The Happiness of People Depending on Different Aspects of Life Using Categorical Data Analysis.

Presented by:

Shahd Hesham ElSayed Hassan

Code:

5200384

Presented to:

Dr. Sally Abdelfattah

Teaching Assistant:

Dr. Menna Medhat

Introduction:

This paper is concerned with studying the happiness of people. Through the lens of categorical data analysis, we aim to reveal the effect of different life aspects including security, religion, health, economy, corruption, and demographics on the happiness of Japan's citizens. This study's primary goal is not only to shed light on how these factors affect happiness but also to make informed decisions and implement strategies to help them foster happier lives.

Therefore, our research questions that we are willing to answer after analyzing the data are:

1. Are people living in Japan happy or not?
2. Does the perception of corruption in public institutions impact the happiness levels of individuals in Japan?
3. Does the perception of security influence the happiness levels of individuals in Japan?
4. Does the general health status affect people's happiness?
5. To what extent does social class influence people's happiness levels?
6. How does the perceived importance of God correlate with the reported levels of happiness?

Methodology:

In the examination ahead, we'll use a dataset from the **World Values Survey (WVS – Wave 7).** A sample of 1353 individuals living in Japan was drawn. We'll encompass six categorical variables (qualitative ordinal), one categorical response variable, and five explanatory variables each representing a distinct aspect of peoples' lives. To analyze our data, we will be using **Excel** and **STATA software.**

List of questions used from the survey:

- Q46: Taking all things together, would you say you are (read the categories).
- Q47: All in all, how would you describe your state of health these days? Would you say it is (read the categories).
- Q131: Could you tell me how secure do you feel these days?
- Q287: People sometimes describe themselves as belonging to the working class, the middle class, or the upper or lower class. Would you describe yourself as belonging to the (read the categories)?
- Q164: How important is God in your life?

- Q112: Now I'd like you to tell me your views on corruption – when people pay a bribe, give a gift, or do a favor to other people to get the things they need to be done or the services they need. How would you place your views on corruption in your country on a 10-point scale?

Variables:

1. Happiness:

This is the response variable in our study, the interviewer asked participants about their level of happiness and asked them to say in which of the following four categories they fall in, (Very happy, rather happy, not very happy, and not at all happy), these four categories are coded as (1,2,3, and 4) respectively. The following frequency table shows the number of individuals in each category.

Row Labels	Count of Q46: Feeling of happiness	Count of Q46: Feeling of happiness2
1	405	30.59%
2	790	59.67%
3	120	9.06%
4	9	0.68%
Grand Total	**1324**	**100.00%**

2. State of Health:

This is an explanatory variable that shows the effect of health on the happiness of individuals. It has 5 categories (Very good, good, fair, poor, and very poor), these four categories are coded as (1,2,3,4, and 5) respectively.

Row Labels	Count of Q47: State of health (subjective)	Count of Q47: State of health (subjective)2
1	187	13.94%
2	506	37.73%
3	488	36.39%
4	145	10.81%
5	15	1.12%
Grand Total	**1341**	**100.00%**

3. Feeling secure in the neighborhood:

This explanatory variable highly relates to the security aspect, as respondents say whether they feel secure in their neighborhood or not. The responses are divided into four categories (Very secure, quite secure, not very secure, not at all secure), these four categories are coded as (1,2,3, and 4) respectively.

Row Labels	Count of Q131: Secure in neighborhood	Count of Q131: Secure in neighborhood2
1	39	2.93%
2	753	56.62%
3	476	35.79%
4	62	4.66%
Grand Total	**1330**	**100.00%**

4. Social Class:

In this question, respondents were asked to classify themselves according to which social class they think they belong to. This is related to measuring whether individuals lie in a high or low socioeconomic class can affect their happiness. There are 5 categories that respondents may choose to place themselves in (Upper class, upper middle class, lower middle class, working class, and lower class) and these categories were coded as (1,2,3,4, and 5) respectively.

Row Labels	Count of Q287: Social class (subjective)	Count of Q287: Social class (subjective)2
1	20	1.56%
2	203	15.85%
3	571	44.57%
4	357	27.87%
5	130	10.15%
Grand Total	**1281**	**100.00%**

5. Importance of God:

Participants were asked to rank how important God is in their lives on a scale from (1 to 10) where 1 means “not important at all” and 10 means “very important”. This variable represents the religious aspect. I re-coded this to convert the scale from 1 to 5 to make it easier to interpret. The following categories are the ones resulting from recoding (Not important, slightly important, fairly important, important, and very important) and they were coded as (1,2,3,4, and 5) respectively.

Row Labels	Count of Q164: Importance of God	Count of Q164: Importance of God2
1	356	28.57%
2	241	19.34%
3	345	27.69%
4	207	16.61%
5	97	7.78%
Grand Total	**1246**	**100.00%**

6. Perceptions of corruption in the country:

Respondents were asked to place their views on corruption in Japan on a 10-point scale, where “1” means “there is no corruption in my country” and “10” means “there is abundant corruption in my country”. This 10-point scale was converted to a 5-point scale where the following categories were assigned after mixing the old categories, (Not at all corrupted, slightly corrupted, moderately corrupted, very corrupted, extremely corrupted) and were coded as follows (1,2,3,4, and 5) respectively.

Row Labels	Count of Q112: Perceptions of corruption in the country	Count of Q112: Perceptions of corruption in the country2
1	25	2.02%
2	134	10.85%
3	317	25.67%
4	505	40.89%
5	254	20.57%
Grand Total	**1235**	**100.00%**

Noting that all frequencies don’t sum up to the sample size (1353), as there exist missing values because people either refused to give an answer or don’t know the answer to the question.

Testing independence and measuring association

This part of the report investigates the relationship between the ordinal response variable "Happiness" and various explanatory variables in Japan. The exploration of these associations is conducted primarily through the construction of 2x2 contingency tables, offering a comprehensive overview of how these factors interplay with the subjective well-being of individuals in Japan.

As a key focus, we employ the Goodman and Kruskal gamma as the measure of association, to quantify and assess the strength and directionality of the monotone relationships between the ordinal response variable and the explanatory variables.

To enhance the depth of analysis, this report delves into the investigation of a 3x3 contingency table specifically examining the impact of the state of health on the well-being/ happiness of individuals living in Japan and introducing the socioeconomic level as the control variable (Z).

I. Two-Way Contingency Tables:

This part includes constructing contingency tables to study the association between our response variable (happiness) and each explanatory categorical variable using suitable measures according to the types of categorical variables whether it's nominal or ordinal, in our study, we only included ordinal variables.

- **Testing the relationship between feeling happy and the state of health:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Very good	137	44	5	0	186
Good	177	315	10	0	502
Fair	76	339	58	3	476
Poor	13	83	37	5	138
Very poor	0	7	7	1	15
Sum	403	788	117	9	1317

According to the above table, it is significant that the frequency of **happy** people living in Japan with a **fair** state of health is the highest among all frequencies. On the other hand, the frequency of those **not at all happy** and are in a **very good/ good** state of health is the lowest, as those who are **very happy** and in a **very poor** state of health.

We'll start by testing the monotone relationship between feeling happy and the state of health, since both variables are ordinal then it's better not to use Pearson's Chi-Square test of independence and instead use **Z=** $\frac{\hat{\gamma}}{s.e(\hat{\gamma})}$ **,**this test statistic is for large sample sizes and since our sample size is **(1317)** then we can apply this test.

$$\boldsymbol{H_0}: \gamma = 0$$

$$\boldsymbol{H}_1: \gamma \neq 0$$

When applying the code in R to get the values for the Goodman and Kruskal Gamma measure of association of the variables under consideration, the following results were displayed:

```
> gkgamma(contingency_table1)

        Goodman-Kruskal's gamma for ordinal categorical data

data:  contingency_table1
Z = 18.15, p-value < 2.2e-16
95 percent confidence interval:
 0.5887047 0.7015528
sample estimates:
Goodman-Kruskal's gamma
              0.6451288
```

The test statistic of independence for ordinal variables is **Z= 18.15**.

Also, since the p-value is very small **<0.05**, we reject the null hypothesis and conclude that there is a monotonic relationship between the two variables.

Since the feeling of happiness and the state of health has a monotone relation, we can move forward to measuring the degree to which the relationship between them is monotone using Goodman and Kruskal Gamma.

The estimated Gamma = **0.645.** We can conclude that there is a strong positive relationship between the feeling of happiness and the state of health, more healthy people tend to be more happy.

✓ **The conditional probability of feeling happy given the state of health:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Very good	0.7365591	0.2365591	0.02688172	0.000000000	1
Good	0.3525896	0.6274900	0.01992032	0.000000000	1
Fair	0.1596639	0.7121849	0.12184874	0.006302521	1
Poor	0.0942029	0.6014493	0.26811594	0.036231884	1
Very poor	0.0000000	0.4666667	0.46666667	0.066666667	1
Sum	1.3430155	2.6443500	0.90343339	0.109201072	5

- **Testing the relationship between feeling happy and feeling secure in the neighborhood:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Very secure	19	17	1	0	37
Quite secure	242	444	57	2	745
Not very secure	122	286	51	4	463
Not at all secure	19	29	8	3	59
Sum	402	776	117	9	1304

According to the above table, the frequency of **happy** people who feel **quite secure** in their neighborhood is the highest. While the lowest frequency lies in those who are **not at all happy** and **very secure**, which is equal to zero.

When checking for the existence of a monotone relationship between the feeling of happiness and feeling secure the following results were obtained:

```
> gkgamma(contingency_table2)

        Goodman-Kruskal's gamma for ordinal categorical data

data:  contingency_table2
Z = 3.5616, p-value = 0.0003686
95 percent confidence interval:
 0.07830079 0.26469584
sample estimates:
Goodman-Kruskal's gamma
              0.1714983
```

Since the p-value associated is **less than 0.05** then we conclude that there is a monotone relationship between feeling happy and feeling secure.
Then we can calculate the degree of monotonicity between them using Gamma.

The estimated Gamma = **0.171,** indicates that there is a positive weak monotone relationship between being happy and feeling secure in the neighborhood among people living in Japan, which means that people who feel more secure tend to be happier.

✓ **The conditional probability of feeling happy given feeling secure:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Very secure	0.5135135	0.4594595	0.02702703	0.000000000	1
Quite secure	0.3248322	0.5959732	0.07651007	0.002684564	1
Not very secure	0.2634989	0.6177106	0.11015119	0.008639309	1
Not at all secure	0.3220339	0.4915254	0.13559322	0.050847458	1
Sum	1.4238785	2.1646686	0.34928150	0.062171330	4

- **Testing the relationship between feeling happy and the social class:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Upper class	11	8	0	0	19
Upper middle class	103	95	5	0	203
Lower middle class	174	362	25	1	562
Working class	72	220	56	1	349
Lower class	19	69	32	6	126
Sum	379	754	118	8	1259

From the above table, we can notice that those who are **happy** and belong to the **lower middle class** are those with the highest frequency.

Furthermore, we need to check the existence of a monotone relationship between social class and feeling happy.

```
> gkgamma(contingency_table3)

        Goodman-Kruskal's gamma for ordinal categorical data

data:  contingency_table3
Z = 11.073, p-value < 2.2e-16
95 percent confidence interval:
 0.3714430 0.5132675
sample estimates:
Goodman-Kruskal's gamma
              0.4423553
```

Since the p-value **< 0.05**, there exists a monotone relationship between the two ordinal variables under consideration (social class and happiness).

By assessing the strength of this monotone relation, we'll check for the value of Gamma.

The estimated value of gamma = **0.442,** indicates a moderate positive relationship between the socioeconomic class one belongs to and his feeling of happiness, as the socioeconomic class increases the level of happiness tends to increase.

✓ **The conditional probability of feeling happy given the socioeconomic level:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Upper class	0.5789474	0.4210526	0.00000000	0.000000000	1
Upper middle class	0.5073892	0.4679803	0.02463054	0.000000000	1
Lower middle class	0.3096085	0.6441281	0.04448399	0.001779359	1
Working class	0.2063037	0.6303725	0.16045845	0.002865330	1
Lower class	0.1507937	0.5476190	0.25396825	0.047619048	1
Sum	1.7530424	2.7111526	0.48354123	0.052263737	5

- **Testing the relationship between happiness and the importance of God in someone's life:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Not important	98	211	40	2	351
Slightly important	64	156	18	0	238
Fairly important	105	200	31	3	339
Important	63	122	13	1	199
Very important	40	44	11	0	95
Sum	370	733	113	6	1222

It's clear that those who are **happy** and the level of importance of God in their life is **fair** have the highest frequency among all groups.

When testing the monotone relation between happiness and the importance of God the following results were obtained:

```
> gkgamma(contingency_table4)

        Goodman-Kruskal's gamma for ordinal categorical data

data:  contingency_table4
Z = -2.2748, p-value = 0.02292
95 percent confidence interval:
 -0.16780649 -0.01271044
sample estimates:
Goodman-Kruskal's gamma
            -0.09025846
```

Since the p-value is **0.022< 0.05**, we can conclude that there is a monotone relation between both variables. Moving forward we'll use the Gamma measure to test the degree to which the relation is monotone.
The estimated gamma = **-0.0902,** then there exists a positive weak relation between the importance of God in one's life and happiness, as the importance of God increases, happiness tends to increase.

✓ **The conditional probability of feeling happy given the importance of God:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Not important	0.2792023	0.6011396	0.11396011	0.005698006	1
Slightly important	0.2689076	0.6554622	0.07563025	0.000000000	1
Fairly important	0.3097345	0.5899705	0.09144543	0.008849558	1
Important	0.3165829	0.6130653	0.06532663	0.005025126	1
Very important	0.4210526	0.4631579	0.11578947	0.000000000	1
Sum	1.5954799	2.9227955	0.46215190	0.019572689	5

- **Testing the relationship between happiness and perceptions of corruption in the country:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Not at all corrupted	11	11	2	1	25
Slightly corrupted	50	77	5	0	132
Moderately corrupted	83	199	27	0	309
Very corrupted	138	312	44	2	496
Extremely corrupted	80	133	32	4	249
Sum	362	732	110	7	1211

The highest frequency lies among those who are **happy** and see that Japan is **very corrupt.**

Moreover, when assessing the relationship between being happy and corruption in Japan, the following results were obtained:

```
> gkgamma(contingency_table5)

        Goodman-Kruskal's gamma for ordinal categorical data

data:  contingency_table5
Z = 1.6786, p-value = 0.09324
95 percent confidence interval:
 -0.01166951  0.15518113
sample estimates:
Goodman-Kruskal's gamma
             0.07175581
```

Since the p-value **0.09>0.05** then there exists NO monotone relationship between happiness and the perception of corruption in Japan.

✓ **The conditional probability of feeling happy given the perception of corruption:**

	Very happy	Happy	Not very happy	Not at all happy	Sum
Not at all corrupted	0.4400000	0.4400000	0.08000000	0.040000000	1
Slightly corrupted	0.3787879	0.5833333	0.03787879	0.000000000	1
Moderately corrupted	0.2686084	0.6440129	0.08737864	0.000000000	1
Very corrupted	0.2782258	0.6290323	0.08870968	0.004032258	1
Extremely corrupted	0.3212851	0.5341365	0.12851406	0.016064257	1
Sum	1.6869072	2.8305151	0.42248116	0.060096515	5

II. Three-Way Contingency Table:

		Happiness				
Upper Class		**Very Happy**	**Happy**	**Not very happy**	**Not at all happy**	**Total**
	Very good	6	0	0	0	**6**
	Good	2	4	0	0	**6**
	Fair	2	4	0	0	**6**
	Poor	1	0	0	0	**1**
	Very poor	0	0	0	0	**0**
Upper Middle						
	Very good	47	4	0	0	**51**
	Good	36	42	0	0	**78**
	Fair	20	40	4	0	**64**
	Poor	0	8	1	0	**9**
	Very poor	0	0	0	0	**0**
Lower Middle						
	Very good	42	21	2	0	**65**
	Good	93	166	2	0	**261**
	Fair	33	145	11	0	**189**
	Poor	6	28	7	1	**42**
	Very poor	0	2	3	0	**5**
Working Class						
	Very good	25	15	2	0	**42**
	Good	26	71	6	0	**103**
	Fair	15	100	31	0	**146**
	Poor	4	30	11	1	**46**
	Very poor	0	4	3	0	**7**
Lower Class						
	Very good	7	4	1	0	**12**
	Good	6	24	2	0	**32**
	Fair	5	26	12	2	**45**
	Poor	1	14	16	3	**34**
	Very poor	0	1	1	1	**3**
TOTAL		**377**	**753**	**115**	**8**	***1253***

In this section, we're going to study the effect of the state of health on the feeling of happiness in Japan, taking into consideration the socioeconomic level as the control variable using the three-way contingency table. Logically, the socioeconomic level affects the state of health and therefore we considered it as the control variable.

We can see that we've five partial tables. The first table shows a two-way contingency table between feeling happy (Y) and the state of health (X) for those who belong to the **upper class** in Japan, the second for those who belong to the **upper middle class**, the third for those who belong to the **lower middle class**, the fourth is for those who belong to the **working class**, and the last table is for those who belong to the **lower class**.

- **Cochran-Mantel-Haenszel Test for Conditional Independence:**

$$\boldsymbol{H_0}: \text{there is conditional independence}$$
$$\boldsymbol{H_1}: \text{no conditional independence}$$

When performing the test, the following results were obtained:

```
> mantelhaen.test(three_way)

        Cochran-Mantel-Haenszel test

data:  three_way
Cochran-Mantel-Haenszel M^2 = 259.28, df = 12, p-value < 2.2e-16
```

Since the **p-value is less than 0.05** then we'll reject the assumption of conditional independence. This means that the association between the state of health and happiness is not consistent across different socioeconomic classes.

In addition, we know that after this step we must calculate the odds ratio but since we don't have a (2 x 2 x K) table we can't calculate it, and we can't decide on the association.

Moreover, since we rejected the assumption of conditional independence in the 3x3 contingency table, there is no necessity to examine mutual and joint independence separately. Conditional independence is considered the weakest form of independence among them. Thus, if we reject conditional independence, it inherently implies the rejection of both mutual and joint independence.

Reminder of the variables' coding:

Happiness: (Very happy, rather happy, not very happy, and not at all happy), these four categories are coded as (1,2,3, and 4) respectively.

State of health: It has 5 categories (Very good, good, fair, poor, and very poor), these five categories are coded as (1,2,3,4, and 5).

Feeling secure in the neighborhood: (Very secure, quite secure, not very secure, not at all secure), these four categories are coded as (1,2,3, and 4.)

Social class: (Upper class, upper middle class, lower middle class, working class, and lower class), and these categories were coded as (1,2,3,4, and 5) respectively.

Importance of God: (Not important, slightly important, fairly important, important, and very important), and they were coded as (1,2,3,4, and 5) respectively.

Perception of corruption: (Not at all corrupted, slightly corrupted, moderately corrupted, very corrupted, extremely corrupted) and were coded as follows (1,2,3,4, and 5) respectively.

Gender: It has two categories: "male" coded as 1 and "female" coded as 2.

Modeling

I. Binary Logistic Model:

To build the binary logistic model, I combined the "very happy" and "happy" categories into a single category called "happy," and the "not very happy" and "not at all happy" categories into another category called "not happy" for the response variable "Happiness." This simplification was done to create a binary response variable with two categories: "happy" and "not happy" recoded as 1 and 0, respectively.

Before proceeding with the analysis, I added a new variable, gender, as the sixth explanatory variable. It has two categories: "male" coded as 1 and "female" coded as 2.

Continuing with the analysis, when fitting the full model, only the explanatory variables "state of health" and "gender" showed significance. Consequently, I re-fitted the full model, introducing an **interaction term** that involves both significant variables.

✓ Backward stepwise regression:

After experimenting with three variable selection techniques—backward stepwise, forward stepwise, and the enter method—it became evident that the backward stepwise regression model had the lowest AIC (Akaike Information Criterion) among them. Consequently, I decided to continue the analysis with the model chosen through backward stepwise regression specifying the p-value of removal with **0.15**.

The model was fitted with the convention of having the first category as the reference category for each of the three variables. Specifically, for the variable "state of health," the reference category is "Very good." For the variable "social class," the reference category is "Upper class." Lastly, for the variable "gender," the reference category is "male."

$$\begin{aligned} \widehat{logit\pi} = & 3.545077e+07 + 8.666449e-01\, D_{12} + 1.653344e-01\, D_{13} + 8.970595e \\ & -02\, D_{14} + 4.330883e-02\, D_{15} + 2.696735e-06\, D_{22} + 1.660440e \\ & -06\, D_{23} + 4.443066e-07\, D_{24} + 2.619833e-07\, D_{25} + 2.201009e+00\, D_{32} \end{aligned}$$

The coding for the four dummy variables associated with the first variable, “state of health,” is represented as D_{1i} . Similarly, the four dummy variables corresponding to the “social class” are coded as D_{2i} , and for the “gender,” and they are coded as D_{3i}.

Variables	Coef	OR	Z-value	Pr(> |z|)
Intercept	17.3836556	3.545077e+07	0.030	0.976129
Health(Good)	-0.1431259	8.666449e-01	-0.212	0.831937
Health (Fair)	-1.7997854	1.653344e-01	-2.941	**0.003272**
Health (Poor)	-2.4112182	8.970595e-02	-3.795	**0.000148**
Health (Very poor)	-3.1393988	4.330883e-02	-3.732	**0.000190**
Social class (Upper middle)	-12.823468	2.696735e-06	-0.022	0.982390
Social class (Lower middle)	-13.308427	1.660440e-06	-0.023	0.981724
Social class (Working)	-14.626751	4.443066e-07	-0.025	0.979914
Social class (Lower)	-15.154985	2.619833e-07	-0.026	0.979188
Gender (Female)	0.7889157	2.201009e+00	3.319	**0.000904**

As shown from the above output produced from R software, there exist only two significant explanatory variables namely the "state of health" and "gender." Moreover, the **interaction term** for the previously mentioned variables was **removed** from the fitted model. After the interaction term was removed, I tried adding several interaction terms between other variables in the full model, but they were also removed from the backward model.

✓ Interpreting the significant coefficients:

1) $e^{\beta_{13}} = 0.1653344$: The estimated odds of being happy among those who have a fair state of health are lower than the odds of being happy among those with a very good state of health by about **83.5%** holding other variables in the model constant.
2) $e^{\beta_{14}} = 0.08970595$: The estimated odds of being happy among those who have a poor state of health are lower than the odds of being happy among those with a very good state of health by about **91%** holding other variables in the model constant.
3) $e^{\beta_{15}} = 0.04330883$: The estimated odds of being happy among those who have a very poor state of health are lower than the odds of being happy among those with a very good state of health by about **95.67%** holding other variables in the model constant.
4) $e^{\beta_{32}} = 2.201009$: The estimated odds of being happy among females are higher than the odds of being happy among males by **120%** holding other variables in the model constant.

As for the rest coefficients, they're all insignificant as their corresponding p-values are very large.

✓ Assessing the goodness of fit:

The following output indicates that the model is not a good fit according to the very small p-value. Despite experimenting with various combinations of variables, the conclusion remained the same. This may be attributed to the sample size and the possibility of not including all relevant explanatory variables that influence the response variable (happiness).

```
Pearson's Chi-squared test with Yates' continuity correction

                data:  conf_matrix1
X-squared = 138.36, df = 1, p-value < 2.2e-16
```

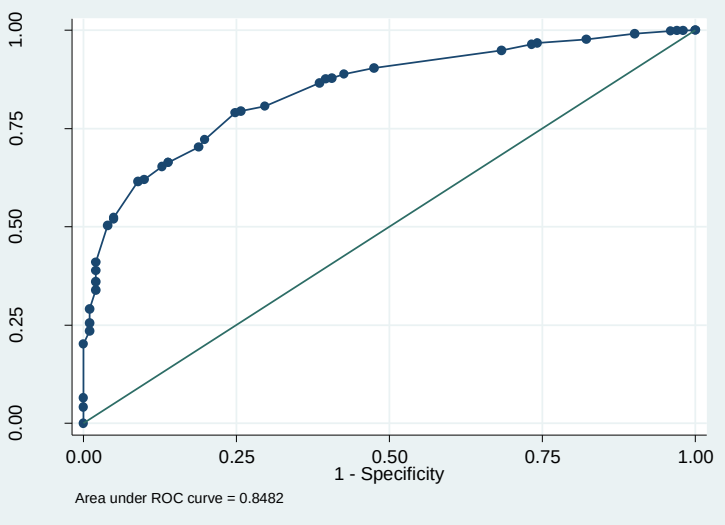
✓ Multicollinearity:

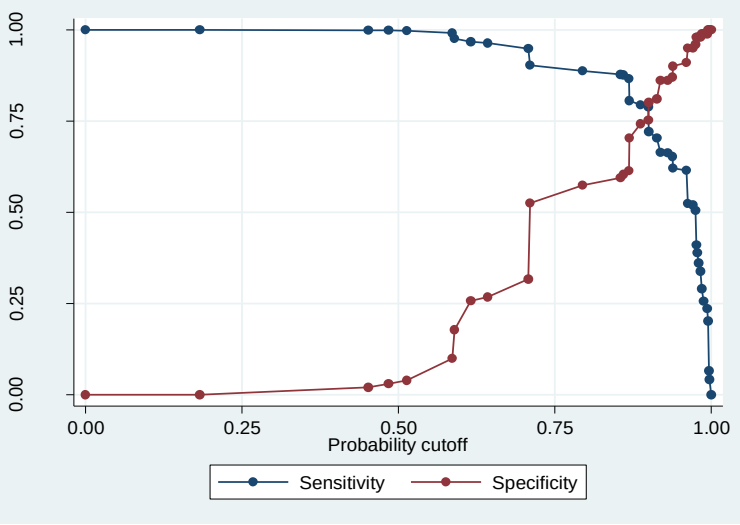
Upon assessing multicollinearity in the data using the (GVIF), a measure used to assess multicollinearity in multiple regression analysis. The results indicated the absence of multicollinearity. All GVIF values were found to be very low (less than 5, as a rule of thumb.) Therefore, we can proceed with the analysis without concerns related to multicollinearity.

Variable	**GVIF**	**Df**	**GVIF^(1/(2*Df))**
Health	1.062871	4	1.007651
Social Class	1.051814	4	1.006334
Gender	1.014269	1	1.007109

✓ The predictive power of the model:

First, using the "pROC" package in R software, we employed the ROC curve to identify the optimal cutoff point for reaching the optimal classification percentages. It was found out that the best cutoff point is **0.9**, and the area under the curve is **0.84**, which is greater than 0.5. Then according to the area under the curve, the model has a very good predictive power. Using STATA software, we're able to plot the following graphs.





Moving forward to the classification table with the optimal cutoff point, and using the "caret" package in R, the model demonstrates a high sensitivity of **79.2%**, indicating its effectiveness in correctly identifying individuals who consider themselves happy. Nevertheless, the specificity stands at **75.2%**, exceeding the 60% threshold. This indicates a high accuracy in correctly identifying individuals who consider themselves not happy. Moreover, the overall accuracy of the model stands at **78.8%**, representing the percentage of correctly classified instances across both positive and negative outcomes.

	True		
Classified	**D**	**~D**	**Total**
+	769	25	794
-	202	76	278
Total	971	101	**1072**

II. Multicategory logit models:

✓ Ordinal Logistic Model:

Since our response variable is ordinal, we tried to fit the ordinal logistic regression model. Using "omodel" package in STATA software, we were able to test the proportional odds assumption and it was shown that the assumption is violated since the p-value is less than 0.05. Therefore, it's not appropriate to fit the ordinal logistic model.

```
Approximate likelihood-ratio test of proportionality of odds
across response categories:
          chi2(32) =    288.89
        Prob > chi2 =    0.0000
```

✓ Multinomial Logistic Regression:

Since the proportional odds assumption is violated, we'll fit the multinomial logistic regression model, treating our response variable as a nominal one. Using STATA software, we're able to fit the model. Noting that STATA takes the most frequent category of the response variable as the base category. In our case, it considered the second category (rather happy) of our response variable, happiness, as the baseline category.

Happiness	**Freq.**	**Percent**	**Cumulative**
Very happy	318	29.66	29.66
Rather happy	653	60.91	90.58
Not very happy	97	9.05	99.63
Not at all happy	4	0.37	100

Moving forward to testing the model's significance. It was shown that the model as a whole is significant compared to the null model as the **p-value< 0.05**.

```
Multinomial logistic regression                 Number of obs     =      1,072
                                                LR chi2(60)       =     392.42
                                                Prob > chi2       =     0.0000
Log likelihood = -769.33906                     Pseudo R2         =     0.2032
```

Happiness	RRR	Standard_Errors	Z_Values	P_Values
Very Happy(1)				
Intercept	23.71267	21.66416	3.47	**0.001**
Health (Good)	0.1930819	0.0438461	-7.24	**0**
Health (Fair)	0.0951807	0.0233671	-9.58	**0**
Health (Poor)	0.0610599	0.0238048	-7.17	**0**
Health (Very poor)	1.05e(-08)	0.0000267	-0.01	0.994
Security (Quite secure)	0.4758441	0.2032306	-1.74	0.082
Security (Not very secure)	0.5044132	0.221956	-1.56	0.12
Security (Not at all secure)	0.3464375	0.2116675	-1.73	0.083
Social class (Upper middle)	0.5005952	0.3270656	-1.06	0.29
Social class (Lower middle)	0.2375696	0.1518994	-2.25	**0.025**
Social class (Working)	0.1533178	0.099673	-2.88	**0.004**
Social class (Lower)	0.1183592	0.0838439	-3.01	**0.003**
Importance of God (Slightly important)	0.8819213	0.2049966	-0.54	0.589
Importance of God (Fairly important)	1.304322	0.2752522	1.26	0.208
Importance of God (Important)	1.495706	0.3597557	1.67	0.094
Importance of God (Very important)	2.437217	0.7630044	2.85	**0.004**
Corruption (Slightly)	0.71233	0.39278	-0.62	0.538
Corruption (Moderately)	0.5283812	0.2822015	-1.19	0.232
Corruption (Very)	0.5491019	0.2891458	-1.14	0.255
Corruption (Extremely)	0.8541453	0.4606186	-0.29	0.77
Gender (Female)	1.431882	0.2305247	2.23	**0.026**
Not very happy (3)				
Intercept	7.39e(-09)	0.0000223	-0.01	0.955
Health (Good)	0.5950247	0.4167547	-0.74	0.459
Health (Fair)	2.69583	1.728975	1.55	0.122
Health (Poor)	4.337902	2.894483	2.2	**0.028**
Health (Very poor)	8.98566	7.746958	2.55	**0.011**
Security (Quite secure)	2.822993	3.41372	0.86	0.391
Security (Not very secure)	3.105526	3.781464	0.93	0.352
Security (Not at all secure)	4.197468	5.397881	1.12	0.265
Social class (Upper middle)	3462224	1.04e(+10)	0	0.996
Social class (Lower middle)	4501462	1.36e(10)	0.01	0.996
Social class (Working)	1.51e(+07)	4.55e(10)	0.01	0.996
Social class (Lower)	2.16e(+07)	6.53e(10)	0.01	0.996
Importance of God (Slighlty important)	0.77248	0.2713815	-0.73	0.462
Importance of God (Fairly important)	0.7400865	0.2309	-0.96	0.335
Importance of God (Important)	0.5138239	0.2139348	-1.6	0.11

Importance of God (Very important)	1.095637	0.5061379	0.2	0.843
Corruption (Slightly)	0.3270398	0.3207422	-1.14	0.254
Corruption (Moderately)	0.5914094	0.5239099	-0.59	0.553
Corruption (Very)	0.589483	0.5155358	-0.6	0.546
Corruption (Extremely)	0.7853552	0.6979905	-0.27	0.786
Gender (Female)	0.4715941	0.1187033	-2.99	**0.03**
Not at all happy (4)				
Intercept	6.94e(-25)	7.44e(-21)	-0.01	0.996
Health (Good)	0.5726752	1228.333	-0.00	1
Health (Fair)	523723.3	9.36e(08)	0.01	0.994
Health (Poor)	4740033	8.47e(09)	0.01	0.99
Health (Very poor)	0.4309145	4074.308	-0.00	1
Security (Quite secure)	33397.14	1.01e(+08)	0	0.997
Security (Not very secure)	8028.434	2.42e(+07)	0	0.998
Security (Not at all secure)	123704.3	3.73e(+08)	0	0.997
Social class (Upper middle)	0.6430965	5561.914	-0.00	1
Social class (Lower middle)	0.6212497	5295.888	-0.00	1
Social class (Working)	778188.8	6.56e(+09)	0	0.999
Social class (Lower)	6836218	5.76e(+10)	0	0.999
Importance of God (Slightly important)	1.76e(-07)	0.0003422	-0.01	0.994
Importance of God (Fairly important)	2.681613	3.751469	0.71	0.481
Importance of God (Important)	1.815951	3.253786	0.33	0.739
Importance of God (Very important)	2.63e(-07)	0.0005637	-0.01	0.994
Corruption (Slightly)	0.4338968	2607.216	-0.00	1
Corruption (Moderately)	0.1710914	983.9012	-0.00	1
Corruption (Very)	790019.9	4.43e(+09)	0	0.998
Corruption (Extremely)	439356.1	2.46e(+09)	0	0.998
Gender (Female)	0.6664503	0.8020175	-0.34	0.736

We can notice that STATA takes the first category in each explanatory variable as the reference category.

✓ Interpreting the significant coefficients:

1) $e^{\beta_{H12}} = 0.1930819$: The estimated relative risk of being very happy between people living in Japan who have a good state of health compared to those who have a very good state of health is about **81%** lower than the corresponding relative risk of being rather happy, holding other variables constant.

2) $e^{\beta_{H13}} = 0.0951807$: The estimated relative risk of being very happy between people living in Japan who have a fair state of health compared to those who have a very good state of health is about **90%** lower than the corresponding relative risk of being rather happy, holding other variables constant.

3) $e^{\beta_{H14}} = 0.0610599$: The estimated relative risk of being very happy between people living in Japan who have a poor state of health compared to those who have a very good state of health is about **94%** lower than the corresponding relative risk of being rather happy, holding other variables constant.

4) $e^{\beta_{S13}} = 0.2375696$: The estimated relative risk of being very happy between people who consider themselves to belong to the lower middle class compared to those who consider themselves to belong to the upper class is about **76%** lower than the corresponding relative risk of being rather happy, holding other variables constant.

5) $e^{\beta_{S14}} = 0.1533178$: The estimated relative risk of being very happy between people who consider themselves to belong to the working class compared to those who consider themselves to belong to the upper class is about **85%** lower than the corresponding relative risk of being rather happy, holding other variables constant.

6) $e^{\beta_{S15}} = 0.1183592$: The estimated relative risk of being very happy among people who consider themselves to belong to the lower class compared to those who consider themselves to belong to the upper class is about **88%** lower than the corresponding relative risk of being rather happy, holding other variables constant.

7) $e^{\beta_{IG15}} = 2.437217$: The estimated relative risk of being very happy among people living in Japan who consider God very important in their lives compared to those who consider God not important in their lives is **144%** higher than the corresponding relative risk of being rather happy, holding other variables constant.

8) $e^{\beta_{G12}} = 1.431882$: The estimated relative risk of being very happy among females in Japan compared to males is **43%** higher than the corresponding relative risk of being rather happy, holding other variables constant.

9) $e^{\alpha_1} = 23.71267$: The estimated relative risk of being very happy between males living in Japan, who have a very good state of health, feel very secure in their neighborhood and consider themselves belonging to the upper class, consider God not important in their lives, and see that there is not at all corruption in their country is **23.7**.

10) $e^{\beta_{H34}} = 4.337902$: The estimated relative risk of being not very happy between people living in Japan who have a poor state of health compared to those who have a very good state of health is about **334%** higher than the corresponding relative risk of being rather happy, holding other variables constant.

11) $e^{\beta_{H35}} = 8.98566$: The estimated relative risk of being not very happy between people living in Japan who have a very poor state of health compared to those who have a very good state of health is about **799%** higher than the corresponding relative risk of being rather happy, holding other variables constant.

12) $e^{\beta_{G32}} = 0.4715941$: The estimated relative risk of being not very happy between females in Japan compared to males is **53%** lower than the corresponding relative risk of being rather happy, holding other variables constant.

As for the rest coefficients, they're all insignificant as their corresponding p-values are greater than **0.05.**

✓ Assessing the goodness of fit for the multinomial model:

By employing the "mlogitgof" package in STATA, we obtained results assessing the goodness of fit for our full multinomial logistic model. The calculated p-value of **0.874** surpasses the 0.05 threshold, suggesting that the model fits the data well.

```
Goodness-of-fit test for a multinomial logistic regression model
Dependent variable: Happiness

  number of observations =   1072
number of outcome values =      4
      base outcome value =      2
        number of groups =     10
   chi-squared statistic =     16.383
      degrees of freedom =     24
      Prob > chi-squared =      0.874
```

Appendix:

- Backward model:

		odds_ratio	z value	Pr(>|z|)	
(Intercept)	17.3836556	3.545077e+07	0.030	0.976129	
data$Q47_factor2	-0.1431259	8.666449e-01	-0.212	0.831937	
data$Q47_factor3	-1.7997854	1.653344e-01	-2.941	0.003272	**
data$Q47_factor4	-2.4112182	8.970595e-02	-3.795	0.000148	***
data$Q47_factor5	-3.1393988	4.330883e-02	-3.732	0.000190	***
data$Q287_factor2	-12.8234687	2.696735e-06	-0.022	0.982390	
data$Q287_factor3	-13.3084278	1.660440e-06	-0.023	0.981724	
data$Q287_factor4	-14.6267510	4.443066e-07	-0.025	0.979914	
data$Q287_factor5	-15.1549850	2.619833e-07	-0.026	0.979188	
data$Q260_factor2	0.7889157	2.201009e+00	3.319	0.000904	***

- Frequency table for the response variable:

Happiness	Freq.	Percent	Cum.
1	318	29.66	29.66
2	653	60.91	90.58
3	97	9.05	99.63
4	4	0.37	100.00
Total	1,072	100.00	

.

- Multinomial Model:

Happiness	RRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
State_of_health						
2	.1930819	.0438461	-7.24	0.000	.1237218	.3013263
3	.0951807	.0233671	-9.58	0.000	.0588269	.1540002
4	.0610599	.0238048	-7.17	0.000	.0284385	.1311008
5	1.05e-08	.0000267	-0.01	0.994	0	.
Feeling_secure						
2	.4758441	.2032306	-1.74	0.082	.2060254	1.099028
3	.5044132	.221956	-1.56	0.120	.2129285	1.194921
4	.3464375	.2116675	-1.73	0.083	.1046056	1.147347
Social_class						
2	.5005952	.3270656	-1.06	0.290	.1391077	1.801449
3	.2375696	.1518994	-2.25	0.025	.0678488	.8318398
4	.1533178	.099673	-2.88	0.004	.0428766	.5482322
5	.1183592	.0838439	-3.01	0.003	.0295273	.474439
Importance_of_God						
2	.8819213	.2049966	-0.54	0.589	.5592085	1.390868
3	1.304322	.2752522	1.26	0.208	.8624916	1.972489
4	1.495706	.3597557	1.67	0.094	.9334914	2.396525
5	2.437217	.7630044	2.85	0.004	1.319514	4.50168
Perception_of_corruption						
2	.7123306	.39278	-0.62	0.538	.2417286	2.09911
3	.5283812	.2822015	-1.19	0.232	.1854948	1.505092
4	.5491019	.2891458	-1.14	0.255	.1956264	1.541268
5	.8541453	.4606186	-0.29	0.770	.2968251	2.457892
2.Gender	1.431882	.2305247	2.23	0.026	1.044404	1.963115
_cons	23.71267	21.66416	3.47	0.001	3.956518	142.1176

2	(base outcome)					
3						
State_of_health						
2	.5952047	.4167547	-0.74	0.459	.1508919	2.34783
3	2.69583	1.728975	1.55	0.122	.7669607	9.475714
4	4.337902	2.894483	2.20	0.028	1.173037	16.04159
5	8.98566	7.746958	2.55	0.011	1.6584	48.68674
Feeling_secure						
2	2.822993	3.41372	0.86	0.391	.2638699	30.20159
3	3.105526	3.781464	0.93	0.352	.2855383	33.77583
4	4.197468	5.397881	1.12	0.265	.3375631	52.1939
Social_class						
2	3462224	1.04e+10	0.00	0.996	0	.
3	4501462	1.36e+10	0.01	0.996	0	.
4	1.51e+07	4.55e+10	0.01	0.996	0	.
5	2.16e+07	6.53e+10	0.01	0.996	0	.
Importance_of_God						
2	.772458	.2713815	-0.73	0.462	.3879976	1.537874
3	.7400865	.2309	-0.96	0.335	.4015278	1.36411
4	.5138239	.2139348	-1.60	0.110	.2272007	1.162034
5	1.095637	.5061379	0.20	0.843	.4430449	2.709479
Perception_of_corruption						
2	.3270398	.3207422	-1.14	0.254	.0478399	2.235687
3	.5914094	.5239099	-0.59	0.553	.1041927	3.356907
4	.589483	.5155358	-0.60	0.546	.1061813	3.272612
5	.7853552	.6979905	-0.27	0.786	.1375796	4.483096
2.Gender	.4715941	.1187033	-2.99	0.003	.2879489	.7723626
_cons	7.39e-09	.0000223	-0.01	0.995	0	.

4						
State_of_health						
2	.5726752	1228.333	-0.00	1.000	0	.
3	523723.3	9.36e+08	0.01	0.994	0	.
4	4740033	8.47e+09	0.01	0.993	0	.
5	.4309145	4074.308	-0.00	1.000	0	.
Feeling_secure						
2	33397.14	1.01e+08	0.00	0.997	0	.
3	8028.434	2.42e+07	0.00	0.998	0	.
4	123704.3	3.73e+08	0.00	0.997	0	.
Social_class						
2	.6430965	5561.914	-0.00	1.000	0	.
3	.6212497	5295.888	-0.00	1.000	0	.
4	778188.8	6.56e+09	0.00	0.999	0	.
5	6836218	5.76e+10	0.00	0.999	0	.
Importance_of_God						
2	1.76e-07	.0003422	-0.01	0.994	0	.
3	2.681613	3.751469	0.71	0.481	.172822	41.60955
4	1.815951	3.253786	0.33	0.739	.0541924	60.85138
5	2.63e-07	.0005637	-0.01	0.994	0	.
Perception_of_corruption						
2	.4338968	2607.216	-0.00	1.000	0	.
3	.1710914	983.9012	-0.00	1.000	0	.
4	790019.9	4.43e+09	0.00	0.998	0	.
5	439356.1	2.46e+09	0.00	0.998	0	.
2.Gender	.6664503	.8020175	-0.34	0.736	.0630112	7.048845
_cons	6.94e-25	7.44e-21	-0.01	0.996	0	.