FA2 Applied Multivariate Analysis

Bayquen, Christopher Gilbert A.

2024-09-09

Two-Way ANOVA (2B) | Political Interest : Gender and Educational Attainment

```
df = read.csv("Political Interest.csv")
df$gender <- factor(df$gender, levels = c(1, 2), labels = c("male",
    "female"))
df$education_level <- factor(df$education_level, levels = c(1,
    2, 3), labels = c("school", "college", "university"))

kable(head(df, 10), caption = " Political Interest : Gender and Educational Attainment")
```

Table 1: Political Interest : Gender and Educational Attainment

gender	education_level	political_interest
male	school	38.0
male	school	39.0
male	school	35.0
male	school	38.0
male	school	41.0
male	school	40.0
male	school	36.0
male	school	37.0
male	school	33.0
male	college	41.5

Checking for Assumptions

Assumption 1: You have one dependent variable that is measured at the continuous level

> Remark : The dependent variable, **Political Interest**, is measured at the continuous level

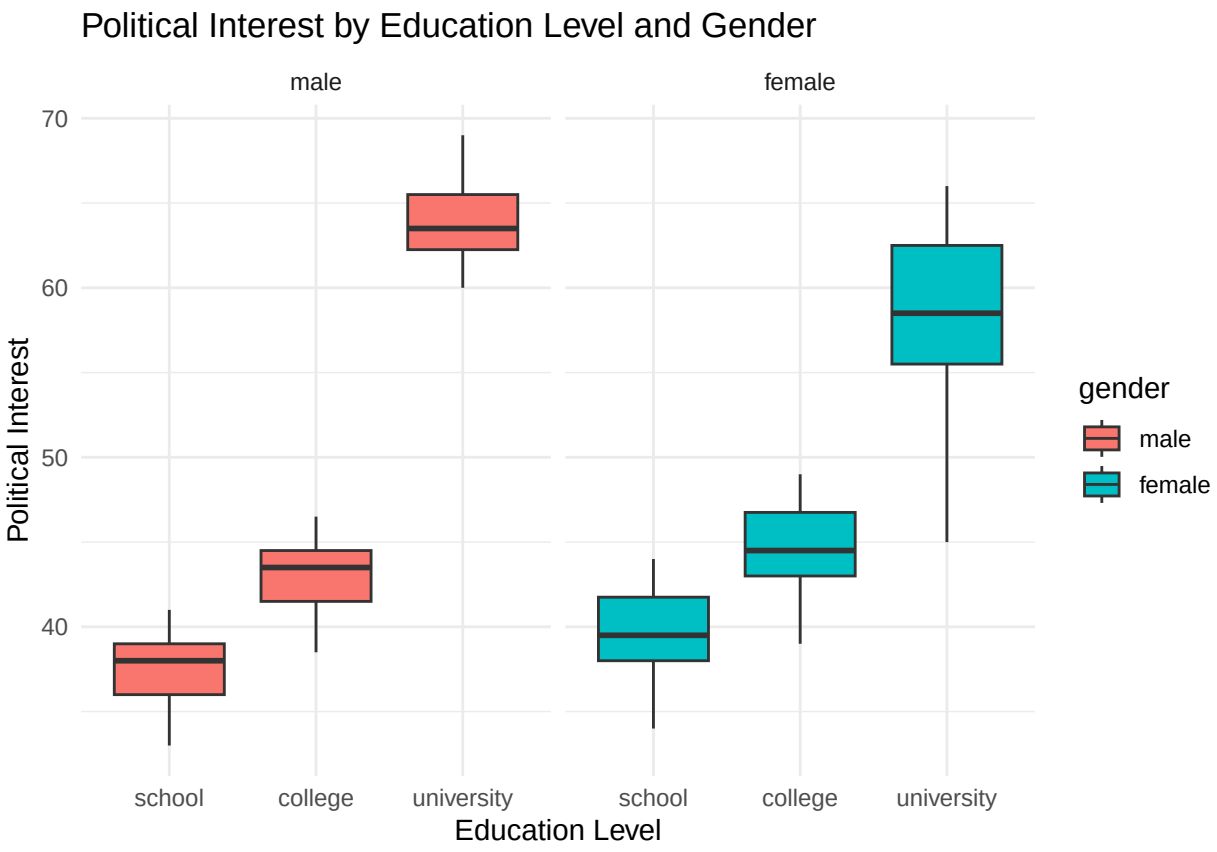
Assumption 2: You have two independent variables where each independent variable consists of two or more categorical, independent groups. An independent variable with only two groups is known as a dichotomous variable whereas an independent variable with three or more groups is referred to as a polytomous variable.

> Remark : We have two independent variables in this dataset, **gender** and **educational level**, where both of the independent variables are categorical. **Gender** contains only two groups (*male* and *female*), while **educational level** contains three groups (*school*, *college*, and *university*).

Assumption 3: You should have independence of observations, which means that there is no relationship between the observations in each group of the independent variable or between the groups themselves.

Remark : There is no relationship between the observations in each group of the independent variable or between the groups themselves.

Assumption 4: There should be no significant outliers in any cell of the design



Remark : As we can observe with the boxplots above, there are no significant outliers for each of the genders, under each of the educational levels.

Assumption 5: The distribution of the dependent variable (residuals) should be approximately normally distributed in every cell of the design.

Table 2: Shapiro-Wilk Test

gender	education_level	shapiro_p_value
male	school	0.9708070
male	college	0.7610941
male	university	0.3197307
female	school	0.8189494
female	college	0.8189494

gender	education_level	shapiro_p_value
female	university	0.6683785

Remark: The Political interest of each group is normally distributed as assessed by the Shapiro-Wilk's test, p > 0.05.

Assumption 6: The variance of the dependent variable (residuals) should be equal in every cell of the design.

```
leveneTest(political_interest ~ gender * education_level, data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  5  2.2054 0.06765 .
##       52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Remarks : There are homogeneity of variances in the dependent variables for all of the groups, as assessed by Levene's test of homogeneity of variances, p = 0.06765.

Implementing the Two-Way ANOVA

```r
# Fitting the two-way ANOVA model
anova_model <- aov(political_interest ~ gender * education_level,
    data = df)

summary(anova_model)
```

```
##                         Df Sum Sq Mean Sq F value  Pr(>F)
## gender                   1     26    25.7   1.788 0.18704
## education_level          2   5410  2705.0 188.136 < 2e-16 ***
## gender:education_level   2    210   105.2   7.315 0.00159 **
## Residuals               52    748    14.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Conduct Tukey's Honest Significant Difference test
TukeyHSD(anova_model)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = political_interest ~ gender * education_level, data = df)
##
## $gender
##                   diff       lwr      upr     p adj
## female-male -1.332143 -3.331504 0.667218 0.1870433
##
## $education_level
##                          diff      lwr       upr    p adj
## college-school       5.236842  2.26881  8.204874 0.000253
## university-school   22.435996 19.50530 25.366693 0.000000
## university-college  17.199154 14.26846 20.129851 0.000000
##
## $`gender:education_level`
##                                         diff          lwr       upr     p adj
## female:school-male:school           2.155556  -2.99895749  7.310069 0.8165366
## male:college-male:school            5.500000   0.21158120 10.788419 0.0370678
## female:college-male:school          7.155556   2.00104251 12.310069 0.0018818
## male:university-male:school        26.655556  21.50104251 31.810069 0.0000000
## female:university-male:school      20.555556  15.40104251 25.710069 0.0000000
## male:college-female:school          3.344444  -1.81006860  8.498957 0.4021433
## female:college-female:school        5.000000  -0.01703459 10.017035 0.0512608
## male:university-female:school      24.500000  19.48296541 29.517035 0.0000000
## female:university-female:school    18.400000  13.38296541 23.417035 0.0000000
## female:college-male:college         1.655556  -3.49895749  6.810069 0.9312003
## male:university-male:college       21.155556  16.00104251 26.310069 0.0000000
## female:university-male:college     15.055556   9.90104251 20.210069 0.0000000
## male:university-female:college     19.500000  14.48296541 24.517035 0.0000000
## female:university-female:college   13.400000   8.38296541 18.417035 0.0000000
## female:university-male:university -6.100000 -11.11703459 -1.082965 0.0088708
```

Analysis and Conclusion

A two-way ANOVA was conducted to examine the effects of gender and education level on political interest using the provided dataset. The model was specified as **political_interest ~ (gender)(education_level)**, and the summary of the ANOVA results indicated that education level had a significant effect on political interest `(F(2,52)=188.136, p < 2e-16)`, whereas the main effect of gender was not statistically significant `(F(1,52)=1.788, p=0.187)`. However, the interaction between gender and education level was significant `(F(2,52)=7.315, p=0.00159)`, suggesting that the effect of education level on political interest varies by gender.

Residual analysis confirmed that the assumptions of the ANOVA were met: there were no outliers, residuals were normally distributed, and there was homogeneity of variances with a p-value of p = 0.06765.

For females, political interest scores varied significantly with education level. The mean scores were 39.60 +- 3.27 for school-educated, 44.60 +- 3.27 for college-educated, and 58.00 +- 6.46 for university-educated females. For males, the mean scores were 37.44 +- 2.51 for school-educated, 42.94 +- 2.34 for college-educated, and 64.10 +- 3.07 for university-educated males. The pairwise comparisons revealed significant differences between education levels, with both school-educated and college-educated individuals showing notably lower political interest compared to their university-educated counterparts. Notably, there were significant differences between:

Female: College vs. Male: School

Male: University vs. Male: School

Female: University vs. Male: School

Various other education level and gender combinations.

In conclusion, while gender alone does not significantly influence political interest, education level has a strong impact, with higher levels of education correlating with increased political interest. The significant interaction indicates that the effect of education level on political interest is different for males and females, highlighting the need to consider both factors in understanding political engagement.