Assignment 2

GEO865 Advanced Quantitative Methods. – SS24

Leo Baldiga

2024-02-21

Selected Environmental Attitudes Variables

ISSP Environment 2020 For this assignment, I selected Thailand as my country and the following variables from the ISSP dataset on attitudes towards the environment:

- **threxg20**: Threats to the environment are greatly exaggerated.
- **prghrm20**: People worry too much about human progress harming the environment.
- **worent20**: We worry too much about the future of the environment and not enough about prices and jobs today.
- **morimp20**: There are more important things to do in life than protect the environment.
- **others20**: There is no point in doing what I can for the environment unless others do the same.
- **TH_RINC:**: Respondent's income Monthly gross income in Thai Baht (THB)

Descriptive Statistics, Missing Data, Visuals The Thailand sample size is (n=1498). Selected variables were recoded to remove -8 and -9 values using the following function:

```
remove_out_of_range <- function(data, likert_variables) {
  data %>%
    mutate(across(all_of(likert_variables), ~ifelse(as.numeric(as.character(.)) > 5 | as.numeric(as.cha
}

load(".RData")

Assignment_2_vars <- remove_out_of_range(
  data = Postcoded,
  likert_variables= c("threxg20", "worent20","prghrm20", "morimp20", "others20")) %>%
  mutate(TH_RINC = ifelse(TH_RINC < 0 , NA, TH_RINC))
```

Then, the following code was used to generate tables of descriptive statistics to make sure there are no outliers remaining in the data, count the NA values, and to see the central tendency, skew, etc. The tables and plots were generated using the following code:

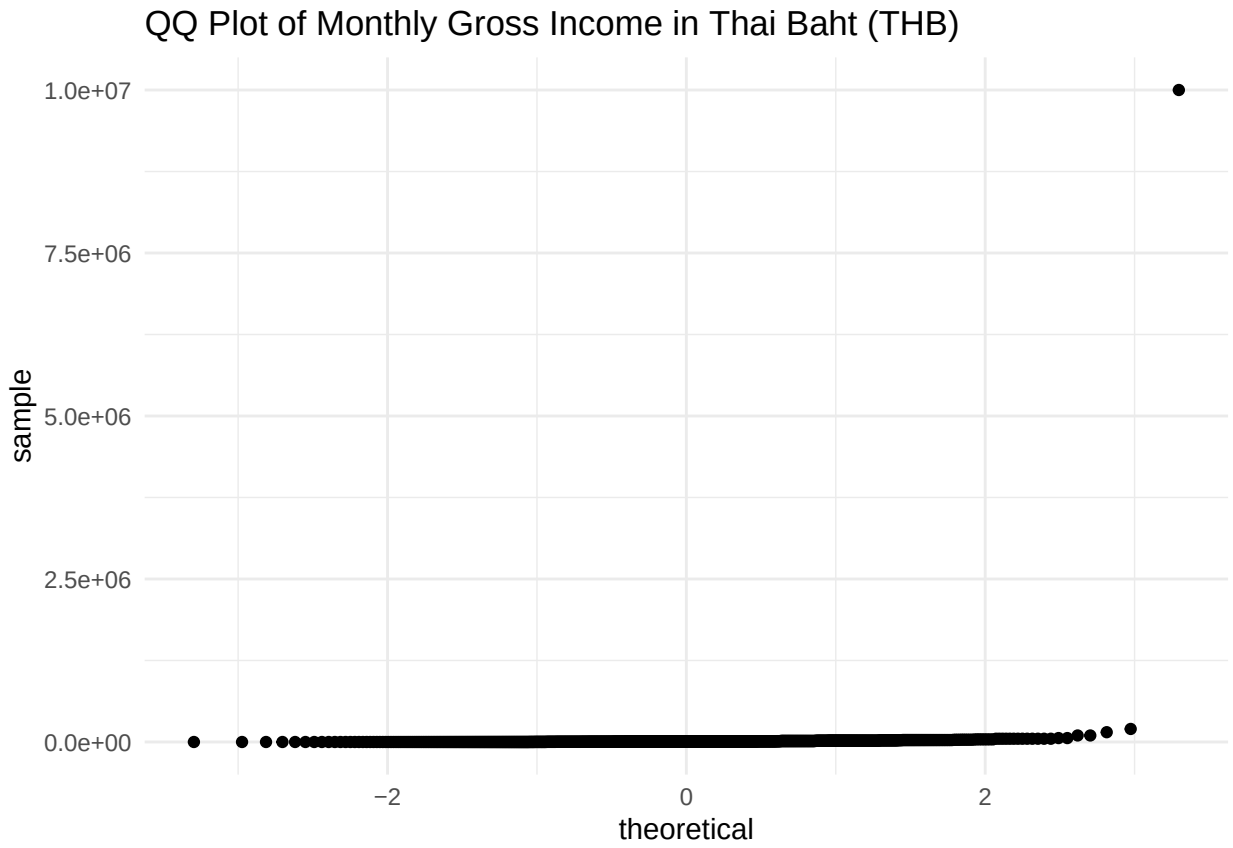
```
environmental_concern_desc_th <- Assignment_2_vars %>%
  select(worent20, prghrm20, morimp20, others20, threxg20, TH_RINC , country) %>%
  filter(country == "Thailand") %>%
  select(-country) %>% #removes the country variable from the table
  describe(na.rm = TRUE)
```

```r
environmental_concern_desc_th_NA <- Assignment_2_vars %>%
  select(worent20, prghrm20, morimp20, others20, threxg20, TH_RINC, country) %>%
  filter(country == "Thailand") %>%
  summarise_all(~sum(is.na(.))) %>%
  select(-country) #removes the country variable from the table

qq_th_rinc <- ggplot(Assignment_2_vars, aes(sample = TH_RINC)) +
  stat_qq() +
  ggtitle("QQ Plot of Monthly Gross Income in Thai Baht (THB)") +
  theme_minimal()

#a light blue histogram and bins of 3600 THB, or Approximately 100 USD
hist_th_rinc <- ggplot(Assignment_2_vars, aes(x = TH_RINC)) +
  geom_histogram(binwidth = 3600, fill = "lightblue") +
  ggtitle("Histogram of Monthly Gross Income in Thai Baht (THB)") +
  labs(x = "Monthly Gross Income in Thai Baht (THB)", y = "Frequency") +
  theme_minimal()
```

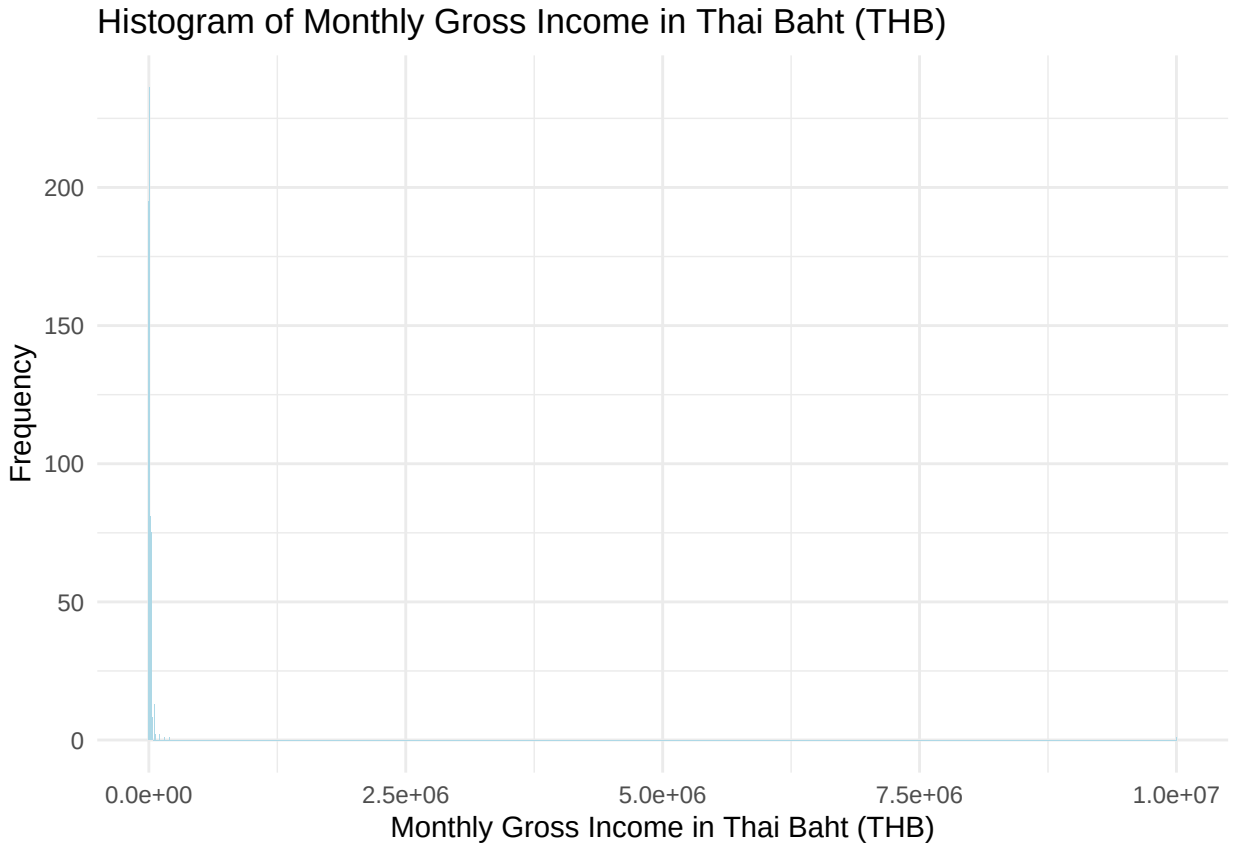
```
##
## \begin{longtable}[t]{lrrrrrrrrrrrr}
## \caption{\label{tab:print_desc_table}Environmental Concerns in Thailand}\\
## \toprule
##  & vars & n & mean & sd & median & trimmed & mad & min & max & range & skew & kurtosis & se\\
## \midrule
## worent20 & 1 & 1394 & 2.682927 & 9.807383e-01 & 2 & 2.664875 & 1.4826 & 1 & 5e+00 & 4e+00 & 0.356604
## prghrm20 & 2 & 1373 & 2.658412 & 9.699551e-01 & 2 & 2.651501 & 1.4826 & 1 & 5e+00 & 4e+00 & 0.313253
## morimp20 & 3 & 1387 & 2.672675 & 9.903555e-01 & 3 & 2.682268 & 1.4826 & 1 & 5e+00 & 4e+00 & 0.217875
## others20 & 4 & 1384 & 3.031069 & 9.775736e-01 & 3 & 3.032491 & 1.4826 & 1 & 5e+00 & 4e+00 & -0.02956
## threxg20 & 5 & 1339 & 3.150859 & 9.499927e-01 & 3 & 3.149115 & 1.4826 & 1 & 5e+00 & 4e+00 & -0.08457
## \addlinespace
## TH\_RINC & 6 & 1020 & 20117.901961 & 3.130596e+05 & 7000 & 8263.186274 & 7413.0000 & 0 & 1e+07 & 1e+
## \bottomrule
## \end{longtable}
```

```
##
## \begin{longtable}[t]{rrrrr}
## \caption{\label{tab:print_desc_table}Missing Values for Environmental Concerns in Thailand}\\
## \toprule
## worent20 & prghrm20 & morimp20 & others20 & threxg20\\
## \midrule
## 104 & 125 & 111 & 114 & 159\\
## \bottomrule
## \end{longtable}
```

```
## Warning: Removed 43080 rows containing non-finite values (`stat_qq()`).
```




```
## Warning: Removed 43080 rows containing non-finite values (`stat_bin()`).
```



The mean values for all variables range from 2.7 to 3.1, which suggests that on average, respondents' environmental concerns are moderate, and the trimmed means are very close to the actual means, which implies that the influence of outliers on the mean is minimal. The skewness values for worent20, prghrm20, and morimp20 are positive but not very high, indicating a slight tendency for respondents to rate their environmental concerns towards the higher end of the scale. threxg20 has a moderate negative skewness, suggesting that responses for this variable are more frequently on the lower end of the scale.

The skew and kurtosis for TH_RINC are very high, indicating that the distribution of income is not normal and there are some sever outliers. The QQ plot for TH_RINC also shows that the distribution of income is not normal, as the line deviates from the 45-degree line at the higher end of the scale. The histogram of TH_RINC shows a very left-skewed distribution of income. Let's remove those observations and re-run the descriptive statistics and QQ plot, and histogram.

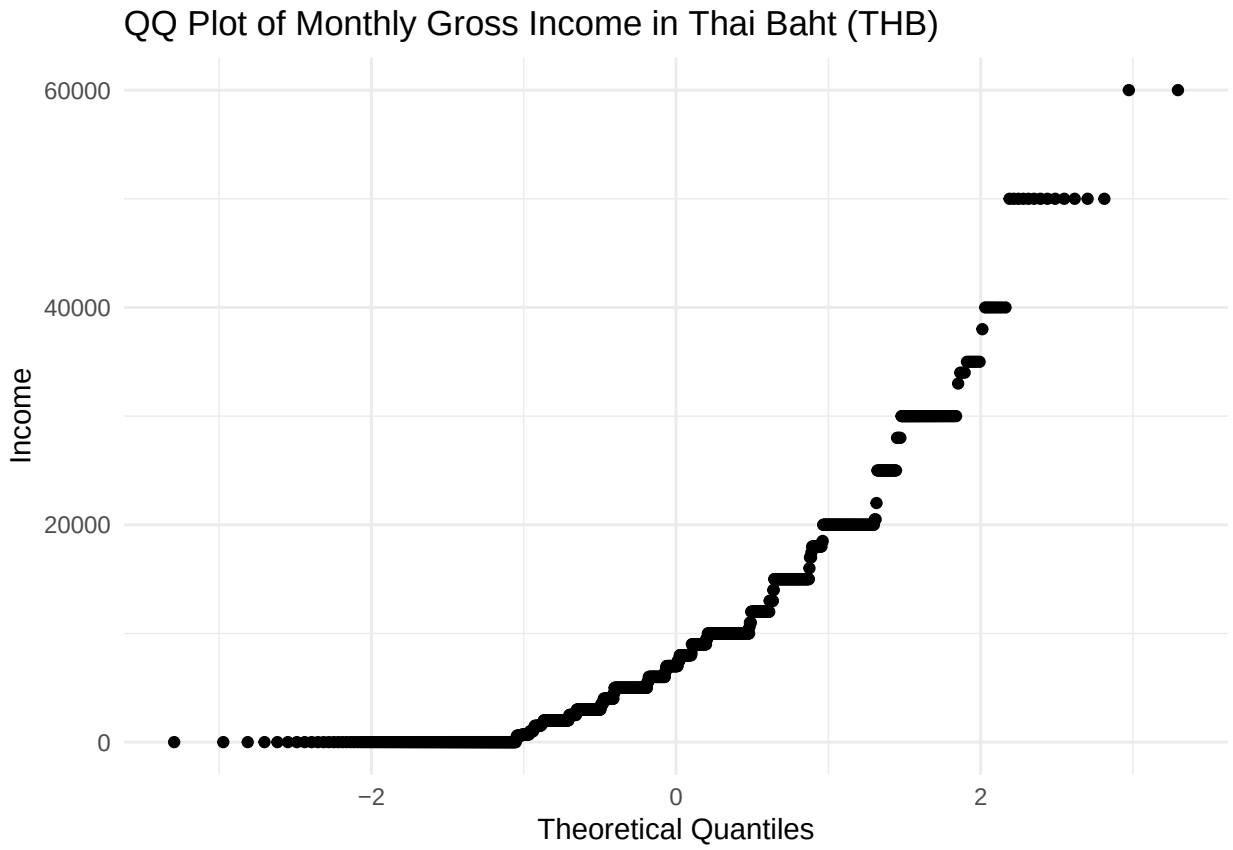
```
#Recode incomes greater than 100000 THB per month to NA
Assignment_2_vars <- Assignment_2_vars %>%
  mutate(TH_RINC = ifelse(TH_RINC >= 100000 , NA, TH_RINC))
```

```
##          vars    n    mean      sd median trimmed     mad min   max range  skew
## worent20    1 1394    2.68    0.98      2    2.66    1.48   1     5     4  0.36
## prghrm20    2 1373    2.66    0.97      2    2.65    1.48   1     5     4  0.31
## morimp20    3 1387    2.67    0.99      3    2.68    1.48   1     5     4  0.22
## others20    4 1384    3.03    0.98      3    3.03    1.48   1     5     4 -0.03
## threxg20    5 1339    3.15    0.95      3    3.15    1.48   1     5     4 -0.08
## TH_RINC     6 1015 9822.92 9971.91   7000 8194.66 7413.00   0 60000 60000  1.72
##          kurtosis     se
## worent20    -0.63   0.03
## prghrm20    -0.62   0.03
## morimp20    -0.63   0.03
```

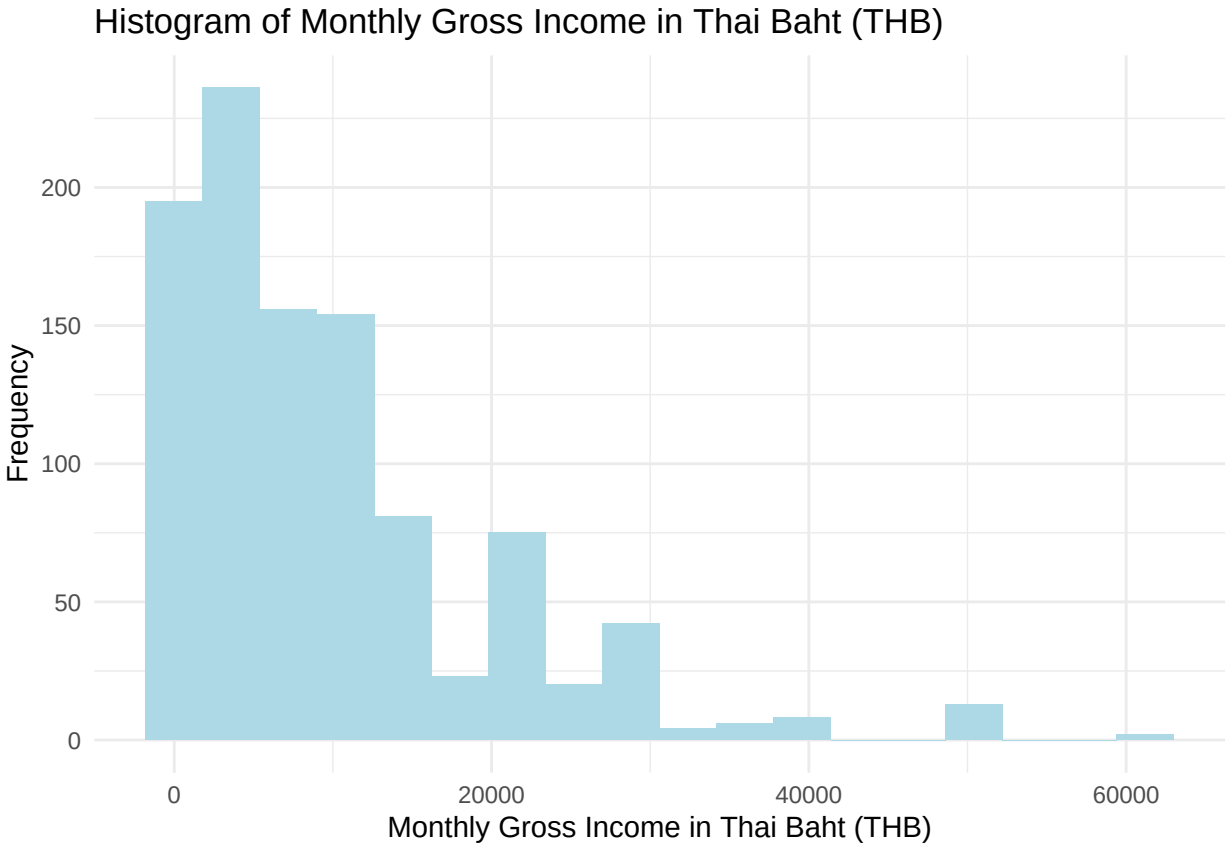
```
## others20    -0.69   0.03
## threxg20    -0.51   0.03
## TH_RINC      3.73 313.00
```

```
## # A tibble: 1 x 6
##   worent20 prghrm20 morimp20 others20 threxg20 TH_RINC
##      <int>    <int>    <int>    <int>    <int>   <int>
## 1      104      125      111      114      159     483
```

```
## Warning: Removed 43085 rows containing non-finite values (`stat_qq()`).
```




```
## Warning: Removed 43085 rows containing non-finite values (`stat_bin()`).
```



Seems a bit more reasonable now, though we lost 5 more observations. The QQ plot shows a slightly more linear relationship between the observed and expected values, and the histogram shows a skill very left skewed, albeit slightly more normal distribution of income. The skewness and kurtosis values for TH_RINC are still high, but not as high as before.

Let's look at the actual distribution of the likert variables: #"'{r likert_dist} max_levels <- max(sapply(Assignment_2_vars %>% select(worent20, prghrm20, morimp20, others20, threxg20), function(x) length(unique(x))))

Now, filter the data frame for "Thailand", omit rows with NAs, and ensure all factors have the same levels

items <- Assignment_2_vars %>% filter(country == "Thailand") %>% select(worent20, prghrm20, morimp20, others20, threxg20) %>% na.omit() %>% mutate(across(everything(), ~factor(., levels = c(1, 2, 3, 4, 5), labels = c("agree strongly", "agree", "neither agree nor disagree", "disagree", "disagree strongly"))))

Perform Likert analysis specifying the maximum number of levels

l <- likert(items, grouping = "country") summary(l) plot(l)

"'