Unit 11 Lab Report

Christopher Byrd

2024-12-12

Contents

Section 1: Hypothesis

The goal of this research paper is to determine how age, income, and regular attendance of church services affects prayer rates.

Null: There will be no significant difference in prayer frequency based upon age, income range, or church attendance
Research$_1$: There will be a significant difference in prayer frequency based upon age, income range, or church attendance.
Research$_2$: Median age, income range, and church attendance will result in a prayer rate of a few times a week.

H_0: $\rho_{xy} = 0$
H_1: $r_{xy} \neq 0$
H_2: $Y_{x\_median} = 3$

Section 2: Data

The data being utilized is from the Baylor Survey of Religious Life. Participants asked to respond to the question listed in the following code book:

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
About how often do you spend time alone praying outside of religious services?	0) Never 1) Only on certain occasions 2) Once a week or less 3) A few times a week 4) Once a day 5) Several times a day
Attends a worship service about once a week or more.	1) No 2) Yes

Attendance was originally a nominal variable. To allow for a linear regression, it was recoded as a dummy variable.

Section 3: Method

To evaluate the data, the data set was first re-coded with shorter variable names, mainly because of my hate for long variable names. The new data set was then coerced to be a `data.frame` to make later actions easier (particularly with `ggplot2`, since data needs to be in tidy format).

```
# Re-coded with shorter names because idk I don't like long variable
# names. ¯\_( )_/¯

pf <- Module_11_Lab_Data$`Pray Frequency`
inc <- Module_11_Lab_Data$`Income Intervals`
age <- Module_11_Lab_Data$AGE
regat <- Module_11_Lab_Data$`Regular Attendance`

df <- data.frame(pf, inc, age, regat)
```

The next evaluation consisted of finding the median of income interval, age, and regular attendance. To cut down on the amount of code I needed to write, I made this a `for` loop. From here, the linear model (AKA regression) was calculated with prayer frequency regressed over income interval, age, and regular attendance of church services. A multiple regression was utilized to show how each variable interacts with prayer frequency together.

```
for (i in 2:4) {
  print(paste(median(df[[i]])))
}

lmresult <- lm(formula = pf ~ inc + age + regpr, df)
```

After running a summary to find the exact results of the `lm()`, a ggplot2 plot was created based upon the fitted values.

```
df$fitted <- lmresult$fitted.values
df$residuals <- lmresult$residuals

ggplot(df, aes(x = fitted, y = pf)) +
  geom_point() +
  geom_smooth(method = 'lm', color = '#679df5') +
  labs(title = "Prayer Frequency Based Upon Fitted Values") +
  xlab("Fitted values (see section 2-4 for more info)") +
  ylab("Prayer Frequency") +
  stat_poly_eq(use_label(c("eq", "R2")), rr.digits = 4,vjust = 1.3) +
  theme_pander() +
  theme(plot.margin = margin(t = 15, r = 15, b = 15, l = 15, unit = "pt")) +
  geom_point(aes(x = x_value, y = y_value), color = "red", size = 2.2) +
  geom_text(aes(x = x_value, y = y_value,
                label = paste("(", round(x_value, 2), ", ",
                              round(y_value, 2), ")", sep = "")),
            vjust = -1, hjust = 1, size = 3.2)
```

I completed the work below thinking I was working with Pearson's r, which I wasn't. I've placed it here to display the work I put into it:

```
for (i in 1:length(lmresult$coefficients)) {
  if (lmresult$coefficients[i] > 0) {
    print(paste("The coefficient for", names(lmresult$coefficients[i]),
                "is", lmresult$coefficients[i], "and direct"))
  }   else {
    print(paste("The coefficient for", names(lmresult$coefficients[i]),
                "is", lmresult$coefficients[i], "and indirect"))
  }
  if (abs(lmresult$coefficients[i]) > 0.2) {
    print("There is a weak to strong relationship.")
  }   else {
    print("There is little to no relationship")
  }
}
```

Section 4: Results

Full results can be found in figure 2 within section 6. Figure 2 shows that each individual independent variable is statistically significant in relation to how they affect the dependent. Furthermore, regular prayer and age display positive relationships, while income displays a negative relationship. Regular prayer is the largest predictor and the most statistically significant, at a p-value of $2E-16$.

Medians for the 3 dependent variables are as follows:

```
## [1] "5"
## [1] "57"
## [1] "0"
```

…while median for the fitted values is *2.241326*, based upon the following string: `median(df$fitted)`. To find the median of prayer frequency, the following code snippet was used:

```
x_value <- 2.241326
y_value <- (2.72e-14) + 1 * x_value
print(y_value)
```

```
## [1] 2.241326
```

As is shown, the value is 2.24, which, when rounded, means median prayer frequency falls within once a week or less. Fitted values are plotted in figure 1.

Section 5: Discussion and Conclusion

From these results, the null is rejected and research$_1$ is accepted. Research$_2$ is rejected, as the actual value of Y is 2.

Section 6: Graphics and Visuals

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

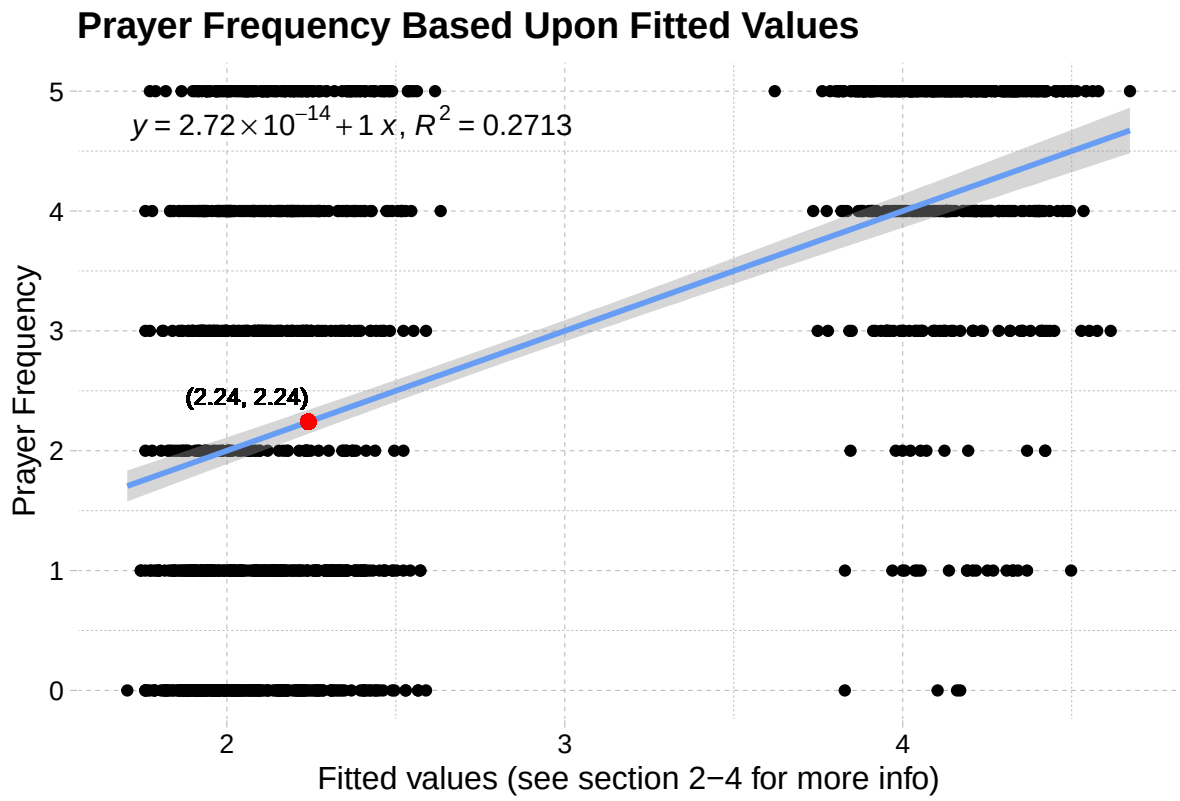

Figure 1: Scatter plot with linear regression based upon fitted values. Median value of x labeled.

```
##
## Call:
## lm(formula = pf ~ inc + age + regat, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1701 -1.2637 -0.0147  1.0431  3.2278
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.17812    0.19763  11.021  < 2e-16 ***
## inc         -0.09048    0.02632  -3.438 0.000604 ***
## age          0.00669    0.00270   2.478 0.013355 *
## regat        1.96279    0.09709  20.216  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.6 on 1294 degrees of freedom
## Multiple R-squared:  0.2713, Adjusted R-squared:  0.2696
```

```
## F-statistic: 160.5 on 3 and 1294 DF,  p-value: < 2.2e-16
```

Figure 2: Regression Table. Test statistics included.