

Unit 10 Lab Report

Christopher Byrd

2024-11-28

Contents

Section 1: Hypothesis	1
Section 2: Data	2
Section 3: Method	2
Section 4: Results	2
Section 5: Discussion and Conclusion	3
Section 6: Graphics and Visuals	3

Section 1: Hypothesis

The goal of this research paper is to determine if age of those who pray regularly has a statistically significant effect on income. The following are the hypothesizes examined:

Null: There is no significant correlation between age of those who pray regularly and their income.
Research₁: There is a significant correlation between age of those who pray regularly and their income.

$H_0: \rho_{xy} = 0$
 $H_1: r_{xy} \neq 0$

Section 2: Data

The data being utilized is from the Baylor Survey of Religious Life. Participants who prayed at least once or more a day were asked to respond with their age and the bracket in which their income falls within. The following code book was used:

Question or Description	Possible Answers
Age range of people who pray once or several times a day as reported in 2017.	Range: 17-95
By your best estimate, what was your total household income last year, before taxes?	1) \$10,000 or less 2) \$10,001 to \$20,000 3) \$20,001 to \$35,000 4) \$35,001 to \$50,000 5) \$50,001 to \$100,000 6) \$100,001 to \$150,000 7) \$150,001 or more

Section 3: Method

To evaluate the data, Pearson's r was calculated using `rcorr`, a function which evaluates Pearson's r and the associated P-value. The data needed to be coerced into a matrix to allow calculation of each variable. The other option that exists is to create two separate data frames for each variable.

```
rcorr(as.matrix(Module_10_Lab_Data), type = "pearson")
```

To determine R^2 , I took a function off the internet to create a function for evaluating R^2 . The original function required you to coerce the variables to be matrices, so I put it in the function to avoid this.

```
rsq <- function (x, y) cor(as.matrix(x), as.matrix(y)) ^ 2
```

```
rsq(Module_10_Lab_Data, Module_10_Lab_Data)
```

Lastly, `ggplot2` was used to produce a scatter plot. This was achieved through the `geom_point()`, while `geom_smooth()` was used to create the trend line.

```
graphdata <- as.data.frame(Module_10_Lab_Data)
```

```
ggplot(data = graphdata, aes(x = Age, y = `Income Interval`)) +  
  geom_point(shape = 18, color = "black") +  
  geom_smooth(method = lm, color = "#f21f1f") +  
  xlim(0, 100) +  
  ylim(0, 8) +  
  theme_bw()
```

Section 4: Results

Due to there being only 2 variables, each individual result within each statistic is identical. This is because evaluating Age-Income is identical to Income-Age for r and R^2 .

As for the results, Pearson's r is -0.11 , showing a weak to nonexistent negative correlation between both variables. The p -value for this test was 0.0055 , showing this result to likely be statistically insignificant.

Lastly, for R^2 , the result was 0.0131723 . This shows that the model visible from the trend line in figure 1 is a poor predictor and does not explain and predict the variability observed.

Section 5: Discussion and Conclusion

From these results, it's clear that age has little to no impact on income.

It stands to reason the null is accepted and H_1 is rejected.

Section 6: Graphics and Visuals

```
## `geom_smooth()` using formula = 'y ~ x'
```

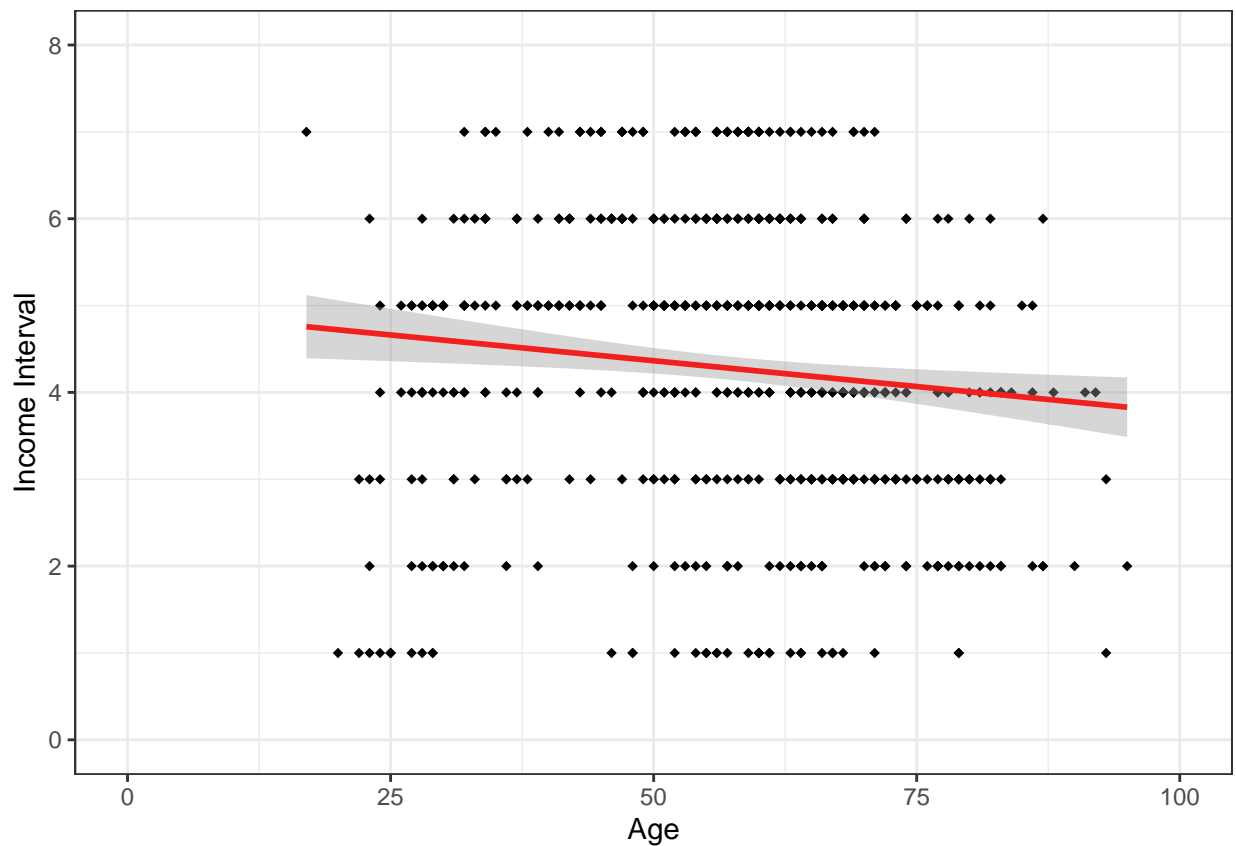


Figure 1: Scatter plot with trend line. Age and income of respondents.

```
##           Age Income Interval
## Age           1.00         -0.11
## Income Interval -0.11         1.00
##
## n= 584
##
```

```
##
## P
##           Age      Income Interval
## Age              0.0055
## Income Interval 0.0055
```

Figure 2: Matrix chart with n value. Pearson's r and p-values of age and income interval.

```
##           Age Income Interval
## Age          1.0000000      0.0131723
## Income Interval 0.0131723      1.0000000
```

Figure 3: Matrix chart. r^2 of age and income interval.