

Attitudes about Science and Faith

Rory Quinlan

2024-09-03

Setup

```
require(tidyverse)
library(car)
library(kableExtra)
library(psych)
require(plotly)
library(dplyr)
library(ggplot2)
```

```
# Load data
ebsurvey <- read.csv('https://uoepsy.github.io/data/science-faith-attitude.csv')

# Select columns to use
ebsurvey0 <- ebsurvey %>%
  select(kstot, age, toomuchscience)

# Rename columns to be more understandable
ebsurvey0 <- ebsurvey0 %>%
  rename(science_score = kstot,
         attitude = toomuchscience)
```

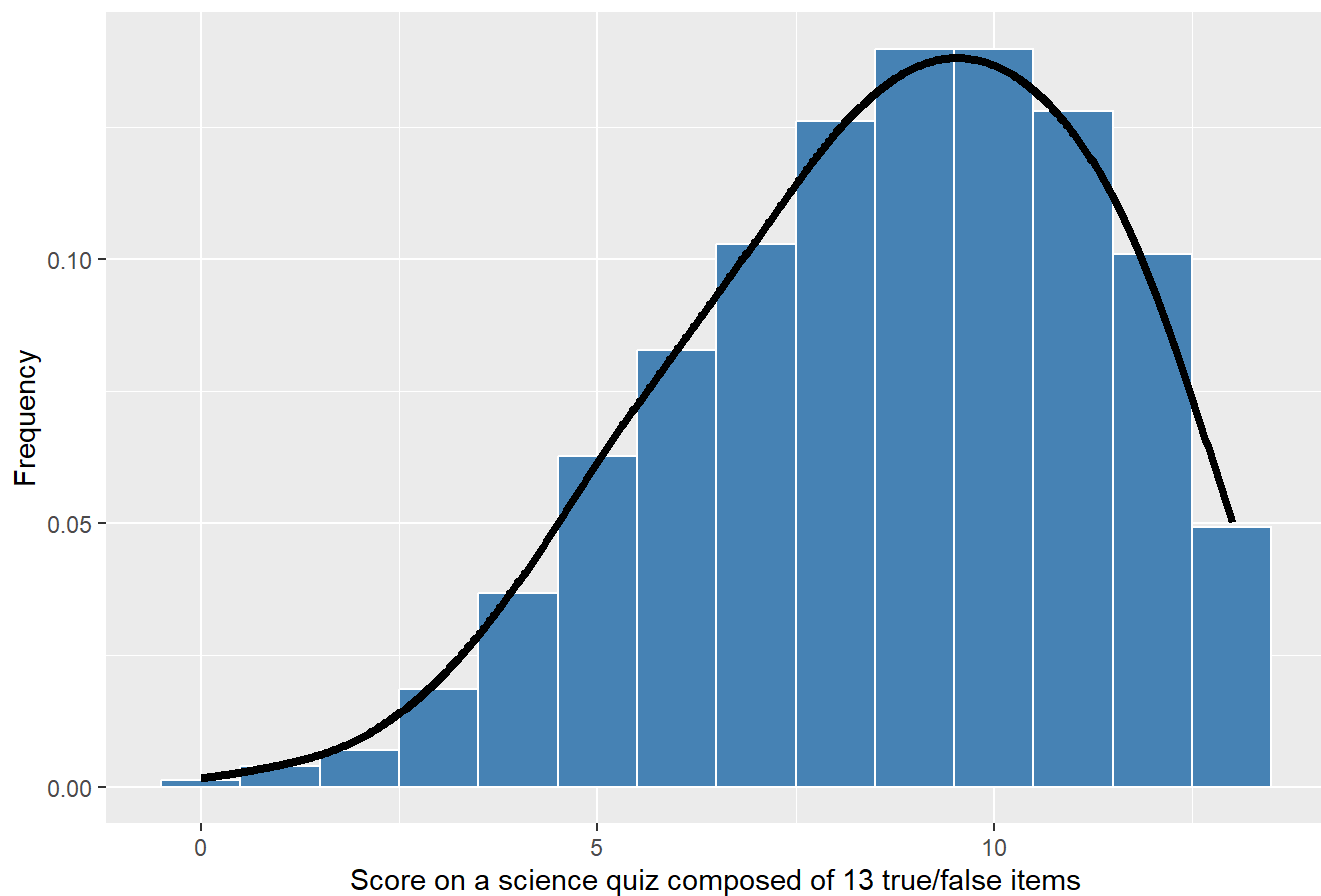
```
# Omit na values
dim(ebsurvey0)
ebsurvey1 <- na.omit(ebsurvey0)
dim(ebsurvey1)
```

Data Exploration

Science Score

```
# Plot Science Score of Respondents
ggplot(ebsurvey1, aes(x = science_score)) +
  geom_histogram(aes(y=..density..), color = 'white', fill='steelblue', binwidth = 1,) +
  labs(x = 'Score on a science quiz composed of 13 true/false items', y = 'Frequency')+ geom_
density(adjust=2, size= 1.5)+
  ggtitle("Science Score of Respondents Frequency")
```

Science Score of Respondents Frequency



Summary Statistics of Science Score

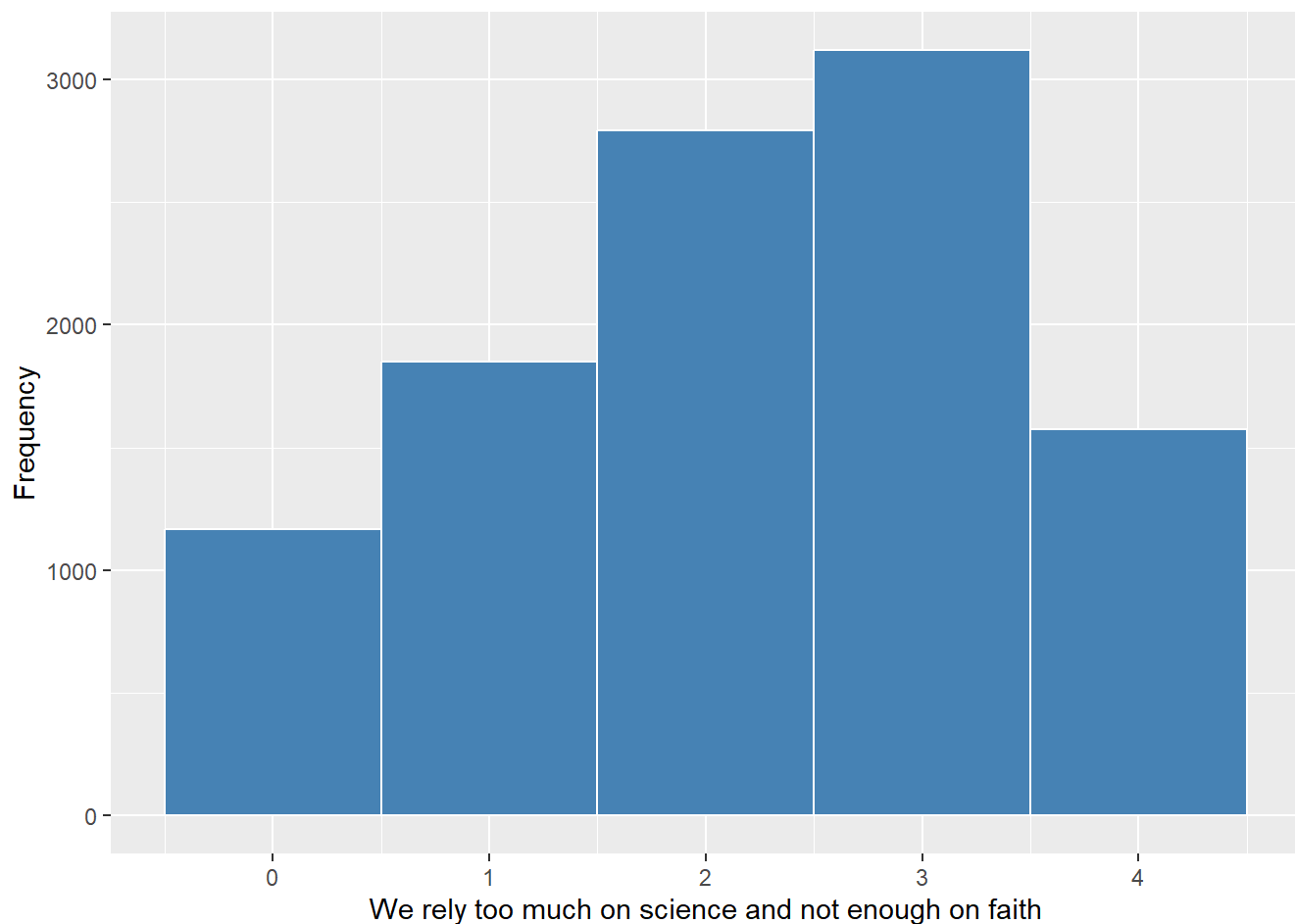
```
# Create Table for summary Stats
ebsurvey1 %>%
  summarise(Min = min(science_score),
            Median = quantile(science_score, 0.5),
            IQR = IQR(science_score),
            Mean = mean(science_score),
            SD = sd(science_score),
            Max = max(science_score)) %>%
  kable(digits = 2,
        caption = 'Descriptive Statistics of Science Score') %>%
  kable_styling(full_width = FALSE)
```

Descriptive Statistics of Science Score

Min	Median	IQR	Mean	SD	Max
0	9	4	8.68	2.61	13

Attitude

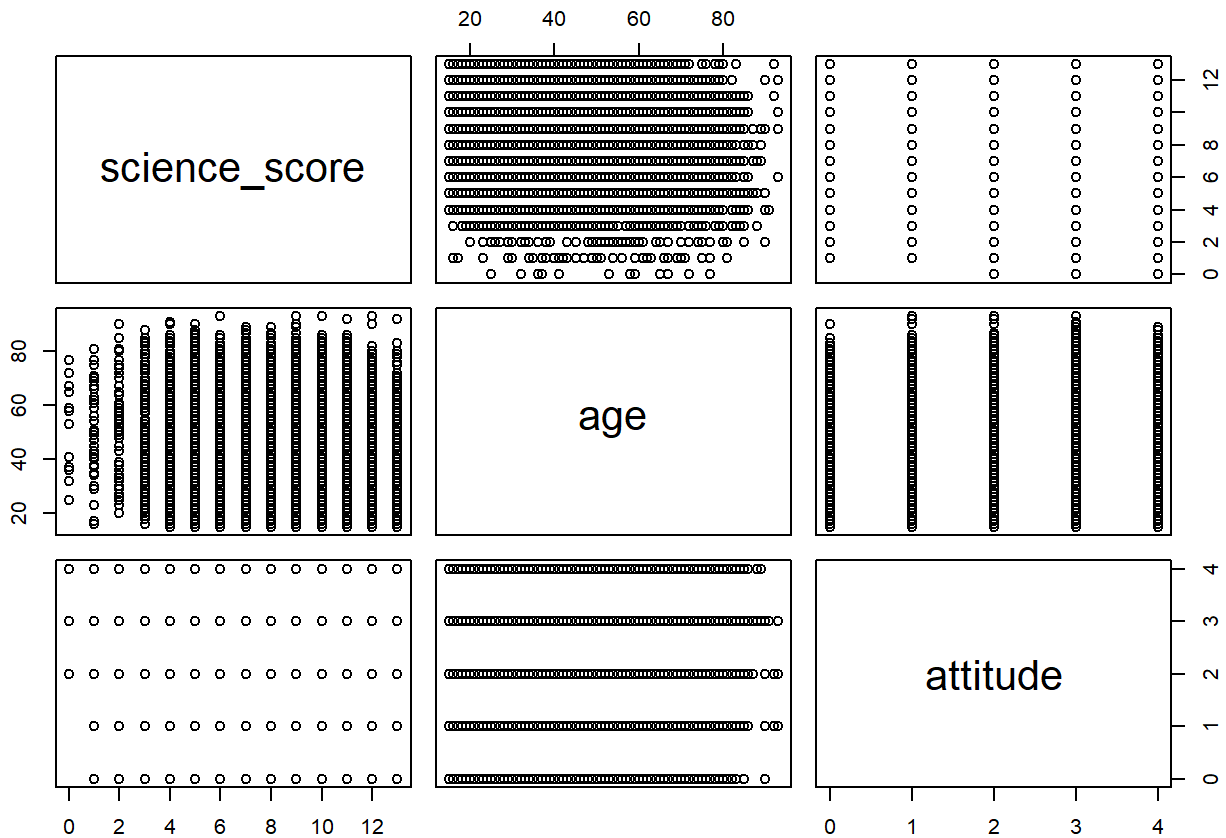
```
ggplot(ebsurvey1, aes(x = attitude)) +
  geom_histogram(color = 'white', fill='steelblue', binwidth = 1) +
  labs(x = 'We rely too much on science and not enough on faith',
       y = 'Frequency')
```



- The mean score on the science and faith attitude variable is just over 2. There are only 5 discrete values possible in the distribution, based on the response options available, but the distribution looks approximately normal, with a slight negative skew.

Visualize the pairwise relationships between your variables and explore the possible correlations

```
plot(ebsurvey1)
```



```
cor(ebsurvey1)
```

```
##           science_score      age    attitude
## science_score      1.0000000 -0.12239434 -0.17634507
## age                -0.1223943  1.00000000  0.05496583
## attitude           -0.1763451  0.05496583  1.00000000
```

- From the pairwise scatterplots, it does not seem like there is a strong linear dependence of attitude to science and faith on a person's age and science knowledge
- In this case, the Pearson correlation coefficient between age and science knowledge is -0.12. The correlation is relatively small in absolute terms, and we therefore have little concern about multicollinearity influencing this regression analysis.
- The correlation between attitude and age is 0.05, while with science knowledge it is -0.17. So, overall there is a very weak linear relationship.

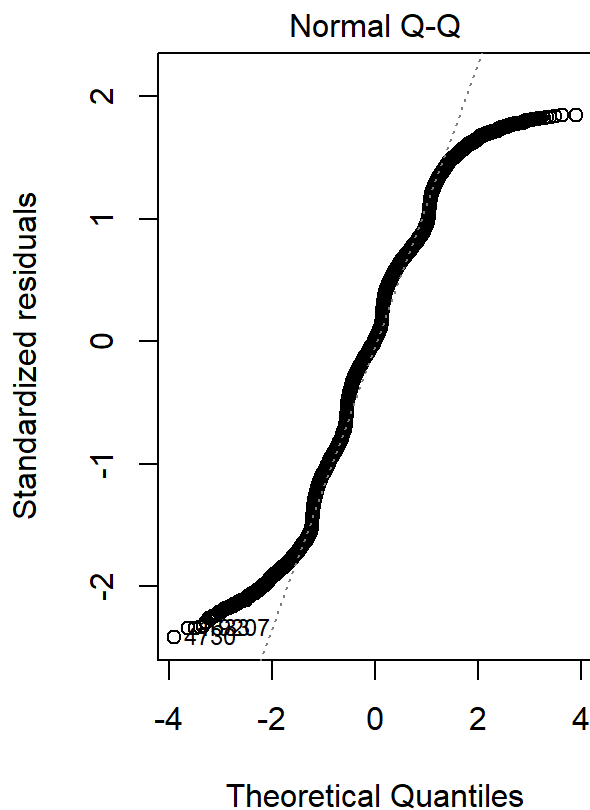
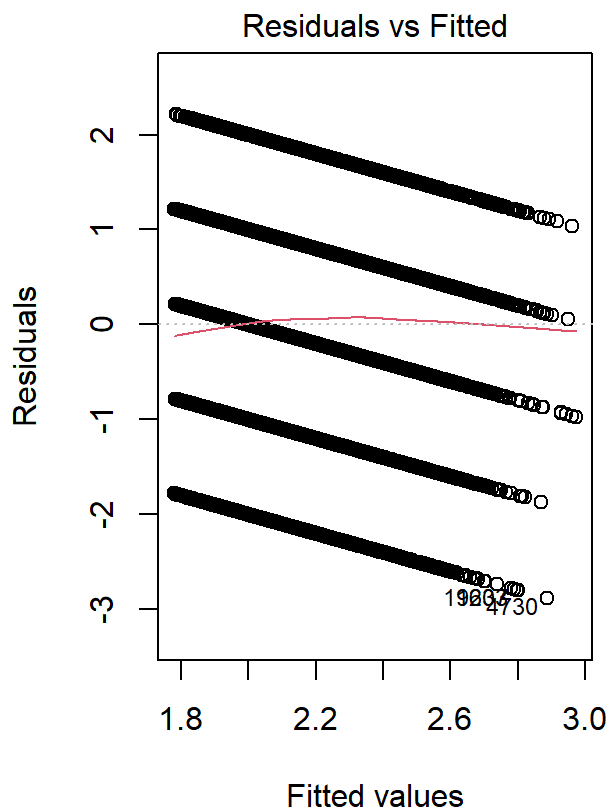
Linear Regression

```
reg <- lm(attitude ~ science_score + age, data=ebsurvey1)
summary(reg)
```

```
##
## Call:
## lm(formula = attitude ~ science_score + age, data = ebsurvey1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.88643 -0.97484  0.00913  0.88427  2.21489
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.7882487   0.0537759   51.849 < 2e-16 ***
## science_score -0.0802763   0.0045095  -17.802 < 2e-16 ***
## age           0.0023794   0.0006792    3.503 0.000461 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.196 on 10500 degrees of freedom
## Multiple R-squared:  0.03223,    Adjusted R-squared:  0.03204
## F-statistic: 174.8 on 2 and 10500 DF,  p-value: < 2.2e-16
```

Check Regression Assumptions

```
par(mfrow = c(1,2))
plot(reg, which=c(1,2))
```



- It appears that normality and heteroskedasticity are violated in our model

Bootstrap

```
# Bootstrap
boot_reg <- Boot(reg, R = 1000)
# CI for bootstrap coeff
Confint(boot_reg, level = 0.95, type = "perc")
```

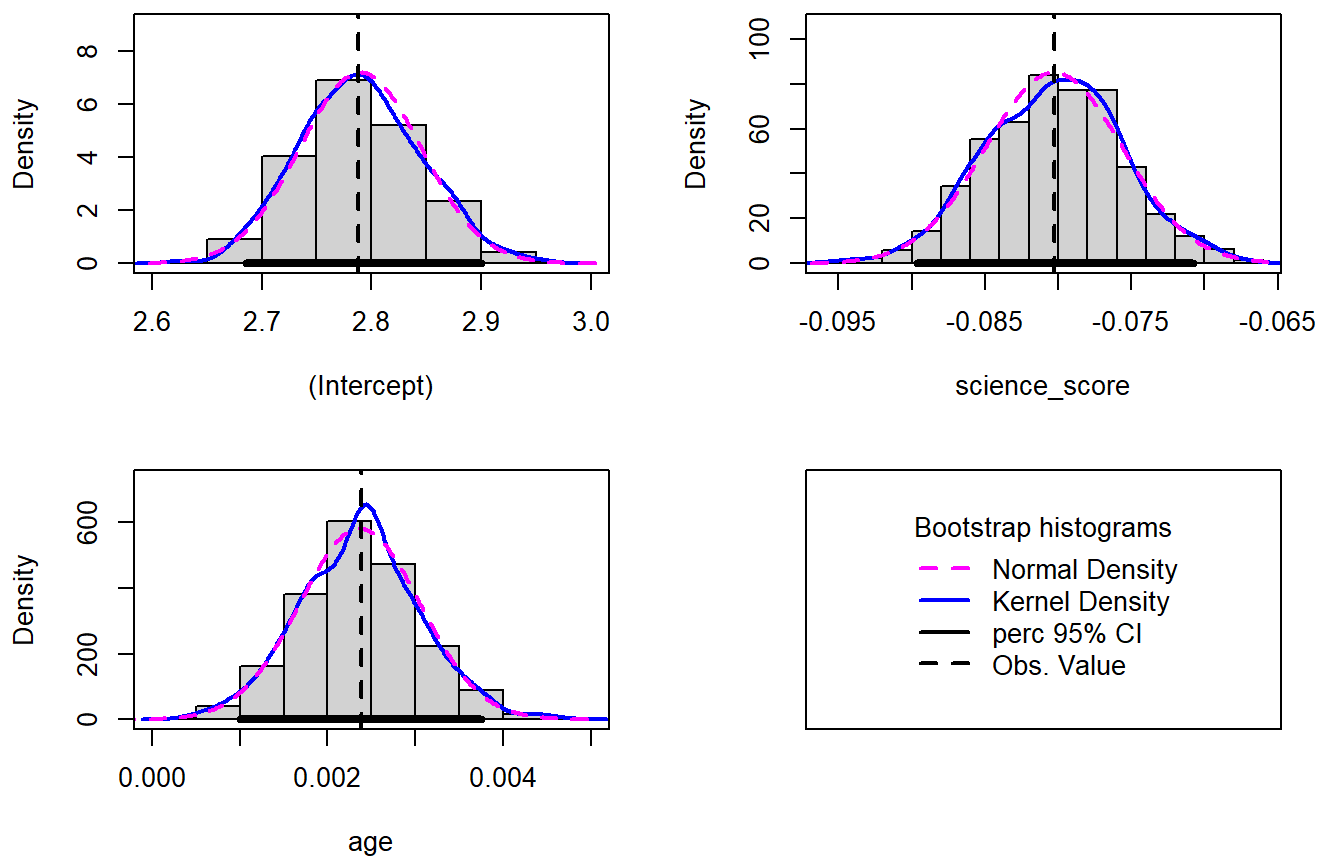
```
## Bootstrap percent confidence intervals
##
##           Estimate      2.5 %      97.5 %
## (Intercept)  2.788248712  2.685667291  2.899992788
## science_score -0.080276256 -0.089652341 -0.070774961
## age          0.002379446  0.001010217  0.003744146
```

```
Confint(boot_reg, type = "perc") %>%
  kable(digits = 3, caption = 'Bootstrap 95% CIs') %>%
  kable_styling(full_width = FALSE)
```

Bootstrap 95% CIs

	Estimate	2.5 %	97.5 %
(Intercept)	2.788	2.686	2.900
science_score	-0.080	-0.090	-0.071
age	0.002	0.001	0.004

```
# Plot distribution of slope values for each variable from bootstrap
plot_data <- as_tibble(boot_reg$t)
hist(boot_reg, ci = "perc", legend = "separate")
```



Intercept

The results in the table report an estimate of the intercept (or constant) as equal to approximately 2.79. The constant of a multiple regression model can be interpreted as the average expected value of the dependent variable when all of the independent variables equal zero. In this case, the independent variable science knowledge has only a handful of respondents that score zero, and no one is aged zero, so the constant by itself does not tell us much. Researchers do not often have predictions based on the intercept, so it often receives little attention.

Science knowledge slope

The estimated value for the slope coefficient linking knowledge to attitude is estimated to be approximately -0.08. This represents the average marginal effect of knowledge on attitude, and can be interpreted as the expected change in the dependent variable on average for a one-unit increase in the independent variable, controlling for age. In this example, every increase in quiz score by one point is associated with a decrease in attitude score of about -0.08, adjusted for age. Bearing in mind the valence of the question wording, this means that those who are more knowledgeable tend to be more favorable towards science - i.e. disagreeing with the statement.

Age slope

The slope coefficient linking age to attitude is estimated to be approximately 0.002. This represents the average marginal effect of each additional year on attitude, and can be interpreted as the expected change in the dependent variable on average for a one-unit increase in the independent variable, controlling for science knowledge. For this example, that means that for every year older a person is, their attitude score is expected to increase by 0.002, controlling for science knowledge. This may seem like a very small effect, but remember that this is the effect of only one additional year. Bearing in mind the valence of the question wording, this means that older people tend to be less favorable towards science - i.e. agreeing with the statement.

Bootstrap Results

The bootstrap confidence intervals table also reports that the 95% confidence intervals for both slope estimates do not include 0. This leads us to reject both null hypotheses at the 5% significance level, and conclude that there appear to be relationships for both age and science knowledge with attitude to science and faith.