

FoodExpenditures

Braden Critchfield and Thomas Olsen

2024-02-15

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
## Warning: package 'purrr' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
## Warning: package 'lubridate' was built under R version 4.2.3
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.2      v readr      2.1.4
```

```
## v forcats   1.0.0      v stringr   1.5.0
```

```
## v ggplot2    3.4.4      v tibble     3.2.1
```

```
## v lubridate  1.9.2      v tidyr      1.3.0
```

```
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
food <- read_delim("FoodExpenses.txt", delim=" ")
```

```
## Rows: 523 Columns: 2
```

```
## -- Column specification -----
```

```
## Delimiter: " "
```

```
## dbl (2): Income, EatingOut
```

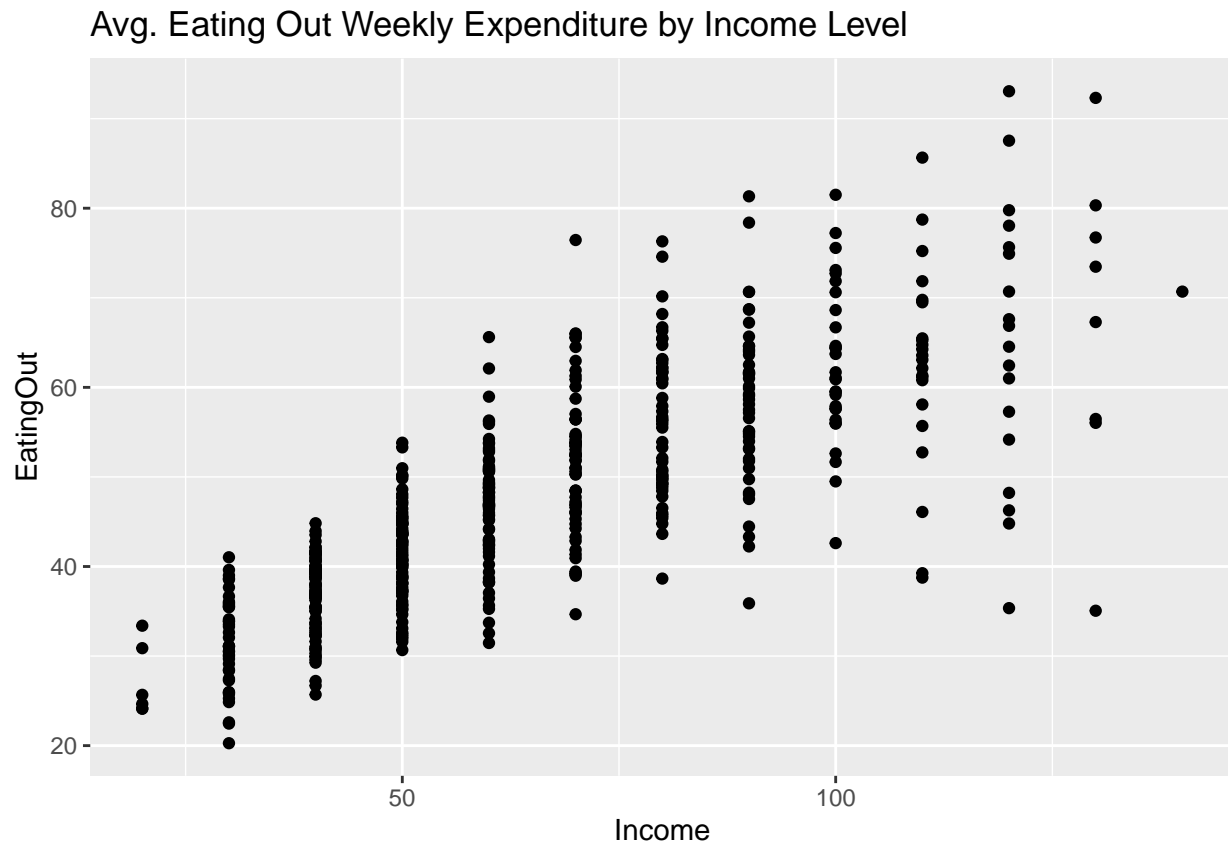
```
##
```

```
## i Use 'spec()' to retrieve the full column specification for this data.
```

```
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

1: Exploratory Plots and Summary Statistics

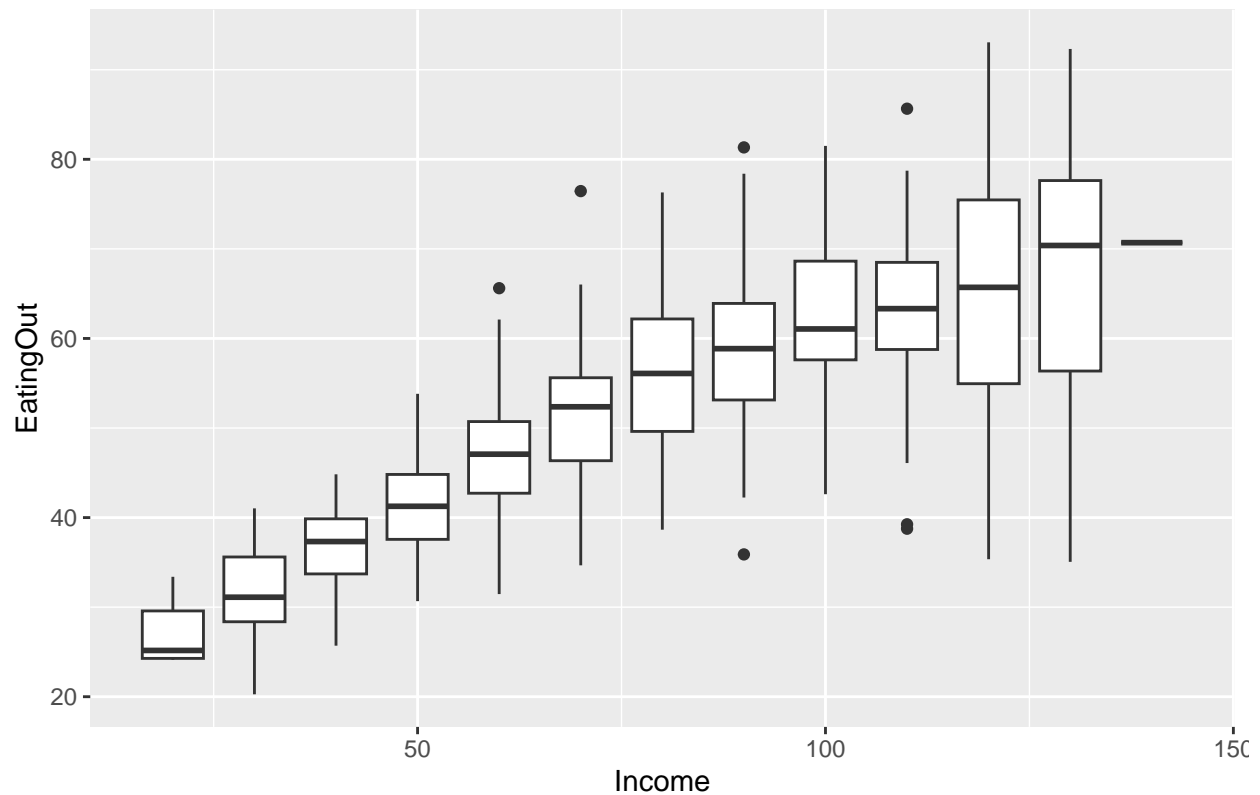
```
ggplot(data = food, aes(y = EatingOut, x = Income)) +  
  geom_point() +  
  ggtitle("Avg. Eating Out Weekly Expenditure by Income Level")
```



There seems to be a positive linear relationship between annual household income and average weekly expenditure on food not cooked at home. However, it looks like the variance grows larger as Income grows larger.

```
ggplot(data = food, aes(y = EatingOut, x = Income, group=Income)) +  
  geom_boxplot() +  
  ggtitle("Distributions of Eating Out Weekly Expenditure by Income Level")
```

Distributions of Eating Out Weekly Expenditure by Income Level



Because Income is a discrete variable, this shows the distributions of Average weekly eating out expenditure per income level. This still shows a positive relationship between the two variables, and again shows how the variance grows as the income increases.

Here are the summary statistics:

```
summary(food)
```

```
##      Income      EatingOut
##  Min.   : 20.00   Min.    :20.27
##  1st Qu.: 45.00   1st Qu.:38.65
##  Median : 60.00   Median :46.53
##  Mean   : 65.97   Mean    :48.04
##  3rd Qu.: 80.00   3rd Qu.:56.40
##  Max.   :140.00   Max.    :93.06
```

- 2: Fit homoskedastic linear model
- 3: Write down heteroskedastic model
- 4: Fit Model from Q3 to Eating Out and check assumptions
- 5: Validate predictions via cross-validation
- 6: Report Beta Hat and Variance Parameters
- 7: Test if the economy is not healthy
- 8: Predict your own restaurant spend