Cairo University

Faculty of Economics & Political Science

Categorical Data Analysis

Mini Project Part 1

Feeling of Happiness

Prepared by: Mariam Amir Hassan Elsayed

Supervised by: Dr\ Sally Farid

2023-2024

Introduction

In our study conducted in Pakistan, we aim to delve into the intricate relationship between the feeling of happiness and various categorical factors, employing suitable measures from categorical data analysis. Recognizing that our dataset primarily consists of categorical variables, we aim to explore how these factors intersect with individuals' sense of happiness.

Objective

We are going to study the relationship between the feeling of happiness and some factors using the suitable measures from the categorical data analysis since all the variables we have are categorical.

Research Questions

1. What is the effect of explanatory variables on the feeling of happiness in case of two-way relationships and case of two-way relationships contorlling other variables?
2. Are there associations between feeling of happiness and the explanatory variables?
3. What are the suitable measures to test the associations if it exist?

Measures

To study the relationships between feeling of happiness and the explanatory variables we will introduce tables that display these relationships. We have two types of relationships: the first type is the two way relationships and the other type is the two way relationships with taking into consideration the control variables which may affect relationships dramatically.

In the case of having control variable → We will control it by studying the relationships between feeling of happiness and the explanatory variables at its fixed levels.

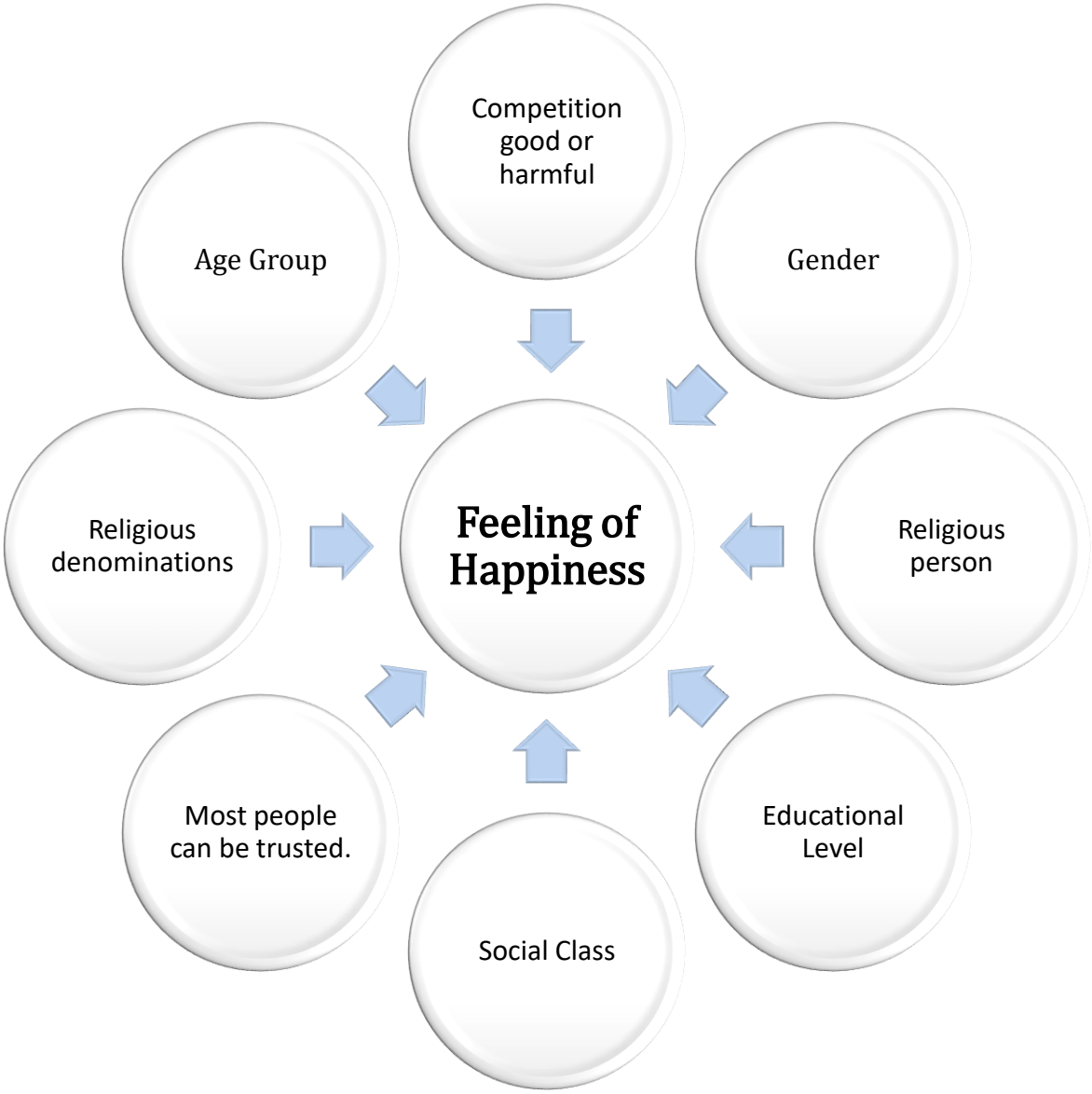
This study will depend on some measures to test the independence of the relationship between feeling of happiness and the explanatory variables and if it doesn’t exist then measure the strength (and the direction if it’s possible) of this relationship.

Response Variable			Explanatory Variable			
Name	No. of Categories	Type of Scale	Name	No. of Categories	Type of Scale	Suitable Measures
			Gender	2	Nominal "Binary"	Gamma
			Age Group	3	Ordinal	Gamma
			Competition good or harmful	11	Ordinal	Gamma
			Religious person	4	Nominal	Chi- Sqaure Monto Cralo Simulation was used
			Educational Level	9	Ordinal	Gamma
			Social Class	7	Ordinal	Gamma
			Most people can be trusted.	3	Nominal	Chi- Sqaure Monto Cralo Simulation was used
Feeling of happiness.	4	Ordinal	Religious denominations	2	Nominal "Binary"	Gamma

General Notes:

- Since Gender and Religion denominations variables are nominal binary, then we will use Gamma measure to test the independence and measure the strength and direction of the association
- When the two variables are ordinal, then we will use Gamma measure to test the independence(specifically we will test if monotone relationship exists or not) and measure the strength and direction of the association
- If we run Chi-Square/ Gamma test and it results not to reject the null hypothesis (i.e. there is independence/ there is no Monotone relationship) then we will stop. But If we reject the null hypothesis (i.e. there is no independence "association exists"/ Monotone relationship exists) then we will use appropriate measures to measure the strength of the association. Goodman Kruskal Gamma was mainly used as most of my variables are ordinal.
- R programming Language was used in conducting the analysis, Missing values were removed

Conceptual Framework



Two-Way Relationships

This case will depend on some measures to test the independence of the relationship between **Feeling of happiness** and

other explanatory variables and if it doesn't exist then measure the strength (and the direction if it's possible) of this relationship.

Feeling Of Happiness and Gender:

Table (1): Two-Way Contingency Table for Happiness & Sex

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Female	25	84	390	440	939
Male	26	73	427	501	1027
Sum	51	157	817	941	1966

Since Gender is nominal binary, then we will use Gamma measure if the variables have monotone relationship and measure the strength and direction of the association if applicable.

Hypotheses to be tested:

$$H_0: \text{No Monotone Relationship}$$

$$H_1: \text{Monotone Relationship Exists}$$

R Output

```
        Goodman-Kruskal's gamma for ordinal categorical data

data:  table_1
Z = 1.149, p-value = 0.2506
95 percent confidence interval:
 -0.03223144  0.12382960
sample estimates:
Goodman-Kruskal's gamma
             0.04579908
```

Interpretation

∵p-value>0.05,

Hence, we don't reject the null hypothesis which means sex and happiness does not have monotone relationship between them

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & Sex

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Female	0.0127162	0.0427263	0.1983723	0.2238047	0.4776195
Male	0.0132248	0.0371312	0.2171923	0.2548321	0.5223805
Sum	0.0259410	0.0798576	0.4155646	0.4786368	1.0000000

Feeling Of Happiness and Age:

Table (1): Two-Way Contingency Table for Happiness & Age

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Less Than 29	18	42	228	333	621
30-49	27	86	481	511	1105
Greater Than 50	6	29	108	97	240
Sum	51	157	817	941	1966

Hypotheses to be tested:

$$H_0: \text{No Monotone Relationship}$$

$$H_1: \text{Monotone Relationship Exists}$$

R Output

```
        Goodman-Kruskal's gamma for ordinal categorical data

data:  table_2
Z = -3.7316, p-value = 0.0001903
95 percent confidence interval:
 -0.20445205 -0.06418212
sample estimates:
Goodman-Kruskal's gamma
             -0.1343171
```

Interpretation

∵p-value<0.05

Hence, we reject the null hypothesis which means age and happiness have monotone relationship between them. Also, we have weak negative relationship between the age of the person and his feeling of happiness. To elaborate, as the person age increases his feeling of happiness decreases.

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & Sex

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Less Than 29	0.0091556	0.0213632	0.1159715	0.1693795	0.3158698
30-49	0.0137335	0.0437436	0.2446592	0.2599186	0.5620549
Greater Than 50	0.0030519	0.0147508	0.0549339	0.0493388	0.1220753
Sum	0.0259410	0.0798576	0.4155646	0.4786368	1.0000000

Feeling Of Happiness and Competition:

Table (1): Two-Way Contingency Table for Happiness & Competition

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
1	18	29	298	373	718
2	3	10	48	61	122
3	1	14	54	60	129
4	1	13	53	48	115
5	4	13	83	85	185
6	2	15	50	49	116
7	4	21	51	34	110
8	3	15	39	28	85
9	4	11	35	27	77
10	11	15	101	163	290
Don't know	0	1	5	13	19
Sum	51	157	817	941	1966

Hypotheses to be tested:

$$H_0: \text{No Monotone Relationship}$$

$$H_1: \text{Monotone Relationship Exists}$$

R Output

```
        Goodman-Kruskal's gamma for ordinal categorical data

data:  table_3
Z = -2.6505, p-value = 0.008037
95 percent confidence interval:
 -0.12512428 -0.01903782
sample estimates:
Goodman-Kruskal's gamma
            -0.07208105
```

Interpretation

∵p-value<0.05

Hence, we reject the null hypothesis which means competition and happiness have monotone relationship between them. Also, we have weak negative relationship between the person perception of competition and his feeling of happiness. To elaborate, as the person who hates the competition more tends to have larger feeling of happiness.

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & Competition

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
1	0.0091556	0.0147508	0.1515768	0.1897253	0.3652085
2	0.0015259	0.0050865	0.0244151	0.0310275	0.0620549
3	0.0005086	0.0071211	0.0274669	0.0305188	0.0656155
4	0.0005086	0.0066124	0.0269583	0.0244151	0.0584944
5	0.0020346	0.0066124	0.0422177	0.0432350	0.0940997
6	0.0010173	0.0076297	0.0254323	0.0249237	0.0590031
7	0.0020346	0.0106816	0.0259410	0.0172940	0.0559512
8	0.0015259	0.0076297	0.0198372	0.0142421	0.0432350
9	0.0020346	0.0055951	0.0178026	0.0137335	0.0391658
10	0.0055951	0.0076297	0.0513733	0.0829095	0.1475076
Don't know	0.0000000	0.0005086	0.0025432	0.0066124	0.0096643
Sum	0.0259410	0.0798576	0.4155646	0.4786368	1.0000000

Feeling Of Happiness and Education:

Table (1): Two-Way Contingency Table for Happiness & Education

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Early childhood education (ISCED 0) / no education	22	72	249	250	593
Primary education (ISCED 1)	3	22	105	134	264
Lower secondary education (ISCED 2)	7	13	90	106	216
Upper secondary education (ISCED 3)	8	30	188	228	454
Post-secondary non-tertiary education (ISCED 4)	2	10	96	126	234
Short-cycle tertiary education (ISCED 5)	1	1	9	9	20
Bachelor or equivalent (ISCED 6)	6	9	61	58	134
Master or equivalent (ISCED 7)	2	0	18	29	49
Doctoral or equivalent (ISCED 8)	0	0	1	1	2
Sum	51	157	817	941	1966

Hypotheses to be tested:

$$H_0: \text{No Monotone Relationship}$$

$$H_1: \text{Monotone Relationship Exists}$$

R Output

```
        Goodman-Kruskal's gamma for ordinal categorical data

data:  table_4
Z = 3.6845, p-value = 0.0002291
95 percent confidence interval:
 0.04864689 0.15876232
sample estimates:
Goodman-Kruskal's gamma
             0.1037046
```

Interpretation

∵p-value<0.05

Hence, we reject the null hypothesis which means educational level and happiness have monotone relationship between them. Also, we have weak Positive relationship between the person educational and his feeling of happiness. To elaborate, the person who has higher educational level tends to have larger feeling of happiness.

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & Education

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Early childhood education (ISCED 0) / no education	0.0111902	0.0366226	0.1266531	0.1271617	0.3016277
Primary education (ISCED 1)	0.0015259	0.0111902	0.0534079	0.0681587	0.1342828
Lower secondary education (ISCED 2)	0.0035605	0.0066124	0.0457782	0.0539166	0.1098678
Upper secondary education (ISCED 3)	0.0040692	0.0152594	0.0956256	0.1159715	0.2309257
Post-secondary non-tertiary education (ISCED 4)	0.0010173	0.0050865	0.0488301	0.0640895	0.1190234
Short-cycle tertiary education (ISCED 5)	0.0005086	0.0005086	0.0045778	0.0045778	0.0101729
Bachelor or equivalent (ISCED 6)	0.0030519	0.0045778	0.0310275	0.0295015	0.0681587
Master or equivalent (ISCED 7)	0.0010173	0.0000000	0.0091556	0.0147508	0.0249237
Doctoral or equivalent (ISCED 8)	0.0000000	0.0000000	0.0005086	0.0005086	0.0010173
Sum	0.0259410	0.0798576	0.4155646	0.4786368	1.0000000

Feeling Of Happiness and Social Class:

Table (1): Two-Way Contingency Table for Happiness & Social Class

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Don't know	0	0	19	21	40
Lower class	15	33	108	91	247
Lower middle class	9	39	255	259	562
Working class	24	61	330	384	799
Upper middle class	3	20	90	155	268
Upper class	0	4	15	31	50
Sum	51	157	817	941	1966

Hypotheses to be tested:

$$H_0: \text{No Monotone Relationship}$$

$$H_1: \text{Monotone Relationship Exists}$$

R Output

```
        Goodman-Kruskal's gamma for ordinal categorical data

data:  table_5
Z = 4.4274, p-value = 9.536e-06
95 percent confidence interval:
 0.0759381 0.1951755
sample estimates:
Goodman-Kruskal's gamma
              0.1355568
```

Interpretation

$\because$p-value$<$0.05

Hence, we reject the null hypothesis which means Social Class and happiness have monotone relationship between them. Also, we have weak Positive relationship between the person social class and his feeling of happiness. To elaborate, the person who has higher social class tends to have larger feeling of happiness.

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & Social Class

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Don't know	0.0000000	0.0000000	0.0096643	0.0106816	0.0203459
Lower class	0.0076297	0.0167854	0.0549339	0.0462869	0.1256358
Lower middle class	0.0045778	0.0198372	0.1297050	0.1317396	0.2858596
Working class	0.0122075	0.0310275	0.1678535	0.1953204	0.4064090
Upper middle class	0.0015259	0.0101729	0.0457782	0.0788403	0.1363174
Upper class	0.0000000	0.0020346	0.0076297	0.0157681	0.0254323
Sum	0.0259410	0.0798576	0.4155646	0.4786368	1.0000000

Feeling Of Happiness and Religious Denominations:

Hindu religion and others where merged as they represent around 1.5%

Table (1): Two-Way Contingency Table for Happiness & Religious denominations

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Muslim	51	156	803	931	1941
Other	0	1	14	10	25
Sum	51	157	817	941	1966

Hypotheses to be tested:

H_0: No Monotone Relationship

H_1: Monotone Relationship Exists

R Output

```
        Goodman-Kruskal's gamma for ordinal categorical data

data:  table_6
Z = -0.37726, p-value = 0.706
95 percent confidence interval:
 -0.3796689  0.2560754
sample estimates:
Goodman-Kruskal's gamma
            -0.06179674
```

Interpretation

∵p-value>0.05,

Hence, we don't reject the null hypothesis which means Religious denominations and happiness does not have monotone relationship between them.

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & Religious denominations

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Muslim	0.025941	0.0793489	0.4084435	0.4735504	0.9872838
Other	0.000000	0.0005086	0.0071211	0.0050865	0.0127162
Sum	0.025941	0.0798576	0.4155646	0.4786368	1.0000000

Feeling Of Happiness and People Trust:

Table (1): Two-Way Contingency Table for Happiness & People Trust

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Don't know	0	1	6	8	15
Most people can be trusted	10	39	193	217	459
Need to be very careful	41	117	618	716	1492
Sum	51	157	817	941	1966

Chi-Square Test of independence should be conducted; however, the conditions of cell counts are not satisfied. Thus, Monto Carlo simulation was used.

Hypotheses to be tested:

H_0: Happiness & People Trust are Independent

H_1: Happiness & People Trust are Not Independent

R Output

```
        Pearson's Chi-squared test with simulated p-value (based on
        10000 replicates)

data:  table_7
X-squared = 1.2259, df = NA, p-value = 0.9845
```

Interpretation

∵p-value>0.05,

Hence, we don't reject the null hypothesis which means happiness and people are trust are independent. (Specifically, We do not have enough evidence to to say that they are dependent)

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & People Trust

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
Don't know	0.0000000	0.0005086	0.0030519	0.0040692	0.0076297
Most people can be trusted	0.0050865	0.0198372	0.0981689	0.1103764	0.2334690
Need to be very careful	0.0208545	0.0595117	0.3143438	0.3641913	0.7589013
Sum	0.0259410	0.0798576	0.4155646	0.4786368	1.0000000

Feeling Of Happiness and Level of religiousty of the person:

Table (1): Two-Way Contingency Table for Happiness & Religious person

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
An atheist	0	2	5	2	9
Not a religious person	2	17	37	34	90
Don't know	0	1	8	6	15
A religious person	49	137	767	899	1852
Sum	51	157	817	941	1966

Chi-Square Test of independence should be conducted; however, the conditions of cell counts are not satisfied. Thus, Monto Carlo simulation was used.

Hypotheses to be tested:

H_0: Happiness & Religous Person are Independent

H_1: Happiness & People Trust are Not Independent

R Output

```
        Pearson's Chi-squared test with simulated p-value (based on
        10000 replicates)

data:  table_8
X-squared = 21.666, df = NA, p-value = 0.025
```

Interpretation

∵p-value<0.05,

Hence, we reject the null hypothesis which means happiness and people are trust are are not independent. Furthermore, the two conditional distributions are not homogenous.

To measure the strength of the association between the two variables, we will calculate Cramer's V

```
Cramer's V          : 0.061
```

This Indicates that there is little(If any associations) between the two variables

Conditional Probability Table

Table (2): Two-Way Contingency (proportions) Table for Happiness & Religious person

	Not at all happy	Not very happy	Quite happy	Very happy	Sum
An atheist	0.0000000	0.0010173	0.0025432	0.0010173	0.0045778
Not a religious person	0.0010173	0.0086470	0.0188199	0.0172940	0.0457782
Don't know	0.0000000	0.0005086	0.0040692	0.0030519	0.0076297
A religious person	0.0249237	0.0696846	0.3901322	0.4572737	0.9420142
Sum	0.0259410	0.0798576	0.4155646	0.4786368	1.0000000

Three-Way Relationships

This case will depend on some measures to test the independence of the relationship between **Feeling of Happiness** and other explanatory variables controlling the effect of other variables.

		happy Not at all happy	Not very happy	Quite happy	Very happy
age	social_class				
Less Than 29	Don't know	0	0	7	6
	Lower class	3	6	27	33
	Lower middle class	6	11	84	84
	Working class	7	16	77	133
	Upper middle class	2	8	28	57
	Upper class	0	1	5	20
30-49	Don't know	0	0	11	15
	Lower class	10	19	66	48
	Lower middle class	2	20	135	146
	Working class	14	34	214	214
	Upper middle class	1	10	46	78
	Upper class	0	3	9	10
Greater Than 50	Don't know	0	0	1	0
	Lower class	2	8	15	10
	Lower middle class	1	8	36	29
	Working class	3	11	39	37
	Upper middle class	0	2	16	20
	Upper class	0	0	1	1

We conducted an analysis examining the correlation between the feeling of happiness (Y) and Social Class (X) in Pakistan, with age as the controlled variable. This investigation involved the construction of three distinct partial tables, each representing a two-way contingency table. The first table pertains to individuals under the age of 29, the second encompasses those aged between 30 and 49, and the third table focuses on individuals aged 50 and above. These tables allowed for a nuanced exploration of the relationship between happiness, social class, and age groups, enabling a more comprehensive understanding of their interplay within the context of Pakistani society.

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

H_0: there is conditional independence

H_1: no conditional independence

```
        Cochran-Mantel-Haenszel test

data:  three_way_table
Cochran-Mantel-Haenszel M^2 = 23.19, df = 10, p-value = 0.01007
```

The obtained p-value, which is below the threshold of 0.05, leads us to reject the assumption of conditional independence. This rejection indicates that the relationship between social class and the feeling of happiness varies significantly across different socioeconomic classes.

However, due to the absence of a (2 x 2 x K) table necessary for calculating the odds ratio, determining the strength of this association becomes challenging, hindering our ability to make a conclusive decision about the association's magnitude.

Furthermore, the rejection of conditional independence in the 3x3 contingency table negates the necessity to examine mutual and joint independence separately. Conditional independence, being the most lenient form of independence, encompasses both mutual and joint independence. Hence, the rejection of conditional independence inherently implies the rejection of both mutual and joint independence