

FA2

Frances Aneth Rosales

2024-09-09

Creating Two-Way Anova by Gender and Education Level

```
data_polinterest$gender <- as.factor(data_polinterest$gender)
data_polinterest$education_level <- as.factor(data_polinterest$education_level)
anova_result <- aov(political_interest ~ gender * education_level, data = data_polinterest)

anova_table <- tidy(anova_result)

kable(anova_table, caption = "Two-Way ANOVA of Gender and Education Level")
```

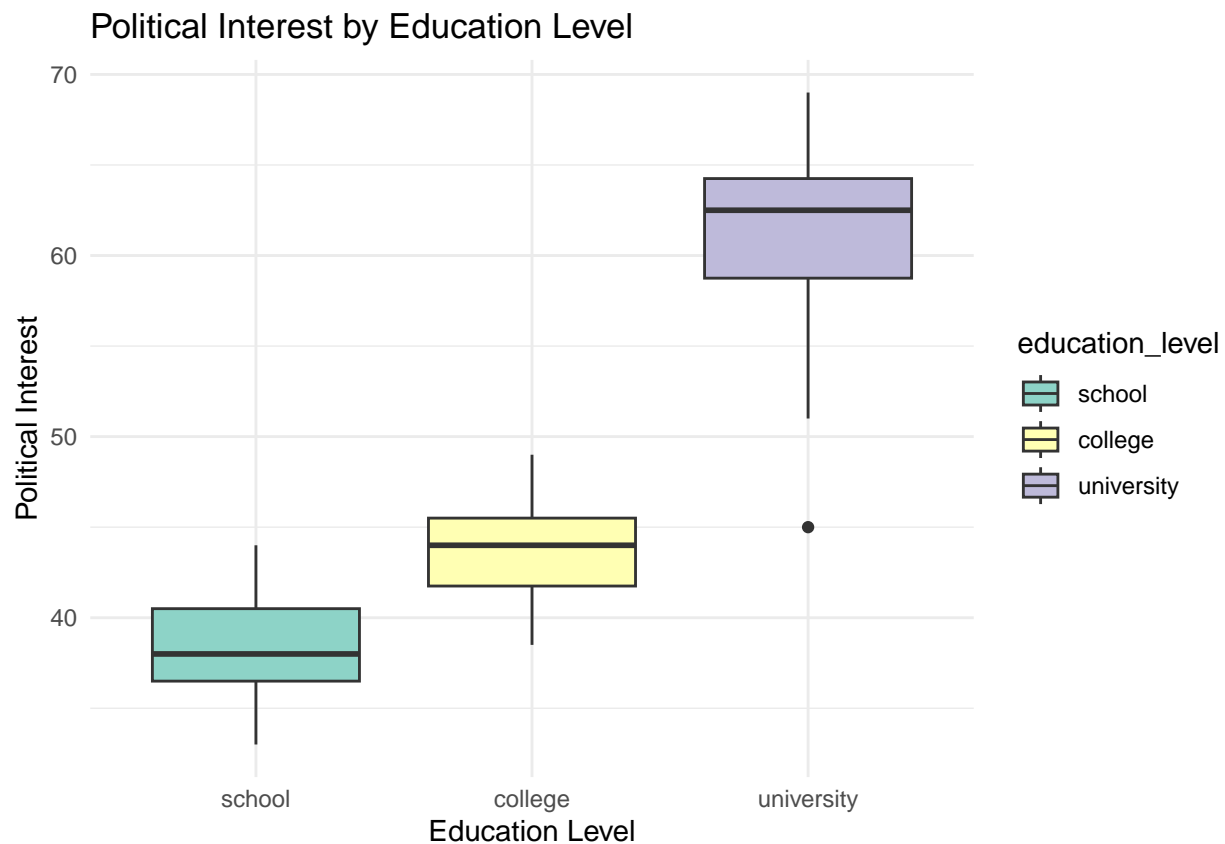
Table 1: Two-Way ANOVA of Gender and Education Level

| term | df | sumsq | meansq | statistic | p.value |
|------------------------|----|------------|------------|------------|-----------|
| gender | 1 | 25.70117 | 25.70117 | 1.787562 | 0.1870433 |
| education_level | 2 | 5409.95897 | 2704.97948 | 188.136131 | 0.0000000 |
| gender:education_level | 2 | 210.33766 | 105.16883 | 7.314679 | 0.0015877 |
| Residuals | 52 | 747.64444 | 14.37778 | NA | NA |

As shown, we can see a that the P.Value of our ANOVA is significant . Stated that the p.value of gender accumulates 0.1870433, and an approximately 0 in education_level. It is also visible that smaller residual mean square and sum of squares would suggest that a significant amount of the variability is explained by the model.

Two-Way Anova of Gender and Education Level

```
ggplot(data_polinterest, aes(x = education_level, y = political_interest, fill = education_level)) +  
  geom_boxplot() +  
  labs(  
    title = "Political Interest by Education Level",  
    x = "Education Level",  
    y = "Political Interest"  
  ) +  
  theme_minimal() +  
  scale_fill_brewer(palette = "Set3")
```



It was shown that the mean of the Political Interest accumulate the highest with Students in University Education level, this is followed by College, and then School. The University shows an Upper Quantile Boxplot in which its mean is above 60 with respect with the political Interest of the Students. College also shows an Upper Quantile in which on the other hand beyond 40. Lastly, School in which on the other hand has closely Lower Quantile scale below the mean 40.

Descriptives Statistics

```
library(dplyr)

descriptive_stats <- data_polinterest %>%
  group_by(gender, education_level) %>%
  summarise(
    N = n(),
    Mean = mean(political_interest, na.rm = TRUE),
    SD = sd(political_interest, na.rm = TRUE),
    SE = SD / sqrt(N),
    CV = SD / Mean
  )

## 'summarise()' has grouped output by 'gender'. You can override using the
## '.groups' argument.

#print(descriptive_stats)

kable(descriptive_stats, caption = "Descriptive Statistics by Gender and Education Level")
```

Table 2: Descriptive Statistics by Gender and Education Level

| gender | education_level | N | Mean | SD | SE | CV |
|--------|-----------------|----|----------|----------|-----------|-----------|
| male | school | 9 | 37.44444 | 2.505549 | 0.8351831 | 0.0669138 |
| male | college | 9 | 42.94444 | 2.337793 | 0.7792645 | 0.0544376 |
| male | university | 10 | 64.10000 | 3.071373 | 0.9712535 | 0.0479153 |
| female | school | 10 | 39.60000 | 3.272783 | 1.0349450 | 0.0826460 |
| female | college | 10 | 44.60000 | 3.272783 | 1.0349450 | 0.0733808 |
| female | university | 10 | 58.00000 | 6.463573 | 2.0439613 | 0.1114409 |

As also shown, our analysis can show that student in University accumulates the highest in its Mean equivalency. As such, Male-University accumulate 64.10 of mean interest in politics while Female-University shows 58.00 of Mean interest.

By compounding, we have shown that our assumption in above Box plot and Plotting correlate with data shown in our Descriptive Analysis.