Indian Statistical Institute Bangalore

Bachelor of Mathematics

Introduction to Statistics and Computation with Data

Mapping Discrimination in Europe

by
Aishik Sarkar
Anit Das
Gandharv Sawant
Tejas Varma

April 2023

Abstract

As more people immigrate from different nations, societies are becoming more diversified. However, immigrants and members of ethnic minorities frequently experience discrimination in the form of fewer housing and employment options as well as restrictions on relationships in other spheres of society. 24,915 amateur football teams in 23 European nations were contacted using fictitious email addresses with names that had both typical native and foreign sounding names, asking them for an opportunity take part in a practise session. Response rates varied by nation and were, on average, 10% lower for names that were foreign. The current field experiment uncovers bias against ethnic minority groups and reveals structural flaws in a system that is supposed to promote social contact.

Contents

1 Introduction

All modern societies host diverse ethnic and cultural groups in their population. However this doesn't mean that the society is tolerant of the diversity. Social science research examines ethnic and racial discrimination through field experiments. Some examples where discrimination is not directly observed are finding a job, renting an apartment and using the sharing economy. Research shows that the minority group of a society often face discrimination when accessing labour, housing, transportation, politics, and the sharing economy. Most discrimination studies focus on discrimination that has an economic impact but there is evidence of ethnic and racial discrimination in many other activities. We may not notice it but such discrimination has unfortunately turned into a standard-issue in our daily lives. Participation in social activities and gatherings can provide meaningful interactions which may be essential for improvements in the society. However, it requires reciprocity, i.e., the native members must willingly accept the foreign members and also the foreign members must willingly attend such social activities with the native members.

Amateur sports are often considered to be a social activity where people of different ethnic and cultural groups can interact with one-another. However, there are numerous papers which show that ethnic minorities are less likely to participate in sports than the majority. This under representation is more often than not explained with the fact that ethnic minorities prefer to have social activities with familiar people or people of their ethnicity. Nevertheless, researchers have speculated on the possibility of ethnic discrimination within the club-organized sport system. The major problem seems to be the structure of the system of recruitment of new players is not well established and most of the responsibility of inviting new players fall on other players or the coach of the club. A similar system exists in numerous European countries. Football is a sport that has the same rules for everyone so the ethnicity, culture or race of a person must not interfere with the opportunity one has.

The use of qualitative methods and case studies prevents this research from disentangling practices of exclusion from ethnic preference. Data on daily interaction between groups mostly relies on information reported by the members regarding the number of mutual visits but this may overlook the actions of the native members leading to a possible response bias since the data is self-reported. The use of a field experiment seems to be most appropriate to measure ethnic discrimination and focus on the actions of the natives as it avoids the major drawback of self-reporting. An added benefit of using a field experiment is that it eliminated the influence of other factors such as appearance, attitude and accent.

The multi-country study done in Europe provides a picture of discrimination however, it fails to provide a more refined analysis of the mechanisms that play a role in ethnic discrimination. These include mechanisms such as cultural differences and religion. Nonetheless, the research provides us with a solid foundation for future research.

2 Method

Data was gathered only for amateur leagues. The data was collected by starting from the lowest league and then moving up the rankings for each country until 1,500 observations were recorded. However, the intended number of observations weren't reached for every country due to limited number of amateur football clubs. Each amateur club was only contacted once for this research. No respondent was contacted several times, as respondents would modify their behaviour. Some clubs also use social media to receive applications. Clubs that work exclusively with social networks were not contacted for two reasons. First, the guidelines of the ethical commission of the University of Zurich were followed and hence creating fictitious friends were avoided. A profile with no friends is unconventional, suspicious, and, therefore, likely to bias the results of the experiment. Second, some of the clubs use their social network pages exclusively to communicate with players who are already members and do not allow inquiries.

First names were generated for each group in each country: 5 native-sounding and 6 foreign-sounding names (two names for each of the 3 largest foreign groups in each country, except for Switzerland which included more than 3 foreign groups). The largest foreign groups were chosen as they represent the largest share of the population and will represent discrimination against ethnic minority groups that represent a sizable part of the population. Block randomization by the state for each country was used to select the club to contact. One name was assigned to each club contacted such that native and foreign names were equally assigned. Using a random number generator, 330 seven-digit G-mail accounts were created for sending the email. In some instances the same account was used for another country after the name of the alias was changed. This was done only after the previous country on which the account was used to analyse was finished.

When the clubs were contacted, the emails only expressed interest in joining the club for a trial training session. The emails were sent in the language of the homepage of the club. After sending the emails, each account was checked daily in the first week and at least twice a week in the following 3 weeks. No checking continued after 4 weeks. Most of the responses were received after 1 or 2 days. Responses were answered promptly, the fictitious player told the respondents that he is no longer interested in the trial training session and thanked the club for their response. The responses were categorized as follows:

- No response or rejection
- Positive response
- Positive response after further inquiry

No responses and outright rejections were clubbed together as only a small percentage of responses were outright rejections. For some countries (like Hungary, England, Portugal, Romania and Switzerland), not enough data was available to categorize the responses into the 3 categories mentioned above and hence all responses were simplified to just positive or negative.

3 Data

A total of $24,915$ amateur football clubs were contacted from 23 different countries: Austria ($n = 1,840$), Belgium ($n = 663$), Croatia ($n = 447$), Czech Republic ($n = 1,598$), Denmark ($n = 1,135$), England ($n = 1,527$), Finland ($n = 536$), France ($n = 1,847$), Germany ($n = 1,681$), Greece ($n = 437$), Hungary ($n = 798$), Ireland ($n = 308$), Italy ($n = 1,463$), the Netherlands ($n = 715$), Norway ($n = 1,000$), Poland ($n = 1,312$), Portugal ($n = 791$), Romania ($n = 493$), Russia ($n = 1,143$), Serbia ($n = 383$), Spain ($n = 1,410$), Sweden ($n = 1,493$), and Switzerland ($n = 1,895$).

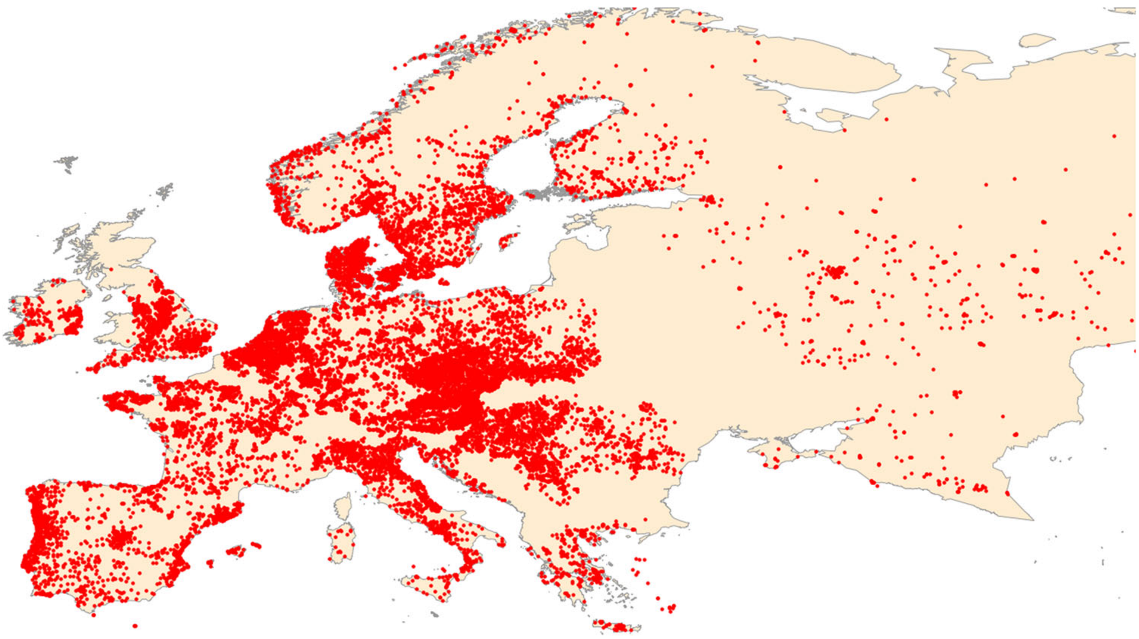

Figure 1: Football clubs contacted

Of the $24,915$ responses, $11,300$ of them were positive and and $13,615$ of them were negative. For $20,930$ of the data entries, the responses were categorized into (1) negative, (2) positive, and (3) positive with further inquires (The data isn't available for Hungary, England, Portugal, Romania and some responses from Switzerland). Of this, $6,194$ of them were positive, $10,901$ were negative and $3,835$ were positive with further inquiries.

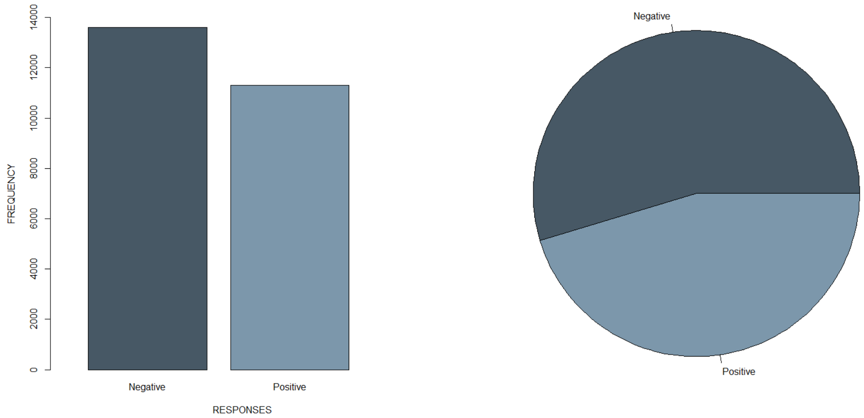

Figure 2: Bar plot and pie chart of responses

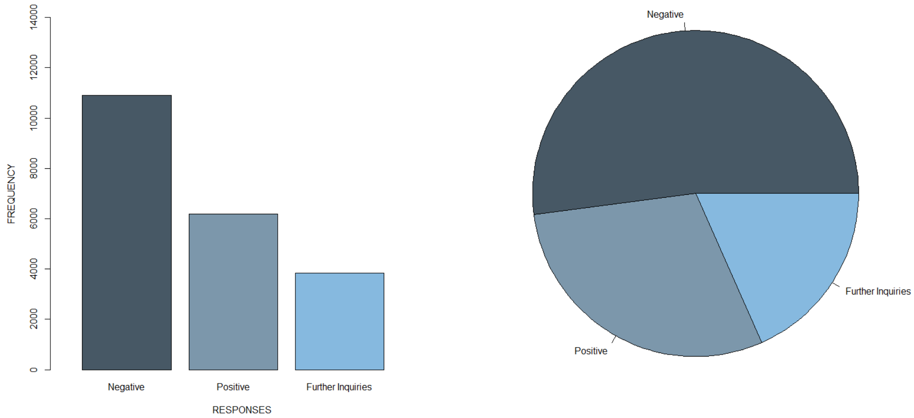

Figure 3: Bar plot and pie chart of responses (limited sample)

On taking a look at the responses to only native applicants, $6,471$ of the $12,787$ turned out to be positive while the remaining $6,316$ were negative. For the limited sample (of size $10,789$), $3,503$ were positive, $5,019$ were negative and $2,267$ were positive with further inquires.

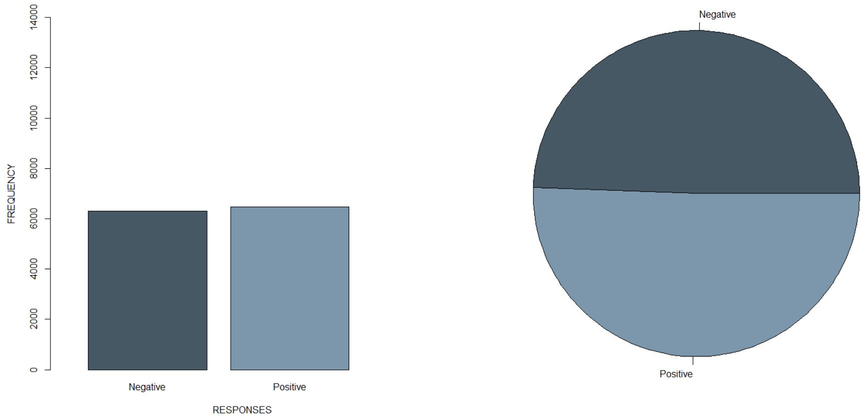

Figure 4: Bar plot and pie chart of responses to native applicants

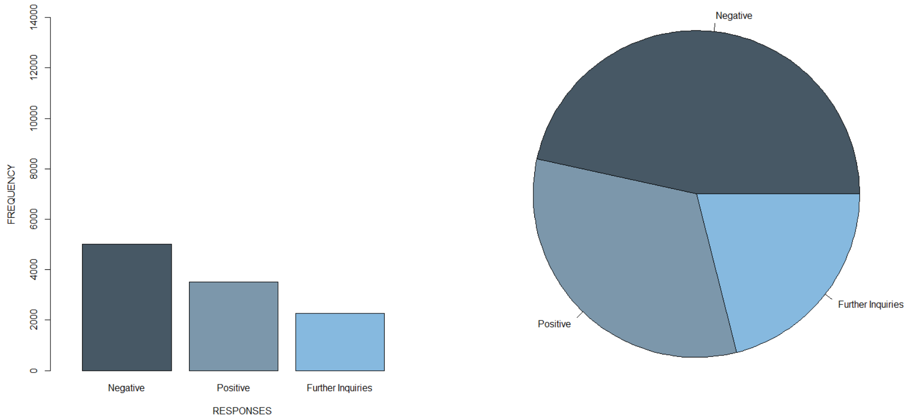

Figure 5: Bar plot and pie chart of responses to native applicants (limited sample)

As for the foreign applicants, 4,829 of the 12,128 responses were positive and 7,299 of the responses were negative. For the limited sample (of size 10,141), 2,691 were positive, 5,882 were negative and 1,568 were positive with further inquires.

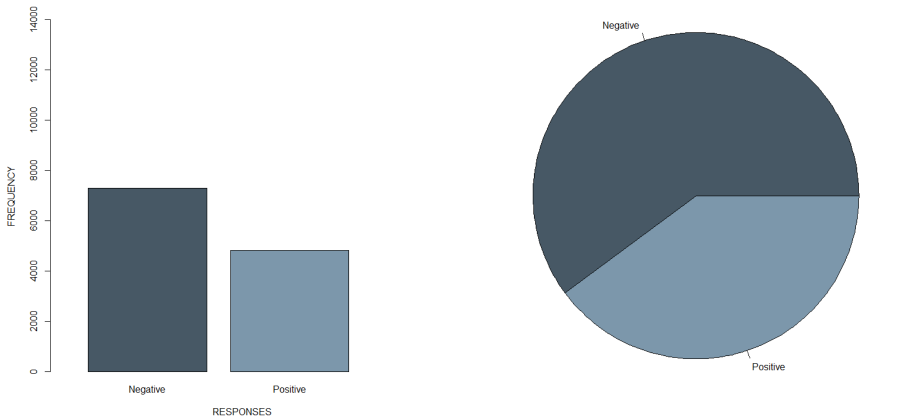

Figure 6: Bar plot and pie chart of responses to foreign applicants

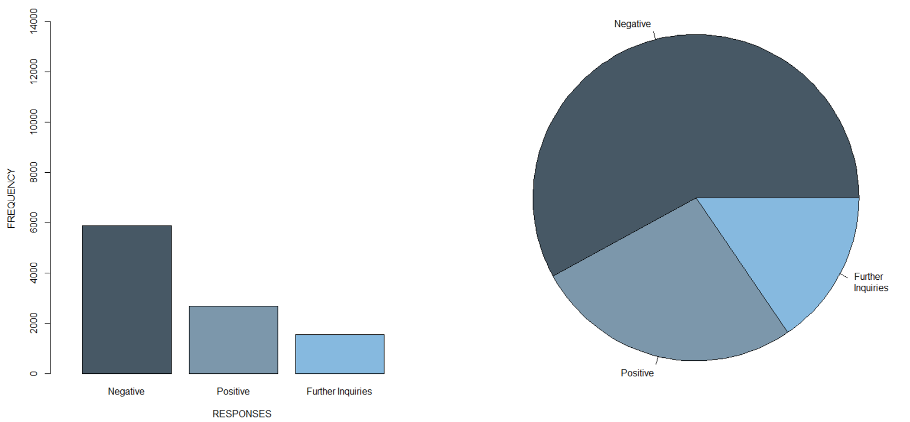

Figure 7: Bar plot and pie chart of responses to foreign applicants (limited sample)

The overall positive response rate is about 45%. For native applicants it is about 51% and for foreign applicants it is about 40%, indicating that foreign-sounding names have received significantly fewer responses than native-sounding names.

4 Responses at the Country Level

Figure 8 shows the positive response rates for the native-sounding and foreign-sounding names categorized by country. There is a clear difference in response rates, with the native-sounding names receiving a higher proportion of positive responses than foreign-sounding names for each country. The Netherlands has the highest response rates (overall = 69.37%, native = 74.29%, foreign = 64.54%) and Serbia has the lowest response rates (overall = 12.01%, native = 15.08%, foreign = 9.31%). These results have been displayed on the European map in Figure 9.

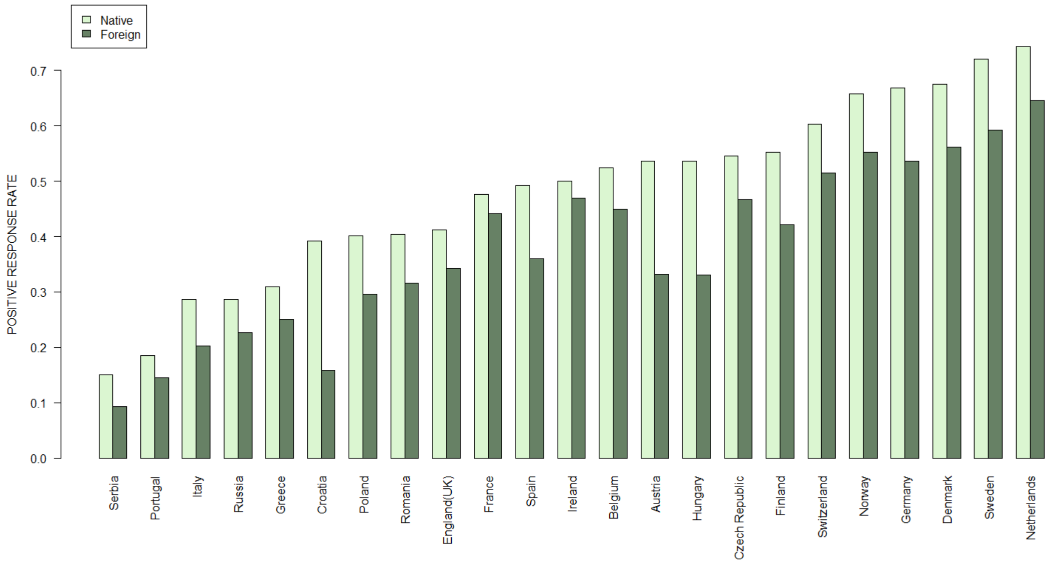

Figure 8: Positive response rates for native-sounding and foreign-sounding names

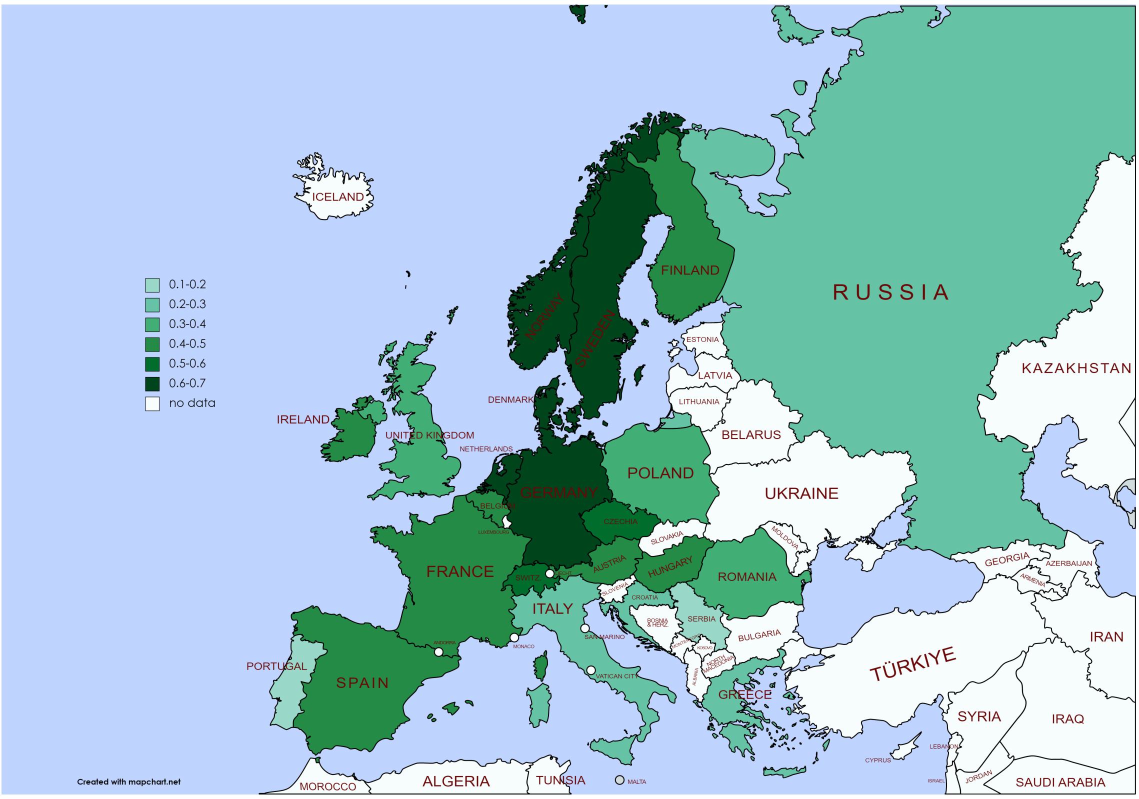

Figure 9: Positive response rates across Europe

4.1 Native vs Foreign Response Rates

We will first perform the chi-squared test of independence for each country separately and then for all countries together to test if there is any difference in the positive response rates between foreign-sounding and native-sounding names. Let X_1 be an indicator random variable taking value 1 if the name is a foreign-sounding name and 0 otherwise. Let X_2 be an indicator random variable taking value 1 if the response was positive and 0 otherwise. Let n_{ij} be the number of data entries with $X_1 = i$ and $X_2 = j$, and let n be the total number of data entries. Arrange the values in a contingency table and let R_i be the sum of values in the i^{th} row and let C_j be the sum of values in the j^{th} column. Then the estimator for the expected frequency is given by

$$\widehat{E_{ij}} = \frac{R_i C_j}{n}$$

The conditions required for the valid application of the χ^2 test are:

- The n observed counts must be sampled randomly from the population of interest.
- For every cell, the expected count $\hat{E}(n_{ij})$ must be at least 5.

Both conditions are satisfied. Our null and alternate hypotheses are

H_0 : The two classifications (native/foreign and positive/negative response) are independent
H_a : The two classifications are dependent

The test statistic is given by

$$\chi_c^2 = \sum \frac{(n_{ij} - \widehat{E_{ij}})^2}{\widehat{E_{ij}}}$$

and the rejection region is $\chi_c^2 > \chi_\alpha^2$, where α is the significance level. The p-value at any point c_0 is given by $P(\chi^2 > \chi_{c_0}^2)$.
The results of the χ^2 test are summarised in Table 1.

Country	**Observations**	**Response Rates (%)**		**χ^2-value**	**p-value**
		Native	**Foreign**		
Austria	1840	53.61	33.15	77.570	0.000
Belgium	663	52.39	44.90	3.177	0.075
Croatia	447	39.25	15.88	29.694	0.000
Czech Republic	1598	54.61	46.72	9.656	0.002
Denmark	1135	67.54	56.11	15.250	0.000
England	1527	41.22	34.29	7.510	0.006
Finland	536	55.19	42.11	8.660	0.003
France	1847	47.69	44.12	2.226	0.136
Germany	1681	66.87	53.61	30.273	0.000
Greece	437	30.95	25.11	1.571	0.210
Hungary	798	53.67	33.00	33.885	0.000
Ireland	308	50.00	46.95	0.176	0.675
Italy	1463	28.69	20.19	13.866	0.000
Netherlands	715	74.29	64.54	7.546	0.006
Norway	1000	65.79	55.27	11.155	0.000
Poland	1312	40.15	29.61	15.610	0.000
Portugal	791	18.48	14.50	1.901	0.168
Romania	493	40.42	31.62	3.768	0.052
Russia	1143	28.70	22.61	5.242	0.022
Serbia	383	15.08	9.31	2.483	0.115
Spain	1410	49.29	36.06	24.687	0.000
Sweden	1493	72.05	59.26	26.477	0.000
Switzerland	1895	60.37	51.50	13.534	0.000
All countries	24915	50.61	39.82	291.910	0.000

Table 1: χ^2 test of independence results

Taking the entire sample into consideration, from the results we can conclude at the 0.01 significance level ($\alpha = 0.01$) that the null hypothesis has to be rejected, i.e, the classifications are not independent. The foreign-sounding names received much fewer responses than the native-sounding names. The same results are replicated in most countries (all countries except Belgium, France, Greece, Ireland, Portugal, Romania and Serbia, at the 0.05 significance level).

4.2 Difference in Response Rates

Now in order to compare the difference between the positive response rates for native-sounding and foreign-sounding names, we will make use of the Wilcoxon signed-rank test for paired differences. Let D_1 and D_2 be the distributions of responses to native and foreign applicants respectively, and let n be the size of the sample taken from the population of differences. The conditions required for a valid signed rank test are:

- The sample of differences has to be taken randomly from the population of differences.
- The distribution from which the sample of paired differences is drawn has to be continuous.

Both conditions are satisfied and hence we can proceed with the test. Rank the absolute value of all the differences in increasing order. The differences equal to 0 must be eliminated and n must be reduced accordingly. If there are ties, those absolute differences receive ranks equal to the average of the ranks they would receive had they not been tied. Let T_+ be the sum of ranks of positive differences and T_- be the sum of ranks of negative differences. Our null and alternate hypotheses are:

$$H_0 : \text{The distributions } D_1 \text{ and } D_2 \text{ are identical}$$
$$H_a : \text{The distribution } D_1 \text{ is shifted either to the left or to the right of } D_2$$

The test statistic is the smaller of T_+ and T_-. Let T_0 be the critical value corresponding to n given in the table of critical values for the Wilcoxon signed-ranked test. The rejection region is $T_+ \leq T_0$ (or $T_- \leq T_0$), and the p-value at any point c is given by $P(T_+ \leq c)$ (or $P(T_- \leq c)$).

Country	Response Rates (%)		Difference	Rank of Absolute Difference
	Native	Foreign		
Austria	53.61	33.15	20.46	21
Belgium	52.39	44.90	7.49	8
Croatia	39.25	15.88	23.37	23
Czech Republic	54.61	46.72	7.89	9
Denmark	67.54	56.11	11.43	16
England	41.22	34.29	6.93	7
Finland	55.19	42.11	13.08	18
France	47.69	44.12	3.57	2
Germany	66.87	53.61	13.26	20
Greece	30.95	25.11	5.84	5
Hungary	53.67	33.00	20.67	22
Ireland	50.00	46.95	3.05	1
Italy	28.69	20.19	8.50	10
Netherlands	74.29	64.54	9.75	13
Norway	65.79	55.27	10.52	14
Poland	40.15	29.61	10.54	15
Portugal	18.48	14.50	3.98	3
Romania	40.42	31.62	8.80	11
Russia	28.70	22.61	6.09	6
Serbia	15.08	9.31	5.77	4
Spain	49.29	36.06	13.23	19
Sweden	72.05	59.26	12.79	17
Switzerland	60.37	51.50	8.87	12

Table 2: Wilcoxon signed-rank test

From the table we get that $T_- = 0$, which has p-value 0.0000002. This gives us enough evidence to reject the null hypothesis and conclude that the two distributions D_1 and D_2 differ. Thus, the foreign-sounding names received significantly fewer responses compared to the native-sounding names. This fact is further depicted in the plots given in Figure 10 and Figure 11.

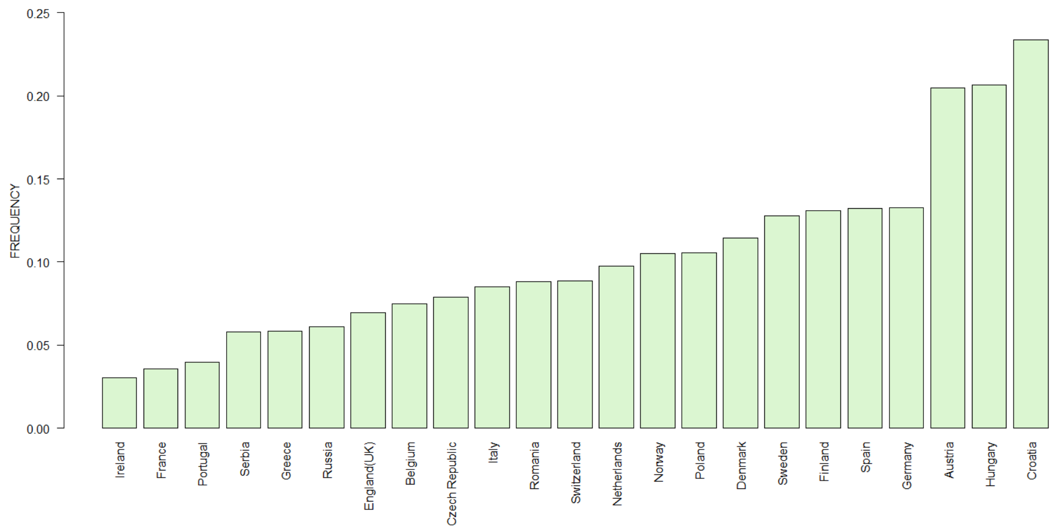

Figure 10: Difference in response rates between native-sounding and foreign-sounding names

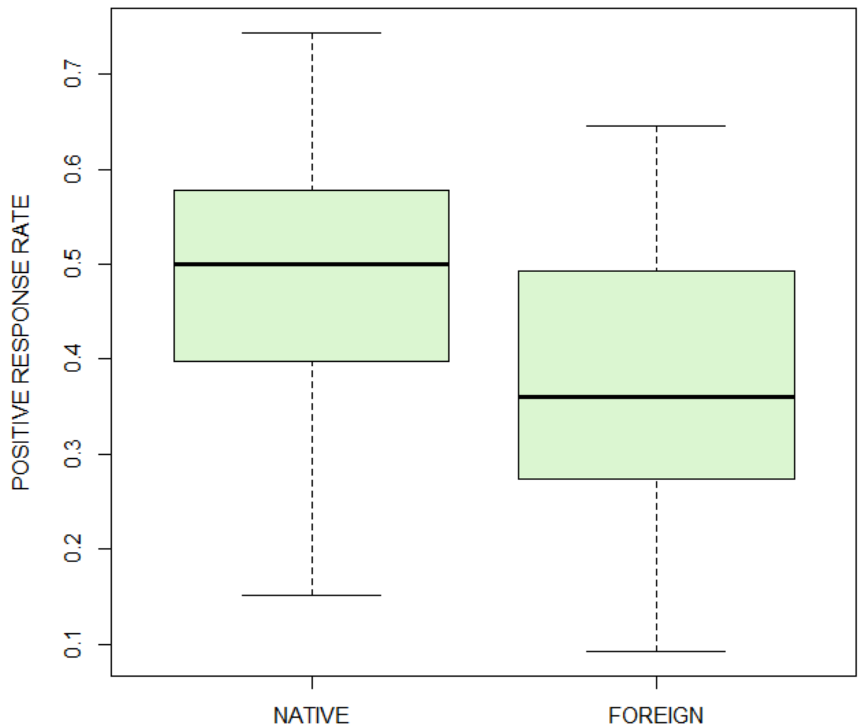

Figure 11: Box plot of response rates for native-sounding and foreign-sounding names

4.3 Preference for Native-Sounding Names

The difference in response rates for each country are displayed on the European map in Figure 12. The colours don't seem to be completely random, i.e, the larger differences (in response rates) seem to be concentrated in certain regions of Europe, suggesting that while there is a preference for native-sounding names in each country, this preference may not be uniformly spread across Europe. We will perform a test of homogeneity to check this fact.

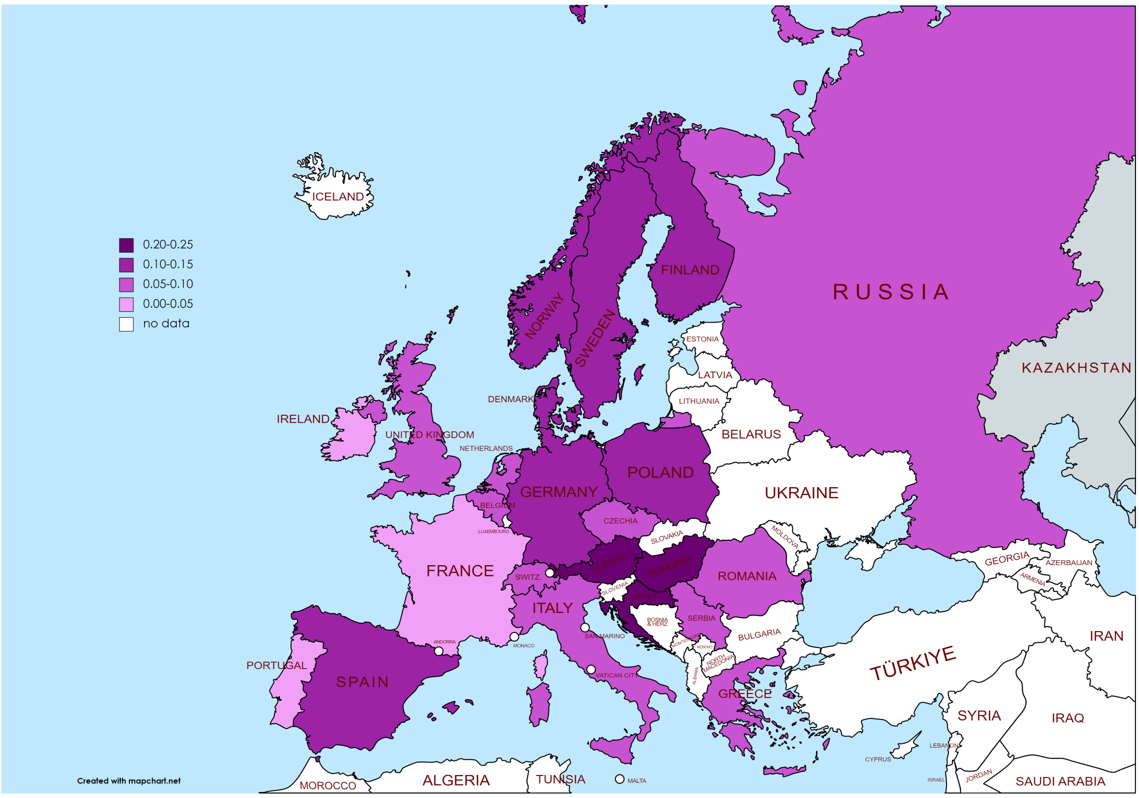

Figure 12: Difference in response rates across Europe

To begin with, we first need to divide Europe into distinct regions. We will be using the division of Europe based on the one in the (CIA) World Factbook. The countries in the study are therefore grouped as follows:

1. Central Europe : Austria, Czech Republic, Germany, Hungary, Poland, Switzerland
2. Eastern Europe : Russia
3. Northern Europe : Denmark, Finland, Norway, Sweden
4. South-eastern Europe : Croatia, Romania, Serbia
5. Southern Europe : Greece, Italy
6. South-western Europe : Portugal, Spain
7. Western Europe : Belgium, England, France, Ireland, Netherlands

We will be using the Breslow-Day test of homogeneity in order to investigate if all 7 regions have the same odds ratio. Consider the k^{th} region. Let X_{1k} be the indicator random variable taking value 1 if the name is a foreign-sounding name and 0 otherwise. Let X_{2k} be the indicator random variable taking value 1 if the response was positive and 0 otherwise. Let n_{ijk} be the number of data entries with $X_{1k} = i$ and $X_{2k} = j$, and let n be the total number of data entries. Then the odds ratio is given as follows,

$$OR_k = \frac{n_{00k}n_{11k}}{n_{01k}n_{10k}}$$

The conditions required for the valid application of the Breslow-Day test are:

- The sample size should be large in each group.
- At least 80% of the cell counts should be greater than 5.

Our data satisfies both conditions. Now, the null and alternate hypotheses are:

$$H_0 : \text{The odds ratio is the same for all seven regions}$$
$$H_a : \text{At least two of the seven regions have different odds ratios}$$

We will be using the Breslow-Day test statistic with Tarone's adjustment, which makes the resulting test statistic asymptotically chi-square. Let n_k be the number of data entries in the k^{th} region. Then the pooled odds ratio

(Cochran-Mantel-Haenszel pooled odds ratio) is calculated as follows:

$$OR = \frac{\sum_{k=1}^{7} \frac{n_{00k} n_{11k}}{n_k}}{\sum_{k=1}^{7} \frac{n_{01k} n_{10k}}{n_k}}$$

The test statistic is given as follows:

$$T = \sum_{k=1}^{7} \frac{(X_{2k} - E(X_{2k}))^2}{Var(X_{2k}|X_k, OR)} - \frac{(X_2 - E(X_2))^2}{\sum_{k=1}^{7} Var(X_{2_k}|X_k, OR)}$$

where $X_2 = \sum_{k=1}^{7} X_{2k}$, $X_k = X_{1k} + X_{2k}$ and $E(X_2) = \sum_{k=1}^{7} E(X_{2k})$. The value of $E(X_{2k})$ can be obtained by solving the following quadratic equation:

$$(OR - 1)E(X_{2k})^2 - ((X_k + n_{11k})OR + (n_{01k} - X_k))E(X_{2k}) + X_k n_{11k} OR = 0$$

The value of $Var(X_{2k}|X_k, OR)$ is given by

$$Var(X_{2k}|X_k, OR) = \left(\frac{1}{E(X_{2k})} + \frac{1}{X_k - E(X_{2k})} + \frac{1}{n_{11k} - E(X_{2k})} + \frac{1}{n_{01k} - X_k + E(X_{2k})} \right)^{-1}$$

Since the test statistic follows a χ^2 distribution with 6 degrees of freedom, the rejection region is given by $T > \chi^2_\alpha$ (with 6 degrees of freedom). The p-value at any point c is $\mathrm{P}(T > c)$.

On performing the test, we get T to be 27.219 (which has p-value 0.0001). This gives us enough evidence to reject null hypothesis at the $\alpha = 0.01$ significance level and hence we conclude that while there is a preference for native-sounding names, it is not spread evenly across Europe.

Country	Response Rates (%)		Difference
	Native	Foreign	
Central Europe	55.99	41.98	14.01
Eastern Europe	28.70	22.61	6.09
Northern Europe	67.12	55.26	11.86
South-eastern Europe	32.86	19.71	13.15
Southern Europe	29.20	21.35	7.85
South-western Europe	37.07	29.20	7.87
Western Europe	50.31	44.30	6.01

Table 3: Group-wise difference in response rates

5 Responses at the Foreign Group Level

5.1 Response Rates by Foreign Group Number

As mentioned before, for this study the responses for foreign-sounding names were collected by taking three different foreign groups into consideration for each country (except Switzerland). We will now attempt to make use of the Kruskal-Wallis test and see if there is any difference in the response rates received by each of the foreign groups. Note that for this section, Switzerland will be excluded because there were more than 3 foreign groups taken for the study done in Switzerland.
Let D_1, D_2 and D_3 be the distributions of responses to applicants from foreign groups 1, 2 and 3 respectively. Let the sizes of samples taken from D_1, D_2 and D_3 be n_1, n_2 and n_3 respectively and let $n = n_1 + n_2 + n_3$. The conditions for the valid application of the Kruskal-Wallis test are:

- The 3 samples must be random and independent.
- There must be five or more observations in each sample.
- The 3 probability distributions from which the samples are drawn must be continuous.

The first two conditions are satisfied but the third one isn't and we will require a correction term. Pool all the observations and rank them in increasing order. If there are tied measurements, assign them the average of the ranks they would receive had they not been tied. Let R_1, R_2 and R_3 be the sum of ranks of the samples taken from D_1, D_2 and D_3 respectively. Our null and alternate hypotheses are:

$$H_0 : \text{The distributions } D_1, D_2 \text{ and } D_3 \text{ are identical}$$
$$H_a : \text{At least two of three distributions differ in location}$$

The test statistic H is given by

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{3} \frac{R_j^2}{n_j} - 3(n+1)$$

But since the distributions are not continuous, in order to account for the large number of ties in the observation, H will have to be divided by the following correction factor

$$1 - \frac{\sum_{i=1}^{G}(t_i^3 - t_i)}{N^3 - N}$$

where G is the number of groupings ($G = 2$ in this case) of different tied ranks, and t_i is the number of tied values within group i that are tied at a particular value.
If the null hypothesis is true, the distribution of H is approximately a χ^2 distribution with 2 degrees of freedom. The rejection region is, therefore, $H > \chi^2_\alpha$ with 2 degrees of freedom (where α is the significance level). The p-value at any point c is given by $P(H > c)$.
The 3 foreign groups corresponding to each country are shown in Table 4.

Country	Foreign Group 1	Foreign Group 2	Foreign Group 3
Austria	Serbian	Turkish	Bosnian-Herzegovinian
Belgium	Italian	Moroccan	Polish
Croatia	German	Italian	Chinese
Czech Republic	Ukrainian	Vietnamese	Russian
Denmark	Turkish	Polish	German
England	Polish	Indian	Italian
Finland	Estonian	Russian	Somalian
France	Portuguese	Algerian	Moroccan
Germany	Turkish	Italian	Polish
Greece	Albanian	Bulgarian	Romanian
Hungary	Romanian	German	Ukrainian
Ireland	Polish	Lithuanian	Latvian
Italy	Romanian	Albanian	Moroccan
Netherlands	Turkish	Moroccan	Indonesian
Norway	Polish	German	Lithuanian
Poland	Ukrainian	German	Belarusian
Portugal	Ukrainian	Angolan	Romanian
Romania	Turkish	Chinese	German
Russia	Kazakh	Uzbek	Azerbaijani
Serbia	Hungarian	Slovakian	Albanian
Spain	Moroccan	Romanian	English
Sweden	Finnish	Iraqi	Polish
Response Rate = 39.14%	Response Rate = 40.31%	Response Rate = 38.84%	Response Rate = 38.28%

Table 4: Foreign groups

On performing the test, we get the value of H to be 3.5293, which has p-value 0.1712. This isn't enough evidence to reject the null hypothesis (at the $\alpha = 0.05$ level). Therefore, we conclude that there isn't any difference in positive response rates between the 3 foreign groups.

5.2 Response Rates by Geographic Location of Foreign Group

We observed in a previous section that the difference in response rates is not evenly spread across Europe. Similarly, we will use the Kruksal-Wallis test again to see if among the different foreign groups used for this study the response rates were the same. We will be grouping the foreign groups based on their geographic locations and then run the test. The division will happen as follows, based on the United Nations geoscheme and the (CIA) World Factbook:

1. Africa : Algeria, Angola, Morocco, Somalia
2. Central Asia : Kazakhstan, Uzbekistan
3. Western Asia : Azerbaijan, Saudi Arabia, Turkey
4. Southern Asia : India, Iraq
5. Eastern/ South-eastern Asia : China, Indonesia, Vietnam
6. Central Europe : Germany, Hungary, Poland, Slovakia
7. Eastern Europe : Belarus, Estonia, Latvia, Lithuania, Russia, Ukraine
8. Northern Europe : England, Finland
9. Southern Europe : Italy, Portugal
10. South-eastern Europe : Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Romania, Serbia

Let $D_1, ..., D_{10}$ be the distribution of responses to applicants from each of the regions above (in the same order). Let the size of the sample taken from each D_i be n_i. The conditions for the valid application of the Kruskal-Wallis test are:

- The 10 samples must be random and independent.
- There must be five or more observations in each sample.
- The 10 probability distributions from which the samples are drawn must be continuous.

The first two conditions are satisfied but the third one isn't and we will require a correction term. Pool all the observations and rank them in increasing order. If there are tied measurements, assign them the average of the ranks they would receive had they not been tied. Let R_i be the sum of ranks of the sample take from D_i. Our null and alternate hypotheses are:

H_0 : The ten distributions are identical

H_a : At least two of the ten distributions differ in location

The test statistic H is given by

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{10} \frac{R_j^2}{n_j} - 3(n+1)$$

But since the distributions are not continuous, in order to account for the large number of ties in the observation, H will have to be divided by the following correction factor

$$1 - \frac{\sum_{i=1}^{G}(t_i^3 - t_i)}{N^3 - N}$$

where G is the number of groupings of different tied ranks ($G = 2$ in this case), and t_i is the number of tied values within group i that are tied at a particular value.
If the null hypothesis is true, the distribution of H is approximately a χ^2 distribution with 9 degrees of freedom. The rejection region is, therefore, $H > \chi^2_\alpha$ with 9 degrees of freedom (where α is the significance level). The p-value at any point c is given by $P(H > c)$.

Group	**Response Rates (%)**	**Group**	**Response Rates (%)**
Africa	36.18	Central Europe	46.55
Central Asia	25.53	Eastern Europe	39.32
Western Asia	41.29	Northern Europe	52.76
Southern Asia	44.99	Southern Europe	43.38
Eastern Asia	43.40	South-eastern Europe	28.59

Table 5: Reponse rates of each group

On performing the test, we get the value of H to be 228.99, (which has a p-value of around 2.682×10^{-44}). This gives us enough evidence to reject the null hypothesis (at the $\alpha = 0.01$ level) and hence conclude that the response rates differ based on the geographic location of the foreign group.

6 Influence of External Factors

In a randomized field experiment, the data should not be influenced by other factors. To control for the validity of the experiment, it is of interest to us to study if and how potential factors could have had an influence on the response rates in any way. For this study, we will try to analyse how a country's population, net migration and share of right-wing voters influence the response rates. The values and results will be clustered at the country level. We will run three binary logistic regression models to analyse the effect of the these external factors. Since we are also interested in seeing if the factors influence the native and foreign response rates differently, we will make use of a dummy variable as well. We will be making the following assumptions with our model:

- The response variable is binary.
- The observations are independent.
- The predictor variables are not highly correlated with each other.
- There are no extreme outliers.
- The sample size is sufficiently large.

6.1 Effect of Population on Response Rates

The binary logistic regression model is of the following form:

$$p(x_1, x_2) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}$$

p denotes the success probability of the dependent variable y, which is the response received (The response is an indicator random variable taking value 0 if the response was negative and 1 if the response was positive). The independent variables are

$$x_1 = \text{Population}$$
$$x_2 = \begin{cases} 1 & \text{if the name is native-sounding} \\ 0 & \text{if the name is foreign-sounding} \end{cases}$$

The likelihood function for the binary logistic regression model is given as follows:

$$L(\beta_0, \beta_1, \beta_2, \beta_3) = \prod_{i=1}^{n} p(x_{1i}, x_{2i})^{y_i} (1 - p(x_{1i}, x_{2i}))^{1-y_i}$$

This yields the log likelihood function:

$$l(\beta_0, \beta_1, \beta_2, \beta_3) = \sum_{i=1}^{n} \left(y_i(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i}) - \log(1 + e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\beta_3 x_{1i} x_{2i}}) \right)$$

Thus an estimate of the regression coefficients is found by maximizing the likelihood (or the log likelihood function).
To assess how good of a fit the model is, we will be using McFadden's pseudo-R-squared. It takes values from 0 to 1, and the closer the value is to 1, the better is the fit. The value is calculated by comparing the log likelihood of the entire model with the log likelihood of the null model (model with the intercept only).

$$\text{McFadden's } R^2 = 1 - \frac{l(\beta_0, \beta_1, \beta_2, \beta_3)}{l(\beta_0)}$$

where $l(\beta_0) = \sum_{i=1}^{n} \left(y_i \beta_0 - \log(1 + e^{\beta_0}) \right)$.
Now, with our model, we are interested in testing whether the values of β_1 and β_3 are 0, because that tells us how the population influences the response rates. A positive value indicates that with an increase in population, the probability of a positive response increases and a negative value indicates that with an increase in population, the probability of a positive response decreases. If the value is 0, there is either no effect on the probability of a positive response, or the effect is not linear. We will make use of the Wald statistic to test this fact. The

standard normal curve is used to determine the rejection region and p-value of this test. The hypothesis tests for β_1 and β_3 are as follows:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \\ \text{Test statistic} &: z = \frac{\hat{\beta}_1}{\sigma_{\hat{\beta}_1}} \\ \text{Rejection region} &: |z| > z_{\alpha/2} \end{aligned} \qquad \begin{aligned} H_0 &: \beta_3 = 0 \\ H_a &: \beta_3 \neq 0 \\ \text{Test statistic} &: z = \frac{\hat{\beta}_3}{\sigma_{\hat{\beta}_3}} \\ \text{Rejection region} &: |z| > z_{\alpha/2} \end{aligned}$$

The regression results are summarised in the table below.

Parameter	**Estimate**	**Standard Error**	**z-value**	**p-value**
β_0	-0.2713	0.0260	-10.326	0.0000
β_1	-0.0041	0.0005	-7.649	0.0000
β_2	0.4875	0.0358	13.610	0.0000
β_3	-0.0016	0.0007	-2.197	0.0281

Table 6: Regression results

From the table, we have sufficient evidence (at the $\alpha = 0.01$ level) to reject the null hypothesis that $\beta_1 = 0$. The fact that β_1 is not 0 and is negative tells us that the population of a country has a negative effect on the response rates. Also, since we cannot reject the hypothesis stating that $\beta_3 = 0$ ($\alpha = 0.01$), this tells that the population of the country has a similar influence on both the native-sounding and foreign-sounding names, or that the difference in influence, if it exists, is not linear. McFadden's R^2 of this model is 0.014, indicating that the model is not a very good fit.

6.2 Effect of Net Migration on Response Rates

Similar to the previous case, a binary logistic model that characterises the relation between the net migration and the positive response rate is of the form:

$$p(x_1, x_2) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}$$

p denotes the success probability of the dependent variable y, which is the response received (The response is an indicator random variable taking value 0 if the response was negative and 1 if the response was positive). The independent variables are

$$x_1 = \text{Net migration}$$
$$x_2 = \begin{cases} 1 & \text{if the name is native-sounding} \\ 0 & \text{if the name is foreign-sounding} \end{cases}$$

We will be performing the same tests as before and therefore our hypothesis tests are as follows:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \\ \text{Test statistic} &: z = \frac{\hat{\beta}_1}{\sigma_{\hat{\beta}_1}} \\ \text{Rejection region} &: |z| > z_{\alpha/2} \end{aligned} \qquad \begin{aligned} H_0 &: \beta_3 = 0 \\ H_a &: \beta_3 \neq 0 \\ \text{Test statistic} &: z = \frac{\hat{\beta}_3}{\sigma_{\hat{\beta}_3}} \\ \text{Rejection region} &: |z| > z_{\alpha/2} \end{aligned}$$

The regression results are summarised in the table below.

Parameter	Estimate	Standard Error	z-value	p-value
β_0	-0.7400	0.0281	-26.290	0.0000
β_1	0.1341	0.0084	15.892	0.0000
β_2	0.4083	0.0391	10.431	0.0000
β_3	0.0126	0.0120	1.048	0.2940

Table 7: Regression results

From the table, we have sufficient evidence (at the $\alpha = 0.01$ level) to reject the null hypothesis stating that $\beta_1 = 0$. The fact that β_1 is not 0 and is positive tells us that the net migration of a country has a positive effect on the response rates. We do not reject the null hypothesis stating that $\beta_3 = 0$, and we therefore conclude that the net migration had a similar influence on both native and foreign response rates, or that the difference in influence, if it exists, is not linear. McFadden's R^2 of this model is 0.025, indicating that the model is not a very good fit.

6.3 Effect of Percentage of Right-Wing Voters on Response Rates

A binary logistic model that characterises the relation between the percentage of right-wing voters and the positive response rate is of the form:

$$p(x_1, x_2) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}$$

p denotes the success probability of the dependent variable y, which is the response received (The response is an indicator random variable taking value 0 if the response was negative and 1 if the response was positive). The independent variables are

$$x_1 = \text{Percentage of right-wing voters}$$
$$x_2 = \begin{cases} 1 & \text{if the name is native-sounding} \\ 0 & \text{if the name is foreign-sounding} \end{cases}$$

We will be performing the same tests as before and therefore our hypothesis tests are as follows:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \\ \text{Test statistic} &: z = \frac{\hat{\beta}_1}{\sigma_{\hat{\beta}_1}} \\ \text{Rejection region} &: |z| > z_{\alpha/2} \end{aligned} \qquad \begin{aligned} H_0 &: \beta_3 = 0 \\ H_a &: \beta_3 \neq 0 \\ \text{Test statistic} &: z = \frac{\hat{\beta}_3}{\sigma_{\hat{\beta}_3}} \\ \text{Rejection region} &: |z| > z_{\alpha/2} \end{aligned}$$

The regression results are summarised in the table below.

Parameter	Estimate	Standard Error	z-value	p-value
β_0	-0.3400	0.0335	-10.153	0.0000
β_1	-0.0039	0.0015	-2.612	0.0090
β_2	0.3408	0.0467	7.295	0.0000
β_3	0.0052	0.0021	2.476	0.0133

Table 8: Regression results

From the table, we have sufficient evidence (at the $\alpha = 0.01$ level) to reject the null hypothesis stating that $\beta_1 = 0$. The fact that β_1 is not 0 and is negative tells us that the percentage of right-wing voters of a country has a negative effect on the response rates. We do not reject the null hypothesis stating that $\beta_3 = 0$ ($\alpha = 0.01$), and we therefore conclude that the percentage of right-wing voters had a similar influence on both native and foreign response rates, or that the difference in influence, if it exists, is not linear. McFadden's R^2 of this model is 0.009, indicating that the model is not a very good fit.

6.4 Effect of Proportion of Foreign Group Population on Response Rates

In this part, we will try to see if there is any relationship between response rates received by a foreign group and the proportion of country's population consisting of that particular foreign group. A binary logistic model that characterises this is of the following form:

$$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

p denotes the success probability of the dependent variable y, which is the response received (The response is an indicator random variable taking value 0 if the response was negative and 1 if the response was positive), and the independent variable x is the percentage of the population consisting of the foreign group.
We will be performing the same test as before and therefore our hypothesis test is as follows:

$$\begin{aligned} H_0 :& \beta_1 = 0 \\ H_a :& \beta_1 \neq 0 \\ \text{Test statistic :}& z = \frac{\hat{\beta_1}}{\sigma_{\hat{\beta_1}}} \\ \text{Rejection region :}& |z| > z_{\alpha/2} \end{aligned}$$

The regression results are summarised in the table below.

Parameter	Estimate	Standard Error	z-value	p-value
β_0	-0.6141	0.0331	-18.568	0.0000
β_1	0.1947	0.0231	8.421	0.0000

Table 9: Regression results

Therefore, we have sufficient evidence to reject the null hypothesis. With the increase in population percentage, the probability of a positive response increases. McFadden's R^2 of this model is 0.005, indicating that the model is not a very good fit.

7 Conclusion

The findings of this study demonstrate that individuals with names that sound foreign encounter difficulties playing amateur football in Europe. We discover evidence of racial discrimination in European amateur football, notwithstanding the possibility that cultural preferences still play a part in the intricate process of integration and sport participation. Evidence of ethnic discrimination in sports aligns with traditional theories of social identity and in-group preferences, and extends prior finding in the labor and housing markets of European countries, including Denmark, Finland, France, Germany, Greece, Ireland, the Netherlands, Norway, Spain, and Sweden, among others.

In the study conducted, we see significantly lower responses for foreign-sounding names. Not just that but we also noticed that the percentage of positive responses was also noticeably less than native-sounding names (52% for native and 40% for foreign). All this was quite clear from the plots.

The χ^2 test of independence and Wilcoxon signed-rank test helped us conclude that the foreign sounding names received much fewer responses than native sounding names which is the same result as the one we get from the descriptive inferences.

We further analysed whether the preference for native sounding names was distributed equally in Europe. We first grouped the data into 7 groups based on geographical location as given by the CIA Factbook and then we performed the Breslow-Day test of homogeneity and were able to conclude that the preference for native sounding names is not evenly distributed across Europe.

Next we analysed the discrimination against each foreign group. For this purpose we divided the groups in 2 different ways. In the first one we divided them as foreign group 1, foreign group 2, and foreign group 3, in accordance with population size. Then we performed the Kruskal-Wallis test to conclude that there is no clear evidence that one particular of the above foreign group is discriminated against more than the other. While in some countries the largest foreign group received the fewest number of responses (e.g., Denmark), in other countries it is the second-largest (e.g., Germany), or the third-largest foreign group (e.g., Croatia).

In the second grouping we used the UN geoscheme to group countries according to their geographical location into 10 groups. This grouping, in most cases, helped to group countries of similar culture or origin together. This helped us test whether certain cultures faced a greater amount of discrimination. When we performed the Kruskal-Wallis test for this grouping we found evidence that certain foreign groups are discriminated against more than the others. This was not unexpected as the percentage of positive responses varied from 25.53% for Central Asians to 52.76% for Northern Europeans, more than double the percentage of Central Asia.

Next we perform an analysis to check the influence of external factors on the field experiment using Logistic Regression. We analysed the influence of the population, net migration and percentage of right-wing voters of the country. We found evidence to conclude that the population and percentage of right-wing voters have a negative impact on the response rates, whereas the net migration has a positive influence on the response rates. We also concluded that the population proportion of foreign groups has a positive influence on the responses received.

The cross-national experiment is also bound to neglect the potential intersections of nationality and ethnicity with religion, race, and perceived status. The limited number of female clubs in Europe did not allow the authors to analyse the influence of gender. Moreover, amateur football is a popular social activity but directed to a certain age group, which excludes older adults. Raising awareness is important to reduce biases against minority groups. Discrimination in a real issue in the modern society and will only be an hindrance to development. Toni Morrison once said “There is no such thing as race. None. There is just a human race-scientifically, anthropologically.”

8 References

1. C. Nesseler, C. Gomez-Gonzalez, H. Dietl, What’s in a name? Measuring access to social activities with a field exper- iment, Palgrave Commun. 5 (2019), doi: 10.1057/s41599- 019- 0372- 0

2. H.M. Dietl, C. Gomez-Gonzalez, P. Moretti, C. Nesseler, Does persistence pay off? Accessing social activities with a foreign-sounding name, Appl. Econ. Lett. (2020) 1–5, doi: 10.1080/13504851.2020.1784381

3. C. Gomez-Gonzalez, C. Nesseler, H.M. Dietl, Mapping discrimination in Europe through a field experiment in amateur sport, Humanit. Soc. Sci. Commun. 8 (2021), doi: 10.1057/s41599- 021- 00773- 2

4. Osama Almalik and Edwin R. van den Heuvel | Yueqing Hu (Reviewing Editor) (2018) Testing homogeneity of effect sizes in pooling 2x2 contingency tables from multiple studies: a comparison of methods, Cogent Mathematics and Statistics, 5:1, doi: 10.1080/25742558.2018.1478698