Mid-term Assignment

B149008

28th February 2024

1 The Data

The data in this study was obtained from the 2002 wave of the European Social Survey (ESS) [3]. The responses were nested within 20 European countries. This report aims to investigate the variation in Universalism scores across different countries, along with how the inclusion of the following explanatory variables may impact the variation of Universalism scores among all the countries:

- `age`: The age of an individual;
- `gender`: The sex of an individual; and
- `eduyrs`: The number of years of education the individual has received.

The dataset contains 36,537 rows. There are 7,499 missing values in the dataset, 6,787 of which are in the column containing information about respondents' incomes. Since income and the scores for values except Universalism are not relevant to this report, they can be excluded from any analysis. As shown in Table 1, since there are not many missing values among the remaining variables, we may omit them from the dataset.

Column Name	**No. of Missing Values**
`country`	0
`gender`	31
`age`	164
`eduyrs`	445
`universalism`	0
Total	**640**

Table 1: No. of missing values in the relevant columns of the dataset.

2 Variation Across Countries

To investigate variation in the Universalism scores across all countries in the data, a Variance Component Model (VCM) was implemented. The universalism score was the response variable and only the variation in the random intercept for each country was considered. The model's estimate for the fixed effect is shown in Table 2. The estimate's large t-value suggests that it is statistically significant, indicating that the model is statistically accurate. The variation of the model's random effects is given in Table 3. The variance at level 1 is 40.287, while the variance at level 2 (between countries) is 1.315. To see what proportion of the variance in universalism scores is associated with country level, we have

$$\% \text{ variance} = \frac{\text{variance at level 2}}{\text{variance at level 2} + \text{variance at level 1}} \times 100 = \frac{1.315}{1.315 + 40.287} \times 100 = \underline{3.16\%}.$$

Term	Estimate	Std. Error	t-value
Intercept	5.8596	0.2587	22.66

Table 2: Summary of the fixed term in the VCM.

Groups	Name	Variance	Std. Dev.
Country	(Intercept)	1.315	1.147
Residual		40.287	6.347

Table 3: Variation in random effects of the VCM fitted across countries.

Furthermore, as shown in Figure 1, countries such as Israel and Greece have the lowest universalism scores on average while France and Switzerland have the highest. This is expected, as at the time of the data collection, Israel had a majorly conservative government. Greece was also struggling with a large inflow of illegal immigrants at the time, possibly causing public opinions to be against those who were different to the native citizens. In addition to this, due to their relative closeness with organisations such as the UN, Western and Central European countries such as France and Switzerland were likely to hold more environmentally conscious and progressive views.

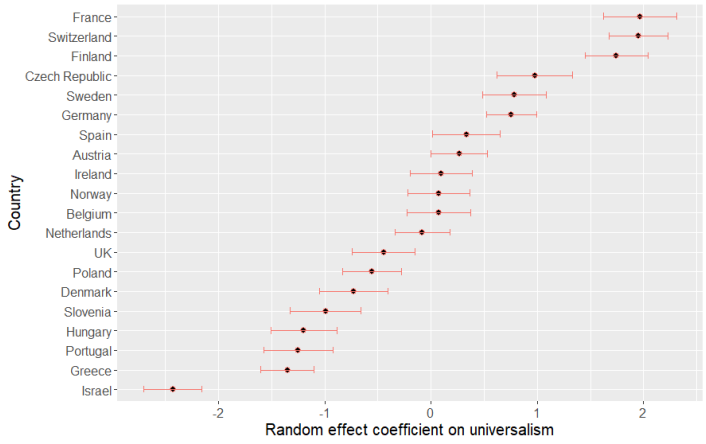

Figure 1: Dotplot illustrating the variance in Universalism at a country-by-country level.

3 Universalism Explained by Other Variables

3.1 Model 1

To study how an individual's Universalism score may be impacted by other variables, we may fit a Random Intercept Model (RIM). The universalism score is the response variable in this model and `age`, `gender`, and `eduyrs` are explanatory variables. The variation between each of the countries is also considered. We may name this model RIM1.

3.1.1 Preliminary hypotheses

- Women are generally likely to be more empathetic than men, possibly causing their universalism scores to be higher [1].
- Older individuals may be more conservative and less socially tolerant than younger individuals, causing their universalism scores to be relatively lower [2].

- It is unlikely that an individual's level of education will have a large impact on their universalism score, though it may still be positive.

3.1.2 Variable transformation

Before implementing the model, it is necessary to explore descriptive statistics for explanatory variables and transform them accordingly. There were 16,909 male respondents and 19,027 female respondents in the dataset. Statistics for `age` and `eduyrs` are shown in Table 4. The age ranges from 14 to 102 and the range for `eduyrs` is 0 to 40. To obtain a meaningful intercept while modelling, we must centre these variables around their respective means. Statistics for the transformed variables are shown in Table 5.

Statistic \ Variable	`age`	`eduyrs`
Minimum value	14	0
1st Quartile	32	9
Median	45	12
Mean	46.1	11.92
3rd Quartile	60	14
Maximum value	102	40

Table 4: Descriptive statistics for the `age` and `eduyrs` variables.

Statistic \ Centred Variable	`age`	`eduyrs`
Minimum value	-32.09609	-11.917826
1st Quartile	-14.09609	-2.917826
Median	-1.09609	0.082174
Mean	0	0
3rd Quartile	13.90391	2.082174
Maximum value	55.90391	28.082174

Table 5: Descriptive statistics for centred `age` and `eduyrs` variables.

3.1.3 Implementation and results

The resulting random effects for RIM1 are shown in Table 6. We see the value of the intercept is not too different from that of the VCM, shown in Table 3. However, it is worth noting that the variation in RIM1 was smaller than in the VCM.

The model's estimates are shown in Table 7. The respondents' age does not have a large impact on their universalism score. A unit increase in age corresponds to an increase of only 0.063 in the universalism score. Hence, though the association is not large, it is positive as hypothesised. The estimate for the `eduyrs` variable is positive but not particularly large, as previously hypothesised. The association is found to be 0.201. Finally, an individual being female is found to have a positive association with their universalism score, with women having an individualism score 1.614 times higher than men.

Additionally, the t-values for all the estimates shown in Table 7 are also statistically significant. Hence, it is reasonable to conclude that each of the relationships described above is statistically significant.

In addition to the numerical summaries obtained from RIM1, we may also inspect the dotplot shown in Figure 2. As with the numerical summaries, the dotplots for the VCM and RIM1 are not

Groups	Name	Variance	Std. Dev.
Country	(Intercept)	1.26	1.122
Residual		38.32	6.190

Table 6: Variance/standard deviation table for the random effects in RIM1.

Term	Estimate	Std. Error	t-value
Intercept	5.010115	0.255547	19.61
`age`	0.063021	0.001903	33.12
`gender` (female)	1.614377	0.065628	24.60
`eduyrs`	0.201241	0.009050	22.24

Table 7: Summary of the fixed terms in RIM1.

very different. The extreme values in Figure 2 are almost the same as those in Figure 1, while the more central values are closer to the average at 0 than in Figure 1.

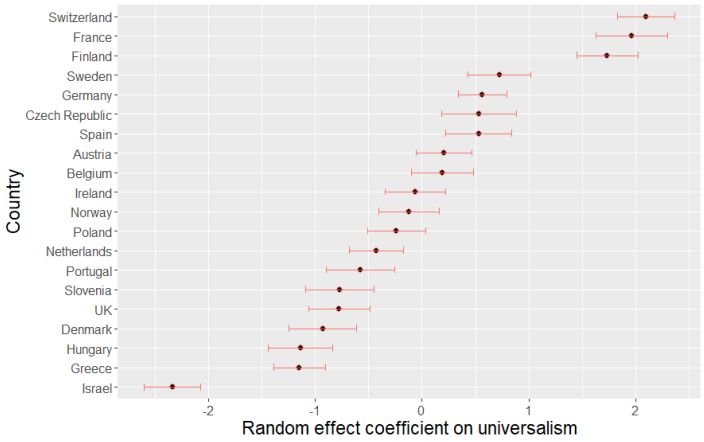

Figure 2: Dotplot showing the variance across countries in RIM1.

Finally, we may perform an ANOVA between RIM1 and the VCM to test whether controlling for age, gender and education improves the fit. The ANOVA results are shown in Table 8. The AIC and BIC values for RIM1 are lower than those for the VCM. Furthermore, the p-value for the likelihood test being almost trivial suggests that the model fits are significantly different from one another, with RIM1 performing better.

Model	npar	AIC	BIC	logLik	deviance	Chisq	D.f.	Pr($>$ Chisq)
VCM	3	234890	234915	-117442	234884			
RIM1	6	233090	233141	-116539	233078	1805.4	3	$< 2.2 \times 10^{-16}$ ***

Table 8: Results of ANOVA between VCM and RIM1.

3.2 Model 2

The second model is similar to RIM1 but with the inclusion of gender in the random effects term. In essence, this model includes a random slope for gender, allowing us to investigate whether the relationship between gender and universalism varies across countries. We may refer to the model as RIM2.

The resulting table of variances is given in Table 9. There are two values of variance between countries, one to represent the variation in the random intercept and one to represent the variation in the relationship between gender and universalism across countries (random slope). There is a positive correlation of 0.14 between the intercept and the slope, suggesting that for higher values of the intercept, the slope is also steeper.

Groups	**Name**	**Variance**	**Std. Dev.**	**Correlation**
Country	(Intercept)	1.066	1.0324	
	Gender (female)	0.361	0.6008	0.14
Residual		38.229	6.1830	

Table 9: Variance table for the random effects in RIM2.

The estimates and their metrics for RIM2 are shown in Table 10. These are not very different from those of RIM1. The estimates for the explanatory variables are slightly smaller, while the intercept is slightly higher. While the t-values for RIM2 are also slightly smaller than for RIM1, they are still statistically significant.

In addition to this, we can inspect the dotplot produced by this model, shown in Figure 3. The plot on the left, representing the variation in the intercept, is very similar to the plots shown for the VCM and RIM1. On the other hand, the plot on the right, representing the variation among the slopes, does not exhibit any obvious patterns or trends. However, countries such as Finland and Germany seem to have slightly higher values of universalism. This may be attributed to them having less conservatism in their societies, causing them to be more accepting and tolerant of others who are different from them.

Term	**Estimate**	**Std. Error**	t**-value**
Intercept	5.030624	0.235893	21.33
`age`	0.062944	0.001901	33.11
`gender` (female)	1.579335	0.150000	10.53
`eduyrs`	0.199154	0.009052	22.00

Table 10: Summary of the fixed terms in RIM1.

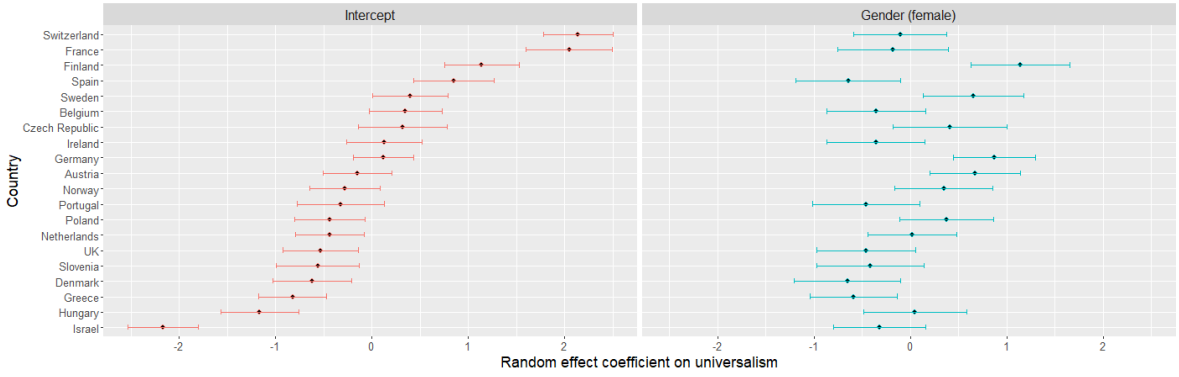

Figure 3: Dotplot showing the variance in the random intercepts across countries and the random slopes across countries in RIM2, also grouped by gender.

3.3 Model 3

A third model is implemented to investigate whether the relationship between age and universalism varies across countries. This means fitting a model similar to RIM2 but with the random effects term including the age variable instead of the gender variable. We may refer to this model as RIM3.

Since age is a continuous variable, it is important to convert it to a categorical variable by binning all the values into different groups. These are defined in Table 11.

Age Group	Age Range (years)
1	14-25
2	26-35
3	36-50
4	51-65
5	66 and over

Table 11: Age groups that the `age` variable is binned into.

The variation among the random effects in RIM3 is shown in Table 12. There is relatively little variation in the random effects of the younger age groups (groups 2 and 3), while the variation among the older age groups (groups 4 and 5) is much higher. Furthermore, the random effect of each of the age groups has a large negative correlation with the random intercept. So, as the random intercept increases, it is likely that the random slopes for each of the age groups decrease across the countries. This results in the correlation between the random slopes for each of the age groups being largely positive.

Groups	**Name**	**Variance**	**Std. Dev.**	**Correlation**		
Country	(Intercept)	2.2111	1.4870			
	Age group 2	0.2091	0.4573	-0.71		
	Age group 3	0.2091	0.4573	-0.51	0.97	
	Age group 4	1.5267	1.2356	-0.67	0.97	0.94
	Age group 5	1.8279	1.3520	-0.76	0.94	0.88
Residual		37.9950	6.1640			

Table 12: Variance table for the random effects in RIM2.

The estimates for each of the variables in RIM3 are given in Table 13. All variables have a positive association with the universalism scores and they are all statistically significant. The older age groups, specifically groups 4 and 5, have the largest association with universalism scores, with individuals from age group 5 likely to have a universalism score over 3 times higher than individuals from age group 1. The estimates for the remaining variables, i.e. gender and education, are very similar to those calculated for RIM1 and RIM2.

Term	**Estimate**	**Std. Error**	***t*-value**
Intercept	2.891506	0.345265	8.375
`age` group 2 (26-35)	0.987667	0.155590	6.348
`age` group 3 (36-50)	2.381364	0.248169	9.596
`age` group 4 (51-65)	2.974765	0.297854	9.987
`age` group 5 (66 and over)	3.414944	0.325645	10.487
`gender` (female)	1.634598	0.065442	24.978
`eduyrs`	0.179510	0.009316	19.268

Table 13: Summary of the fixed terms in RIM3.

In addition to the tables, we may inspect the dotplot for RIM3, shown in Figure 4. There is no discernible pattern in age group 2, the youngest age group in the model. This is expected, as the values are calculated in comparison to the individuals in age group 1, who are very close in age to individuals in group 2, but much younger than those in groups 3, 4, and 5.

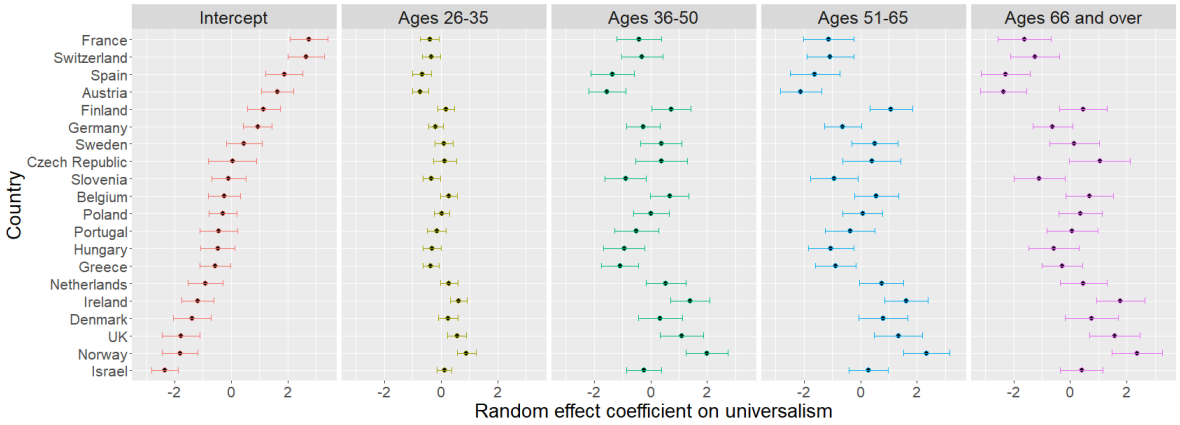

Figure 4: Dotplot showing the variance in the random intercepts across countries and the random slopes across countries in RIM3, also grouped by age brackets.

References

[1] Leonardo Christov-Moore et al. 'Empathy: Gender effects in brain and behavior'. In: *Neuroscience & Biobehavioral Reviews* 46.4 (Oct. 2014), pp. 604–627. DOI: https://doi.org/10.1016/j.neubiorev.2014.09.001. URL: https://www.sciencedirect.com/science/article/pii/S0149763414002164.

[2] Jan G. Janmaat and Avril Keating. 'Are today's youth more tolerant? Trends in tolerance among young people in Britain'. In: *Ethnicities* 19.1 (Aug. 2017), pp. 44–65. DOI: https://doi.org/10.1177/1468796817723682. URL: https://journals.sagepub.com/doi/pdf/10.1177/1468796817723682 (visited on 26/02/2024).

[3] Norwegian Science Data Services. *ESS Round 1: European Social Survey Round 1 Data*. 2002. URL: http://www.europeansocialsurvey.org/data (visited on 26/02/2024).