Afundar, Audrie Lex L. APM1111 - FA 9
Rodillas, Christian Miguel T.

2B ANOVA

Assumptions:

1. You have one dependent variable that is measured at the continuous level (i.e., the interval or ratio level).
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.
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.
4. There should be no significant outliers in any cell of the design.
5. The distribution of the dependent variable (residuals) should be approximately normally distributed in every cell of the design
6. The variance of the dependent variable (residuals) should be equal in every cell of the design.

Null and Alternative Hypotheses

Null hypothesis: There is no significant interaction effect on political interest between gender and education level.

Alternative Hypothesis: There is a significant interaction effect on political interest between gender and education level.

Dataset and Problem

This analysis utilizes Python to explore and investigate the connection of gender and education level to their respective political interests. We aim to determine if there are underlying statistical differences in political interest with each independent group (gender, education level) and if there is any interaction effect between the two groups.

Assumptions:

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

Remark. The political interest dataset has one dependent variable called political interest. The stated variable evaluates the interest of the respondents with regards to politics, dependent on other given variables, and is at a continuous level. This satisfies the assumption #1.

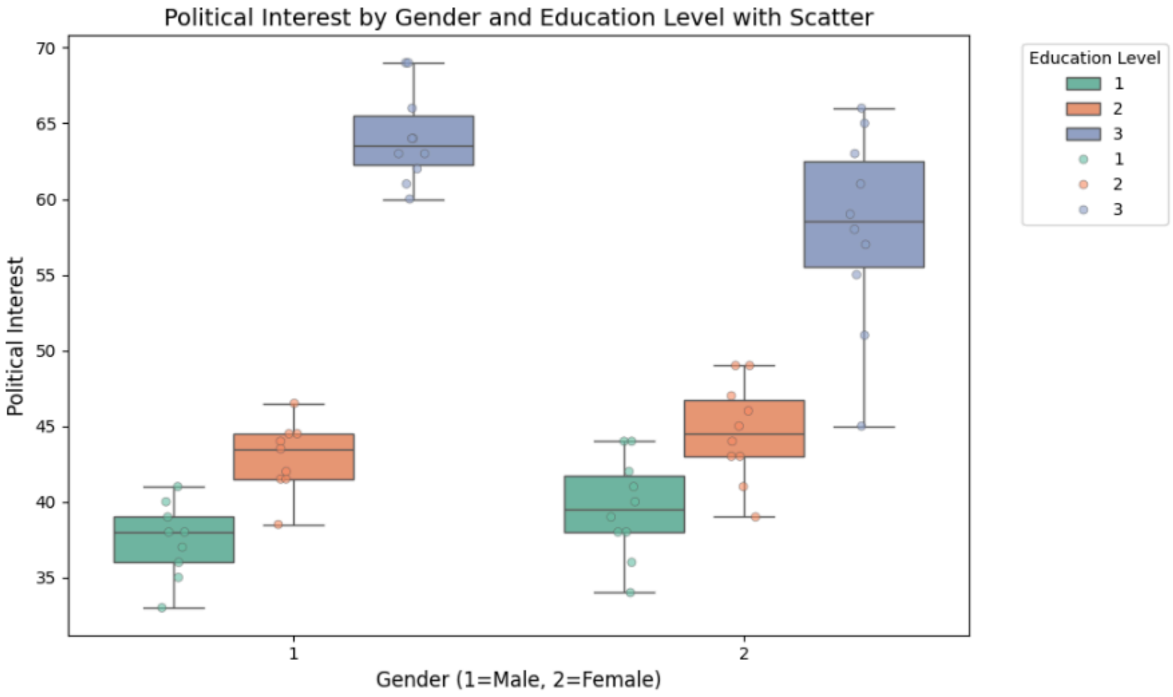
Assumption #2: You have two independent variables where each independent variable consists of two or more categorical, independent groups.

Remark. The dataset consists of two independent variables, named Gender, and Educational Level. The independent variable Gender consists of two independent groups, 1 and 2, making it a dichotomous variable. On the other hand, the independent variable Educational Level is a polytomous variable containing three independent groups 1, 2, and 3. These two independent variables are the basis of the dependent variable political interest. This satisfies the assumption #2.

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. For the dataset given, each observation represents a unique individual value for the two independent variables. There is no indication that each independent group correlates with each other. The two independent groups for gender provide specific values, and are recorded independently with each other. It is the same with the case of educational levels 1, 2, and 3. This satisfies the assumption of independence for 2-way ANOVA.

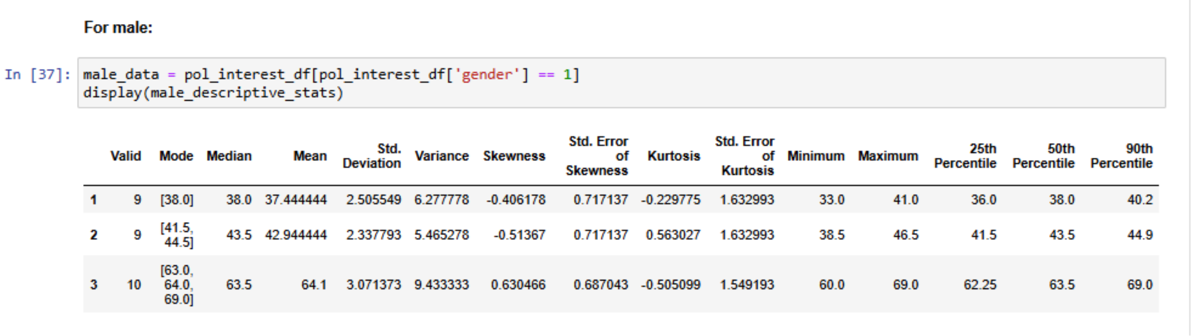
Assumption #4: There should be no significant outliers in any cell of the design.



Remark. Although it may be argued that the Female for level 3 has an outlier, there are no significant outliers between the 6 cells because all fall within the range and are valid.

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

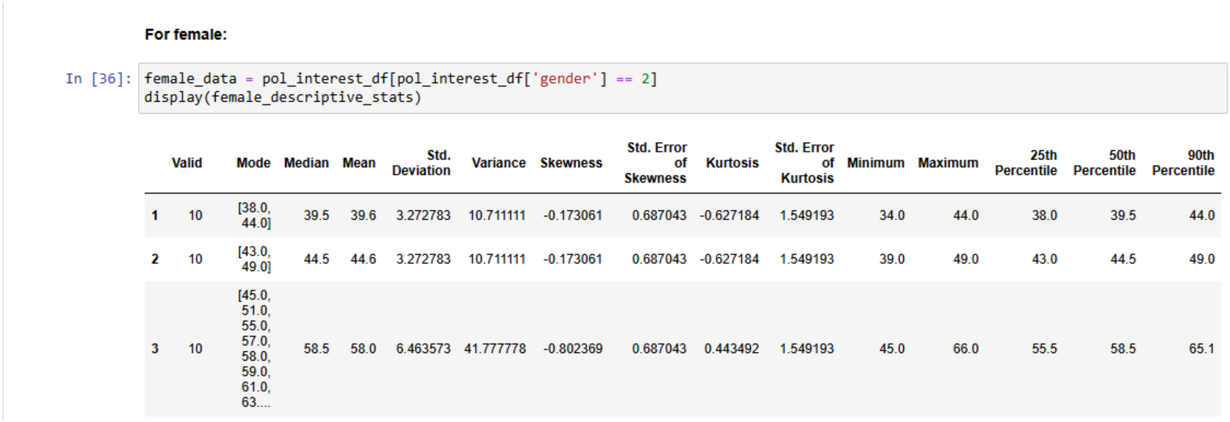
The Descriptive Statistics given are filtered by their gender:



For male:

```
In [37]: male_data = pol_interest_df[pol_interest_df['gender'] == 1]
         display(male_descriptive_stats)
```

	Valid	Mode	Median	Mean	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Minimum	Maximum	25th Percentile	50th Percentile	90th Percentile
1	9	[38.0]	38.0	37.444444	2.505549	6.277778	-0.406178	0.717137	-0.229775	1.632993	33.0	41.0	36.0	38.0	40.2
2	9	[41.5, 44.5]	43.5	42.944444	2.337793	5.465278	-0.51367	0.717137	0.563027	1.632993	38.5	46.5	41.5	43.5	44.9
3	10	[63.0, 64.0, 69.0]	63.5	64.1	3.071373	9.433333	0.630466	0.687043	-0.505099	1.549193	60.0	69.0	62.25	63.5	69.0



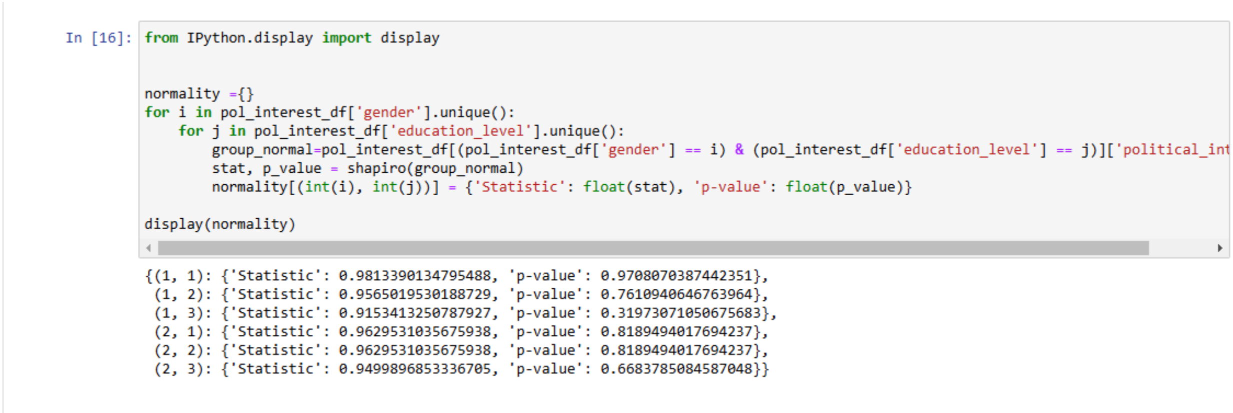
For female:

In [36]:

```
female_data = pol_interest_df[pol_interest_df['gender'] == 2]
display(female_descriptive_stats)
```

	Valid	Mode	Median	Mean	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Minimum	Maximum	25th Percentile	50th Percentile	90th Percentile
1	10	[38.0, 44.0]	39.5	39.6	3.272783	10.711111	-0.173061	0.687043	-0.627184	1.549193	34.0	44.0	38.0	39.5	44.0
2	10	[43.0, 49.0]	44.5	44.6	3.272783	10.711111	-0.173061	0.687043	-0.627184	1.549193	39.0	49.0	43.0	44.5	49.0
3	10	[45.0, 51.0, 55.0, 57.0, 58.0, 59.0, 61.0, 63....	58.5	58.0	6.463573	41.777778	-0.802369	0.687043	0.443492	1.549193	45.0	66.0	55.5	58.5	65.1

For the p-value:



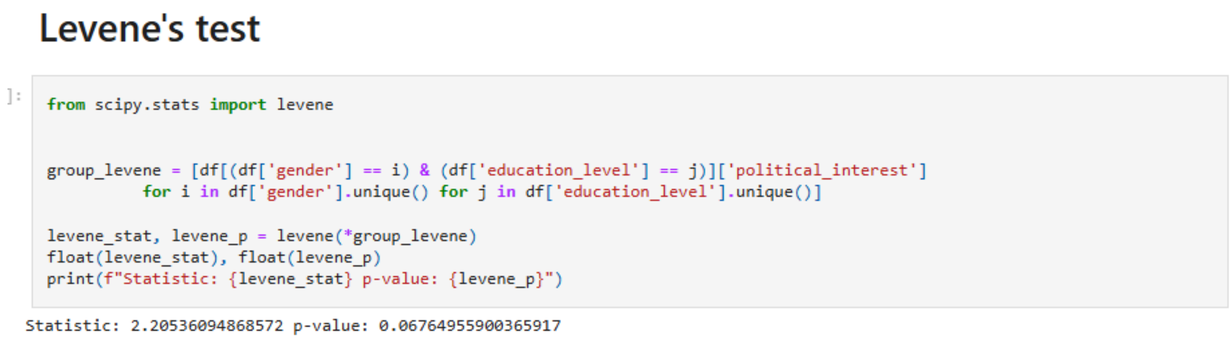
In [16]:

```
from IPython.display import display


normality ={}
for i in pol_interest_df['gender'].unique():
    for j in pol_interest_df['education_level'].unique():
        group_normal=pol_interest_df[(pol_interest_df['gender'] == i) & (pol_interest_df['education_level'] == j)]['political_int
        stat, p_value = shapiro(group_normal)
        normality[(int(i), int(j))] = {'Statistic': float(stat), 'p-value': float(p_value)}

display(normality)
```

```
{(1, 1): {'Statistic': 0.9813390134795488, 'p-value': 0.970807038744235 1},
 (1, 2): {'Statistic': 0.9565019530188729, 'p-value': 0.761094064676396 4},
 (1, 3): {'Statistic': 0.915341325078792 7, 'p-value': 0.31973071050675683},
 (2, 1): {'Statistic': 0.962953103567593 8, 'p-value': 0.8189494017694237},
 (2, 2): {'Statistic': 0.962953103567593 8, 'p-value': 0.8189494017694237},
 (2, 3): {'Statistic': 0.949989685333670 5, 'p-value': 0.6683785084587048}}
```

Remark. The dependent variable, political interest, is approximately normally distributed for each combination of their gender to the respective educational level. And as assessed by the Shapiro-Wilk test of normality, (p > 0.05), all fall within the prescribed basis of greater than 0.05.

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



Levene's test

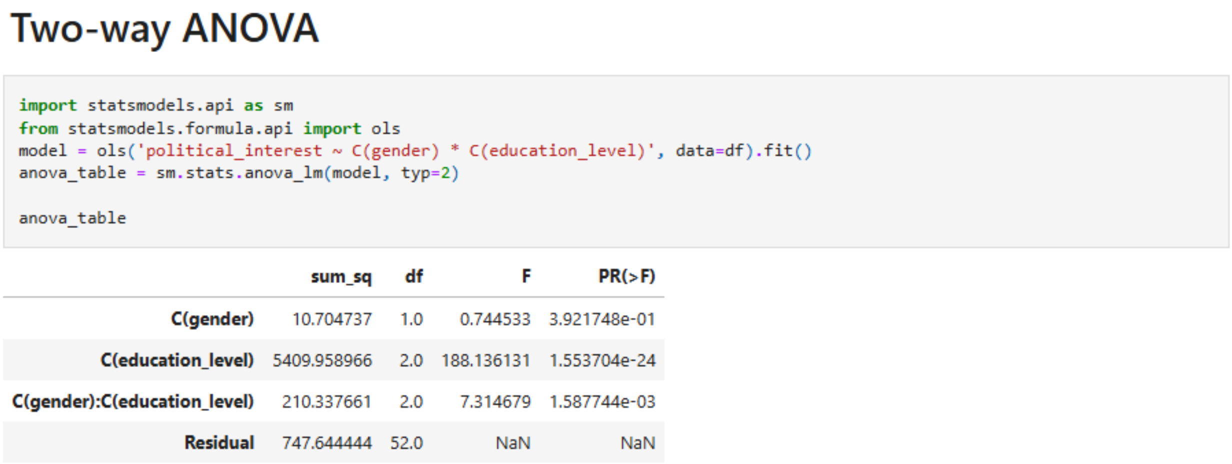
]:

```
from scipy.stats import levene


group_levene = [df[(df['gender'] == i) & (df['education_level'] == j)]['political_interest']
          for i in df['gender'].unique() for j in df['education_level'].unique()]

levene_stat, levene_p = levene(*group_levene)
float(levene_stat), float(levene_p)
print(f"Statistic: {levene_stat} p-value: {levene_p}")
```

```
Statistic: 2.20536094868572 p-value: 0.06764955900365917
```

Remark. As assessed by Levene's test of equality of variances, p = 0.067, then the variances for each combination of the education level and gender are homogenous.

Computation:



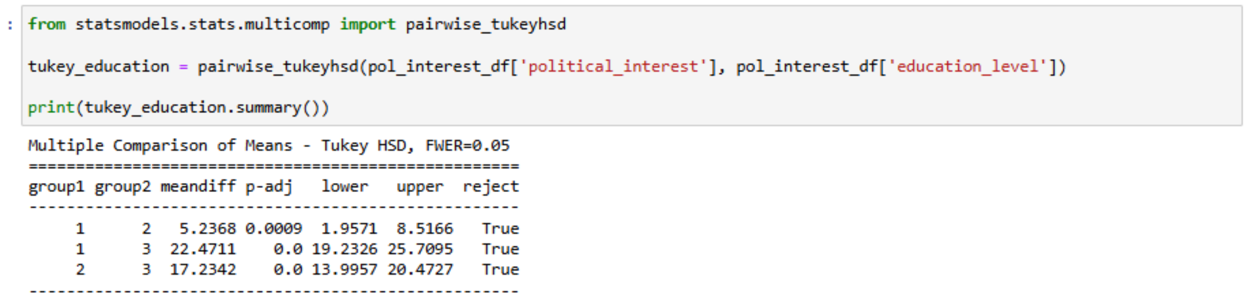
Two-way ANOVA

```
import statsmodels.api as sm
from statsmodels.formula.api import ols
model = ols('political_interest ~ C(gender) * C(education_level)', data=df).fit()
anova_table = sm.stats.anova_lm(model, typ=2)

anova_table
```

	sum_sq	df	F	PR(>F)
C(gender)	10.704737	1.0	0.744533	3.921748e-01
C(education_level)	5409.958966	2.0	188.136131	1.553704e-24
C(gender):C(education_level)	210.337661	2.0	7.314679	1.587744e-03
Residual	747.644444	52.0	NaN	NaN

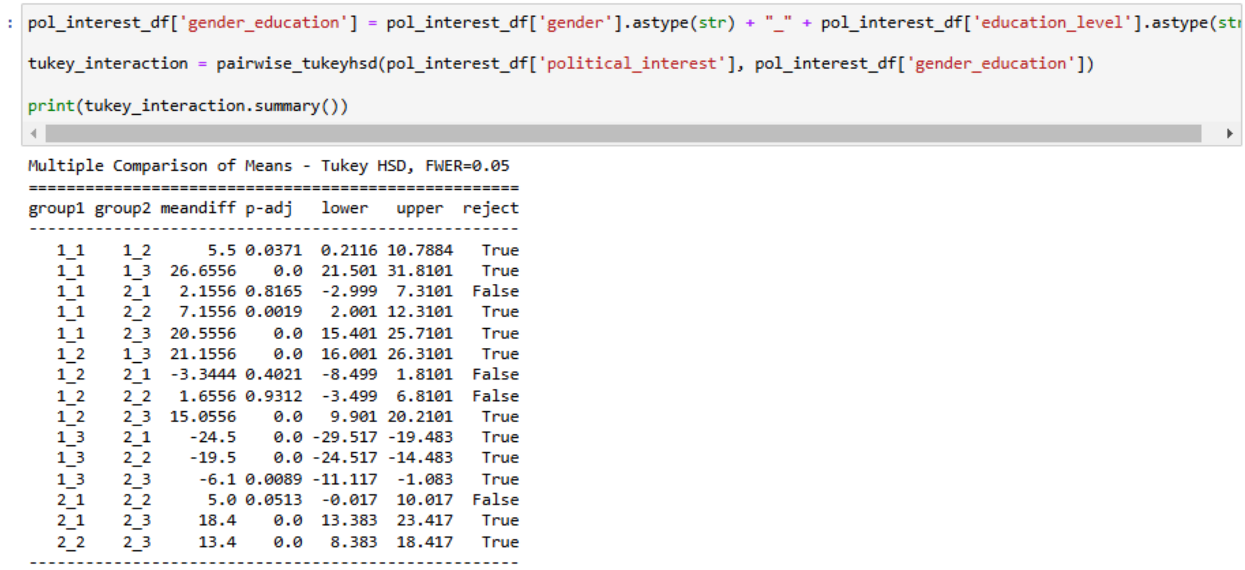
Remark. Upon inspection, there exists a statistically significant interaction between gender with their respective education level on their interest in politics. In the values, F = 7.31, PR(>F) = 1.588×10^{-3}, p = 0.002, indicating that the effect of education level on political interest depends on the year level. Therefore, the main effects of each independent variable were not reported, as they would be biased by this interaction. Instead, an analysis of simple main effects was conducted with statistical significance assessed at the p < .025 level using a Bonferroni adjustment.

POST HOC



```
: from statsmodels.stats.multicomp import pairwise_tukeyhsd

  tukey_education = pairwise_tukeyhsd(pol_interest_df['political_interest'], pol_interest_df['education_level'])

  print(tukey_education.summary())
```

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2 meandiff p-adj   lower   upper  reject
-----------------------------------------------------
     1      2   5.2368 0.0009  1.9571  8.5166   True
     1      3  22.4711    0.0 19.2326 25.7095   True
     2      3  17.2342    0.0 13.9957 20.4727   True
-----------------------------------------------------
```



```
: pol_interest_df['gender_education'] = pol_interest_df['gender'].astype(str) + "_" + pol_interest_df['education_level'].astype(st

  tukey_interaction = pairwise_tukeyhsd(pol_interest_df['political_interest'], pol_interest_df['gender_education'])

  print(tukey_interaction.summary())
```

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2 meandiff p-adj   lower   upper  reject
-----------------------------------------------------
   1_1    1_2      5.5 0.0371  0.2116 10.7884   True
   1_1    1_3  26.6556    0.0  21.501 31.8101   True
   1_1    2_1   2.1556 0.8165  -2.999  7.3101  False
   1_1    2_2   7.1556 0.0019   2.001 12.3101   True
   1_1    2_3  20.5556    0.0  15.401 25.7101   True
   1_2    1_3  21.1556    0.0  16.001 26.3101   True
   1_2    2_1  -3.3444 0.4021  -8.499  1.8101  False
   1_2    2_2   1.6556 0.9312  -3.499  6.8101  False
   1_2    2_3  15.0556    0.0   9.901 20.2101   True
   1_3    2_1    -24.5    0.0 -29.517 -19.483   True
   1_3    2_2    -19.5    0.0 -24.517 -14.483   True
   1_3    2_3     -6.1 0.0089 -11.117  -1.083   True
   2_1    2_2      5.0 0.0513  -0.017  10.017  False
   2_1    2_3     18.4    0.0  13.383  23.417   True
   2_2    2_3     13.4    0.0   8.383  18.417   True
-----------------------------------------------------
```

Note: The groups 1 and 2 display a partnered number (n_m). The n = gender (1,2) and the m = education level (1,2,3).

Remark. Analysis on Post hoc were conducted to examine pairwise comparisons between the levels of education and the apparent interaction with gender.All pairwise comparisons were run for each simple main effect with reported 95% confidence intervals and p-values Tukey-adjusted within each simple main effect. The differences in the interest in politics are apparent in comparing the following: male - level 1 to male - level 3, male - level 2 to male - level 3, and male - level 3 to female - level 1 (with differences at 26.7, 21.2, and 24.5 respectively). Upon inspection, the political interest directly increases along with the year level. Similarly, individuals with education level 2 had significantly lower political interest scores compared to those with university education 3 (mean difference = 17.2342 [95% CI, 13.9957 to 20.4727], p < .0005).

Surprisingly, male - level 3 shows a higher interest in politics in comparison to female - level 3 with 6.1 difference, which could be counter-intuitive since both female levels 1 and 2 displays higher interest than male levels 1 and 2 (differences at 2. 16 and 1.6). On the other hand, for school-educated individuals (1_1 vs. 2_1), there was no significant difference between genders (mean difference = 2.1556 [95% CI, -2.999 to 7.3101], p = .8165).

In conclusion, The provided results shows that the effect of education plays a role on political interest whilst being moderated by their gender, with university-educated males exhibiting the highest scores.