

**CDA Final Project** 

**Submitted by: Nour Mohamed Nashat Eletreby** 

Code: 5200917

**Submitted to: Dr Sally Farid** 

Due to: 9/1/2024

**Importance of Family in Singapore** 

#### **Introduction:**

Family has always been an important value in each person's life but of course as everything its importance varies from one person to another. Hence, my topic of interest is to analyze the importance of family in my assigned country Singapore; I will be using the data from World Values Survey (WVS – Wave 7) with 2012 units.

I want to analyze the effect of importance of leisure time, importance of work, person's goal to make his/her parents proud, age and sex on the importance of family using SPSS software.

Codes and understanding of variables:

My response variable is: family which is question 1 in the questionnaire; a categorical variable and its categories with codes are 1-Very important 2-Rather important 3-Not very important 4-Not at all important

My explanatory variables are:

First: Leisure time which is question 3 in the questionnaire; a categorical variable and its categories with codes are 1-Very important 2-Rather important 3-Not very important 4-Not at all important Second: Work which is question 5 in the questionnaire; a categorical variable and its categories with codes are 1-Very important 2-Rather important 3-Not very important 4-Not at all important Third: One of my goals in life has been to make my parents proud which is statement 27 in the questionnaire; a categorical variable and its categories with codes are 1-Strongly agree 2-Agree 3-Disagree 4-Strongly disagree

Fourth: Sex which is question 260 in the questionnaire in the demographics section; a categorical variable and its categories with codes are 1- Male and 2- Female

Fifth: Age which is question 262 in the questionnaire in the demographics section; a continuous Variable

# **Logistic Binary Analysis**

Since my response variable; importance of family has 4 categories I decided to recode the variables into only 2 categories for logistic binary analysis so they will more generalized categories.

#### The recodes:

0 will be (not important) and will be taken for categories 3 and 4 (3-Not very important 4-Not at all important) which is reference category

While 1 will be (important) and will be taken for categories 1 and 2 (1-Very important 2-Rather important)

Family

					Ourse de tirre
					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	0	20	1.0	1.0	1.0
	1	1990	98.9	99.0	100.0
	Total	2010	99.9	100.0	
Missing	System	2	.1		
Total		2012	100.0		

Figure (1): frequency table for the transformed response variable

Comment: it is clear here that only 20 individuals are within family is not important category while 1990 individuals are within family is important category

Coefficients<sup>a</sup>

		Collinearity Statistics		
Model		Tolerance	VIF	
1	Important in life: Leisure time	.963	1.039	
	Important in life: Work	.878	1.139	
	One of main goals in life has been to make my parents proud	.960	1.042	
	Sex	.984	1.016	
	Age	.905	1.105	

a. Dependent Variable: Family

Figure (2): VIF values for all variables

Comment: I used variance inflation factor (VIF) to check for multicollinearity, here I can see no indication for multicollinearity as VIF is less than 5 for all variables

# **Logistic Binary Model:**

I will define my base category for all variables and the interaction term I will be using as well.

for all categorical variables we have 3 dummies as we have 4 categories (following the rule k-1) to avoid multicollinearity

For Q3 importance of leisure time; the base category is not at all important and the dummies (1-Very important 2-Rather important 3-Not very important)

For Q5 importance of work; the base category is not at all important and the dummies (1-Very important 2-Rather important 3-Not very important)

For Q27 one of my goals in life has been to make my parents proud; the base category is not at all important and the dummies (1-Very important 2-Rather important 3-Not very important)

Q260 sex; the base category is female and the dummy (1-male)

The interaction term: sex (Q 260) and one of my main goals in life has been to make my parents proud (Q 27)

## Choosing the appropriate model

I will try the model with enter method and backward method (check the appendix for further explanation)

#### The enter method

Including all variables importance of leisure time, importance of work, one of main goals has been to make my proud, sex and age as well as the interaction term (sex and one of my main goals has been to make my parents proud)

#### **Model Summary**

,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,							
		Cox & Snell R	Nagelkerke R				
Step	-2 Log likelihood	Square	Square				
1	177.222ª	.023	.218				

Figure (3): model summary for model using enter method

#### The backward method

Including all variables importance of leisure time, importance of work, one of main goals has been to make my proud, sex and age as well as the interaction term (sex and one of my main goals has been to make my parents proud) and the software reached the best variables to enter the model

#### **Model Summary**

		Cox & Snell R	Nagelkerke R
Step	-2 Log likelihood	Square	Square
1	177.222ª	.023	.218
2	180.088 <sup>b</sup>	.022	.205
3	180.690 <sup>b</sup>	.021	.202
<mark>4</mark>	<mark>182.402<sup>b</sup></mark>	<mark>.021</mark>	<mark>.194</mark>

Figure (4): model summary for model using backward method

# Since the backward method model is nested within the enter method model (nested models) I will use The likelihood ratio test or the deviance to check which model is better

Consider two models: M0 with fitted values  $\mu$ 0 and M1 with fitted values  $\mu$ 1 with M0 a special case of (i.e. nested within) M1 • Since M0 is simpler (i.e. has fewer parameters than M1), then:  $L(\mu \ 0; y) \le L(\mu 1; y)$ 

• For testing the null hypothesis that M0 has a better fit than M1

H0: M0 has a better fit that M1

The likelihood ratio test statistic is  $-2[L(\mu 0; y) - L(\mu 1; y)]$ 

 $=D(y; \mu 0)-D(y; \mu 1)$ 

Here in our case M0 is the backward method model (simpler) and M1 is the enter method model

Test statistic= 5.18

 $X^2 = 5.991$ 

Do not reject H0, therefore M0 has a better fit than M1

We will use the simpler model that is backward method.

## The model used is the backward method (my final model)

After the software conducted the backward method the variables that entered the model are: Q3 (importance of leisure time), Q5 (importance of work) and Q27 (one of my main goals in life has been to make my parents proud)

# The classification table obtained by the software as default

Classification Table<sup>a</sup>

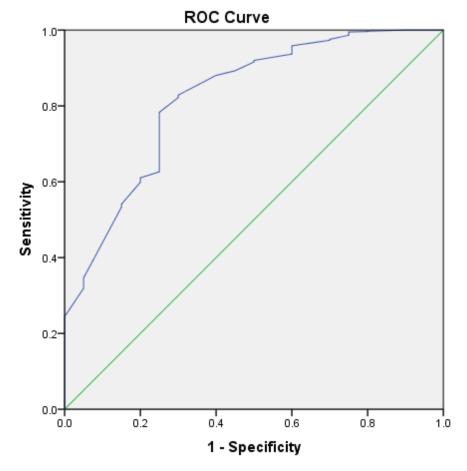
			Predicted				
	<u> </u>		Far	nily	Percentage		
	Observed	I	0	1	Correct		
Step 1	Family	0	0	20	.0		
		1	0	1977	100.0		
	Overall P	ercentage			99.0		
Step 2	Family	0	0	20	.0		
	-	1	0	1977	100.0		
	Overall P	ercentage			99.0		
Step 3	Family	0	0	20	.0		
		1	0	1977	100.0		
	Overall P	ercentage			99.0		
Step 4	<mark>Family</mark>	<mark>0</mark>	1	<mark>19</mark>	<mark>5.0</mark>		
	-	<mark>1</mark>	0	<mark>1977</mark>	<mark>100.0</mark>		
	Overall P	ercentage			<mark>99.0</mark>		

a. The cut value is .500

Figure (5): classification table for my model with cut 0.500

The overall percentage of correct classification in this model is 99%; the model has high accuracy on the other hand focusing on the correct percentage of classification for each category it is found that the percent of correct classification of "family is not important" is 5% which is extremely low but the percent of correct classification in "family is important" is 100% which is an impressive percentage.

Conducting Receiving Operating Characteristic (ROC) Curve to know optimal cut off



Diagonal segments are produced by ties.

Figure (6): ROC for final model

#### **Area Under the Curve**

Test Result Variable(s): Predicted probability

	oot reoder t		calotted probability			
Ī				Asymptotic 95	% Confidence	
				Interval		
	Area	Std. Error <sup>a</sup>	Asymptotic Sig.b	Lower Bound	Upper Bound	
	.821	.049	.000	.724	.917	
_	<u> </u>	<u> </u>	·			

Figure (7): area under curve for ROC for final model

Sensitivity= 0.783, 1-specifity= 0.250

Optimal cut off= 0.988

Hence, I will repeat the classification table with the optimal cut off

#### Classification Table<sup>a</sup>

			Predicted				
	Family		Percentage				
	Observed	d	0	1	Correct		
Step 1	Family	0	15	5	75.0		
		1	347	1630	82.4		
	Overall P	ercentage			82.4		
Step 2	Family	0	12	8	60.0		
	1	1	377	1600	80.9		
	Overall P	ercentage			80.7		
Step 3	Family	0	14	6	70.0		
	1	1	401	1576	79.7		
	Overall P	ercentage			79.6		
Step 4	<mark>Family</mark>	<mark>o</mark>	<mark>15</mark>	<mark>5</mark>	<mark>75.0</mark>		
		1	<mark>429</mark>	<mark>1548</mark>	<mark>78.3</mark>		
	Overall P	ercentage			<mark>78.3</mark>		

Figure (8): classification table with the optimal cut off

It is clear that the overall percentage of correct classification dropped from 99% to 78.3% but on the other hand when focusing on the correct percentage of classification for each category it is found that the percentage correct classification of "family is not important" increased from 5% to 75% which is very high increase also the percent of correct classification of "family is important" dropped from 100% to 78.3% still this classification table is much better as all correct classification percentages are above 70% while the previous one not.

## Interpretations on my final used model

## Variables in the Equation

	Variables in the Equation						
		В	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Q3			14.036	3	.003	
	Q3(1)	3.311	1.060	9.755	1	.002	27.407
	Q3(2)	2.989	.951	9.871	1	.002	19.858
	Q3(3)	1.885	.961	3.846	1	.050	6.587
	Q5			10.571	3	.014	
	Q5(1)	-1.295	1.267	1.045	1	.307	.274

	=	i I	I	Ī		i .	
	Q5(2)	487	1.296	.141	1	.707	.614
	Q5(3)	-2.274	1.205	3.560	1	.059	.103
	Q27			9.529	3	.023	
	Q27(1)	18.964	2150.478	.000	1	.993	172211768.879
	Q27(2)	3.314	1.326	6.249	1	.012	27.493
	Q27(3)	1.295	1.255	1.064	1	.302	3.650
	Q260(1)	18.148	12416.325	.000	1	.999	76119125.466
	Q262	014	.016	.706	1	.401	.986
	Q260 * Q27			.540	3	.910	
	Q260(1) by Q27(1)	-33.867	12601.178	.000	1	.998	.000
	Q260(1) by Q27(2)	-19.115	12416.325	.000	1	.999	.000
	Q260(1) by Q27(3)	-18.369	12416.325	.000	1	.999	.000
	Constant	1.376	1.834	.563	1	.453	3.959
Step 2ª	Q3	1.070	1.004	14.035	3	.003	0.000
0.07 =	Q3(1)	3.160	1.031	9.389	1	.002	23.570
	Q3(2)	2.929	.941	9.698	1	.002	18.717
	Q3(3)	1.761	.945	3.470	1	.062	5.818
	Q5			10.915	3	.012	
	Q5(1)	-1.150	1.244	.855	1	.355	.317
	Q5(2)	348	1.279	.074	1	.786	.706
	Q5(3)	-2.189	1.187	3.400	1	.065	.112
	Q27			16.364	3	.001	
	Q27(1)	3.706	1.481	6.257	1	.012	40.675
	Q27(2)	2.373	1.143	4.308	1	.038	10.726
	Q27(3)	.781	1.137	.472	1	.492	2.184
	Q260(1)	611	.486	1.584	1	.208	.543
	Q262	013	.016	.594	1	.441	.988
	Constant	2.003	1.781	1.265	1	.261	7.409
Step 3 <sup>a</sup>	Q3			13.636	3	.003	
	Q3(1)	3.099	1.035	8.966	1	.003	22.177
	Q3(2)	2.804	.932	9.057	1	.003	16.517
	Q3(3)	1.618	.932	3.012	1	.083	5.045
	Q5	000	4.40=	11.836	3	.008	440
	Q5(1)	886	1.187	.557	1	.456	.412
	Q5(2)	088	1.225	.005	1	.943	.916
	Q5(3) Q27	-2.030	1.160	3.064 15.942	3	.080	.131
	Q27 Q27(1)	3.630	1.478			.001	37.727
	Q21(1)	3.030	1.4/8	6.030	1	.014	31.121

1		ı	[	1		I	
	Q27(2)	2.270	1.136	3.994	1	.046	9.681
	Q27(3)	.710	1.134	.392	1	.531	2.035
	Q260(1)	630	.486	1.684	1	.194	.532
<u> </u>	Constant	1.346	1.567	.737	1	.391	3.840
Step 4 <sup>a</sup>	Q3			<mark>12.982</mark>	<mark>3</mark>	<mark>.005</mark>	
	Q3(1)	3.019	1.035	8.503	1	.004	20.461
	Q3(2)	2.627	.918	8.192	1	.004	<mark>13.835</mark>
	<mark>Q3(3)</mark>	<mark>1.465</mark>	<mark>.923</mark>	<mark>2.520</mark>	1	<mark>.112</mark>	<mark>4.327</mark>
	<mark>Q5</mark>			<mark>10.935</mark>	<mark>3</mark>	<mark>.012</mark>	
	<mark>Q5(1)</mark>	<mark>-1.003</mark>	<mark>1.172</mark>	<mark>.732</mark>	1	<mark>.392</mark>	<mark>.367</mark>
	<mark>Q5(2)</mark>	<mark>170</mark>	<mark>1.213</mark>	<mark>.020</mark>	1	<mark>.889</mark>	<mark>.844</mark>
	<mark>Q5(3)</mark>	<mark>-2.015</mark>	<mark>1.144</mark>	<mark>3.100</mark>	1	<mark>.078</mark>	<mark>.133</mark>
	<mark>Q27</mark>			<mark>15.427</mark>	<mark>3</mark>	<mark>.001</mark>	
	Q27(1)	3.571	1.489	5.749	1	<mark>.016</mark>	35.557
	<mark>Q27(2)</mark>	<mark>2.182</mark>	<mark>1.146</mark>	<mark>3.623</mark>	1	<mark>.057</mark>	<mark>8.862</mark>
	Q27(3)	<mark>.652</mark>	<mark>1.144</mark>	<mark>.325</mark>	1	<mark>.569</mark>	<mark>1.920</mark>
	Constant	<mark>1.291</mark>	<mark>1.576</mark>	<mark>.671</mark>	1	<mark>.413</mark>	<mark>3.638</mark>

a. Variable(s) entered on step 1: Q3, Q5, Q27, Q260, Q262, Q260 \* Q27.

Figure (9): variables in the final model

#### Q3: Importance of leisure time

The base category for Q3 importance of leisure time is not at all important, the estimated odds ratio decreases with the decrease in the importance of leisure time from 20.461 to 4.327, however the odds ratio is only significant for very important (B=3.019) and rather important (B=2.627)

For 
$$B=3.019$$
, exp  $B=20.461$ 

The estimated odds ratio of choosing the importance of leisure time "very important" is higher by 1946.1 % than choosing the importance of leisure time "not at all important" holding other variables constant

For 
$$B=2.627$$
, exp  $B=13.835$ 

The estimated odds ratio of choosing the importance of leisure time "rather important" is higher by 1283.5% than choosing the importance of leisure time "not at all important" holding other variables constant

#### Q5: Importance of work

All dummies are insignificant

## Q27: One of main goals in life has been to make my parents proud

The base category for Q27 one of my goals in life has been to make my parents proud is not at all important, the estimated odds ratio decreases with the decrease in the importance of one of my goals in life has been to make my parents proud from 35.557 to 1.920, however the odds ratio is only significant for very important (B=3.571)

For 
$$B=3.571$$
, exp  $B=35.557$ 

The estimated odds ratio of choosing one of main goals in life has been to make my parents proud "very important" is higher by 3455.7% than choosing one of main goals in life has been to make my parents proud "not at all important" holding other variables constant

# Asses Goodness of fit for my final model

As the I have a continuous explanatory variable I will use Hosmer and Lemeshow Test

**Hosmer and Lemeshow Test** 

Step	Chi-square	df	Sig.
1	17.302	8	.027
2	12.261	8	.140
3	5.203	8	.736
4	4.280	7	<mark>.747</mark>

Figure (10): Goodness of fit for final model

H0: the model is good fit, since p-value for Hosmer and Lemeshow Test is more than 0.05

Therefore, do not reject H0. Hence, the model is good fit

# **Ordinal Logistic Regression Model**

Test the proportional odds assumption:

Test of Parallel Lines<sup>a</sup>

	-2 Log			
Model	Likelihood	Chi-Square	df	Sig.
Null Hypothesis	997.647			
General	.000 <sup>b</sup>	997.647	22	.000

Figure (11): proportional odds assumptions

I used here the response variable with all the 4 categories, and all of my explanatory variables H0: The coefficients are equal across categories; as highlighted p-value is equal to 0 (less than 0.05). Therefore, reject H0

Hence, the proportional odds assumption is violated, then it is advised to use the multinomial logistic regression model.

# **Multinomial Logistic Regression Model**

I will be using the response variable with the 4 categories, all of my explanatory variables (4 categorical and the continuous variable age)

Important in lif	e: Family <sup>a</sup>	В	Std. Error	Wald	df	Sig.	Exp(B)
Rather	Intercept	-15.969	.712	502.842	1	.000	
important	Q262	011	.006	3.322	1	.068	.989
	[Q3=1]	15.989	.293	2985.663	1	.000	<mark>8790481.294</mark>
	[Q3=2]	16.159	.254	4033.645	1	.000	10421426.625
	[Q3=3]	16.409	.000		1	.000	13374421.040
	[Q3=4]	0 <sub>p</sub>			0		
	[Q5=1]	-1.097	.435	6.356	1	<mark>.012</mark>	<mark>.334</mark>
	[Q5=2]	374	.404	.859	1	.354	.688
	[Q5=3]	.055	.422	.017	1	.897	1.056
	[Q5=4]	0 <sub>p</sub>			0		
	[Q27=1]	-2.980	.530	31.581	1	.000	<mark>.051</mark>
	[Q27=2]	-1.951	.476	16.774	1	.000	<mark>.142</mark>
	[Q27=3]	-1.651	.509	10.524	1	.001	<mark>.192</mark>
	[Q27=4]	Op			0		
	[Q260=1]	.628	.183	11.781	1	.001	<mark>1.873</mark>
	[Q260=2]	O <sub>p</sub>			0		
Not very	Intercept	-34.343	2173.440	.000	1	.987	
important	Q262	.020	.017	1.305	1	.253	1.020
	[Q3=1]	14.667	.726	408.274	1	<mark>.000</mark>	<mark>2342449.197</mark>
	[Q3=2]	14.913	.537	769.834	1	<mark>.000</mark>	<mark>2995184.841</mark>
	[Q3=3]	16.135	.000		1	<mark>.000</mark>	<mark>10169093.889</mark>
	[Q3=4]	O <sub>p</sub>			0		
	[Q5=1]	15.934	2173.440	.000	1	.994	8321597.384
	[Q5=2]	15.162	2173.440	.000	1	.994	3844338.586

	[Q5=3]	16.871	2173.440	.000	1	.994	21240974.394
	[Q5=4]	O <sub>p</sub>			0		
	[Q27=1]	-4.067	1.477	7.578	1	<mark>.006</mark>	<mark>.017</mark>
	[Q27=2]	-2.719	1.150	5.587	1	<mark>.018</mark>	<mark>.066</mark>
	[Q27=3]	-1.352	1.163	1.351	1	.245	.259
	[Q27=4]	O <sub>p</sub>			0		
	[Q260=1]	.489	.498	.963	1	.326	1.630
	[Q260=2]	0 <sub>p</sub>			0		
Not at all	Intercept	42.364	30922.565	.000	1	.999	
important	Q262	734	92.434	.000	1	.994	.480
	[Q3=1]	-43.986	3166.749	.000	1	.989	7.893E-20
	[Q3=2]	-36.991	3010.604	.000	1	.990	8.610E-17
	[Q3=3]	-27.370	4979.315	.000	1	.996	1.298E-12
	[Q3=4]	Op			0		
	[Q5=1]	-26.609	5461.432	.000	1	.996	2.778E-12
	[Q5=2]	-14.691	4343.702	.000	1	.997	4.166E-7
	[Q5=3]	-8.611	4320.352	.000	1	.998	.000
	[Q5=4]	Op	•		0		
	[Q27=1]	-18.125	30025.837	.000	1	1.000	1.344E-8
	[Q27=2]	-20.753	30083.466	.000	1	.999	9.705E-10
	[Q27=3]	-3.078	29947.006	.000	1	1.000	.046
	[Q27=4]	Op	-		0		
	[Q260=1]	12.347	2281.904	.000	1	.996	230312.852
	[Q260=2]	O <sub>p</sub>			0		

Figure (12): multinomial model

The base category for the importance of family (response variable) is very important For rather important in response variable

Q3: Importance of leisure time, the base category is not at all important B=15.989 ,  $exp\ B=8790481.294$ 

The estimated odds ratio of choosing "rather important" in importance of family between individuals choosing "very important" in importance of leisure time compared to those chose "not

at all important" in importance of leisure time is 879048029.4% higher than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

# B=16.159, exp B=10421426.625

The estimated odds ratio of choosing "rather important" in importance of family between individuals choosing "rather important" in importance of leisure time compared to those chose "not at all important" in importance of leisure time is 1042142563% higher than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

#### B=16.409, exp B=13374421.040

The estimated odds ratio of choosing "rather important" in importance of family between individuals choosing "not very important" in importance of leisure time compared to those chose "not at all important" in importance of leisure time is 1337442004% higher than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

## Q5: Importance of work, the base category is not at all important

#### B = -1.097, $\exp B = 0.334$

The estimated odds ratio of choosing "rather important" in importance of family between individuals choosing "very important" in importance of work compared to those chose "not at all important" in importance of work is 66.6% lower than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

# Q27: one of my main goals in life has been to make my parents proud, the base category is not at all important

#### B = -2.980, exp B = 0.051

The estimated odds ratio of choosing "rather important" in importance of family between individuals choosing "very important" in one of my main goals in life has been to make my parents proud compared to those chose "not at all important" in one of my main goals in life has been to make my parents proud is 94.9% less than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

#### B=-1.951, exp B=0.142

The estimated odds ratio of choosing "rather important" in importance of family between individuals choosing "rather important" in one of my main goals in life has been to make my parents proud compared to those chose "not at all important" in one of my main goals in life has

been to make my parents proud is 85.8% less than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

# B=-1.651, exp B=0.192

The estimated odds ratio of choosing "rather important" in importance of family between individuals choosing "not very important" in one of my main goals in life has been to make my parents proud compared to those chose "not at all important" in one of my main goals in life has been to make my parents proud is 80.8% less than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

#### Q260 Sex, female is base category

## B = 0.628, exp B = 1.873

The estimated odds ratio of choosing "rather important" in importance of family between male individuals compared to females is 87.3% higher than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

#### For not very important in response variable

# Q3: Importance of leisure time, the base category is not at all important

#### B= 14.667, exp B= 2342449.197

The estimated odds ratio of choosing "not very important" in importance of family between individuals choosing "very important" in importance of leisure time compared to those chose "not at all important" in importance of leisure time is 234244819.7% higher than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

#### B=14.913, exp B=2995184.841

The estimated odds ratio of choosing "not very important" in importance of family between individuals choosing "rather important" in importance of leisure time compared to those chose "not at all important" in importance of leisure time is 299518384.1% higher than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

#### B=16.135, exp B=10169093.889

The estimated odds ratio of choosing "not very important" in importance of family between individuals choosing "not very important" in importance of leisure time compared to those chose "not at all important" in importance of leisure time is 1016909289% higher than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

Q27: one of my main goals in life has been to make my parents proud, the base category is not at all important

B = -4.067, exp B = 0.017

The estimated odds ratio of choosing "not very important" in importance of family between individuals choosing "very important" in one of my main goals in life has been to make my parents proud compared to those chose "not at all important" in one of my main goals in life has been to make my parents proud is 98.3% less than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

B=-2.719, exp B=0.066

The estimated odds ratio of choosing "not very important" in importance of family between individuals choosing "rather important" in one of my main goals in life has been to make my parents proud compared to those chose "not at all important" in one of my main goals in life has been to make my parents proud is 93.4% less than the corresponding odds ratio of choosing "very important" in importance of family, holding other variables constant

# Assess goodness of fit for the model

As the I have a continuous explanatory variable I will use Hosmer and Lemeshow Test p-value for the test: 1.00

H0: the model is good fit, since p-value for Hosmer and Lemeshow Test is more than 0.05 Therefore, do not reject H0. Hence, the model is good fit

#### Conclusion

Throughout the report I have tried first binary logistic model using response variable more general with only 2 categories, I have included an interaction term and I run the model one time with enter method and the other with backward method. The backward method model was better so I chose this model and contributed my analysis and the significant variables as well.

Then, I tried to compute ordinal binary logistic model but the assumption was violated so I moved forward to the multinomial logistic model which was a good model and I interpreted my findings and the significant variables.

# Appendix

# **Enter method**

**Omnibus Tests of Model Coefficients** 

		Chi-square	df	Sig.
Step 1	Step	46.724	14	.000
	Block	46.724	14	.000
	Model	46.724	14	<mark>.000</mark>

Variables in the Equation

_	Variables in the Equation							
		В	S.E.	Wald	df	Sig.	Exp(B)	
Step 1ª	Q3			14.036	3	.003		
	Q3(1)	3.311	1.060	9.755	1	.002	27.407	
	Q3(2)	2.989	.951	9.871	1	.002	19.858	
	Q3(3)	1.885	.961	3.846	1	.050	6.587	
	Q5			10.571	3	.014		
	Q5(1)	-1.295	1.267	1.045	1	.307	.274	
	Q5(2)	487	1.296	.141	1	.707	.614	
	Q5(3)	-2.274	1.205	3.560	1	.059	.103	
	Q27			9.529	3	.023		
	Q27(1)	18.964	2150.478	.000	1	.993	172211768.879	
	Q27(2)	3.314	1.326	6.249	1	.012	27.493	
	Q27(3)	1.295	1.255	1.064	1	.302	3.650	
	Q260(1)	18.148	12416.325	.000	1	.999	76119125.466	
	Q262	014	.016	.706	1	.401	.986	
	Q260 * Q27			.540	3	.910		
	Q260(1) by Q27(1)	-33.867	12601.178	.000	1	.998	.000	
	Q260(1) by Q27(2)	-19.115	12416.325	.000	1	.999	.000	
	Q260(1) by Q27(3)	-18.369	12416.325	.000	1	.999	.000	
	Constant	1.376	1.834	.563	1	.453	3.959	

Class	ificatio.	a Tablaa
Class	ification	า Table <sup>a</sup>

	Observed	Predicted

			Far	nily	Percentage
			0	1	Correct
Step 1	Family	0	0	20	.0
		1	0	1977	100.0
	Overall P	ercentage			<mark>99.0</mark>

It is clear here from the overall correct percentage that it is 99 percent which seems that the model has high accuracy as highlighted

Also the p-value of the model is less than 0.05 (equal zero) which means that the model is significant as highlighted

# **Backward method**

**Omnibus Tests of Model Coefficients** 

		Chi-square	df	Sig.
Step 1	Model	46.724	14	.000
Step 2ª	Step	-2.866	3	.413
	Block	-2.866	3	.413
	Model	43.858	11	.000
Step 3ª	Step	602	1	.438
	Block	-3.468	4	.483
	Model	43.256	10	.000
Step 4 <sup>a</sup>	Step	-1.712	1	.191
	Block	-5.180	5	.394
	Model	41.544	9	<mark>.000</mark>

# Variables in the Equation

		В	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Q3			14.036	3	.003	
	Q3(1)	3.311	1.060	9.755	1	.002	27.407
	Q3(2)	2.989	.951	9.871	1	.002	19.858
	Q3(3)	1.885	.961	3.846	1	.050	6.587
	Q5			10.571	3	.014	
	Q5(1)	-1.295	1.267	1.045	1	.307	.274

	-	i i	I	İ			Ī
	Q5(2)	487	1.296	.141	1	.707	.614
	Q5(3)	-2.274	1.205	3.560	1	.059	.103
	Q27			9.529	3	.023	
	Q27(1)	18.964	2150.478	.000	1	.993	172211768.879
	Q27(2)	3.314	1.326	6.249	1	.012	27.493
	Q27(3)	1.295	1.255	1.064	1	.302	3.650
	Q260(1)	18.148	12416.325	.000	1	.999	76119125.466
	Q262	014	.016	.706	1	.401	.986
	Q260 * Q27			.540	3	.910	
	Q260(1) by Q27(1)	-33.867	12601.178	.000	1	.998	.000
	Q260(1) by Q27(2)	-19.115	12416.325	.000	1	.999	.000
	Q260(1) by Q27(3)	-18.369	12416.325	.000	1	.999	.000
	Constant	1.376	1.834	.563		.453	
Step 2ª	Q3	1.376	1.034	14.035	1	.003	3.959
Step 2	Q3(1)	3.160	1.031	9.389	1	.003	23.570
	Q3(2)	2.929	.941	9.698	1	.002	18.717
	Q3(3)	1.761	.945	3.470	1	.062	5.818
	Q5			10.915	3	.012	
	Q5(1)	-1.150	1.244	.855	1	.355	.317
	Q5(2)	348	1.279	.074	1	.786	.706
	Q5(3)	-2.189	1.187	3.400	1	.065	.112
	Q27			16.364	3	.001	
	Q27(1)	3.706	1.481	6.257	1	.012	40.675
	Q27(2)	2.373	1.143	4.308	1	.038	10.726
	Q27(3)	.781	1.137	.472	1	.492	2.184
	Q260(1)	611	.486	1.584	1	.208	.543
	Q262	013	.016	.594	1	.441	.988
	Constant	2.003	1.781	1.265	1	.261	7.409
Step 3 <sup>a</sup>	Q3			13.636	3	.003	
	Q3(1)	3.099	1.035	8.966	1	.003	22.177
	Q3(2)	2.804	.932	9.057	1	.003	16.517
	Q3(3)	1.618	.932	3.012	1	.083	5.045
	Q5			11.836	3	.008	
	Q5(1)	886	1.187	.557	1	.456	.412
	Q5(2)	088	1.225	.005	1	.943	.916
	Q5(3)	-2.030	1.160	3.064	1	.080	.131
	Q27	2 620	1 470	15.942	3	.001	27 727
	Q27(1)	3.630	1.478	6.030	1	.014	37.727

	ı			1	1	1	
	Q27(2)	2.270	1.136	3.994	1	.046	9.681
	Q27(3)	.710	1.134	.392	1	.531	2.035
	Q260(1)	630	.486	1.684	1	.194	.532
	Constant	1.346	1.567	.737	1	.391	3.840
Step 4 <sup>a</sup>	Q3			12.982	3	.005	
	Q3(1)	3.019	1.035	8.503	1	.004	20.461
	Q3(2)	2.627	.918	8.192	1	.004	13.835
	Q3(3)	1.465	.923	2.520	1	.112	4.327
	Q5			10.935	3	.012	
	Q5(1)	-1.003	1.172	.732	1	.392	.367
	Q5(2)	170	1.213	.020	1	.889	.844
	Q5(3)	-2.015	1.144	3.100	1	.078	.133
	Q27			15.427	3	.001	
	Q27(1)	3.571	1.489	5.749	1	.016	35.557
	Q27(2)	2.182	1.146	3.623	1	.057	8.862
	Q27(3)	.652	1.144	.325	1	.569	1.920
	Constant	1.291	1.576	.671	1	.413	3.638

# Classification Table<sup>a</sup>

			Predicted				
			Far	nily	Percentage		
	Observed	d	0	1	Correct		
Step 1	Family	0	0	20	.0		
		1	0	1977	100.0		
	Overall P	ercentage			99.0		
Step 2	Family	0	0	20	.0		
	1	1	0	1977	100.0		
	Overall P	ercentage			99.0		
Step 3	Family	0	0	20	.0		
	-	1	0	1977	100.0		
	Overall P	ercentage			99.0		
Step 4	Family	0	1	19	5.0		
		1	0	1977	100.0		
	Overall P	ercentage			<mark>99.0</mark>		

It is clear here from the overall correct percentage that it is 99 percent which seems that the model has high accuracy as highlighted Also the p-value of the model is less than 0.05 (equal zero) which means that the model is significant as highlighted **References:** Data from: https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp