Categorical data analysis project

Part 2

Student info: Haneen Ahmed Elgabry – 5200199 – English section

Submitted to: Dr. Sally Farid Abdelfattah.

Deadline: 4/1

Unraveling the Determinants of Happiness in USA: A Logistic Regression Approach

Introduction:

In this analysis, my focus is on understanding the factors influencing the level of happiness among individuals in the United States. Happiness, a subjective and complex concept, is shaped by various demographic and personal factors that may vary across different cultural and societal contexts. By employing logistic regression on a dataset sourced from the final wave of the World Values Survey (WVS-WAVE 7) specifically for the USA, we aim to unravel the specific determinants that impact an individual's reported level of happiness.

This analysis delves into the unique dynamics of happiness within the United States, examining the role of demographic and personal variables using logistic regression. The dataset, carefully extracted from the World Values Survey, includes key variables such as gender, age, freedom of choice and control, satisfaction with the financial situation of the household, and trust in people personally known. By leveraging logistic regression, our goal is to model the probability of an individual falling into distinct happiness categories (e.g., "Very Happy," "Happy," "Not Very Happy") based on these explanatory variables.

My variables:

-Response Variable:

1. Level of Happiness(Q46): This is the variable I am trying to predict or understand. It is categorical and has five levels:

- "1" Very Happy
- "2" Happy
- "3" Quiet Happy
- "4" Not Very Happy
- "5" Not at All Happy

the categories were merged to transform it into a binary variable to be able to model it using binary logistic regression.

-Explanatory Variables:

1. Gender (Categorical, Q260):

- "1" Male
- "2" Female

2. Age (Continuous, Q262):

- Age is a continuous variable representing the age of the individual.

3. Freedom of Choice and Control (Ordinal, Q48):

- This variable is measured on a scale where 1 means "no choice at all" and 10 means "a great deal of choice." It indicates how much freedom of choice and control a person feels they have over the way their life turns out.

4. Satisfaction with Financial Situation of Household (Ordinal, Q50):

- This variable is measured on a scale where 1 is completely dissatisfied and 10 is completely satisfied. It reflects the individual's satisfaction with the financial situation of their household.

5. Trust in People Personally Known (Categorical, Q60):

- "1" Trust Completely

- "2" Trust Somewhat

- "3" Do Not Trust Very Much

- "4" Do Not Trust at All

Binary Logistic Regression Model:

First: checking the assumptions

1- Binary Dependent Variable:

As mentioned, the categories of the response variable were merged to transform it into a binary variable to be able to model it using binary logistic regression.

The variables were merged as follows:

-The categories (very happy – happy – quite happy) were merged under the code 1 (feeling happy).

- The categories (Not very happy &None at all happy) were merged under the code 0 (not happy).

Levelhappiness

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not-happy	33	1.3	1.3	1.3
	feeling-happy	2555	98.4	98.7	100.0
	Total	2588	99.7	100.0	
Missing	System	8	.3		
Total		2596	100.0		

According to the new transformed variable, there are 33 people in the USA that feels not happy and 2555 feels happy.

2- Multi-Collinearity:

Model	Collinearity Statistics	
	Tolerance	VIF
(Constant)		
How much freedom of choice and control	.845	1.183
Satisfaction with financial situation of household	.832	1.201
Trust: People you know personally	.954	1.049
Sex	.941	1.063

Age	.919	1.089

All the VIF values are comfortably below 5, indicating a low risk of multicollinearity. This is a positive sign, suggesting that the predictor variables (freedom of choice, satisfaction with financial situation, trust, sex, and age) are relatively independent of each other.

Binary Logistic Model1:

First, for all categorical variables we have made (k-1) dummies to avoid multicollinearity as follows:

- Gender (Categorical, Q260): the base category is female, and we have 1 dummy.
- Freedom of Choice and Control(Q48): the base category is the 10^{th} scale which is great deal of choice and control, and we have 9 dummies.
- Satisfaction with Financial Situation of Household (Q50): the base category is the 10^{th} scale which is complete satisfaction, and we have 9 dummies.
- Trust in People Personally Known (Categorical, Q60): the base category is don't trust at all, and we have 4 dummies.

We have an interaction term which is between sex and trust in people you personally know.

Model1 is a model which contains all the explanatory variables (dummies) along with an interaction term between the 2 dummy variables made for sex and trust.
After computing Model1 by using enter method, I will compute Model0 which is another binary logistic model, but it is computed by using backward regression and the two models will then be compared.

From the output of model1(look at the appendix), we can see that only 2 (Q48 , Q60 and only certain categories of them) variables are significant and Q50, Q260, and Q262 and all of their categories are not statistically significant as their p-values are greater than 0.05. and the interaction term Q260 * Q60 and its sub-interaction terms are also not statistically significant.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	239.847[a]	.043	.335

Classification Table[a]

Observed			Predicted		
			levelhappiness		
			Not-happy	feeling-happy	Percentage Correct
Step 1	levelhappiness	Not-happy	1	32	3.0
		feeling-happy	4	2509	99.8
	Overall Percentage				98.6

a. The cut value is .500

We can see that the overall percentage of correct classification in this model is 98.6% which indicates a high accuracy, but we can notice that the percentage of correct classification for first category is 3% which is very low.

Binary Logistic Model0:

Model0 which is another binary logistic model, but it is nested within Model1 that is computed by using backward regression. And it also contains all the explanatory variables (dummies) along with an interaction term between the 2 dummy variables made for sex and trust.

From the output of the backward method with cutoff point 0.5 (look at the appendix) emphasizes the importance of Q48 and Q60, particularly their different levels, in predicting level of happiness. These variables consistently contribute to the model's significance across steps. Other variables may have varying degrees of significance.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	239.847[a]	.043	.335
2	240.783[a]	.043	.332
3	241.445[a]	.043	.330
4	242.899[a]	.042	.326

Classification Table[a]

			Predicted		
			levelhappiness		Percentage
	Observed		Not-happy	feeling-happy	Correct
Step 1	levelhappiness	Not-happy	1	32	3.0
		feeling-happy	4	2509	99.8
	Overall Percentage				98.6
Step 2	levelhappiness	Not-happy	2	31	6.1
		feeling-happy	2	2511	99.9
	Overall Percentage				98.7
Step 3	levelhappiness	Not-happy	2	31	6.1
		feeling-happy	3	2510	99.9
	Overall Percentage				98.7
Step 4	levelhappiness	Not-happy	2	31	6.1
		feeling-happy	2	2511	99.9
	Overall Percentage				98.7

a. The cut value is .500

We can see that the overall percentage of correct classification in this model is 98.7% which indicates a high accuracy, but we can notice the same problem as the previous model that the correct classification of the first category is very small which is 6.1%.

Comparing between Model 1 and Model0:

Since the backward method model is nested within the enter method model (nested models), we will compare between the two models using the likelihood ratio test as follows:

For testing the null hypothesis that M0 has a better fit than M1, the likelihood ratio test statistic is -2[L (M0, Y)-L(M1,Y)]= D(M0,Y) – D(M1,Y) , The difference between the two deviances is large when M0 fits poorly compared to M1.

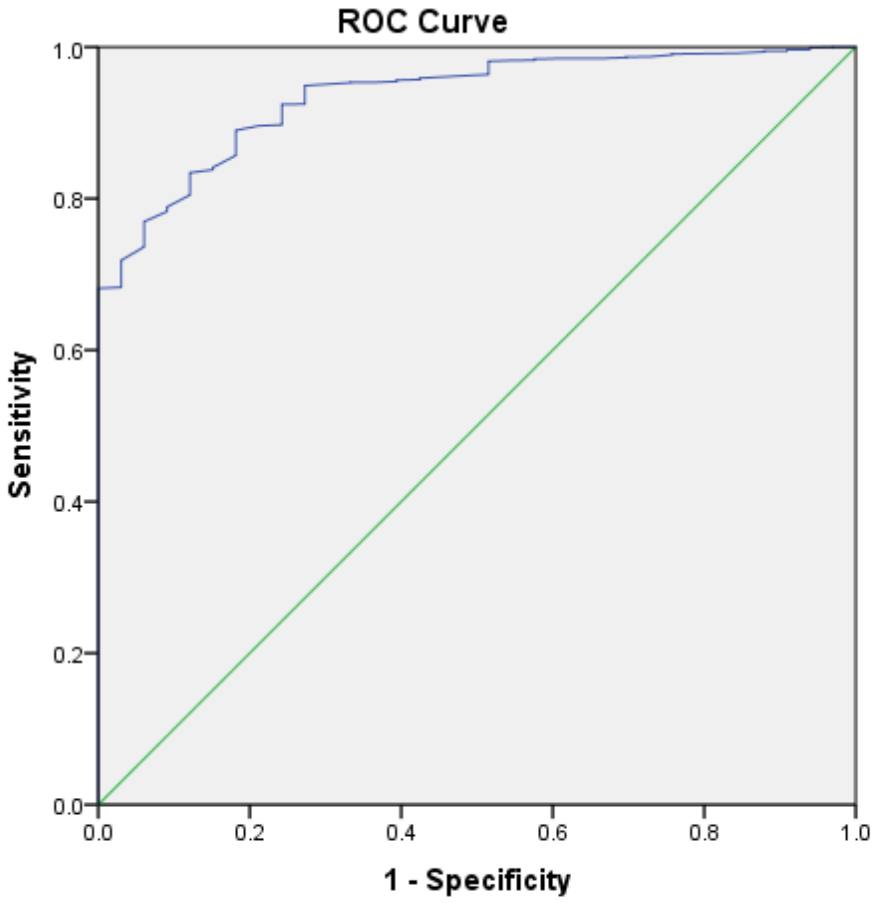
HO: model 0 is better fit

-2 log likelihood of Model1 = 239.847, -2 log likelihood of Model0 = 242.899
Since that: Chi-square value =11.071 > test statistics = 3.052

Therefore: At 95% confidence level we will fail to reject Ho which indicates that the Model0 (backward method) which is simpler is a better fit for the data.

ROC curve to conduct optimal cutoff:

But by looking at the classification table of model0 we can see that it isn't efficient so we will plot to ROC curve to find a more optimal cut off point.



Diagonal segments are produced by ties.

Sensitivity = 0.896 ,**1- specify**= 0.212, **optimal cut off** = 0.982

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
.850	.039	.000	.772	.927

The area under the curve is equal to 0.85 which shows that our model has very good ability to discriminate and can separate between the two classes of level happiness (Not-happy, feeling-happy) with high accuracy and better than by chance as it is significantly different than 0.5.

Classification Table[a]

			Predicted		
			levelhappiness		Percentage
	Observed		Not-happy	feeling-happy	Correct
Step 1	levelhappiness	Not-happy	28	5	84.8
		feeling-happy	269	2244	89.3
	Overall Percentage				89.2
Step 2	levelhappiness	Not-happy	28	5	84.8
		feeling-happy	266	2247	89.4
	Overall Percentage				89.4
Step 3	levelhappiness	Not-happy	27	6	81.8
		feeling-happy	258	2255	89.7
	Overall Percentage				89.6
Step 4	levelhappiness	Not-happy	25	8	75.8
		feeling-happy	257	2256	89.8
	Overall Percentage				89.6

a. The cut value is .982

- We can notice that after changing the cutoff from 0.5 to 0.982 that the overall accuracy of classification of the model is 89.6% which decreased, but the percent of correct classification in group 1 (not happy) increased very much from 6.1% to 75.8% as well .

Binary Logistic Model – Model2 (final model):

In this final model it contains all of the explanatory variables (dummies) and an interaction term between sex and trust (dummies) but with cutoff point equal 0.982.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Q48			33.112	9	.000	
	Q48(1)	-2.161	.727	8.826	1	.003	.115
	Q48(2)	-3.394	.864	15.427	1	.000	.034

	Q48(3)	-1.957	.703	7.748	1	.005	.141
	Q48(4)	-.960	.817	1.381	1	.240	.383
	Q48(5)	-.732	.714	1.052	1	.305	.481
	Q48(6)	16.653	2650.566	.000	1	.995	17069718.760
	Q48(7)	.308	.757	.166	1	.684	1.361
	Q48(8)	1.031	.841	1.503	1	.220	2.804
	Q48(9)	16.459	1985.399	.000	1	.993	14068518.548
	Q50			10.620	9	.303	
	Q50(1)	-.576	.764	.569	1	.451	.562
	Q50(2)	-.422	.947	.199	1	.656	.656
	Q50(3)	17.072	2967.101	.000	1	.995	25946834.467
	Q50(4)	1.047	1.033	1.027	1	.311	2.849
	Q50(5)	1.243	1.000	1.546	1	.214	3.466
	Q50(6)	.372	.857	.188	1	.664	1.450
	Q50(7)	.801	.978	.671	1	.413	2.228
	Q50(8)	1.064	.986	1.165	1	.281	2.899
	Q50(9)	-.039	.970	.002	1	.968	.962
	Q60			7.043	3	.071	
	Q60(1)	2.968	1.400	4.491	1	.034	19.448
	Q60(2)	2.145	1.024	4.390	1	.036	8.543
	Q60(3)	1.083	1.049	1.067	1	.302	2.953
	Q260(1)	.646	1.384	.218	1	.640	1.909
	Q262	.012	.014	.736	1	.391	1.012
	Q260 * Q60			.928	3	.819	
	Q260(1) by Q60(1)	-1.562	1.842	.720	1	.396	.210
	Q260(1) by Q60(2)	-1.147	1.504	.582	1	.446	.318
	Q260(1) by Q60(3)	-1.419	1.574	.813	1	.367	.242
	Constant	1.837	1.233	2.222	1	.136	6.281
Step 2[a]	Q48			31.819	9	.000	
	Q48(1)	-2.196	.723	9.223	1	.002	.111
	Q48(2)	-3.285	.870	14.243	1	.000	.037
	Q48(3)	-1.946	.703	7.654	1	.006	.143
	Q48(4)	-1.058	.802	1.740	1	.187	.347
	Q48(5)	-.767	.713	1.158	1	.282	.464
	Q48(6)	16.600	2660.846	.000	1	.995	16188896.923
	Q48(7)	.280	.757	.137	1	.711	1.324
	Q48(8)	.985	.839	1.379	1	.240	2.678

	Q48(9)	16.448	1991.372	.000	1	.993	13902051.121
	Q50			10.739	9	.294	
	Q50(1)	-.536	.758	.499	1	.480	.585
	Q50(2)	-.353	.940	.141	1	.707	.702
	Q50(3)	17.142	2967.698	.000	1	.995	27854116.975
	Q50(4)	1.140	1.030	1.225	1	.268	3.126
	Q50(5)	1.294	.994	1.697	1	.193	3.649
	Q50(6)	.422	.850	.246	1	.620	1.525
	Q50(7)	.836	.974	.736	1	.391	2.306
	Q50(8)	1.104	.979	1.272	1	.259	3.017
	Q50(9)	.028	.961	.001	1	.977	1.028
	Q60			11.562	3	.009	
	Q60(1)	2.097	.875	5.748	1	.017	8.144
	Q60(2)	1.580	.757	4.358	1	.037	4.857
	Q60(3)	.395	.790	.251	1	.617	1.485
	Q260(1)	-.526	.401	1.724	1	.189	.591
	Q262	.011	.014	.646	1	.422	1.011
	Constant	2.416	1.096	4.864	1	.027	11.204
Step 3[a]	Q48			31.955	9	.000	
	Q48(1)	-2.197	.723	9.247	1	.002	.111
	Q48(2)	-3.315	.869	14.539	1	.000	.036
	Q48(3)	-1.975	.705	7.852	1	.005	.139
	Q48(4)	-1.106	.801	1.904	1	.168	.331
	Q48(5)	-.792	.711	1.239	1	.266	.453
	Q48(6)	16.607	2662.551	.000	1	.995	16297389.627
	Q48(7)	.262	.758	.120	1	.730	1.300
	Q48(8)	.976	.839	1.353	1	.245	2.653
	Q48(9)	16.476	1990.113	.000	1	.993	14298777.901
	Q50			10.754	9	.293	
	Q50(1)	-.586	.761	.593	1	.441	.556
	Q50(2)	-.462	.936	.244	1	.621	.630
	Q50(3)	17.122	2963.831	.000	1	.995	27276551.462
	Q50(4)	1.092	1.037	1.109	1	.292	2.980
	Q50(5)	1.240	.997	1.547	1	.214	3.456
	Q50(6)	.347	.851	.167	1	.683	1.415
	Q50(7)	.770	.976	.623	1	.430	2.161
	Q50(8)	1.039	.980	1.124	1	.289	2.826
	Q50(9)	.005	.964	.000	1	.996	1.005
	Q60			12.889	3	.005	
	Q60(1)	2.145	.870	6.074	1	.014	8.538

	Q60(2)	1.620	.753	4.625	1	.032	5.056
	Q60(3)	.366	.785	.218	1	.641	1.443
	Q260(1)	-.473	.396	1.429	1	.232	.623
	Constant	2.889	.926	9.721	1	.002	17.968
Step 4[a]	Q48			32.315	9	.000	
	Q48(1)	-2.232	.720	9.612	1	.002	.107
	Q48(2)	-3.361	.864	15.128	1	.000	.035
	Q48(3)	-2.011	.703	8.180	1	.004	.134
	Q48(4)	-1.113	.800	1.937	1	.164	.329
	Q48(5)	-.855	.708	1.456	1	.228	.425
	Q48(6)	16.536	2687.919	.000	1	.995	15180949.602
	Q48(7)	.226	.756	.090	1	.765	1.254
	Q48(8)	.930	.837	1.236	1	.266	2.535
	Q48(9)	16.419	2003.258	.000	1	.993	13504525.581
	Q50			10.487	9	.313	
	Q50(1)	-.543	.763	.507	1	.476	.581
	Q50(2)	-.419	.939	.199	1	.655	.658
	Q50(3)	17.163	2972.408	.000	1	.995	28418307.761
	Q50(4)	1.070	1.033	1.073	1	.300	2.914
	Q50(5)	1.277	1.001	1.629	1	.202	3.586
	Q50(6)	.319	.849	.141	1	.707	1.376
	Q50(7)	.735	.978	.564	1	.453	2.085
	Q50(8)	1.065	.986	1.166	1	.280	2.899
	Q50(9)	-.035	.968	.001	1	.971	.965
	Q60			12.293	3	.006	
	Q60(1)	2.093	.861	5.908	1	.015	8.112
	Q60(2)	1.605	.743	4.671	1	.031	4.979
	Q60(3)	.402	.775	.270	1	.604	1.495
	Constant	2.663	.903	8.694	1	.003	14.339

a. Variable(s) entered on step 1: Q48, Q50, Q60, Q260, Q262, Q260 * Q60 .

Interpretations:

The highlighted variables are the significant variables (p-value <0.05):

For Q48(freedom of choice and control)

- Q48 Overall (freedom of choice): The coefficient for Q60 is 32.315, indicating the change in the log-odds of the response variable for a one-unit change in the predictor variable (level of happiness).The p-value associated with Q60 is 0.00, suggesting that the variable Q48 is statistically significant.
- -For Q48(1) (no choice at all), the coefficient is -2.232. This means that, holding other variables constant, the log-odds of choosing "no choice at all" decrease by 2.232 for each unit increase in the level of happiness.

  -the odds ratio is e ^-2.232≈0.107 This means that, holding other variables constant, the odds of choosing "no choice at all" are about 0.107 times the odds of choosing "a great deal of choice.

  - The estimated odds ratio of choosing no choice at all is lower by 89.3% than choosing a great deal of choice holding other variables constant.
- -For Q48(2) (2nd scale), the coefficient is -3.361. This implies that, holding other variables constant, the log-odds of choosing "2nd scale" of choice and control decrease by 3.361 for each unit increase in the level of happiness.

  - The odds ratio for Q48(2) is e^−3.361≈0.034 This means that, holding other variables constant, the odds of choosing "2nd scale" are about 0.034 times the odds of choosing "a great deal of choice." Since that the odds ratios are less than 1, this indicates a decrease in the odds of choosing 2nd scale compared to the reference category "a great deal of choice."

  - The estimated odds ratio of choosing the 2nd scale of choice and control is lower by 96.6% than choosing a great deal of choice holding other variables constant.
- -for Q48(3) (3rd scale): the coefficient is -2.011. This indicates that, holding other variables constant, the log-odds of choosing "3rd scale of choice and control" decrease by 2.011 for each unit increase in the level of happiness.

  -The odds ratio for Q48(3) is e^- 2.011≈0. 134.This means that, holding other variables constant, the odds of choosing "3rd scale" are about 0.134 times the odds of choosing "a great deal of choice." the odds ratios are less than 1, indicating a decrease in the odds of

choosing the respective category 3rd scale compared to the reference category "a great deal of choice."

- The estimated odds ratio of choosing the 3rd scale of choice and control is lower by 86.6% than choosing a great deal of choice holding other variables constant.

For Q60(trust in people you personally know):

- Q60 Overall (Trust in People Personally Known): The coefficient for Q60 is 12.293, indicating the change in the log-odds of the response variable for a one-unit change in the predictor variable (level of happiness). The p-value associated with Q60 is 0.006, suggesting that the variable Q60 is statistically significant.
- -Q60(1) (Trust Completely): The coefficient for Q60(1) is 2.093. This means that, holding the other variables constant, the log-odds of choosing "Trust Completely" increase by 2.093 for each unit increase in level of happiness.

  -The odds ratio for Q60(1) is e^2.093≈8.107. This implies that the odds of choosing "Trust Completely" are approximately 8.107 times the odds of choosing "Do Not Trust at All".

  - The estimated odds ratio of choosing the trust completely is higher by 710.7% than choosing no trust at all holding other variables constant.
- -Q60(2) (Trust Somewhat): The coefficient for Q60(2) is 1.605. This means that, holding all variables constant, the log-odds of choosing "Trust Somewhat" increase by 1.605 for every unit increase in level of happiness.

  -The odds ratio for Q60(2) is e^1.605≈4.974. This implies that the odds of choosing "Trust Somewhat" are approximately 4.974 times the odds of choosing "Do Not Trust at All" when the level of happiness increases by one unit.

  - The estimated odds ratio of choosing trust somewhat is higher by 397.9% than choosing no trust at all holding other variables constant.

Assess Goodness of fit of model 2(final model):

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.085	8	.748
2	3.879	8	.868
3	6.829	8	.555
4	2.877	8	.942

H0: the model is good fit.

p-value more than 0.05, Therefore, do not reject H0. Hence, the model is good fit.

Ordinal Logistic Regression Model

Testing the Proportional Odds Assumption

the response variable here is used with all the 4 categories, and all the explanatory variables.

H0: The coefficients are equal across categories.

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	3967.927			
General	3837.307[b]	130.619[c]	46	.000

- The p-value of the test is less than 0.05 therefore the assumption isn't satisfied, and we will use another type of model which is the multinomial logistic regression model.

Multinomial Logistic Regression Model

The response variable is the level of happiness with its 4 categories along with all of the explanatory variables mentioned before.

Feeling of happiness[a]		B	Std. Error	Wald	df	Sig.	Exp(B)
Very happy	Intercept	.755	.918	.675	1	.411	
	Q262	.006	.011	.356	1	.551	1.006
	[Q48=1]	-5.531	.951	33.818	1	.000	.004
	[Q48=2]	13.082	.000	.	1	.	480091.948

	[Q48=3]	-2.992	.645	21.511	1	.000	.050
	[Q48=4]	-1.406	.753	3.486	1	.062	.245
	[Q48=5]	-.874	.660	1.753	1	.186	.417
	[Q48=6]	.067	.832	.007	1	.936	1.070
	[Q48=7]	-.382	.600	.405	1	.525	.683
	[Q48=8]	-.125	.546	.053	1	.819	.882
	[Q48=9]	.142	.666	.046	1	.831	1.153
	[Q48=10]	0[b]	.	.	0	.	.
	[Q50=1]	-2.027	.730	7.714	1	.005	.132
	[Q50=2]	-1.201	.939	1.635	1	.201	.301
	[Q50=3]	.028	1.056	.001	1	.979	1.028
	[Q50=4]	-.244	.816	.089	1	.765	.784
	[Q50=5]	-.065	.777	.007	1	.933	.937
	[Q50=6]	-.322	.760	.180	1	.672	.725
	[Q50=7]	-.125	.770	.026	1	.871	.882
	[Q50=8]	.082	.764	.011	1	.915	1.085
	[Q50=9]	-.049	.837	.003	1	.953	.952
	[Q50=10]	0[b]	.	.	0	.	.
	[Q60=1]	3.021	.704	18.406	1	.000	20.519
	[Q60=2]	2.798	.652	18.418	1	.000	16.420
	[Q60=3]	1.727	.714	5.850	1	.016	5.626
	[Q60=4]	0[b]	.	.	0	.	.
	[Q260=1]	-.325	.327	.985	1	.321	.723
	[Q260=2]	0[b]	.	.	0	.	.
Quite happy	Intercept	.154	.908	.029	1	.865	
	Q262	.006	.011	.273	1	.602	1.006
	[Q48=1]	-6.101	.939	42.239	1	.000	.002
	[Q48=2]	9.827	.000	.	1	.	18521.666
	[Q48=3]	-2.627	.574	20.933	1	.000	.072
	[Q48=4]	-1.007	.715	1.981	1	.159	.365
	[Q48=5]	-.417	.642	.422	1	.516	.659
	[Q48=6]	.512	.821	.389	1	.533	1.668
	[Q48=7]	.190	.591	.104	1	.747	1.210
	[Q48=8]	.311	.541	.331	1	.565	1.365
	[Q48=9]	.413	.661	.390	1	.533	1.511
	[Q48=10]	0[b]	.	.	0	.	.

	[Q50=1]	-1.297	.711	3.333	1	.068	.273
	[Q50=2]	-.231	.906	.065	1	.799	.794
	[Q50=3]	.943	1.045	.815	1	.367	2.569
	[Q50=4]	.776	.806	.927	1	.336	2.173
	[Q50=5]	.774	.772	1.007	1	.316	2.169
	[Q50=6]	.483	.755	.409	1	.522	1.621
	[Q50=7]	.535	.767	.486	1	.486	1.707
	[Q50=8]	.527	.763	.478	1	.490	1.694
	[Q50=9]	.137	.836	.027	1	.869	1.147
	[Q50=10]	0[b]	.	.	0	.	.
	[Q60=1]	3.430	.682	25.293	1	.000	30.891
	[Q60=2]	3.356	.628	28.593	1	.000	28.682
	[Q60=3]	2.158	.683	9.985	1	.002	8.656
	[Q60=4]	0[b]	.	.	0	.	.
	[Q260=1]	-.211	.321	.433	1	.511	.809
	[Q260=2]	0[b]	.	.	0	.	.
Not very happy	Intercept	-1.496	1.119	1.788	1	.181	
	Q262	.004	.011	.145	1	.703	1.004
	[Q48=1]	-4.519	.936	23.302	1	.000	.011
	[Q48=2]	48.735	.000	.	1	.	14635592916234 51400000.000
	[Q48=3]	-1.339	.597	5.036	1	.025	.262
	[Q48=4]	-.005	.737	.000	1	.995	.995
	[Q48=5]	.317	.668	.226	1	.634	1.374
	[Q48=6]	.964	.843	1.310	1	.252	2.623
	[Q48=7]	.464	.617	.565	1	.452	1.590
	[Q48=8]	.338	.571	.352	1	.553	1.403
	[Q48=9]	.564	.690	.666	1	.414	1.757
	[Q48=10]	0[b]	.	.	0	.	.
	[Q50=1]	-.009	.753	.000	1	.990	.991
	[Q50=2]	.866	.941	.846	1	.358	2.377
	[Q50=3]	1.158	1.085	1.139	1	.286	3.183
	[Q50=4]	.566	.858	.435	1	.509	1.761
	[Q50=5]	.501	.823	.371	1	.543	1.651
	[Q50=6]	.086	.813	.011	1	.916	1.089
	[Q50=7]	.229	.821	.077	1	.781	1.257
	[Q50=8]	.233	.817	.081	1	.776	1.262

	[Q50=9]	-.068	.900	.006	1	.939	.934
	[Q50=10]	0[b]	.	.	0	.	.
	[Q60=1]	3.059	.912	11.237	1	.001	21.298
	[Q60=2]	3.162	.867	13.308	1	.000	23.627
	[Q60=3]	2.416	.908	7.079	1	.008	11.199
	[Q60=4]	0[b]	.	.	0	.	.
	[Q260=1]	-.105	.334	.099	1	.753	.900
	[Q260=2]	0[b]	.	.	0	.	.

Its worth mentioning that the base category for my response variable is Not at all happy.

- **For the "Very Happy" Group:**

1. Freedom of Choice and Control (Q48):

- [Q48=1] (No Choice at All): B=-5.531, expB=.004

- The estimated odds ratio of choosing “very happy” in level of happiness individuals choosing “no choice at all” in freedom of choice and control compared to those chose “a great deal of choice” is 99.6% lower than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q48=2] (2nd Scale): B=13.082, expB=480091.948

- The estimated odds ratio of choosing “very happy” level of happiness individuals choosing “2nd scale” in freedom of choice and control compared to those chose “a great deal of choice” is 48009094.8% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q48=3] (3rd Scale): B=-2.992 ,expB=.050

- The estimated odds ratio of choosing “very happy” level of happiness individuals choosing “3rd scale” in freedom of choice and control compared to those chose “a great deal of choice” is 50% lower than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

2. Satisfaction with Financial Situation of Household (Q50):

- [Q50=1] (Completely Dissatisfied): B=-2.027 expB= .132

- The estimated odds ratio of choosing “very happy” level of happiness individuals choosing “completely dissatisfied” in financial situation of household compared to those chose

“completely satisfied” is 86.8% lower than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

3. Trust in People Personally Known (Q60):

- [Q60=1] (Trust Completely): B=3.021 expB= 20.519

- The estimated odds ratio of choosing “very happy” level of happiness individuals choosing “trust completely” in people personally known compared to those chose “don’t trust at all” is 1951.9% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q60=2] (Trust Somewhat): B=2.798 expB= 16.420

- The estimated odds ratio of choosing “very happy” level of happiness individuals choosing “trust somewhat” in people personally known compared to those chose “don’t trust at all” is 1542% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q60=3] (Do Not Trust Very Much): B=1.727 expB= 5.626

- The estimated odds ratio of choosing “very happy” level of happiness individuals choosing “don’t trust very much” in people personally known compared to those chose “don’t trust at all” is 462.6% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- **For the "Quite Happy" Group:**

1. Freedom of Choice and Control (Q48):

- [Q48=1] (No Choice at All): B=-6.101 expB= .002

- The estimated odds ratio of choosing “quite happy” level of happiness individuals choosing “no choice at all” Freedom of Choice and Control compared to those chose “great deal of choice” is 99.8% lower than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q48=2] (2nd Scale): B=9.827 expB= 18521.666

- The estimated odds ratio of choosing “quite happy” level of happiness individuals choosing “2nd scale” Freedom of Choice and Control compared to those chose “great deal of choice” is 1852066.6% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q48=3] (3rd Scale): B=-2.627expB= .072

- The estimated odds ratio of choosing “quite happy” level of happiness individuals choosing “3rd scale” Freedom of Choice and Control compared to those chose “great deal of choice” is 28% lower than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

2. Trust in People Personally Known (Q60):

- [Q60=1] (Trust Completely): B= 3.430 expB=30.891

- The estimated odds ratio of choosing “quite happy” level of happiness individuals choosing “trust completely” in people personally known compared to those chose “don’t trust at all” is 2989.1`% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q60=2] (Trust Somewhat): B= 3.356 expB=28.682

- The estimated odds ratio of choosing “quite happy” level of happiness individuals choosing “trust somewhat” in people personally known compared to those chose “don’t trust at all” is 2768.2% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q60=3] (Do Not Trust Very Much): B= 2.158 expB=8.656

- The estimated odds ratio of choosing “quite happy” level of happiness individuals choosing “don’t trust very much” in people personally known compared to those chose “don’t trust at all” is 765.6% higher than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- **For the "Not Very Happy" Group:**

1. Freedom of Choice and Control (Q48):

- [Q48=1] (No Choice at All): B= -4.519 expB=.011

- The estimated odds ratio of choosing “not very happy” level of happiness individuals choosing “no choice at all” Freedom of Choice and Control compared to those chose “great deal of choice” is 98.9% lower than the corresponding odds ratio of choosing “Not at all happy” level of happiness, holding other variables constant.

- [Q48=2] (2nd Scale): B=48.735 expB=14635592916234514

- The estimated odds ratio of choosing “not very happy” level of happiness individuals choosing “2nd scale” Freedom of Choice and Control compared to those chose “great deal of

choice" is 1.463559291623e18% higher than the corresponding odds ratio of choosing "Not at all happy" level of happiness, holding other variables constant.

- [Q48=2] (3rd Scale): B=-1.339 expB=.262

- The estimated odds ratio of choosing "not very happy" level of happiness individuals choosing "3rd scale" Freedom of Choice and Control compared to those chose "great deal of choice" is 73.8% lower than the corresponding odds ratio of choosing "Not at all happy" level of happiness, holding other variables constant.

2. Trust in People Personally Known (Q60):

- [Q60=1] (Trust Completely): B= 3.059 expB=21.298

- The estimated odds ratio of choosing "not very happy" level of happiness individuals choosing "trust completely" in people personally known compared to those chose "don't trust at all" is 2029.8% higher than the corresponding odds ratio of choosing "Not at all happy" level of happiness, holding other variables constant.

- [Q60=2] (Trust Somewhat): B= 3.162 expB=23.627

- The estimated odds ratio of choosing "not very happy" level of happiness individuals choosing "trust somewhat" in people personally known compared to those chose "don't trust at all" is 2267% higher than the corresponding odds ratio of choosing "Not at all happy" level of happiness, holding other variables constant.

- [Q60=3] (Do Not Trust Very Much): B= 2.416 expB=11.199

- The estimated odds ratio of choosing "not very happy" level of happiness individuals choosing "don't trust very much" in people personally known compared to those chose "don't trust at all" is 1019.9% higher than the corresponding odds ratio of choosing "Not at all happy" level of happiness, holding other variables constant.

Assess goodness of fit for the multinominal model:

As the I have a continuous explanatory variable, I will use Hosmer and Lemeshow Test p-value for the test: 0.000.

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

Therefore, reject H0. this indicated that the model isn't good fit.

Conclusion:

Throughout the analysis, various models were employed to understand the determinants of happiness levels among individuals in the United States. The analysis involved both binary logistic regression and multinomial logistic regression models. Here is a summary of the findings:

Binary Logistic Model:

Initially, a binary logistic model was developed using a more general response variable with only two categories. An interaction term was included, and two different methods (enter and backward) were used to build the model. After comparing the models, the backward method was found to be better in terms of model fit and significance of variables. The chosen model provided valuable insights into the significant predictors of happiness levels.

Ordinal Binary Logistic Model:

An attempt was made to compute an ordinal binary logistic model. However, it was discovered that the assumptions underlying this model were violated. Consequently, the analysis moved forward to explore a multinomial logistic regression model.

Multinomial Logistic Model:

The multinomial logistic regression model proved to be not q good fit for the data. But it allowed for the examination of the different categories of happiness levels, providing a more comprehensive understanding of the factors influencing happiness. The significant variables were identified, and their interpretations were provided, shedding light on their impact on reported happiness levels.

In conclusion, the analysis encompassed both binary and multinomial logistic regression models. The binary logistic model with the backward method was chosen as the preferred model for the binary response variable. The ordinal binary logistic model was not pursued due to violated assumptions, leading to the adoption of the multinomial logistic regression model.

Appendix :

Model1:

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	112.544	26	.000
	Block	112.544	26	.000
	Model	112.544	26	.000

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Q48			33.112	9	.000	
	Q48(1)	-2.161	.727	8.826	1	.003	.115
	Q48(2)	-3.394	.864	15.427	1	.000	.034
	Q48(3)	-1.957	.703	7.748	1	.005	.141
	Q48(4)	-.960	.817	1.381	1	.240	.383
	Q48(5)	-.732	.714	1.052	1	.305	.481
	Q48(6)	16.653	2650.566	.000	1	.995	17069718.761
	Q48(7)	.308	.757	.166	1	.684	1.361
	Q48(8)	1.031	.841	1.503	1	.220	2.804
	Q48(9)	16.459	1985.399	.000	1	.993	14068518.548
	Q60			7.043	3	.071	
	Q60(1)	2.968	1.400	4.491	1	.034	19.448
	Q60(2)	2.145	1.024	4.390	1	.036	8.543
	Q60(3)	1.083	1.049	1.067	1	.302	2.953
	Q50			10.620	9	.303	
	Q50(1)	-.576	.764	.569	1	.451	.562
	Q50(2)	-.422	.947	.199	1	.656	.656
	Q50(3)	17.072	2967.101	.000	1	.995	25946834.467
	Q50(4)	1.047	1.033	1.027	1	.311	2.849
	Q50(5)	1.243	1.000	1.546	1	.214	3.466
	Q50(6)	.372	.857	.188	1	.664	1.450
	Q50(7)	.801	.978	.671	1	.413	2.228
	Q50(8)	1.064	.986	1.165	1	.281	2.899
	Q50(9)	-.039	.970	.002	1	.968	.962

	Q260(1)	.646	1.384	.218	1	.640	1.909
	Q262	.012	.014	.736	1	.391	1.012
	Q260 * Q60			.928	3	.819	
	Q260(1) by Q60(1)	-1.562	1.842	.720	1	.396	.210
	Q260(1) by Q60(2)	-1.147	1.504	.582	1	.446	.318
	Q260(1) by Q60(3)	-1.419	1.574	.813	1	.367	.242
	Constant	1.837	1.233	2.222	1	.136	6.281

a. Variable(s) entered on step 1: Q48, Q60, Q50, Q260, Q262, Q260 * Q60 .

Model0 with 0.5 cutoff point:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Q48			33.112	9	.000	
	Q48(1)	-2.161	.727	8.826	1	.003	.115
	Q48(2)	-3.394	.864	15.427	1	.000	.034
	Q48(3)	-1.957	.703	7.748	1	.005	.141
	Q48(4)	-.960	.817	1.381	1	.240	.383
	Q48(5)	-.732	.714	1.052	1	.305	.481
	Q48(6)	16.653	2650.566	.000	1	.995	17069718.761
	Q48(7)	.308	.757	.166	1	.684	1.361
	Q48(8)	1.031	.841	1.503	1	.220	2.804
	Q48(9)	16.459	1985.399	.000	1	.993	14068518.548
	Q60			7.043	3	.071	
	Q60(1)	2.968	1.400	4.491	1	.034	19.448
	Q60(2)	2.145	1.024	4.390	1	.036	8.543
	Q60(3)	1.083	1.049	1.067	1	.302	2.953
	Q50			10.620	9	.303	
	Q50(1)	-.576	.764	.569	1	.451	.562
	Q50(2)	-.422	.947	.199	1	.656	.656
	Q50(3)	17.072	2967.101	.000	1	.995	25946834.467
	Q50(4)	1.047	1.033	1.027	1	.311	2.849
	Q50(5)	1.243	1.000	1.546	1	.214	3.466
	Q50(6)	.372	.857	.188	1	.664	1.450

	Q50(7)	.801	.978	.671	1	.413	2.228
	Q50(8)	1.064	.986	1.165	1	.281	2.899
	Q50(9)	-.039	.970	.002	1	.968	.962
	Q260(1)	.646	1.384	.218	1	.640	1.909
	Q262	.012	.014	.736	1	.391	1.012
	Q260 * Q60			.928	3	.819	
	Q260(1) by Q60(1)	-1.562	1.842	.720	1	.396	.210
	Q260(1) by Q60(2)	-1.147	1.504	.582	1	.446	.318
	Q260(1) by Q60(3)	-1.419	1.574	.813	1	.367	.242
	Constant	1.837	1.233	2.222	1	.136	6.281
Step 2[a]	Q48			31.819	9	.000	
	Q48(1)	-2.196	.723	9.223	1	.002	.111
	Q48(2)	-3.285	.870	14.243	1	.000	.037
	Q48(3)	-1.946	.703	7.654	1	.006	.143
	Q48(4)	-1.058	.802	1.740	1	.187	.347
	Q48(5)	-.767	.713	1.158	1	.282	.464
	Q48(6)	16.600	2660.846	.000	1	.995	16188896.923
	Q48(7)	.280	.757	.137	1	.711	1.324
	Q48(8)	.985	.839	1.379	1	.240	2.678
	Q48(9)	16.448	1991.372	.000	1	.993	13902051.121
	Q60			11.562	3	.009	
	Q60(1)	2.097	.875	5.748	1	.017	8.144
	Q60(2)	1.580	.757	4.358	1	.037	4.857
	Q60(3)	.395	.790	.251	1	.617	1.485
	Q50			10.739	9	.294	
	Q50(1)	-.536	.758	.499	1	.480	.585
	Q50(2)	-.353	.940	.141	1	.707	.702
	Q50(3)	17.142	2967.698	.000	1	.995	27854116.978
	Q50(4)	1.140	1.030	1.225	1	.268	3.126
	Q50(5)	1.294	.994	1.697	1	.193	3.649
	Q50(6)	.422	.850	.246	1	.620	1.525
	Q50(7)	.836	.974	.736	1	.391	2.306
	Q50(8)	1.104	.979	1.272	1	.259	3.017
	Q50(9)	.028	.961	.001	1	.977	1.028
	Q260(1)	-.526	.401	1.724	1	.189	.591
	Q262	.011	.014	.646	1	.422	1.011
	Constant	2.416	1.096	4.864	1	.027	11.204
Step 3[a]	Q48			31.955	9	.000	
	Q48(1)	-2.197	.723	9.247	1	.002	.111

	Q48(2)	-3.315	.869	14.539	1	.000	.036
	Q48(3)	-1.975	.705	7.852	1	.005	.139
	Q48(4)	-1.106	.801	1.904	1	.168	.331
	Q48(5)	-.792	.711	1.239	1	.266	.453
	Q48(6)	16.607	2662.551	.000	1	.995	16297389.627
	Q48(7)	.262	.758	.120	1	.730	1.300
	Q48(8)	.976	.839	1.353	1	.245	2.653
	Q48(9)	16.476	1990.113	.000	1	.993	14298777.901
	Q60			12.889	3	.005	
	Q60(1)	2.145	.870	6.074	1	.014	8.538
	Q60(2)	1.620	.753	4.625	1	.032	5.056
	Q60(3)	.366	.785	.218	1	.641	1.443
	Q50			10.754	9	.293	
	Q50(1)	-.586	.761	.593	1	.441	.556
	Q50(2)	-.462	.936	.244	1	.621	.630
	Q50(3)	17.122	2963.831	.000	1	.995	27276551.462
	Q50(4)	1.092	1.037	1.109	1	.292	2.980
	Q50(5)	1.240	.997	1.547	1	.214	3.456
	Q50(6)	.347	.851	.167	1	.683	1.415
	Q50(7)	.770	.976	.623	1	.430	2.161
	Q50(8)	1.039	.980	1.124	1	.289	2.826
	Q50(9)	.005	.964	.000	1	.996	1.005
	Q260(1)	-.473	.396	1.429	1	.232	.623
	Constant	2.889	.926	9.721	1	.002	17.968
Step 4[a]	Q48			32.315	9	.000	
	Q48(1)	-2.232	.720	9.612	1	.002	.107
	Q48(2)	-3.361	.864	15.128	1	.000	.035
	Q48(3)	-2.011	.703	8.180	1	.004	.134
	Q48(4)	-1.113	.800	1.937	1	.164	.329
	Q48(5)	-.855	.708	1.456	1	.228	.425
	Q48(6)	16.536	2687.919	.000	1	.995	15180949.602
	Q48(7)	.226	.756	.090	1	.765	1.254
	Q48(8)	.930	.837	1.236	1	.266	2.535
	Q48(9)	16.419	2003.258	.000	1	.993	13504525.581
	Q60			12.293	3	.006	
	Q60(1)	2.093	.861	5.908	1	.015	8.112
	Q60(2)	1.605	.743	4.671	1	.031	4.979
	Q60(3)	.402	.775	.270	1	.604	1.495

Q50			10.487	9	.313	
Q50(1)	-.543	.763	.507	1	.476	.581
Q50(2)	-.419	.939	.199	1	.655	.658
Q50(3)	17.163	2972.408	.000	1	.995	28418307.762
Q50(4)	1.070	1.033	1.073	1	.300	2.914
Q50(5)	1.277	1.001	1.629	1	.202	3.586
Q50(6)	.319	.849	.141	1	.707	1.376
Q50(7)	.735	.978	.564	1	.453	2.085
Q50(8)	1.065	.986	1.166	1	.280	2.899
Q50(9)	-.035	.968	.001	1	.971	.965
Constant	2.663	.903	8.694	1	.003	14.339

a. Variable(s) entered on step 1: Q48, Q60, Q50, Q260, Q262, Q260 * Q60 .

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	112.544	26	.000
	Block	112.544	26	.000
	Model	112.544	26	.000
Step 2[a]	Step	-.936	3	.817
	Block	111.608	23	.000
	Model	111.608	23	.000
Step 3[a]	Step	-.662	1	.416
	Block	110.946	22	.000
	Model	110.946	22	.000
Step 4[a]	Step	-1.454	1	.228
	Block	109.492	21	.000
	Model	109.492	21	.000

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.