Problem 1

The first goal of the current research was to identify whether there's a statistically significant relationship between participation in Erasmus exchange and language skills. For this task the crosstabulation of those two variables were constructed with the corresponding χ^2 test. Here is the complete table showing the number of observations and the column percentages for different group interactions of variables participation in Erasmus exchange and language skills.

Language skills * Participation in Erasmus exchange Crosstabulation

			Participation in Erasmus exchange			
			Before the exchange	During the exchange	After the exchange	Total
Language skills	Poor to moderate	Count	22	28	28	78
		Expected Count	17,3	33,4	27,2	78,0
		% within Participation in Erasmus exchange	52,4%	34,6%	42,4%	41,3%
	Good	Count	18	44	26	88
		Expected Count	19,6	37,7	30,7	88,0
		% within Participation in Erasmus exchange	42,9%	54,3%	39,4%	46,6%
	Fluent	Count	2	9	12	23
		Expected Count	5,1	9,9	8,0	23,0
		% within Participation in Erasmus exchange	4,8%	11,1%	18,2%	12,2%
Total		Count	42	81	66	189
		Expected Count	42,0	81,0	66,0	189,0
		% within Participation in Erasmus exchange	100,0%	100,0%	100,0%	100,0%

The first step was to determine whether such a test could be applied.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	7,988[a]	4	,092

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,11.

According to the footnote in the table, there's 0% of cells containing less than 5 observations, which means that χ^2 test will produce reliable results. The next step concerns the null hypothesis of χ^2 test. As p value of 0.092 is below the significance level of 0.1 (we can use this level of alpha due to total number of observations =189 <200) we reject the H_0. It means that those difference in percentages shown in the table above is statistically significant.

Symmetric Measures

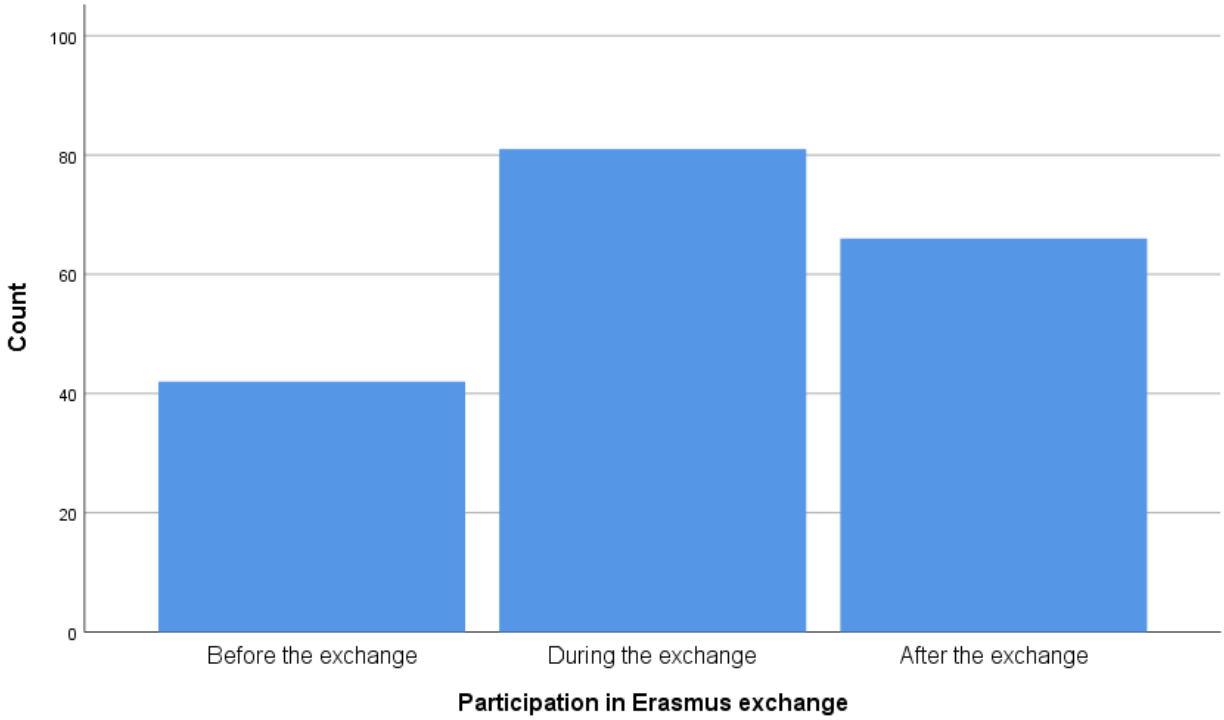
	Value	Approximate Significance
Cramer's V	,145	,092

Cramer's V is seemed to be not meaningful cause it is lease than 0.3, but his particular metric requires more complex analysis.

Problem 2

The second problem concerns examination the difference between group means of willingness to emigrate to Poland based on participation of Erasmus exchange.

Here we can see on the bar plot that there's 3 group of independent variable, so wee need to use ANOVA instead of T-test to compare group means.



First of all we should check the assumptions of normality of errors. Just for simplification task we can explore the skewness and kurtosis of original variable which still will present a meaningful result.

Descriptives

		Statistic	Std. Error
Willingness to emigrate within one year after graduation	Skewness	-,329	,177
	Kurtosis	-,967	,352

We can see that both skewness and kurtosis is within [-1.5;1.5] range, which means that we do not expect non-normal residuals.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Willingness to emigrate within one year after graduation	Based on Mean	2,263	2	186	,107

As we see, we do not reject null hypothesis in homogeneity test (p value = 0.107 > 0.1 and 0.05). Thus, the second assumption is also fulfilled and we can interpret our coefficients.

Multiple Comparisons

Dependent Variable: Willingness to emigrate within one year after graduation

LSD

(I) Participation in Erasmus exchange	(J) Participation in Erasmus exchange	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Before the exchange	During the exchange	,087	,289	,763	-,48	,66
	After the exchange	-,251	,300	,404	-,84	,34
During the exchange	Before the exchange	-,087	,289	,763	-,66	,48
	After the exchange	-,338	,252	,181	-,84	,16
After the exchange	Before the exchange	,251	,300	,404	-,34	,84
	During the exchange	,338	,252	,181	-,16	,84

However, we see that none of the means is significantly different from each other.

Problem 3

Before building a regression we need to check variables for multicollinearity. In case of attitudes towards political and economic situation in Poland, religiosity, family and tradition, we see that they are correlated (with the help of correlation matrix). Thus, we need to compare factor analysis.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,887
Bartlett's Test of Sphericity	Approx. Chi-Square	2377,699
	df	210
	Sig.	,000

First of all we examine KMO and Barlett's statistics. We see that KMO is way above cutoff point of 0.5 and p value of Barlett's test is left than alpha. It means that correlation matrix is significant from Identity matrix and factor analysis is appropriate.

Total Variance Explained

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
1	7,578	36,087	36,087	4,982	23,723	23,723
2	2,525	12,025	48,111	4,347	20,698	44,421
3	1,879	8,948	57,060	2,361	11,244	55,665
4	1,647	7,841	64,901	1,940	9,236	64,901

Extraction Method: Principal Component Analysis.

Then we examine the total variance explained by our common factors. We see that 4 components explains about 65% of all variance, which is pretty good result.

Communalities

	Initial	Extraction
The current government is acting on behalf of all Poles.	1,000	,512
The government is just.	1,000	,649
I'm prouder of being Polish under the current government than before.	1,000	,794
The current government gives Poland a good image abroad.	1,000	,732

I feel safe in Poland under the current government.	1,000	,665
Overall, I think the government makes good choices.	1,000	,846
Religion is important in my life.	1,000	,828
Church should be supported by state.	1,000	,649
Religion should be taught in public schools.	1,000	,653
I attend religious services regularly.	1,000	,800
I pray daily.	1,000	,751
One cannot achieve happiness in life without religion.	1,000	,693
Religious people are more reliable than not religious.	1,000	,592
I cannot afford to go out once a week.	1,000	,431
My job is secured and stable.	1,000	,404
I believe that my living conditions are good.	1,000	,563
My current financial situation is better now than it was a year ago.	1,000	,546
I often participate in local events in my local community.	1,000	,522
In everything I do, I always have in mind, first and foremost, the good of Poland.	1,000	,664
I'm attached to Poland's traditions and values (family, religion).	1,000	,679
I am proud of being Polish/ have Polish decent.	1,000	,658

Extraction Method: Principal Component Analysis.

Then I'd like to show the table which shows how much of variance was explained in each variable by those factors. The cutoff point here is 0.3 so each result is significant.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
Religion is important in my life.	,881			
I attend religious services regularly.	,865			
I pray daily.	,849			
One cannot achieve happiness in life without religion.	,818			
Religion should be taught in public schools.	,732			
Religious people are more reliable than not religious.	,718			
Church should be supported by state.	,703	,352		
Overall, I think the government makes good choices.		,889		
The current government gives Poland a good image abroad.		,837		
I'm prouder of being Polish under the current government than before.		,832		
I feel safe in Poland under the current government.		,770		
The government is just.		,764		
The current government is acting on behalf of all Poles.		,701		
I am proud of being Polish/ have Polish decent.			,742	
I often participate in local events in my local community.			,720	

In everything I do, I always have in mind, first and foremost, the good of Poland.		,346	,706	
I'm attached to Poland's traditions and values (family, religion).	,421		,690	
I believe that my living conditions are good.				,742
My current financial situation is better now than it was a year ago.				,715
My job is secured and stable.				,616
I cannot afford to go out once a week.				-,603

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

The next matrix shows which factors explains which variables best. I decided to create 4 components and call them religion, political, tradition ,finance.

Then we needed to transform dependent variable cause it consisted of more than 2 groups.

Statistics

Willingness to emigrate within one year after graduation

N	Valid	189
	Missing	0

Willingness to emigrate within one year after graduation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not willing at all	11	5,8	5,8	5,8
	2	26	13,8	13,8	19,6
	3	32	16,9	16,9	36,5
	4	37	19,6	19,6	56,1
	5	44	23,3	23,3	79,4
	I'll surely emigrate.	39	20,6	20,6	100,0
	Total	189	100,0	100,0	

I decided to recode 5 and 6 as 0 and all other groups as 1.

willingness_bin

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	83	43,9	43,9	43,9
	1,00	106	56,1	56,1	100,0
	Total	189	100,0	100,0	

Next I held regression analysis. In a result we first obtain classification table of the original model with no independent variables. We see that it's accuracy of predicting is 56,5%.

Classification Table[a,b]

			Predicted		
			willingness_bin		Percentage
	Observed		,00	1,00	Correct
Step 0	willingness_bin	,00	0	81	,0
		1,00	0	105	100,0
	Overall Percentage				56,5

a. Constant is included in the model.

b. The cut value is ,500

Then I included a table with pseudo-R-squared variables (Nagelkerke). We see that the last models shows the best result so I decided to investigate it. The final models shows 0.521 of reduction in the prediction error as compared to the baseline model.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	193,840[a]	,279	,374
2	187,813[b]	,302	,405
3	182,965[b]	,320	,429
4	177,436[b]	,340	,456
5	172,617[b]	,357	,479
6	168,034[b]	,373	,500
7	163,301[b]	,388	,521

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

b. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 7		4,734	1	,030
		35,570	7	,000
		91,445	18	,000

The next test is about whether the final model is more accurate than the previous and we see that is actually true due to low sig.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8,483	8	,388
2	10,360	8	,241
3	4,327	8	,827
4	5,628	8	,689
5	7,676	8	,466
6	11,028	8	,200
7	11,995	8	,151

Classification Table[a]

			Predicted		
			willingness_bin		Percentage
	Observed		,00	1,00	Correct
Step 1	willingness_bin	,00	59	22	72,8
		1,00	22	83	79,0
	Overall Percentage				76,3
Step 2	willingness_bin	,00	59	22	72,8
		1,00	23	82	78,1
	Overall Percentage				75,8
Step 3	willingness_bin	,00	61	20	75,3
		1,00	24	81	77,1
	Overall Percentage				76,3
Step 4	willingness_bin	,00	61	20	75,3
		1,00	21	84	80,0
	Overall Percentage				78,0
Step 5	willingness_bin	,00	61	20	75,3
		1,00	23	82	78,1
	Overall Percentage				76,9
Step 6	willingness_bin	,00	62	19	76,5

		1,00	21	84	80,0
	Overall Percentage				78,5
Step 7	willingness_bin	,00	64	17	79,0
		1,00	21	84	80,0
	Overall Percentage				79,6

a. The cut value is ,500

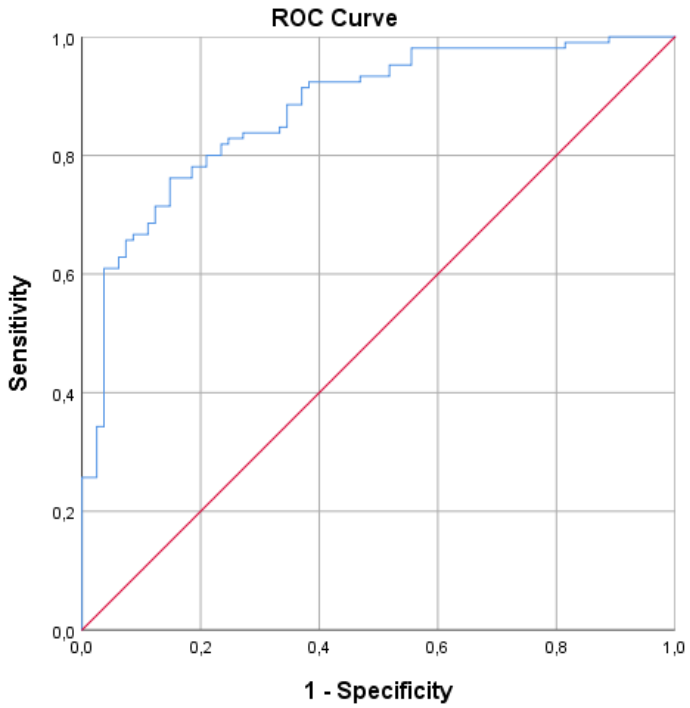
The next table shows us how the accuracy has improved. It's actually pretty well (79,6-56.5)/56.5

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 7[g]	religion	1,082	,285	14,376	1	,000	2,952
	political	,088	,225	,153	1	,696	1,092
	tradition	,474	,240	3,891	1	,049	1,606
	finance	1,083	,339	10,198	1	,001	2,955
	Q.4=Poor to moderate	3,239	1,461	4,914	1	,027	25,518
	Q.4=Good	3,345	1,467	5,202	1	,023	28,368
	Q.1=Before the exchange	-,372	,575	,420	1	,517	,689
	Q.1=During the exchange	-3,006	1,210	6,175	1	,013	,049
	Gender	-1,840	,927	3,939	1	,047	,159
	Work status	-1,027	,487	4,456	1	,035	,358
	Family members living abroad	-,265	,471	,316	1	,574	,767
	religion * erasmus_6	-,917	,451	4,145	1	,042	,400
	tradition * erasmus_6	,967	,466	4,298	1	,038	2,630
	finance * erasmus_5	-,995	,446	4,973	1	,026	,370
	D.1=Male * Q.5.2=No	20,633	7833,746	,000	1	,998	913795831,278
	D.1=Female * erasmus_5	2,554	1,203	4,506	1	,034	12,863
	Q.2=Yes * language_3	4,023	1,753	5,268	1	,022	55,885
	language_3 * erasmus_4	25,136	22750,794	,000	1	,999	82518689283,783
	Constant	-,462	1,711	,073	1	,787	,630

g. Variable(s) entered on step 7: tradition * erasmus_6.

1 standard deviation in political corresponds to increased odds of being classified as group 1 by 0.092%.



ROC curve is wery well with an AUC more about 88%.

Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,877	,025	,000	,828	,926

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s): Predicted probability

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
,0000000	1,000	1,000
,0026134	1,000	,988
,0050948	1,000	,975
,0074446	1,000	,963
,0125766	1,000	,951
,0195934	1,000	,938

,0248917	1,000	,926
,0308946	1,000	,914
,0361368	1,000	,901
,0524169	1,000	,889
,0686412	,990	,889
,0706882	,990	,877
,0725605	,990	,864
,0756587	,990	,852
,0789790	,990	,840
,0952870	,990	,827
,1104733	,990	,815
,1141772	,981	,815
,1204976	,981	,802
,1256496	,981	,790
,1334678	,981	,778
,1424748	,981	,765
,1512251	,981	,753
,1578583	,981	,741
,1687144	,981	,728
,1797840	,981	,716
,1820780	,981	,704
,1854166	,981	,691
,1941543	,981	,679
,2010107	,981	,667
,2024015	,981	,654
,2061662	,981	,642
,2112939	,981	,630
,2153036	,981	,617
,2194111	,981	,605
,2255063	,981	,593
,2340913	,981	,580
,2400720	,981	,568
,2415917	,981	,556
,2437973	,971	,556
,2518394	,962	,556
,2617217	,952	,556
,2684064	,952	,543
,2730237	,952	,531
,2751706	,952	,519

,2836277	,943	,519
,2920122	,933	,519
,2976603	,933	,506
,3031815	,933	,494
,3034701	,933	,481
,3049730	,933	,469
,3159087	,924	,469
,3319195	,924	,457
,3387891	,924	,444
,3396396	,924	,432
,3421197	,924	,420
,3454254	,924	,407
,3478373	,924	,395
,3498523	,924	,383
,3527039	,914	,383
,3566512	,914	,370
,3592691	,905	,370
,3670792	,895	,370
,3745935	,886	,370
,3815034	,886	,358
,3929576	,886	,346
,3982741	,876	,346
,4045556	,867	,346
,4125273	,857	,346
,4156635	,848	,346
,4187046	,848	,333
,4235633	,838	,333
,4273216	,838	,321
,4311384	,838	,309
,4377019	,838	,296
,4448452	,838	,284
,4509959	,838	,272
,4536094	,829	,272
,4541772	,829	,259
,4563357	,829	,247
,4580374	,819	,247
,4634603	,819	,235
,4708325	,810	,235
,4765426	,800	,235

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Cutoff point is ,4765426.