SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
WARSAW SCHOOL OF ECONOMICS

Master's Studies

Authors: Aleksandra Romanowicz,
Ireneusz L

Binary logistic regression

Final paper in
"Logistic Regression with SAS"

Warsaw 2019

Table of contents

Introduction

The objective of the project was to prepare binary logistic regression model on the data chosen from the ESS (European Social Survey) carried out in 2016. As the objective the following problem was chosen – to check the propensity of person to vote in national elections basing on interest in political situation, trust in government institutions and basic demographic statistics. Study was prepared on the data from Poland.

The methodology of the study is as follows.

Basing on the own knowledge and general studies on the willingness to vote the explanatory variables where chosen:

- trstprl - Trust in country's parliament
- trstlgl - Trust in the legal system
- trstplt - Trust in politicians
- trstprt - Trust in political parties
- trstep - Trust in the European Parliament
- polintr - How interested in politics
- agea - Age of respondent, calculated
- badge - Worn or displayed campaign badge/sticker last 12 months
- pbldmn - Taken part in lawful public demonstration last 12 months

The “trust variables” are discrete and based on the 11-point scale (from no trust to complete trust) and required transformation to achieve better results and improve readability of the model results. Polintr was represented on smaller number of dimensions (4) but also has been transformed. The binning process is detailed in the forthcoming sections. Yes/No variables are just encoded as 1 for “yes” and 0 for “no”. It affects vote and badge variables. The responses such as “refusal”, “don’t know”, “not eligible” etc. were not taken into consideration.

The intermediary step between variable selection and modelling was to check collinearity and investigate if there are any variables to be removed from the set.

As binning can be achieved in many ways to choose the most effective way, different bin approaches (2, 3, and 5) were prepared with subsequent models created with the same parameters. Models were tuned to achieve best combination of hyper parameters such as: elimination method, confidence level with following assessment of model fit, relevancy of variables and coefficients and number of concordant pairs.

Basing on the research results the model with the most optimal connecting method and parameters was chosen.

Variable binning

Model I - two categories

The following binning methodology was applied to receive transformed variables for the first model:

- “Trust answers” are highly detailed where 11 “non-missing” options are available. It was split in a half, where sixth (numbered as 5 in data) is associated with the first group
- Interest in politics (polintr) is the exception as due to very good results in both models when it is split in three. Split is carried out by generating separate groups for the most radical answers (Very Interested and Not at all interested) and less confident answers (Quite interested, Hardly interested) which are closely related are blended together into one group
- Age, according to the analysis has the most influence if person is below 30, set is splitted regarding that findings. In group of 30 year old people, only half of them went voting while in other groups around or more than 70% took part, it can be said that participation rate increases with the age, that conclusion is used in second model.
- The rule behind encoding variables as 1 or 0 is that 1 is presumably considered as positively influencing vote probability whereas 0 shall work in opposite direction

Model II - three categories

Depending on the source different binning methods are applied:

- “Trust answers” are highly detailed where 11 “non-missing” options are available. They are split in a one thirds, where fourth value (3) is associated with the first group
- Split for polintr is carried out by generating separate groups for the most radical answers (Very Interested and Not at all interested) and less confident answers (Quite interested, Hardly interested) which are closely related are blended together into one group
- Age, according to the analysis has the most influence if person is below 30, set is splitted regarding that findings. In group of 30 year old people, only half of them went voting while in other groups around or more than 70% took part, it can be said that participation rate increases with the age. That made the base for the third split for the most frequent voters that are aged above 60 years old.
- The rule behind encoding variables as 1 or 0 is that 1 is presumably considered as positively influencing vote probability whereas 0 shall work in opposite direction

Model III - five categories

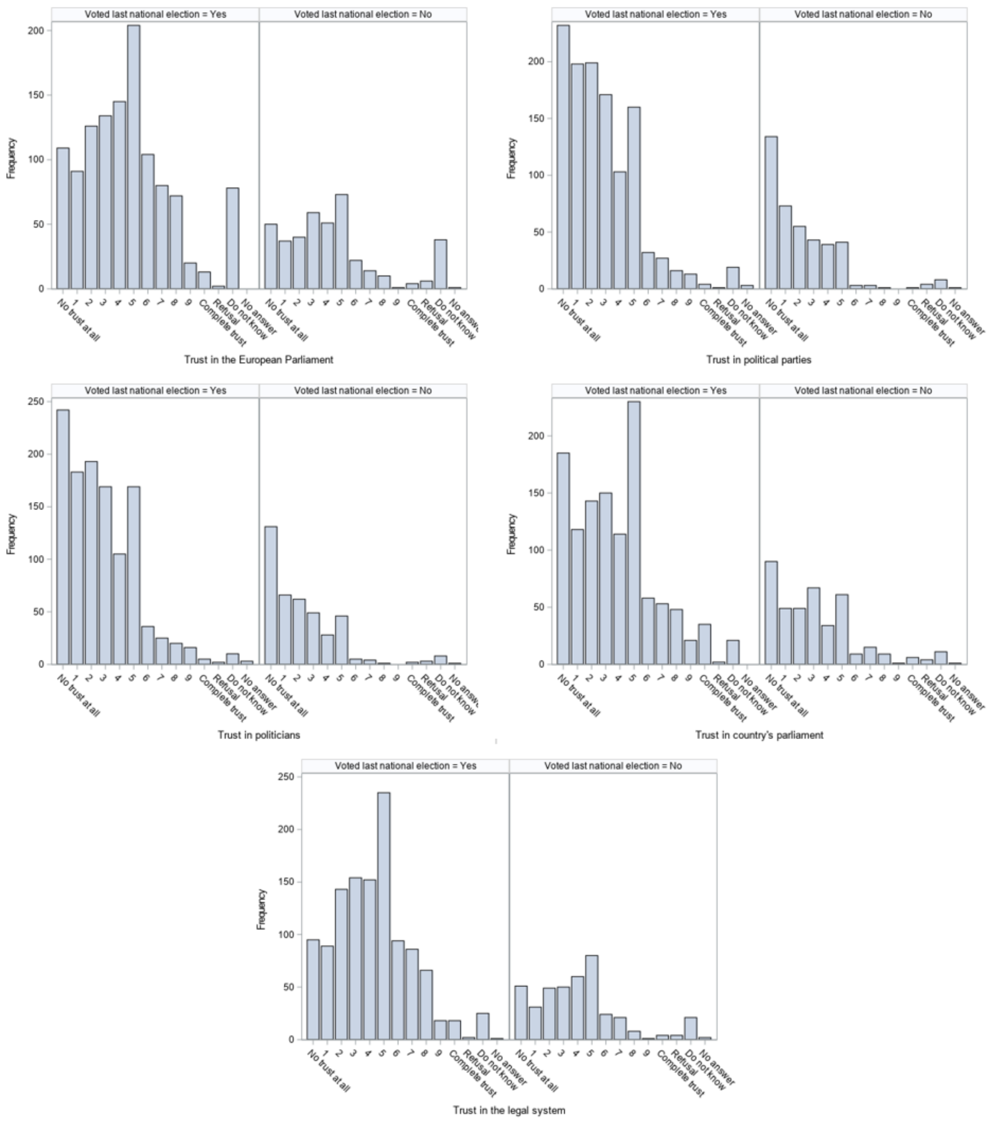
In this section split into categories is done on the basis of visual and statistical analysis. We take into consideration frequency of answers and their significance for the research and divide them into five logical groups. In the first step we observe number of answers for dependent variable - vote in last national election.

Table 1. Frequencies for dependent variable - vote

Voted last national election				
vote	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	1178	69.54	1178	69.54
No	406	23.97	1584	93.51
Not eligible for vote	92	5.43	1676	98.94
Refusal	12	0.71	1688	99.65
Don't know	4	0.24	1692	99.88
No answer	2	0.12	1694	100.00

We see that the total percent of those who refused the answer or did not know is less than 1% in each group and share of people who are not eligible to vote (which in our case has the same meaning as those who refused the answer) is less than 6%. We will remove those records from our research. Moving further, basing on the distribution, we want to group the results of explanatory variables in less categories (there is over 10 of them in the source data).

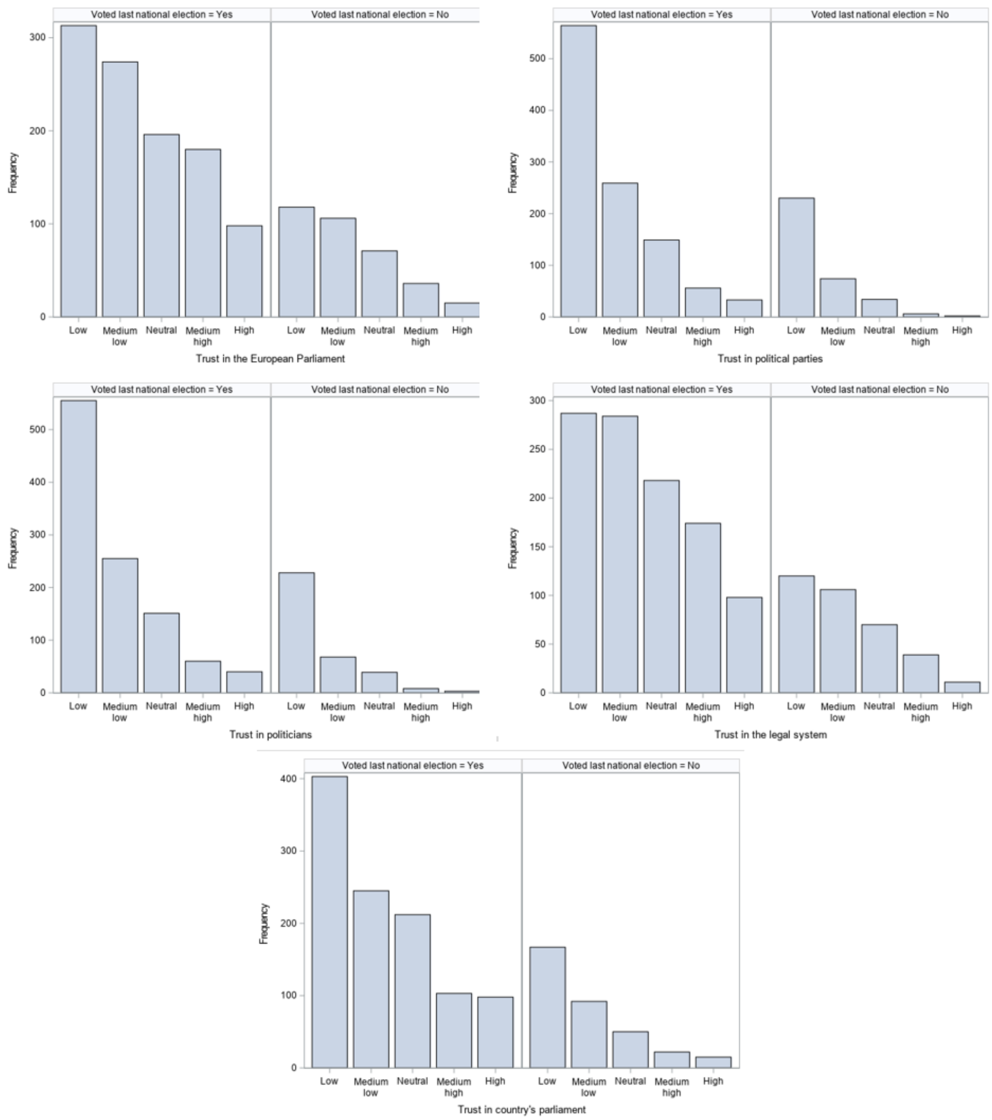
Chart 1. Frequencies among explanatory variables



First observation is a significant peak for the value 5 in each group. 5, as a completely neutral value, is probably chosen by those who have no thought about the topic. For the research purpose it can be important if somebody chose 4 instead of 6 (we can say that even though he or she is not strongly on the one side, he or she is not completely neutral). As a result, we decide to split the results in at least 3 categories "<5", '5' and ">5". However, we cannot underestimate radical answers such as 1 and 2 instead of more neutral 4 hence it is justified to split “<5” and “>5” categories into two.

Moreover, we can also see that responses of those who refused/did not answer and did not know are not important for our research (we want to examine the influence on voting), so we remove those observations from our dataset.

Chart 2. Frequencies for explanatory variables after aggregation.



Now it is clear that for each variable there is a decreasing dependence between the number of observations and trust - the highest number of people has low trust in authorities.

Final set of variables that will be taken into modelling consists of variables related to trust splitted into five categories presented above, interest in politics, badge and age binned as in the second model.

Collinearity investigation

Model I

Using simple Pearson statistic for correlation we can examine whether variables used in the model are correlated with each other.

Table 2 . Assessment of correlation coefficients

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	trstprl	trstlgl	trstplt	trstprt	trstep	polintr	agea	badge	pbldmn
trstprl	1.00000	0.34616	0.50424	0.45629	0.11486	0.09797	0.05277	0.00729	-0.05381
Trust in country's parliament		<.0001	<.0001	<.0001	<.0001	<.0001	0.0323	0.7678	0.0293
	1645	1606	1629	1621	1541	1642	1645	1641	1641
trstlgl	0.34616	1.00000	0.23455	0.26364	0.26313	0.07231	-0.12894	0.03151	0.06150
Trust in the legal system	<.0001		<.0001	<.0001	<.0001	0.0035	<.0001	0.2038	0.0131
	1606	1632	1615	1606	1533	1630	1632	1628	1628
trstplt	0.50424	0.23455	1.00000	0.67202	0.17985	0.08120	0.01421	0.03209	-0.03472
Trust in politicians	<.0001	<.0001		<.0001	<.0001	0.0009	0.5630	0.1919	0.1579
	1629	1615	1660	1641	1551	1658	1660	1656	1656
trstprt	0.45629	0.26364	0.67202	1.00000	0.15762	0.10476	0.01639	0.01212	-0.01878
Trust in political parties	<.0001	<.0001	<.0001		<.0001	<.0001	0.5056	0.6232	0.4463
	1621	1606	1641	1651	1543	1648	1651	1647	1647
trstep	0.11486	0.26313	0.17985	0.15762	1.00000	0.08808	-0.03879	0.05981	0.08473
Trust in the European Parliament	<.0001	<.0001	<.0001	<.0001		0.0005	0.1258	0.0183	0.0008
	1541	1533	1551	1543	1559	1557	1559	1556	1556
polintr	0.09797	0.07231	0.08120	0.10476	0.08808	1.00000	0.11409	0.13361	0.17428
How interested in politics	<.0001	0.0035	0.0009	<.0001	0.0005		<.0001	<.0001	<.0001
	1642	1630	1658	1648	1557	1689	1689	1685	1685
agea	0.05277	-0.12894	0.01421	0.01639	-0.03879	0.11409	1.00000	-0.04559	-0.06130
Age of respondent, calculated	0.0323	<.0001	0.5630	0.5056	0.1258	<.0001		0.0610	0.0117
	1645	1632	1660	1651	1559	1689	1694	1690	1690
badge	0.00729	0.03151	0.03209	0.01212	0.05981	0.13361	-0.04559	1.00000	0.42356
Worn or displayed campaign badge/sticker last 12 months	0.7678	0.2038	0.1919	0.6232	0.0183	<.0001	0.0610		<.0001
	1641	1628	1656	1647	1556	1685	1690	1690	1689
pbldmn	-0.05381	0.06150	-0.03472	-0.01878	0.08473	0.17428	-0.06130	0.42356	1.00000
Taken part in lawful public demonstration last 12 months	0.0293	0.0131	0.1579	0.4463	0.0008	<.0001	0.0117	<.0001	
	1641	1628	1656	1647	1556	1685	1690	1689	1690

Absolute value of Pearson statistic above 0.8 means that correlation between variables is high. In our model none of the variables meets this condition.

We can examine multicollinearity through the Variance Inflation Factor and Tolerance. This can be done by specifying the “vif”, “tol”, and “collin” options after the model statement.

Table 3 Assessment of VIF and Toleration for the two bins model

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	Intercept	1	0.50427	0.03242	15.55	<.0001	.	0
trstprl	Trust in country's parliament	1	0.02278	0.03668	0.62	0.5347	0.64871	1.54151
trstlgl	Trust in the legal system	1	0.06343	0.02969	2.14	0.0328	0.78913	1.26721
trstplt	Trust in politicians	1	0.05281	0.06078	0.87	0.3851	0.45677	2.18929
trstprt	Trust in political parties	1	0.03943	0.06264	0.63	0.5292	0.48676	2.05440
trstep	Trust in the European Parliament	1	0.06664	0.02761	2.41	0.0159	0.90068	1.11027
polintr	How interested in politics	1	0.15910	0.02394	6.65	<.0001	0.93437	1.07023
agea	Age of respondent, calculated	1	0.06777	0.02842	2.38	0.0172	0.96332	1.03808
badge	Worn or displayed campaign badge/sticker last 12 months	1	0.02564	0.04930	0.52	0.6032	0.81714	1.22378
pbldmn	Taken part in lawful public demonstration last 12 months	1	0.12121	0.04995	2.43	0.0154	0.78810	1.26887

Accessing collinearity by reviewing tolerance, we want to make sure that no values fall below 0.1. In the example above, if we split variables into 2 categories, multicollinearity does not occur.

As for variance inflation, the number to look out for is anything above the value of 10. The results are the same as in the case of tolerance - none of the variables needs to be removed from the model.

Table 4 . Collinearity diagnostic for the two bins model

Collinearity Diagnostics												
			Proportion of Variation									
Number	Eigenvalue	Condition Index	Intercept	trstprl	trstlgl	trstplt	trstprt	trstep	polintr	agea	badge	pbldmn
1	4.32632	1.00000	0.00468	0.01268	0.01471	0.00766	0.00774	0.01444	0.00776	0.00676	0.00506	0.00468
2	1.68326	1.60318	0.00346	0.03744	0.00330	0.05876	0.06121	0.00168	0.00622	0.00457	0.05114	0.06639
3	1.15373	1.93646	0.00792	0.00075573	0.00002197	0.01913	0.02293	0.00110	0.00728	0.01679	0.23562	0.19590
4	0.75629	2.39174	0.00455	0.00408	0.23900	0.01118	0.01829	0.43696	0.00951	0.01819	0.02700	0.00265
5	0.60721	2.66926	0.00007953	0.18661	0.34693	0.04660	0.04684	0.34575	0.00025552	0.00106	0.00483	0.00020084
6	0.53922	2.83254	0.00002178	0.00908	0.00107	0.00004896	0.01927	0.04309	0.00145	0.00003034	0.64923	0.64128
7	0.41022	3.24752	0.00176	0.69218	0.33426	0.00259	0.12742	0.12770	0.00224	0.00013237	0.01933	0.05709
8	0.26826	4.01586	0.00123	0.05199	0.01800	0.85152	0.68740	0.02570	3.889979E-9	0.00004257	0.00425	0.00447
9	0.17640	4.95233	0.00109	0.00162	0.01260	0.00002828	0.00063267	0.00037953	0.64940	0.44862	0.00309	0.02593
10	0.07910	7.39574	0.97521	0.00356	0.03011	0.00248	0.00827	0.00320	0.31589	0.50380	0.00045545	0.00141

In review of Collinearity Diagnostics results, our focus is going to be on the relationship of the eigenvalue column to the condition index column. If one or more of the eigenvalues are small (close to zero) and the corresponding condition number large, then we have an indication of multicollinearity[1]. As we can see from the above results, our eigenvalues and condition index associations rather do not match mentioned description.[2]

Model II and III

Taking into consideration models with three and five categories we get the same results - none of the variables meets the conditions neither of Pearson correlation nor tolerance, variance inflation and eigenvalue for multicollinearity. Outputs are provided in the attachment.

---

[1] Multicollinearity: What Is It, Why Should We Care, and How Can It Be Controlled?, Deanna Naomi Schreiber-Gregory, Henry M, Jackson Foundation / National University, 2017

[2]https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_reg_sect038.htm (access: 2019-05-06)

Modelling

Model I - results

The research below compare different binning methods and its effect on model quality and prediction of response variable - vote in last election. For each dataset there were used 9 variables: trstprl, trstlgl, trstplt, trstprt, trstep, polintr, agea, badge and pbldmn. We used backward selection to choose proper variables in the final model (with the required level of significance of 0.05) and used ROC curve to asses fit of the model.

For the first model, backward selection have led to the model consisting of 5 variables (four of them was removed):

$$\ln(vote) = \beta_0 + \beta_1 * \text{trstprl} + \beta_2 * \text{trstprt} + \beta_3 * \text{polintr} + \beta_4 * \text{agea} + \beta_5 * \text{pbldmn}$$

Fit statistics for first model that will be taken into account while comparing the models are presented below.

Table 5 . Assessment of model fit statistics

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	1117.129	1060.249
SC	1121.964	1089.260
-2 Log L	1115.129	1048.249

Moving further, basing on the R-square statistic, fit of the model is 6.94%. All three provided statistics for testing global null hypothesis that all parameters in the model are equal to zero are less than the level of significance. As a result, we reject the null hypothesis.

Table 6. Beta=0 test and R-square

R-Square	0.0694	**Max-rescaled R-Square**	0.0993

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	66.8800	5	<.0001
Score	54.5321	5	<.0001
Wald	45.7435	5	<.0001

Table 7. Concordance statistics

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	48.8	**Somers' D**	0.285
Percent Discordant	20.4	**Gamma**	0.411
Percent Tied	30.8	**Tau-a**	0.117
Pairs	177021	**c**	0.642

In analysis of concordance presented above, value c represents the concordance index and is the percent concordant adjusted for ties. The concordance index also happens to be equivalent to the AUC. The Somers' D statistic is similar to the concordance index, except it resides on a -1 to 1 scale (rather than a 0 to 1 scale) and is equal to 2(c - 0.5). If Sommers' D statistic is closer to 1, the model is better. The values for Gamma and Tau-a also indicate how much better the model is compared to random chance, but they handle ties differently and Tau-a ranges from 0 to 0.5 for the range of concordance values for which the concordance index ranges from 0.5 to 1.[3] In our example results are rather good - there are 48.8% of concordant pairs and Sommers' D statistic is above 0 (equals to 0.285). AUC, main model evaluator, equals to 0.642.

Model II - results

Second model, also after backward variable selection, consists of five different variables splitted into categories (there are nine binary variables in the final model). Variables related to trust are splitted into three categories where medium level is a reference value.

$$\ln(vote) = \beta_0 + \beta_1 * \text{trstprl\_high} + \beta_2 * \text{trstprl\_low} + \beta_3 * \text{trstlgl\_high} + \beta_4 * \text{trstlgl\_low} + \beta_5 * \text{polintr\_high} + \beta_6 * \text{polintr\_low} + \beta_7 * \text{agea0} + \beta_8 * \text{agea1} + \beta_9 * \text{pbldmn\_low}$$

Model fit statistics are worse for three variable categories. In the previous model neither of values below exceeded 1100.

Table 8. Assessment of model fit statistics

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1561.532	1463.228
SC	1566.777	1515.677
-2 Log L	1559.532	1443.228

Basing on the R-square statistic, fit of the model is 7.97% (1.03 pp more than the first one). All three provided statistics for testing global null hypothesis, saying that all parameters in the model are equal to zero, are less than the level of significance. As a result, we reject the null hypothesis.

Table 9. Beta=0 test and R-square

R-Square	0.0797	Max-rescaled R-Square	0.1186

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	116.3045	9	<.0001
Score	109.4010	9	<.0001
Wald	96.2506	9	<.0001

Results for discriminatory performance of the model are similar for the first and second model. We also get Sommers' D statistic above 0 and share of concordant pairs equals to 65.9% which are good results. AUC equals to 0.681.

[3] The Logic and Logistics of Logistic Regression, Lawrence Rasouliyan and Dave P. Miller Ovation Research Group

Table 10. Concordance statistics

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	65.9	Somers' D	0.362
Percent Discordant	29.7	Gamma	0.379
Percent Tied	4.5	Tau-a	0.134
Pairs	362894	c	0.681

Model III - results

Third model, with variables splitted into five categories, has four different variables just like the second one. All variables related to trust has been removed from the model as not meeting the criterion of significance of 0.05.

$$\ln(vote) = \beta_0 + \beta_1 * \text{pbldmn\_no} + \beta_2 * \text{polintr\_high} + \beta_3 * \text{polintr\_low} + \beta_4 * \text{agea0} + \beta_5 * \text{agea1}$$

Table 11. Assessment of model fit statistics

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1201.424	1133.050
SC	1206.358	1162.657
-2 Log L	1199.424	1121.050

Fit statistics are much lower than for the second model (for example, AIC has decreased from 1463 to 1133) which means better fitness.

Table 12. Beta=0 test and R-square

R-Square	0.0735	Max-rescaled R-Square	0.1066

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	78.3734	5	<.0001
Score	74.6204	5	<.0001
Wald	65.3542	5	<.0001

The R-square statistic provide worse results - its value decreased by 0.62 pp. from 7.97% to 7.35% compared to the second model. As in previous models, we reject the null hypothesis that all estimates are equal to zero.

Table 13. Concordance statistics

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	56.5	Somers' D	0.332
Percent Discordant	23.3	Gamma	0.416
Percent Tied	20.3	Tau-a	0.131
Pairs	208222	c	0.666

Percent of concordant pairs has decreased from 65.9 to 56.5 as well as AUC value - to 0.67.

Conclusion

Researched effect is how "High values" (encoded with the logic explained in binning part) affect the vote participation in comparison to "Low values". Confidence level is assumed to be 0.05, there is no high differences in model quality when using different selection methods. The base for assessment is ROC and AUC which verges on the edges of 0.66-0.69 depending on minor model changes. Models have passed each test when taken into account Hosmer-Lemeshow test, global 0-Beta test and comparing to only intercept.

There are no significant changes in the final model quality when interchanging between two, three and five bins but the best results measured by AUC was achieved by second model (with three variable categories). The final form of the model is presented as follows:

$$\ln(vote) = 3.164 - 0.224 * \mathrm{trstprl}_{\mathrm{high}} - 0.567 * \mathrm{trstprl}_{\mathrm{low}} + 0.663 * \mathrm{trstlgl}_{\mathrm{high}} + 0.020 * \mathrm{trstlgl}_{\mathrm{low}} + 1.038 * \mathrm{polintr}_{\mathrm{high}} - 0.891 * \mathrm{polintr}_{\mathrm{low}} - 0.578 * \mathrm{agea0} - 0.239 * \mathrm{agea1} - 1.463 * \mathrm{pbldmn\_low}$$

Table 14. Estimates for the best performing model (with three bins)

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
trstprl High value vs Medium value	0.799	0.487	1.313
trstprl Low value vs Medium value	0.567	0.416	0.773
trstlgl High value vs Medium value	1.941	1.204	3.130
trstlgl Low value vs Medium value	1.020	0.766	1.359
polintr High value vs Medium value	2.824	1.335	5.972
polintr Low value vs Medium value	0.410	0.302	0.558
agea 0 vs 2	0.561	0.383	0.822
agea 1 vs 2	0.787	0.575	1.078
pbldmn Low value vs Medium value	0.232	0.098	0.545

Interpretation of odds ratio estimates is quite surprising. Basing on the outputs above, people that has both high and low trust in country's parliament has less chance of voting compared to the medium trust what does not seem intuitive (by 21.1% and 43.3% respectively). High trust in the legal system increases chance of voting by 94.1% comparing to medium trust and low trust in the legal system increases this chance by 2%. High interest in politics increases chance of voting by 182.4% and low interest - decreases it by 59%, which is in line with expectations and intuition.

Summary

There is observed correlation between trust in government institutions, simple demographic statistics and interests in political situation that can improve evaluates if someone is eager to vote in elections or not. Less important is binning method which was assessed on small differences in AUC values between models that had the same parameters but different transformation of explanatory variables.

Eventually, the model with three bins has turned to out to be the top performing with AUC verging on the edge of 0.67. The highest influence on the model had interest in politics followed by trust in government and trust in parliament and age. Displaying a badge, contrary to initial assumptions, there is no high correlation here. Rest of the variables were removed during the elimination process as they did not meet the 0.05 confidence criteria. Common conclusion from the models is that interest in politics and trust in legal system increases chance to vote.

Indifferently on the criteria and inputs each model passed the Hosmer-Lemeshow test, global 0-Beta test and comparing to only intercept.

The further research can be carried out to check distortion of sample data in comparison to population data as in dataset the group of no voting people is underrepresented. Elections is the process where population data are widely accessible and from the last decade local government, government and presidential election participation rate is varying from the 40%-50% to 55% in extreme cases. Whilst in examined ESS set in total 70% went voting with range of 50% to almost 80% depending on the age group. There is a possibility that in survey took part people that are engaged and like to express their views what led to distorted results on the dependent variable.

Bibliography

- Multicollinearity: What Is It, Why Should We Care, and How Can It Be Controlled?, Deanna Naomi Schreiber-Gregory, Henry M, Jackson Foundation / National University, 2017
- The Logic and Logistics of Logistic Regression, Lawrence Rasouliyan and Dave P. Miller Ovation Research Group
- Logistic Regression, A Self-Learning Text, Second Edition, David G. Kleinbaum, Mitchel Klein, Atlanta, 2002
- ESS Data Documentation, http://nesstar.ess.nsd.uib.no/webview/ (access: 2019-04-15)
- SAS Documentation on Logistic Procedure: https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_logistic_sect011.htm (access: 2019-05-11)
- https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_reg_sect038.htm (access: 2019-05-06)

Attachments

Code

Code that generated model with the best results basing on the main evaluation statistics.

```
ods graphics on;
proc logistic data=work.pasted_3_all plots = (ROC EFFECT) simple;
      class trstprl trstlgl trstplt trstprt trstep polintr agea badge
pbldmn vote / param=reference ref=LAST;
      model vote(event='Yes')= trstprl trstlgl trstplt trstprt trstep
polintr agea badge pbldmn /
      outroc=work.roc
      details
      lackfit rsquare
      corrb covb
      expb
      selection=backward slstay=0.05
      ctable pprob=0.3;
      format vote vote_binned_three. trstprl trstlgl trstplt trstprt
trstep polintr badge pbldmn trust_binned_three.;
      output out=trebles predprobs=(cumulative) predicted=p;
      ods output parameterestimates=est covb=cov;
run;
ods graphics off;
```

Outputs for Model I

Table 15. Maximum likelihood estimates for the model with two bins

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	0.6874	0.0936	53.9792	<.0001	1.988
trstprl	High value	1	0.5587	0.2648	4.4517	0.0349	1.748
trstlgl	High value	1	0.5774	0.2147	7.2346	0.0072	1.781
polintr	High value	1	1.3837	0.3840	12.9832	0.0003	3.990
agea	0	1	-0.4871	0.1796	7.3538	0.0067	0.614
pbldmn	High value	1	1.8071	0.7464	5.8613	0.0155	6.093

Outputs for Model II

Table 16. Correlation assessment for the model with three bins

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	trstprl	trstlgl	trstplt	trstprt	trstep	polintr	agea	badge	pbldmn
trstprl	1.00000	0.42545	0.53799	0.52298	0.21131	0.06561	-0.00905	-0.03105	-0.08226
Trust in country's parliament		<.0001	<.0001	<.0001	<.0001	0.0078	0.7136	0.2087	0.0009
	1645	1606	1629	1621	1541	1642	1645	1641	1641
trstlgl	0.42545	1.00000	0.35089	0.36154	0.31549	0.03041	-0.11188	0.01884	0.01370
Trust in the legal system	<.0001		<.0001	<.0001	<.0001	0.2199	<.0001	0.4474	0.5806
	1606	1632	1615	1606	1533	1630	1632	1628	1628
trstplt	0.53799	0.35089	1.00000	0.77561	0.28196	0.08949	0.02510	0.01579	0.00488
Trust in politicians	<.0001	<.0001		<.0001	<.0001	0.0003	0.3067	0.5207	0.8426
	1629	1615	1660	1641	1551	1658	1660	1656	1656
trstprt	0.52298	0.36154	0.77561	1.00000	0.30351	0.08480	-0.00892	0.00832	-0.00967
Trust in political parties	<.0001	<.0001	<.0001		<.0001	0.0006	0.7172	0.7359	0.6949
	1621	1606	1641	1651	1543	1648	1651	1647	1647
trstep	0.21131	0.31549	0.28196	0.30351	1.00000	0.04861	-0.02773	0.02433	0.04573
Trust in the European Parliament	<.0001	<.0001	<.0001	<.0001		0.0552	0.2739	0.3376	0.0713
	1541	1533	1551	1543	1559	1557	1559	1556	1556
polintr	0.06561	0.03041	0.08949	0.08480	0.04861	1.00000	0.11457	0.13361	0.17428
How interested in politics	0.0078	0.2199	0.0003	0.0006	0.0552		<.0001	<.0001	<.0001
	1642	1630	1658	1648	1557	1689	1689	1685	1685
agea	-0.00905	-0.11188	0.02510	-0.00892	-0.02773	0.11457	1.00000	-0.05329	-0.07384
Age of respondent, calculated	0.7136	<.0001	0.3067	0.7172	0.2739	<.0001		0.0285	0.0024
	1645	1632	1660	1651	1559	1689	1694	1690	1690
badge	-0.03105	0.01884	0.01579	0.00832	0.02433	0.13361	-0.05329	1.00000	0.42356
Worn or displayed campaign badge/sticker last 12 months	0.2087	0.4474	0.5207	0.7359	0.3376	<.0001	0.0285		<.0001
	1641	1628	1656	1647	1556	1685	1690	1690	1689
pbldmn	-0.08226	0.01370	0.00488	-0.00967	0.04573	0.17428	-0.07384	0.42356	1.00000
Taken part in lawful public demonstration last 12 months	0.0009	0.5806	0.8426	0.6949	0.0713	<.0001	0.0024	<.0001	
	1641	1628	1656	1647	1556	1685	1690	1689	1690

Table 17. Parameter estimates for the model with three bins

Parameter Estimates								
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Tolerance**	**Variance Inflation**
Intercept	Intercept	1	0.48107	0.03275	14.69	<.0001	.	0
trstprl	Trust in country's parliament	1	0.09538	0.02796	3.41	0.0007	0.63534	1.57396
trstlgl	Trust in the legal system	1	0.00682	0.02580	0.26	0.7916	0.75682	1.32132
trstplt	Trust in politicians	1	0.01903	0.03986	0.48	0.6331	0.36821	2.71587
trstprt	Trust in political parties	1	-0.01190	0.03979	-0.30	0.7649	0.37621	2.65808
trstep	Trust in the European Parliament	1	0.02968	0.02400	1.24	0.2164	0.86706	1.15332
polintr	How interested in politics	1	0.16228	0.02387	6.80	<.0001	0.93999	1.06384
agea	Age of respondent, calculated	1	0.04390	0.01653	2.66	0.0080	0.96512	1.03614
badge	Worn or displayed campaign badge/sticker last 12 months	1	0.03398	0.04926	0.69	0.4904	0.81853	1.22171
pbldmn	Taken part in lawful public demonstration last 12 months	1	0.14611	0.04990	2.93	0.0035	0.78990	1.26598

Table 18. Collinearity assessment for model with three bins

Collinearity Diagnostics												
		Condition	**Proportion of Variation**									
Number	**Eigenvalue**	**Index**	**Intercept**	**trstprl**	**trstlgl**	**trstplt**	**trstprt**	**trstep**	**polintr**	**agea**	**badge**	**pbldmn**
1	5.68677	1.00000	0.00294	0.00722	0.00722	0.00476	0.00481	0.00805	0.00467	0.00562	0.00230	0.00217
2	1.39182	2.02135	0.00030269	0.01053	0.00082558	0.00995	0.01100	0.00000274	0.00231	0.00015724	0.21923	0.22408
3	0.92240	2.48298	0.01340	0.00836	0.00030426	0.06276	0.06810	0.00554	0.02048	0.05118	0.06683	0.03999
4	0.54312	3.23581	0.00010855	0.00429	0.00001094	0.00067909	0.00099307	0.00384	0.00008623	0.00170	0.70308	0.65203
5	0.40297	3.75661	0.00182	0.00649	0.23573	0.04146	0.02637	0.30130	0.02569	0.14776	0.00118	0.00254
6	0.35552	3.99945	0.00019155	0.37015	0.10870	0.01989	0.03555	0.42631	0.00222	0.00848	0.00326	0.02594
7	0.25179	4.75239	0.00209	0.57474	0.47279	0.00690	0.03790	0.20812	0.01601	0.00397	0.00026505	0.01555
8	0.20486	5.26868	0.00816	0.00008283	0.09797	0.00122	0.00562	0.00074321	0.49437	0.54726	0.00352	0.03485
9	0.15609	6.03604	0.00057440	0.00953	0.00367	0.84234	0.80963	0.00108	0.00482	0.00580	0.00014368	0.00139
10	0.08465	8.19646	0.97042	0.00862	0.07279	0.01004	0.00001693	0.04501	0.42936	0.22808	0.00019351	0.00146

Table 19. Maximum likelihood estimates for the model with three bins

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	3.1640	0.4721	44.9090	<.0001	23.666
trstprl	High value	1	-0.2238	0.2533	0.7807	0.3769	0.799
trstprl	Low value	1	-0.5671	0.1577	12.9318	0.0003	0.567
trstlgl	High value	1	0.6632	0.2438	7.4035	0.0065	1.941
trstlgl	Low value	1	0.0201	0.1463	0.0188	0.8909	1.020
polintr	High value	1	1.0380	0.3822	7.3781	0.0066	2.824
polintr	Low value	1	-0.8909	0.1570	32.2172	<.0001	0.410
agea	0	1	-0.5782	0.1948	8.8057	0.0030	0.561
agea	1	1	-0.2392	0.1604	2.2234	0.1359	0.787
pbldmn	Low value	1	-1.4626	0.4364	11.2326	0.0008	0.232

Outputs for Model III

Table 20. Correlation assessment for the model with five bins

Pearson Correlation Coefficients, N = 1407 Prob > |r| under H0: Rho=0	trstprl	trstlgl	trstplt	trstprt	trstep	polintr	agea	badge	pbldmn
trstprl	1.00000	0.54673	0.70360	0.67016	0.18330	-0.10686	0.04894	-0.00254	0.01801
Trust in country's parliament		<.0001	<.0001	<.0001	<.0001	<.0001	0.0665	0.9242	0.4996
trstlgl	0.54673	1.00000	0.52302	0.51272	0.37743	-0.06638	-0.06152	-0.04178	-0.06872
Trust in the legal system	<.0001		<.0001	<.0001	<.0001	0.0128	0.0210	0.1172	0.0099
trstplt	0.70360	0.52302	1.00000	0.87502	0.32720	-0.13191	0.03269	-0.01897	-0.00891
Trust in politicians	<.0001	<.0001		<.0001	<.0001	<.0001	0.2204	0.4771	0.7385
trstprt	0.67016	0.51272	0.87502	1.00000	0.34854	-0.16106	0.02167	-0.02685	0.00584
Trust in political parties	<.0001	<.0001	<.0001		<.0001	<.0001	0.4167	0.3143	0.8269
trstep	0.18330	0.37743	0.32720	0.34854	1.00000	-0.06479	0.00318	-0.04451	-0.02551
Trust in the European Parliament	<.0001	<.0001	<.0001	<.0001		0.0151	0.9052	0.0951	0.3390
polintr	-0.10686	-0.06638	-0.13191	-0.16106	-0.06479	1.00000	-0.06313	0.06403	0.08137
How interested in politics	<.0001	0.0128	<.0001	<.0001	0.0151		0.0179	0.0163	0.0023
agea	0.04894	-0.06152	0.03269	0.02167	0.00318	-0.06313	1.00000	0.02203	0.02880
Age of respondent, calculated	0.0665	0.0210	0.2204	0.4167	0.9052	0.0179		0.4089	0.2804
badge	-0.00254	-0.04178	-0.01897	-0.02685	-0.04451	0.06403	0.02203	1.00000	0.61148
Worn or displayed campaign badge/sticker last 12 months	0.9242	0.1172	0.4771	0.3143	0.0951	0.0163	0.4089		<.0001
pbldmn	0.01801	-0.06872	-0.00891	0.00584	-0.02551	0.08137	0.02880	0.61148	1.00000
Taken part in lawful public demonstration last 12 months	0.4996	0.0099	0.7385	0.8269	0.3390	0.0023	0.2804	<.0001	

Table 21. Parameter estimates for the model with five bins

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	Intercept	1	0.95471	0.07606	12.55	<.0001	.	0
trstprl	Trust in country's parliament	1	-0.00582	0.00644	-0.90	0.3669	0.43647	2.29110
trstlgl	Trust in the legal system	1	-0.00839	0.00606	-1.38	0.1664	0.59315	1.68591
trstplt	Trust in politicians	1	0.00230	0.01103	0.21	0.8348	0.20670	4.83793
trstprt	Trust in political parties	1	-0.01452	0.01102	-1.32	0.1879	0.22006	4.54412
trstep	Trust in the European Parliament	1	-0.00705	0.00512	-1.38	0.1687	0.79846	1.25241
polintr	How interested in politics	1	0.10410	0.01312	7.93	<.0001	0.96219	1.03929
agea	Age of respondent, calculated	1	-0.00062339	0.00036556	-1.71	0.0884	0.98099	1.01938
badge	Worn or displayed campaign badge/sticker last 12 months	1	0.04317	0.03603	1.20	0.2311	0.62343	1.60402
pbldmn	Taken part in lawful public demonstration last 12 months	1	0.03539	0.03425	1.03	0.3016	0.61691	1.62099

Table 22. Collinearity assessment for model with five bins

Collinearity Diagnostics												
		Condition	Proportion of Variation									
Number	Eigenvalue	Index	Intercept	trstprl	trstlgl	trstplt	trstprt	trstep	polintr	agea	badge	pbldmn
1	8.27593	1.00000	0.00028755	0.00186	0.00193	0.00099722	0.00105	0.00259	0.00111	0.00292	0.00031673	0.00034819
2	0.77844	3.26058	0.00160	0.02470	0.00443	0.02750	0.02837	0.00003570	0.01148	0.03094	0.00218	0.00236
3	0.28154	5.42171	0.00017233	0.02488	0.04563	0.00524	0.00287	0.24840	0.00445	0.50966	0.00013854	0.00011828
4	0.22932	6.00742	0.00164	0.07179	0.00245	0.00020293	0.00223	0.44542	0.03248	0.30531	0.00410	0.00447
5	0.16670	7.04591	0.00092331	0.16822	0.42487	0.05413	0.08533	0.00020793	0.01264	0.06369	0.00319	0.00478
6	0.10716	8.78813	0.00004108	0.68597	0.48193	0.01808	0.04038	0.27531	0.00174	0.02340	0.00003360	0.00051303
7	0.07587	10.44424	0.00326	0.00239	0.00486	0.02680	0.01071	0.00609	0.74218	0.01885	0.04905	0.05627
8	0.05264	12.53919	0.00001494	0.01839	0.00012941	0.86263	0.82222	0.00101	0.02970	0.00139	0.00493	0.00121
9	0.01746	21.77353	0.71888	0.00008800	0.03353	0.00008328	0.00003517	0.01176	0.11797	0.03415	0.01263	0.54997
10	0.01495	23.53024	0.27319	0.00171	0.00022976	0.00434	0.00680	0.00917	0.04624	0.00969	0.92344	0.37996

Table 23. Maximum likelihood estimates for the model with five bins

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	1.7667	0.2508	49.6344	<.0001	5.852
agea	0	1	-0.3223	0.1186	7.3804	0.0066	0.724
agea	1	1	-0.0252	0.0978	0.0665	0.7965	0.975
polintr	High value	1	1.0168	0.2941	11.9505	0.0005	2.764
polintr	Low value	1	-1.0145	0.1790	32.1305	<.0001	0.363
pbldmn	No	1	-0.6847	0.2198	9.7002	0.0018	0.504