairy foods from recommended daily

servings (2-3). Consumption of 2 – 3 servings will be counted as 0 difference from the recommendation.

- unHealthyEatingScore(UHES) – This is the total deviation of five food groups' daily consumption from the sum of the recommended servings. We added absolute values of unHealthyStarchyFoodsServings, unHealthyVeggieServings, unHealthyFruitServings, unHealthyProteinServings, and unHealthyDairyServings to compute unHealthyEatingScore (UHES). We used absolute values because we are only interested in measuring the distance from recommended servings and are not factoring in under or over-consumption in this calculation.

*Descriptive statistics* are performed to summarize data using indexes such as mean difference in food consumption from the sum of recommended daily servings of five food groups.

*Inferential statistics* are conducted to draw conclusions from data using multiple linear regression. We estimated the linear relationship between a response variable unhealthy Eating Score (UHES), and five predictors:

- Whether the student is pursuing a health science or social science degree.
- Student's gender recorded as either male or female.
- Student's family socioeconomic status, either low, middle, or high.
- Student's location of the family home, either urban, semi-urban, or rural.
- Whether the student cooks for themselves, either yes or no

*We tested the following statistical hypotheses:* -

- There is no difference in the mean unhealthy Eating Score (UHES) between students pursuing a health science or social science degree.
- There is no difference in the mean UHES between male and female students.
- There is no difference in the mean UHES between students whose family socioeconomic status is either high, middle, or low.
- There is no difference in the mean UHES between students whose family home location is either urban, semi-urban, or rural.
- There is no difference in the mean UHES between students, whether they cook for themselves or not.

We didn't use Age as a predictor because all respondents are University students, mostly in the same cohort. The median age of students in the data set is 19 years.

## RESULTS

### Descriptive Results

- Students with a Health Science major have a lower mean unhealthy Eating Score (UHES) of 7.5 as compared to Social Science major students' mean of 8.6.
- Females have a mean UHES of 8.2, and male students have a mean of 8.6.
- We don't see much difference in Mean UHES scores for students with High, Middle, and Low family socioeconomic status, 8.0, 8.6, and 8.2, respectively.
- There is little difference in Mean UHES scores for students with family home locations, either Urban, Semi-urban, or Rural, 8.3, 8.7, and 8.1, respectively. Note that our sample data comprises more students from urban areas than the combined total of students from both semi-urban and rural areas.
- Similarly, no significant difference is observed in mean UHES scores for students who cook for themselves and otherwise, 8.2 and 8.7, respectively.

*Demographics of the study population: -*

- Degree: Heath Science (142), Social Science (451)
- Gender: Male (249), Female (344)
- Family socioeconomic status: High (105), Middle (238), Low (250)
- Family home location: Urban (339), Semi-urban (162), Rural (92)
- Whether the student cooks for themselves: Yes (420), No (173)

### Analytic Results

- The students pursuing a health science degree have a 1.0 point lower mean unhealthy Eating Score (UHES) than students pursuing a social science degree when gender, family's socioeconomic status, family home location, and whether the student cooks for themselves values are held constant. The p-value is 0.005. There is a 0.5% chance of observing such a difference if there is no difference between the mean UHES in the health and social science students population. The p-value is less than or equal to 0.05; statistically significant evidence at the alpha = 0.05 level exists to reject the null hypotheses. With 95% confidence, the average UHES when gender, family's socioeconomic status, family home location, and whether the student cooks for themselves value are held constant is captured by the interval (-1.79, -0.33).

- The average UHES for males is 0.18 points higher than the females when degree, family's socioeconomic status, family home location, and whether the student cooks for themselves values are held constant. The p-value is 0.58. There is a 58% chance of observing such a difference if there is no difference between the mean UHES in the population of male and female students. The p-value is greater than 0.05; we do not have sufficient evidence to reject the null hypotheses at the alpha = 0.05 level. With 95% confidence, the interval (-0.45, 0.80) captures the average UHES when degree, family's socioeconomic status, family home location, and whether the student cooks for themselves values are held constant.

- The students who do not cook for themselves have 0.6 points higher mean UHES than those who cook for themselves when degree, gender, family's socioeconomic status, and family home location are held constant. The p-value is 0.086. There is an 8.6% chance of observing such a difference if there is no difference between the UHES in the population of students who cook for themselves and those who don't. The p-value is greater than 0.05; we do not have sufficient evidence to reject the null hypotheses at the alpha = 0.05 level. With 95% confidence, the

average UHES when degree, gender, family's socioeconomic status, and family home location are held constant is captured by the interval (-0.09, 1.31).

- The average UHES for students with high family socioeconomic status is 0.13 points lower than those with low family socioeconomic status when degree, gender, family home location, and whether the student cooks for themselves values are held constant. The p-value is 0.76. There is a 76% chance of observing such a difference if there is no difference between the mean UHES in the population of students with family socioeconomic status high and low. The p-value is greater than 0.05; we do not have sufficient evidence to reject the null hypotheses at the alpha = 0.05 level. With 95% confidence, the average UHES when degree, gender, family home location, and whether the student cooks for themselves value are held constant is captured by the interval (-1.01, 0.74).
- The students whose family home location is rural have a 0.45 points lower mean UHES than those whose family home location is urban when degree, gender, family's socioeconomic status, and whether the student cooks for themselves values are held constant. The p-value is 0.33. There is a 33% chance of observing such a difference if there is no difference between the mean UHES in the population of male and female students. The p-value is greater than 0.05; we do not have sufficient evidence to reject the null hypotheses at the alpha = 0.05 level. With 95% confidence, the average UHES when degree, gender, family's socioeconomic status, and whether the student cooks for themselves values are held constant is captured by the interval (-1.36, 0.46).

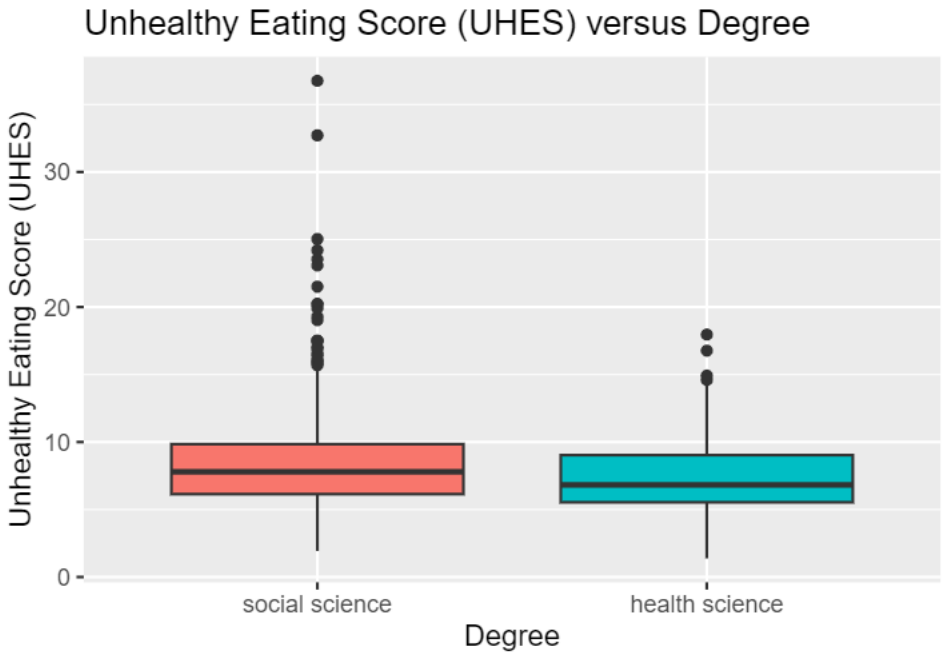
*Assumptions for multiple regression:*

- Linearity: It is not necessary to assess linearity with respect to any of our predictor variables (degree, gender, cook, hisei, or location) since all our predictors are categorical variables. A best-fit line drawn through two points (that is, the mean of the two groups) is necessarily linear.
- Constant variability: The residuals (with relatively few exceptions) exist in a band of constant width across values of predicted Unhealthy Eating Score(UHES). Thus, the constant variance assumption is reasonable for this model.
- Independent observations: The data are food groups consumption from a sample of participants randomly recruited from the inland Spain student population of the University of Castilla-La Mancha in the cities of Albacete, Ciudad Real, Cuenca, Talavera de la Reina, and Toledo, so it is reasonable to assume that the observations are independent within and across groups. Knowing how one student's food group consumption is not informative of another student's food group consumption.
- Normality of residuals: The residuals are approximately normally distributed in the center of the distribution, but the large positive residuals show a departure from normality. This is likely not a serious violation of model assumptions since the sample size is relatively large.

The model explains 2.5% of the variability in observed UHES. Only 2.5% of the variance in UHES can be explained by five variables: degree, gender, family socioeconomic status, family home location, and whether the student cooks for themselves.

Reference: (See the summary of the Linear regression model in the appendix.)

## Main Figures



## DISCUSSION

Food intake varied amongst university students. A substantial proportion of students follow health-promoting diets, which have good nutrient profiles obviating a need for dietary intervention. However, some students consume poor diets, which may have long-term health effects.

Students pursuing a health science degree seem to consume a more balanced diet than social science students. They have a 1.0-point lower mean unhealthy Eating Score (UHES) than students with social science majors, which may not be practically significant.

There is no significant difference in the unhealthy eating score between male and female students. Students' family home location (urban, semi-urban, or rural) or whether the student cooks for themselves have no substantial effect on the daily consumption difference from recommended servings of five food groups in the population of university students.

Additionally, the student's family socioeconomic status (high, middle, low) also doesn't seem to impact their unhealthy eating score significantly. In practice, families with lower gross income find it challenging to afford a balanced diet.

University policy to improve students' diets should integrate initiatives to promote student engagement in cooking and food preparation and increase the availability of affordable, healthier food items targeting young adults. Furthermore, providing tailored seminars and courses on nutritional benefits, cooking skills, time management, and budget organization for students with various backgrounds and characteristics can be helpful. Social networks could be leveraged to promote knowledge about healthy eating by creating social support groups among students who want to adopt a healthier lifestyle.

*Limitations:*

- As there is a set list of foods/food groups named in the survey questionnaire, not all foods may be captured if not presented on the list.
- As the survey questionnaires usually correspond to foods consumed in the previous year, this can be difficult for the participants, as they may forget what or how often they have consumed a certain food.
- Survey questionnaires usually ask for portion size (small, medium, or large), but it may be difficult for the participant to determine the correct portion size.

# APPENDIX

## METHODS

```
#Data is loaded in the Main Figures section.

#fit the model
diets.mod = lm(UHES ~ degree + gender + cook + hisei + location, data = diets_new)
```

## RESULTS

### Descriptive Results (Data Exploration)

```
#summarize age
diets_new %>%
summarise(n = n(), m=mean(age), d=median(age))
```

```
##     n        m  d
## 1 593 20.21079 19
```

```
#summarize by degree
diets_new  %>%
group_by(degree) %>%
summarise(mean = mean(UHES), n = n(), median = median(UHES))
```

```
## # A tibble: 2 x 4
##   degree          mean     n median
##   <fct>          <dbl> <int>  <dbl>
## 1 social science  8.60   451   7.79
## 2 health science  7.49   142   6.82
```

```
#summarize by gender
diets_new  %>%
group_by(gender) %>%
summarise(mean = mean(UHES), n = n(), median = median(UHES))
```

```
## # A tibble: 2 x 4
##   gender  mean     n median
##   <fct>  <dbl> <int>  <dbl>
## 1 female  8.18   344   7.40
## 2 male    8.55   249   7.79
```

```
#summarize by Family socioeconomic status
diets_new  %>%
group_by(hisei) %>%
summarise(mean = mean(UHES), n = n(), median = median(UHES))
```

```
## # A tibble: 3 x 4
##   hisei   mean     n median
##   <fct>  <dbl> <int>  <dbl>
## 1 low     8.24   250   7.54
```

```
## 2 middle  8.56   238   7.86
## 3 high    8.01   105   7.39
```

```
#summarize by location
diets_new  %>%
group_by(location) %>%
summarise(mean = mean(UHES), n = n(), median = median(UHES))
```

```
## # A tibble: 3 x 4
##   location   mean     n median
##   <fct>     <dbl> <int>  <dbl>
## 1 urban      8.32   339   7.54
## 2 semiurban  8.47   162   7.92
## 3 rural      8.13    92   7.27
```

```
#summarize by cook
diets_new  %>%
group_by(cook) %>%
summarise(mean = mean(UHES), n = n(), median = median(UHES))
```

```
## # A tibble: 2 x 4
##   cook   mean     n median
##   <fct> <dbl> <int>  <dbl>
## 1 yes    8.17   420   7.54
## 2 no     8.72   173   7.88
```

### Analytic Results (Data Analysis)

```
#summarize
summary(diets.mod)
```

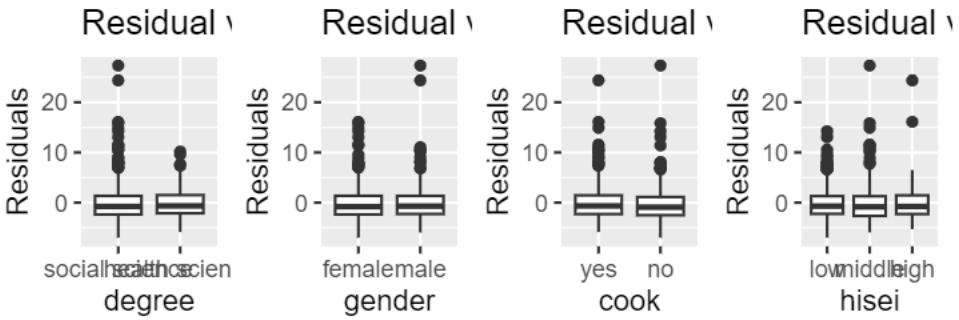
```
##
## Call:
## lm(formula = UHES ~ degree + gender + cook + hisei + location,
##     data = diets_new)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9806 -2.3007 -0.7086  1.3502 27.3140
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            8.29743    0.35117  23.628  < 2e-16 ***
## degreehealth science -1.05598    0.37179  -2.840  0.00466 **
## gendermale             0.17614    0.32030   0.550  0.58258
## cookno                 0.61461    0.35763   1.719  0.08622 .
## hiseimiddle            0.35778    0.34760   1.029  0.30378
## hiseihigh             -0.13334    0.44578  -0.299  0.76496
## locationsemiurban     -0.06105    0.37110  -0.165  0.86938
## locationrural         -0.44734    0.46301  -0.966  0.33437
```

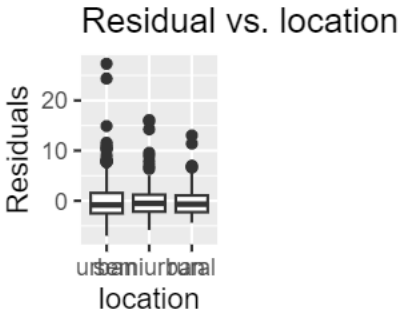
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.754 on 585 degrees of freedom
## Multiple R-squared:  0.02456,    Adjusted R-squared:  0.01289
## F-statistic: 2.105 on 7 and 585 DF,  p-value: 0.04132
```

```
#confidence intervals
confint(diets.mod)
```

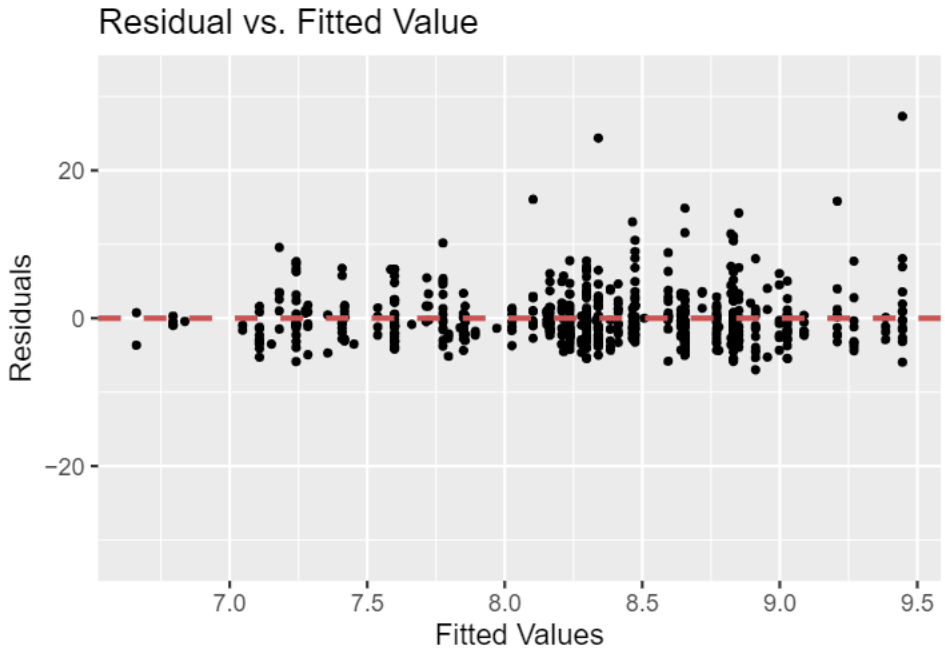
```
##                              2.5 %     97.5 %
## (Intercept)            7.60772107  8.9871315
## degreehealth science -1.78617357 -0.3257766
## gendermale           -0.45294360  0.8052251
## cookno               -0.08778857  1.3170086
## hiseimiddle          -0.32492659  1.0404806
## hiseihigh            -1.00885841  0.7421864
## locationsemiurban    -0.78990679  0.6678016
## locationrural        -1.35669966  0.4620276
```

```
#linearity plots
gg_resX(diets.mod, ncol = 4)
```

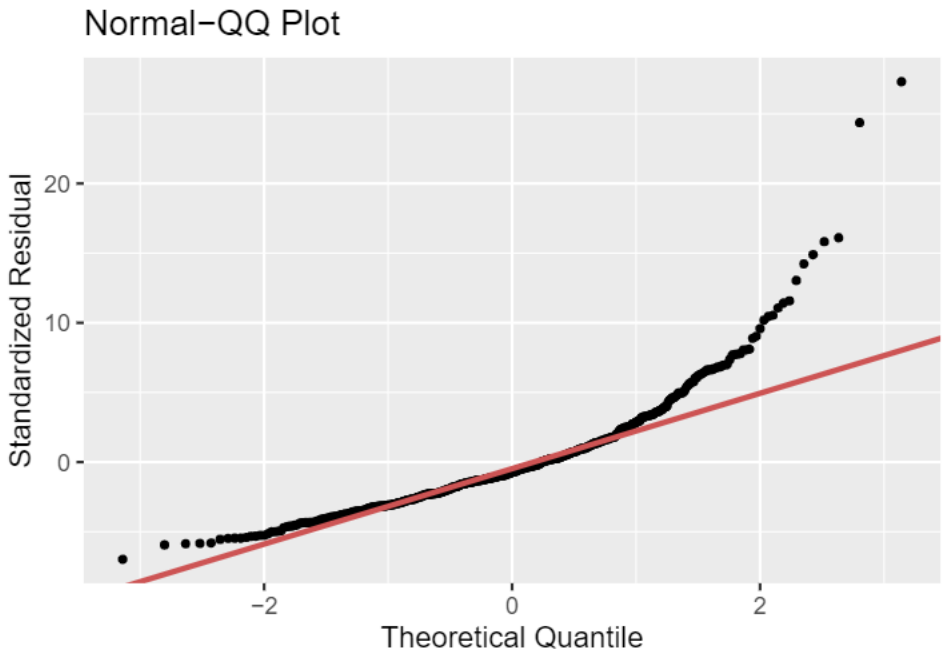



```
#assess constant variance
gg_resfitted(diets.mod)
```




```
#assess normality of residuals
gg_qqplot(diets.mod)
```



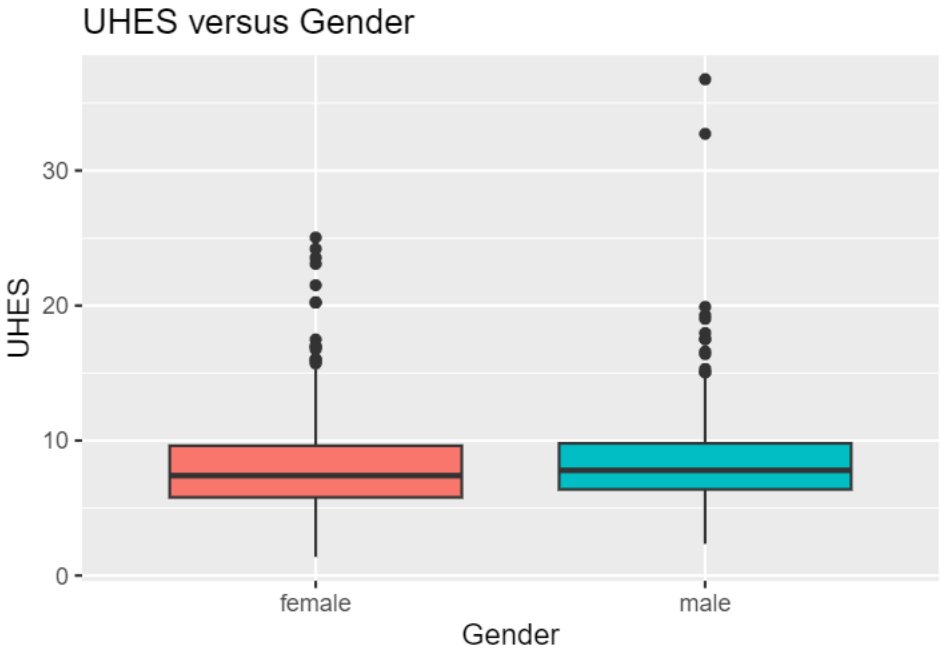

```
summary(diets.mod)$r.squared
```

```
## [1] 0.02456361
```

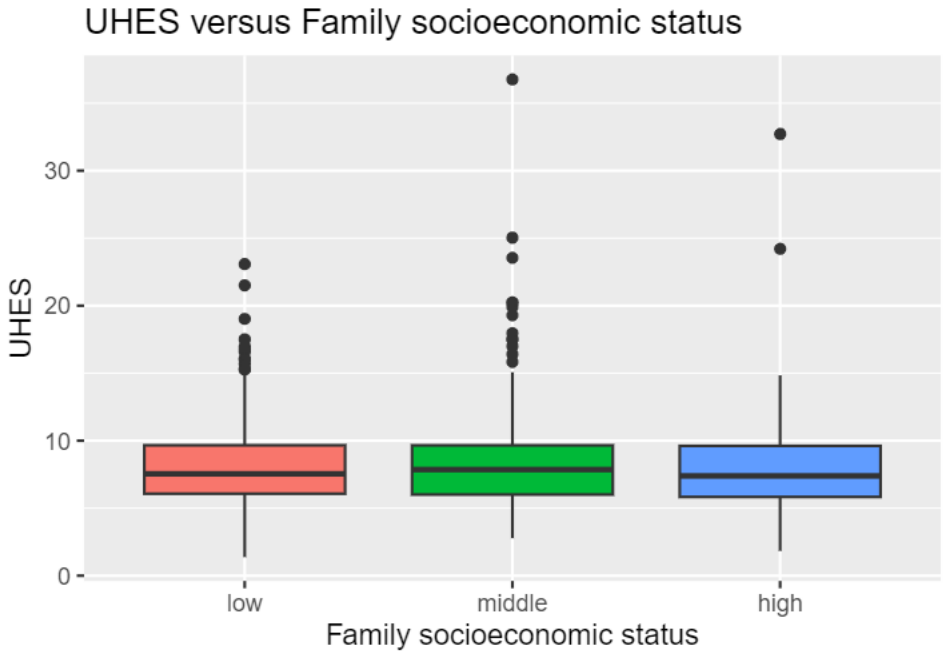
### Code for Main Figures

```
#create a boxplot UHES vs Degree
deg <- diets_new  %>%
```

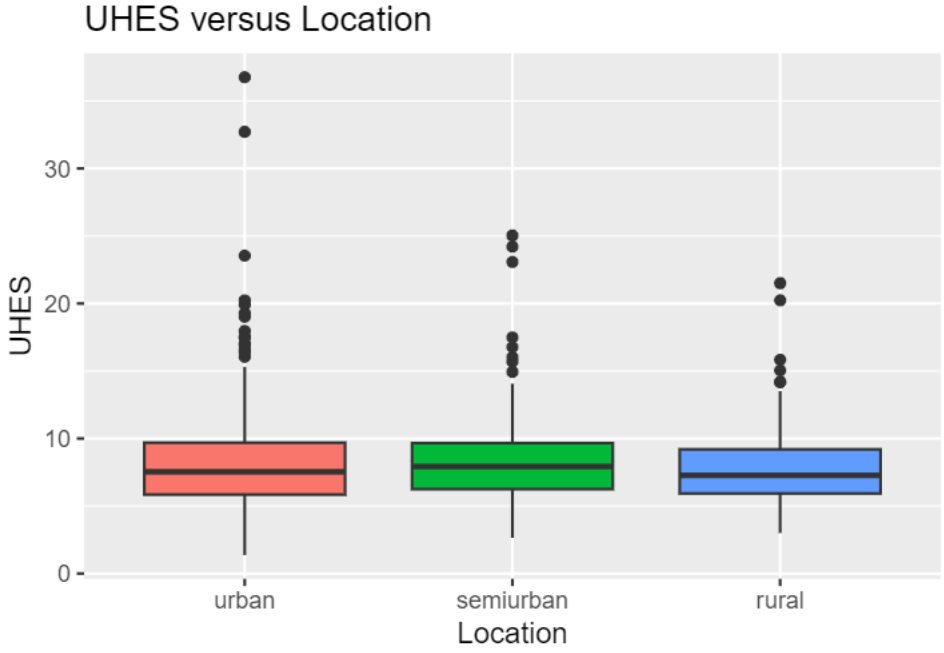
```
ggplot(aes(y = UHES, x = degree, fill = degree)) +
geom_boxplot() +
labs(title = "Unhealthy Eating Score (UHES) versus Degree",
y = "Unhealthy Eating Score (UHES)",
x = "Degree") +
guides(fill = "none")

#create a boxplot UHES vs gender
gnd <- diets_new  %>%
ggplot(aes(y = UHES, x = gender, fill = gender)) +
geom_boxplot() +
labs(title = "UHES versus Gender",
y = "UHES",
x = "Gender") +
guides(fill = "none")
gnd
```




```
#create a boxplot for UHES vs Family socioeconomic status
soc <- diets_new  %>%
ggplot(aes(y = UHES, x = hisei, fill = hisei)) +
geom_boxplot() +
labs(title = "UHES versus Family socioeconomic status",
y = "UHES",
x = "Family socioeconomic status") +
guides(fill = "none")
soc
```




```
#create a boxplot for UHES vs Location
loc <- diets_new  %>%
ggplot(aes(y = UHES, x = location, fill = location)) +
geom_boxplot() +
labs(title = "UHES versus Location",
y = "UHES",
x = "Location") +
guides(fill = "none")
loc
```



```r
#create a boxplot UHES vs cook
cok <- diets_new  %>%
ggplot(aes(y = UHES, x = cook, fill = cook)) +
geom_boxplot() +
labs(title = "UHES versus whether the student cooks for themselves",
y = "UHES",
x = "Student cooks for themselves") +
guides(fill = "none")
cok
```