Children's Nutrition Survey
in Subdivided Flat in To Kwa Wan

Quantitative Methods
For Community Service

Supported By the People Service Center

AMA3S01
GROUP 1A
DATA ANALYSIS

LAM YI NOK, EURUS(19059671D)
LI TSZ WING, CYRUS(19053779D)
LI SIU KWONG, TIM(19053694D)

ABOUT OUR SERVICE LEARNING & PSC

Collaborate with People Service Center
Designed and conducted surveys through phone calling and home visiting in To Kwa Wan
Inputed and analyzed data

People Service Center

- establish in 2002
- focus on
  - responding to the growing disparity between rich and poor in society,
  - providing services for poor families and the elderly in society,
  - promoting social reform, and
  - building an equal and just society

PSC To Kwa Wan Center provide services to

- Families live in subdivided flats
- Elderly

Sep — Design Questionnaire

Oct — Phone calling and home visiting for completing the questionnaires

Nov — Data Analysis and Presentation

RESEARCH PURPOSE

- Gain **the most updated** information on the needy
- Provide **suitable support** to them
- Figure out
  - Are **children and families** living in subdivided flats **with enough nutrition?**
  - Do any significant factors influence them **to eat healthy?**

INTERVIEWEES

- Registered members of the People Service Center
  - Live in subdivided flats in **To Kwa Wan**
  - Families with **3-12 years old** children

Adding a new attributes = Nutrition Score

If the child can accomplish the criteria, then he or she can get 1 mark. Here are the daily criteria per day: (Full mark = 6)

- 3 to 4 bowls of cereal
- 2 servings of vegetables
- 2 servings of fruit
- 3 to 5 taels of meat, fish, eggs or substitutes
- 2 servings of milk or alternatives
- 6 to 8 drinks

=

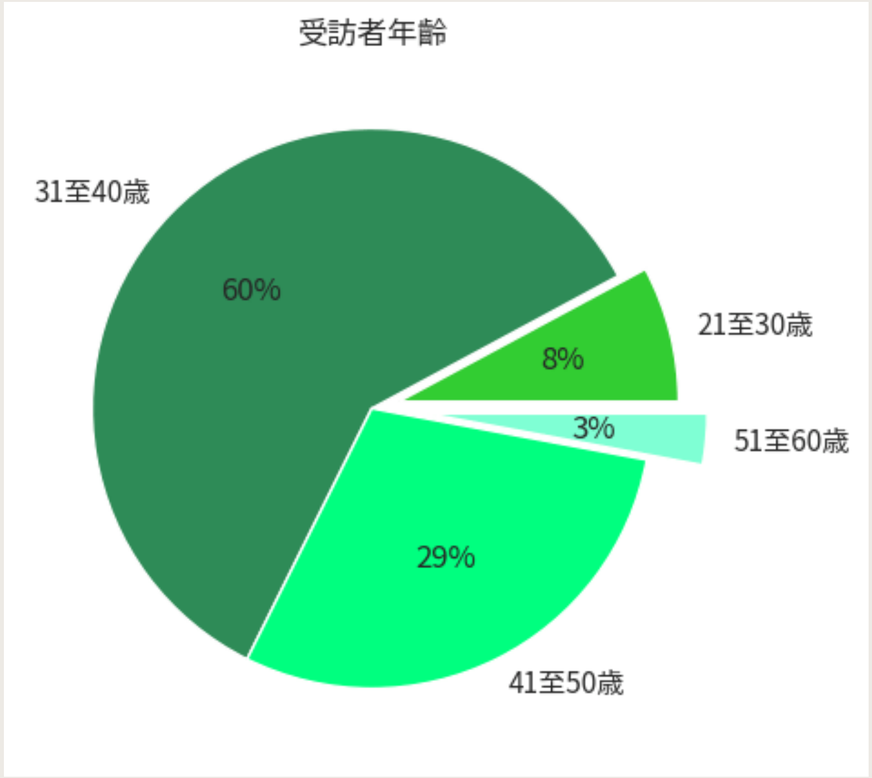
Nutrition Score

	小朋友昨天全日有否攝取以下數量[3至4碗穀物類]	小朋友昨天全日有否攝取以下數量[最少2份蔬菜]	小朋友昨天全日有否攝取以下數量[最少2份水果]	小朋友昨天全日有否攝取以下數量[3至5兩肉、魚、蛋及代餐品]	小朋友昨天全日有否攝取以下數量[2份奶類及代餐品]	小朋友昨天全日有否攝取以下數量[6至8杯流質飲品]	小朋友營養分數
0	0	1	1	1	0	1	4
1	0	0	0	0	0	0	0
2	0	1	0	1	0	1	3
3	1	0	1	1	1	0	4
4	0	0	0	1	0	0	1
...	...	...	...	...	...	...	...
97	0	0	0	0	0	0	0
98	1	1	1	1	0	0	4
99	0	0	0	1	1	1	3
100	0	0	0	1	0	0	1
101	1	1	1	1	0	0	4

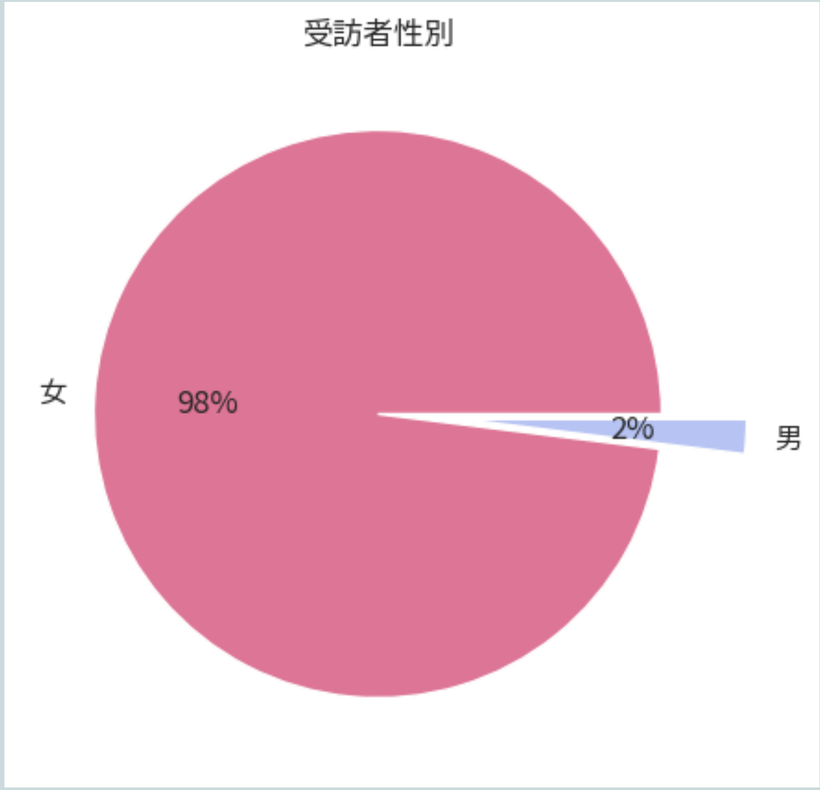
102 rows × 7 columns

DATA DESCRIPTION

RESPONDENTS' DEMOGRAPHIC DATA

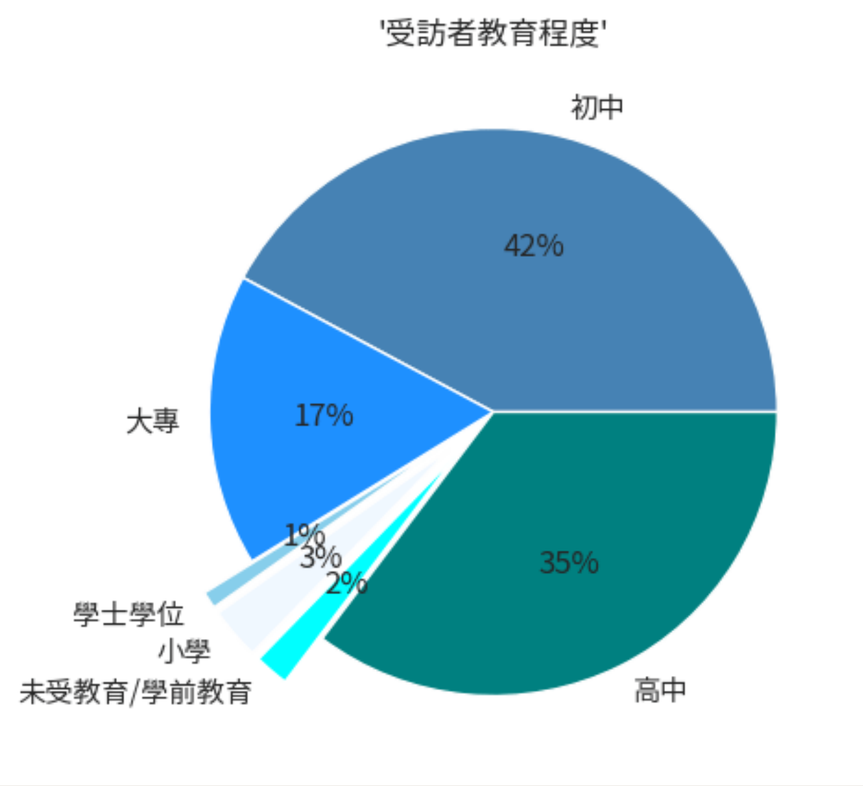


Most of the respondents(60%) were aged 31 to 40.
29% of respondents aged from 41 to 50.

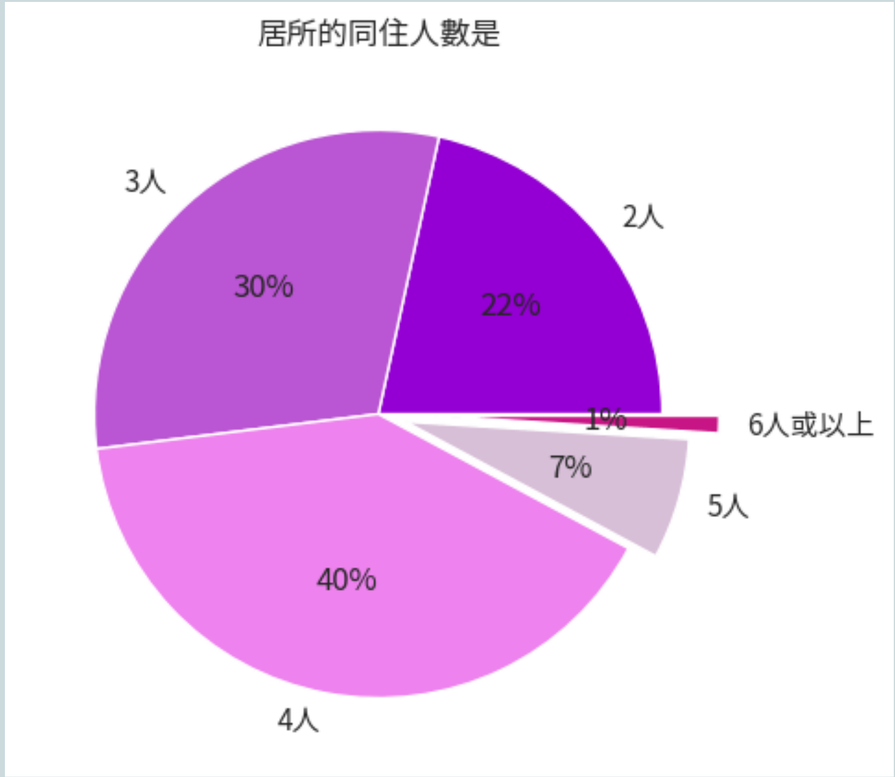


Most of the respondents(98%) are female.
2% of respondents are male.

RESPONDENTS' BACKGROUND

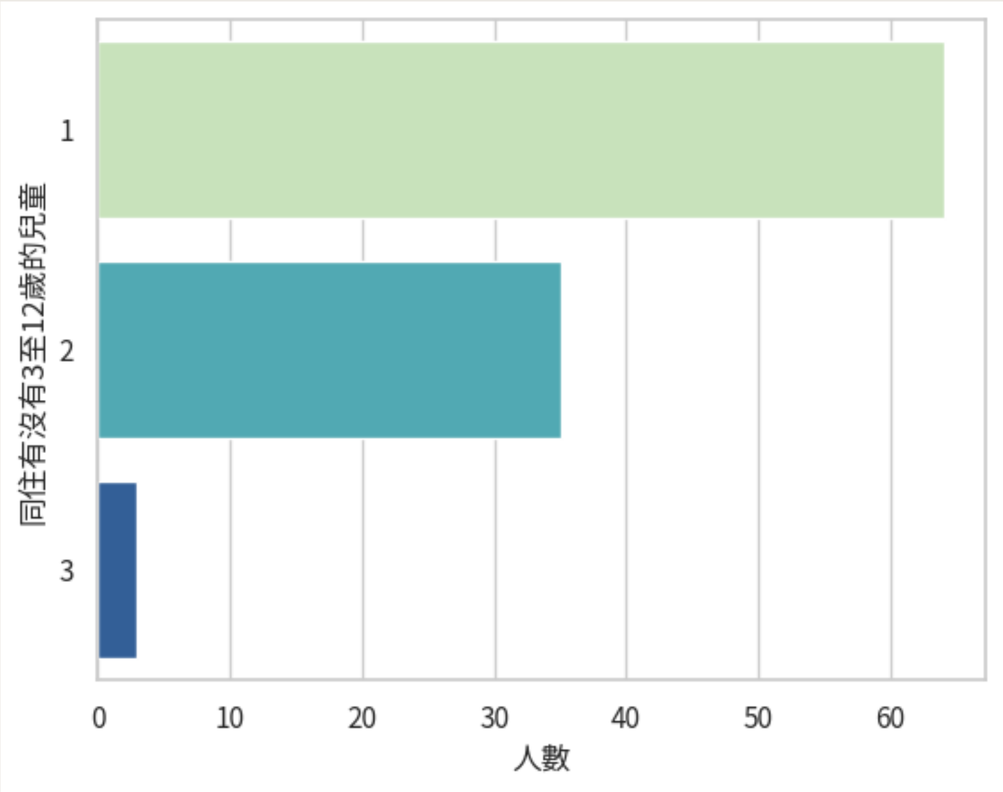


1% of respondents have a bachelor's degree.
2% of respondents have never attended college, but 42% of respondents graduated from junior high school

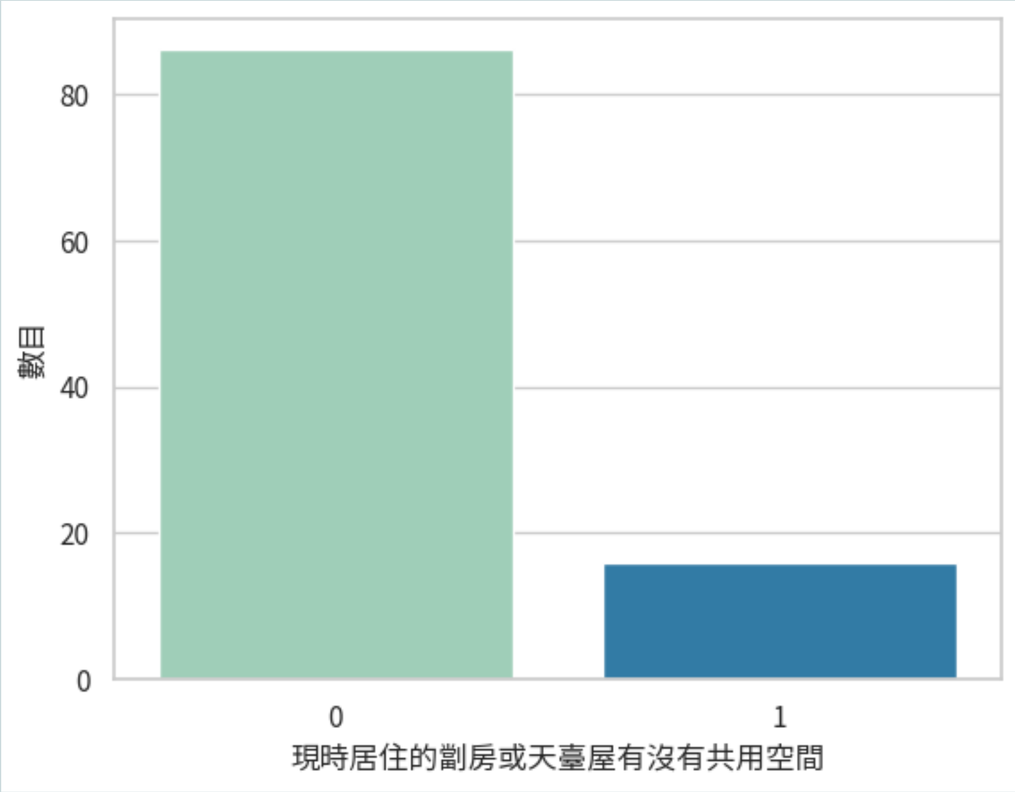


Most of the respondents(40%) are families of four.
30% of respondents are families of three.

RESPONDENTS' BACKGROUND

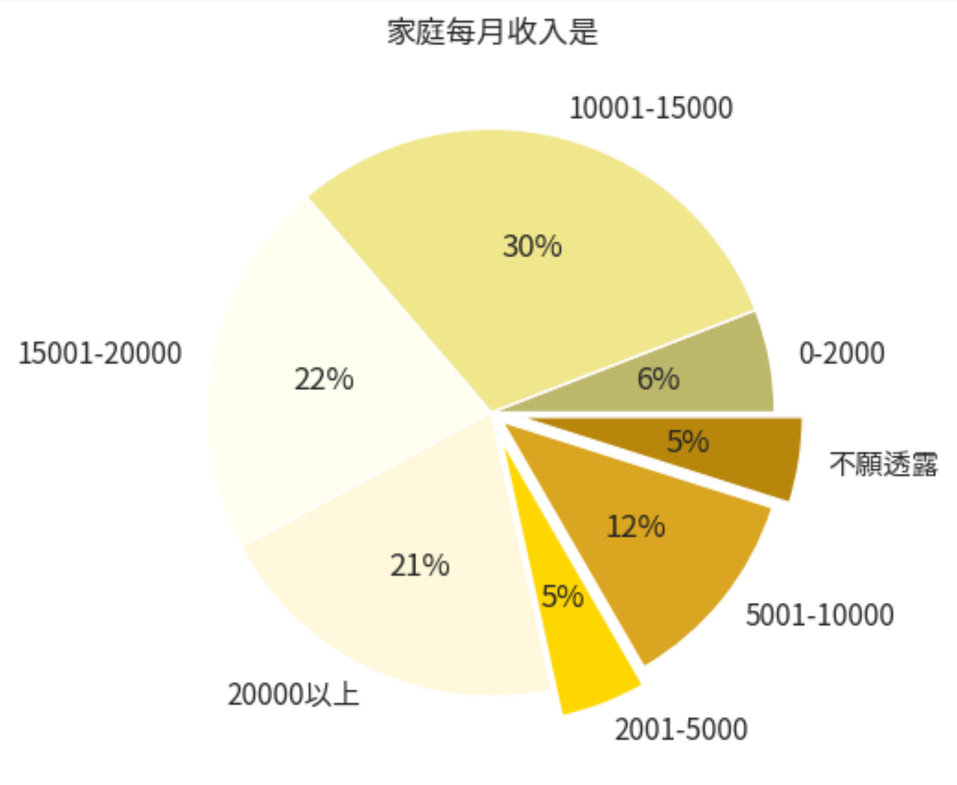


The majority of families only have one child between the ages of 3 and 12; only two families have three children between the ages of 3 and 12

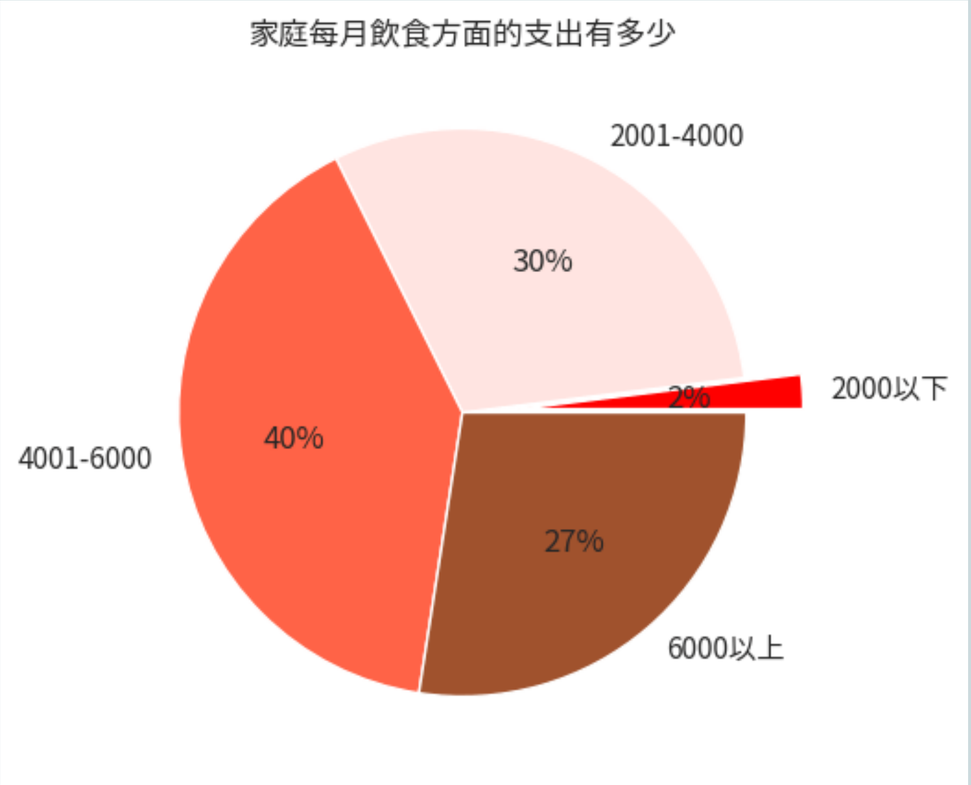


Almost all of them do not share rooms with others

RESPONDENTS' ECONOMIC DATA

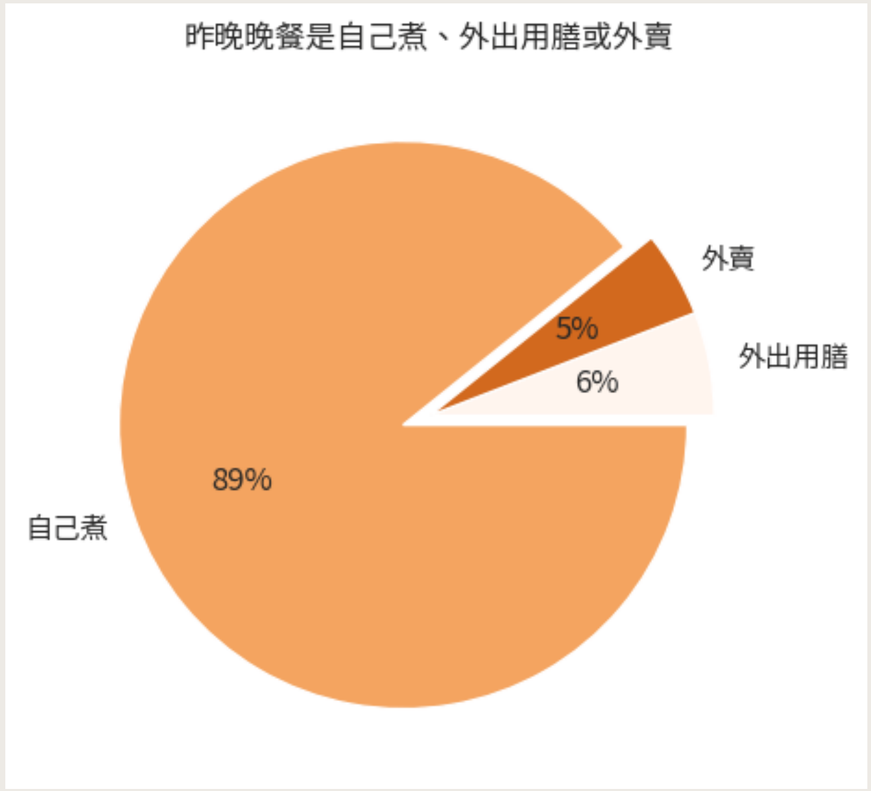


Most of the respondents(30%) had monthly family incomes ranging from HK$10001 to HK$15000. 22% of respondents' monthly family incomes ranged from HK$15001 to HK$20000.



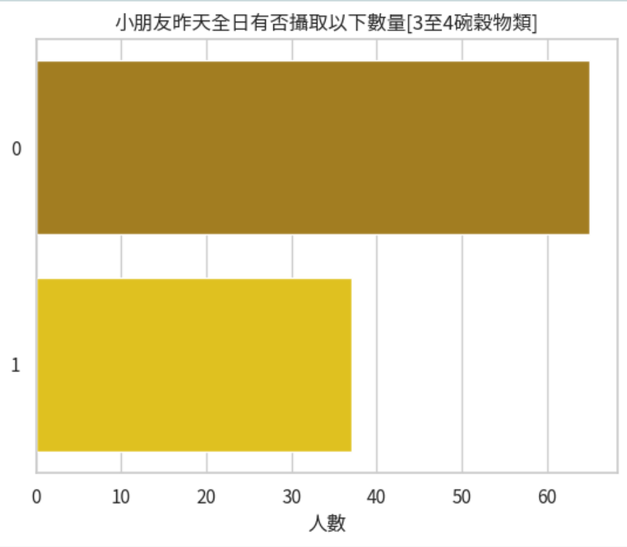
Most of the respondents(40%) had monthly family food expenses ranging from HK$4001 to HK$6000. 30% of respondents' monthly family food expenses ranged from HK$2001 to HK$4000.

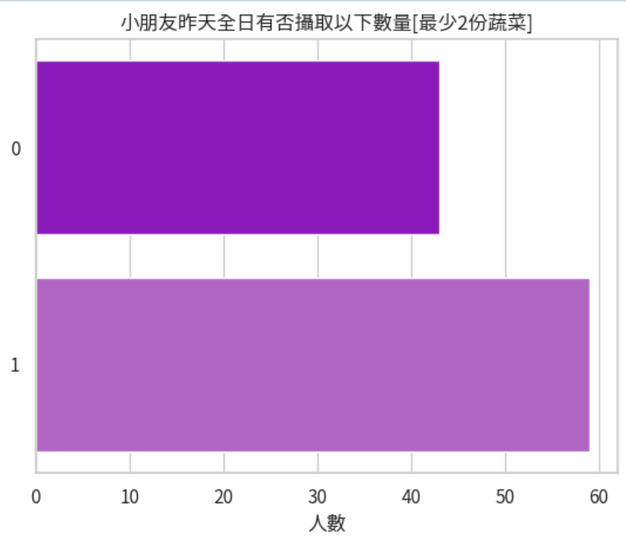
RESPONDENTS' EATING HABITS

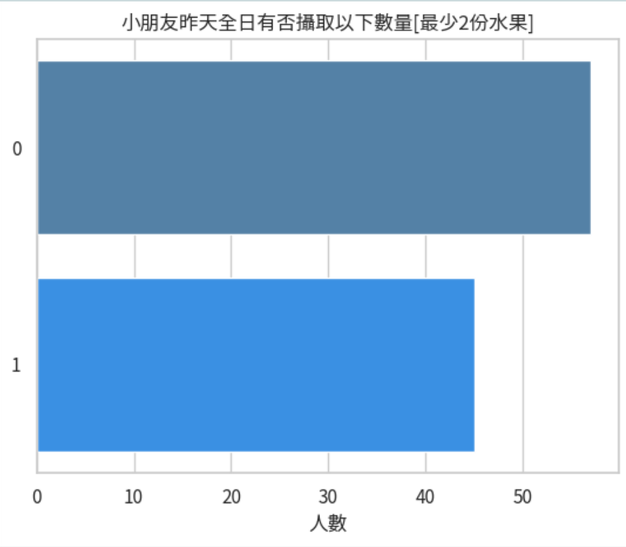


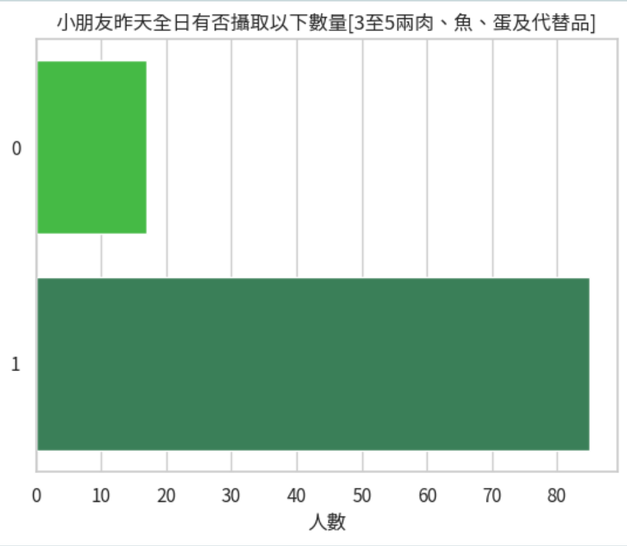
The majority of families (89%) prepare their final meal at home; only 5% and 6% of the families, respectively, takeaway or ate at a restaurant

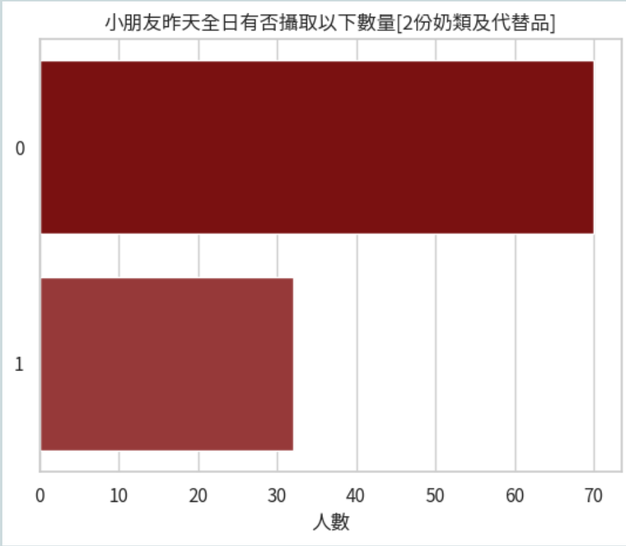
RESPONDENTS' EATING HABITS

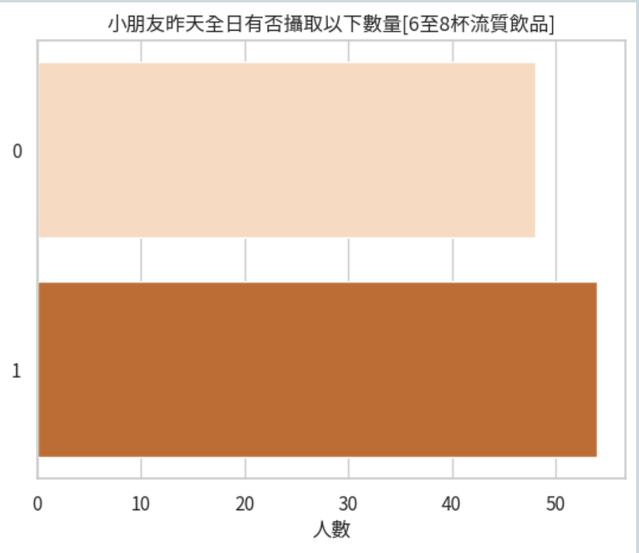




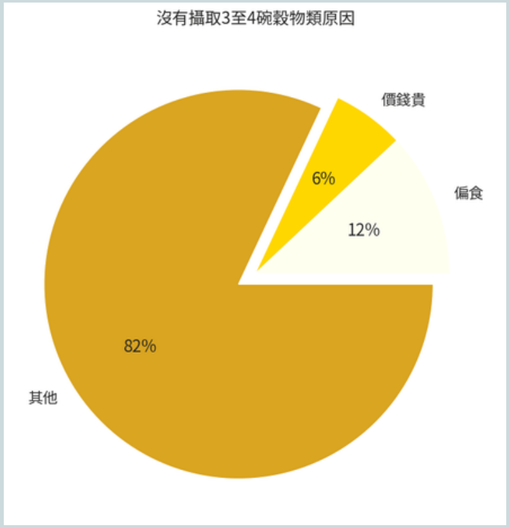


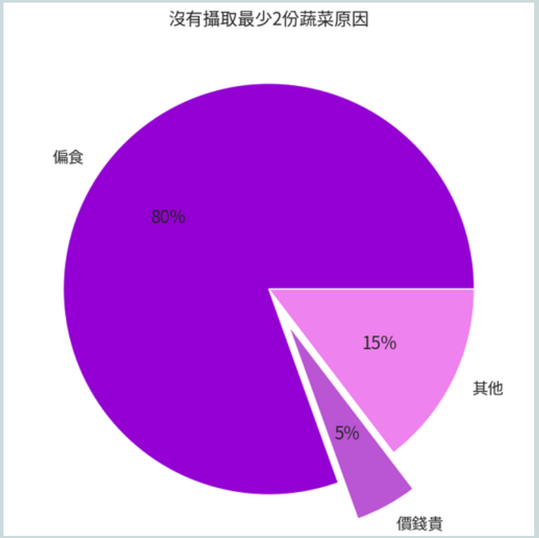


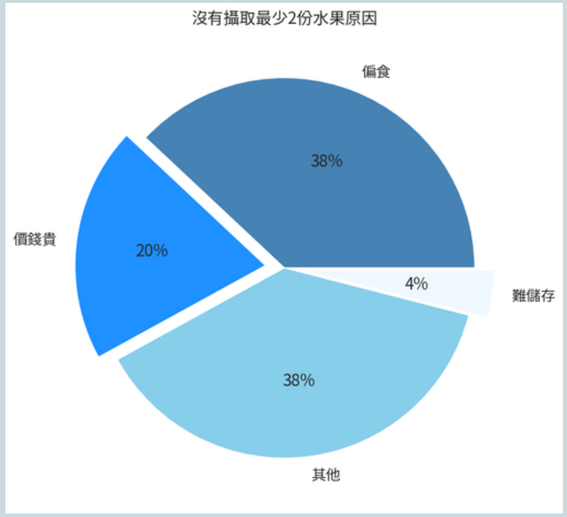


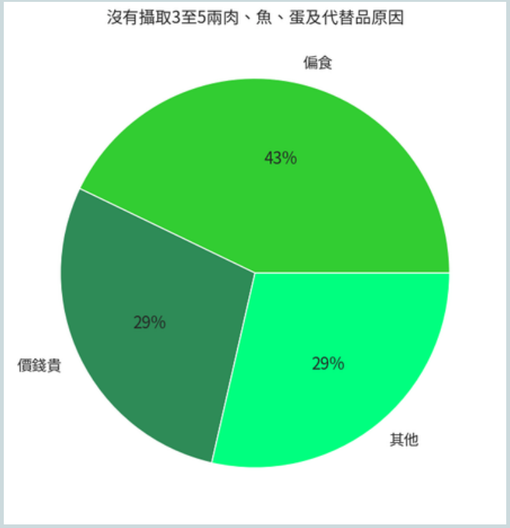


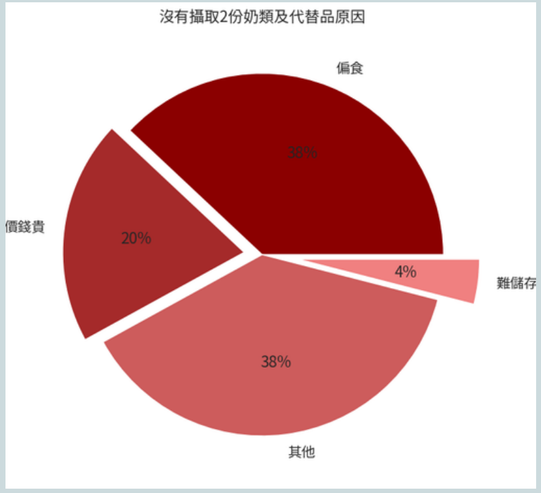
RESPONDENTS' EATING HABITS

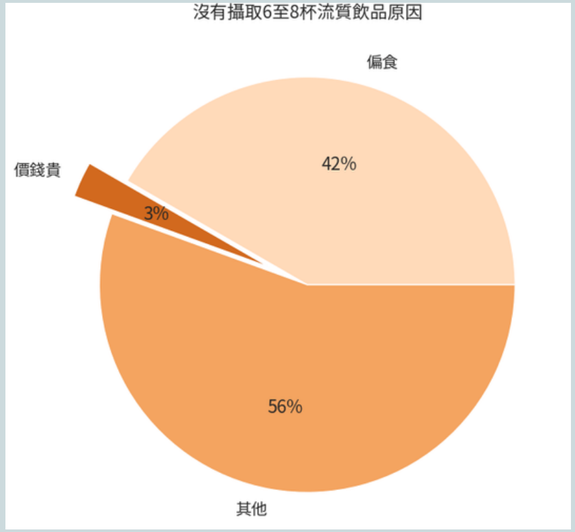




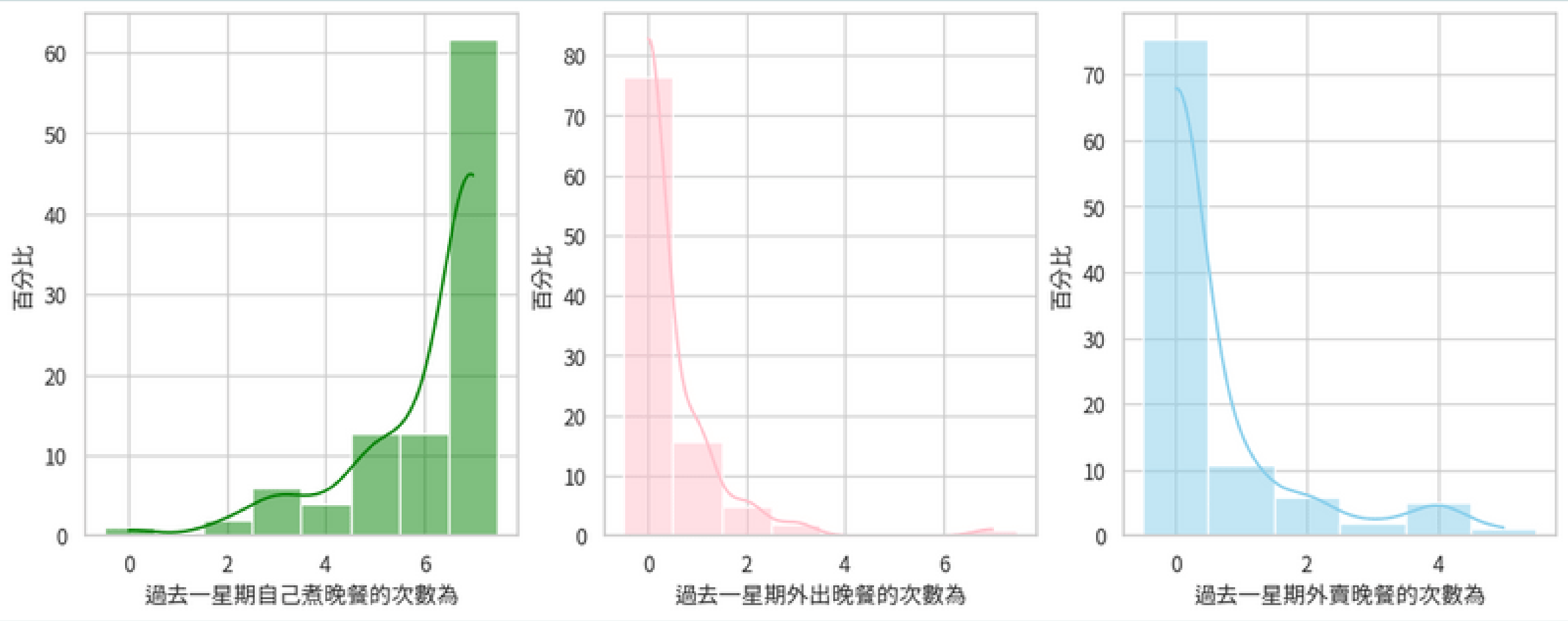






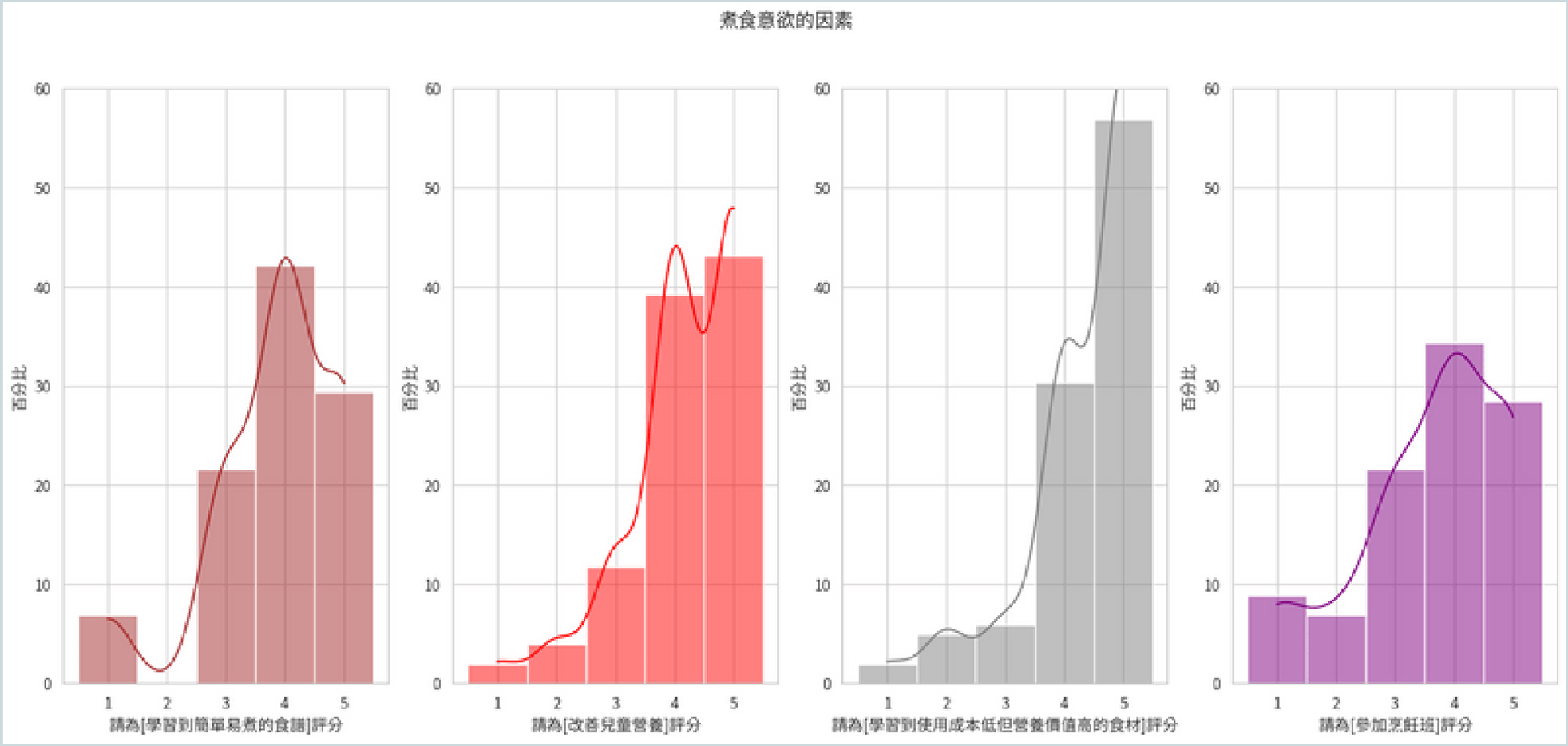


RESPONDENTS' COOKING HABITS



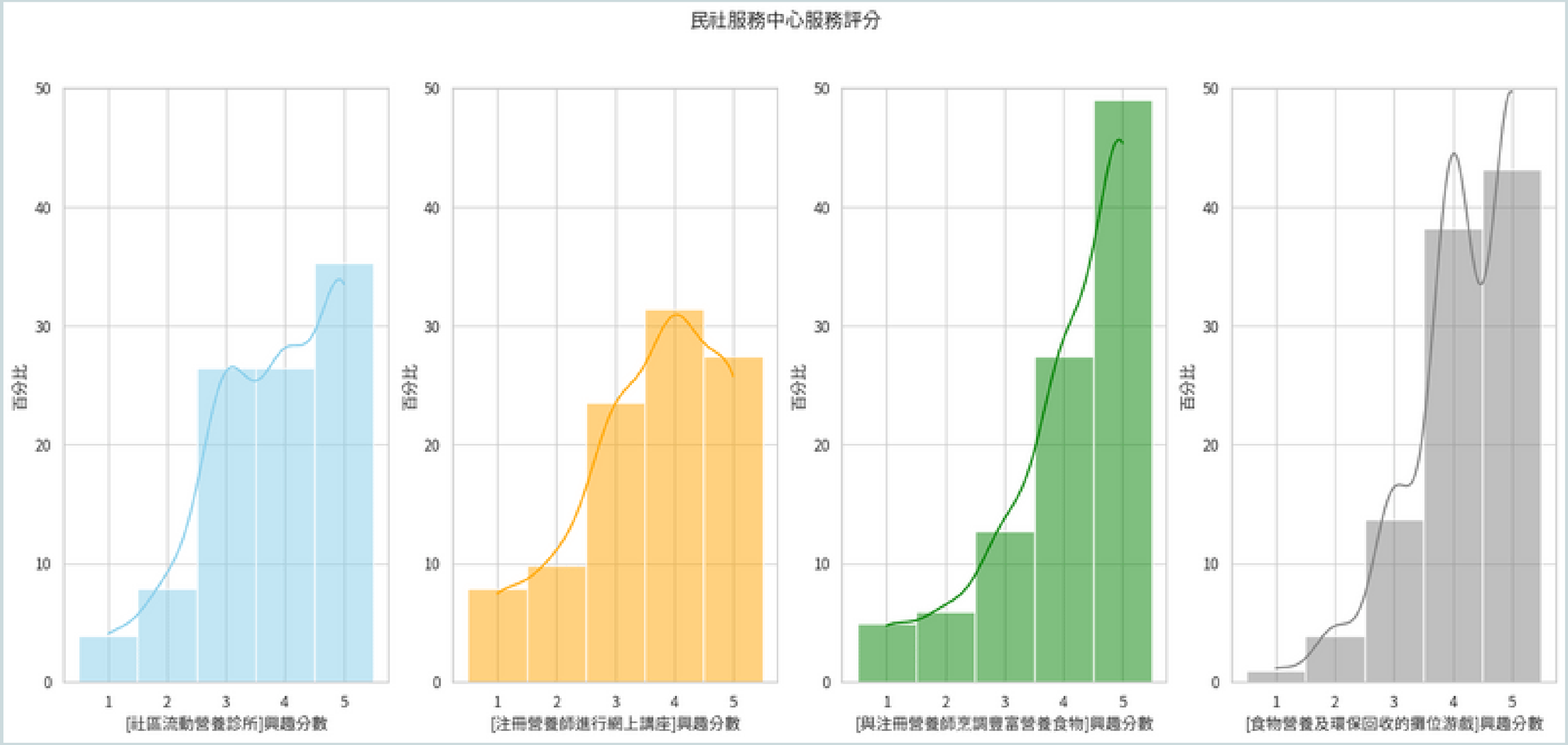
There are more than 60% of respondents cook their dinner at home every night.

WAYS TO INCREASE THE WILLINGNESS TO COOKING AT HOME DATA



More family carers agree that the desire to cook at home can be stoked by learning about inexpensive but nutritious ingredients.

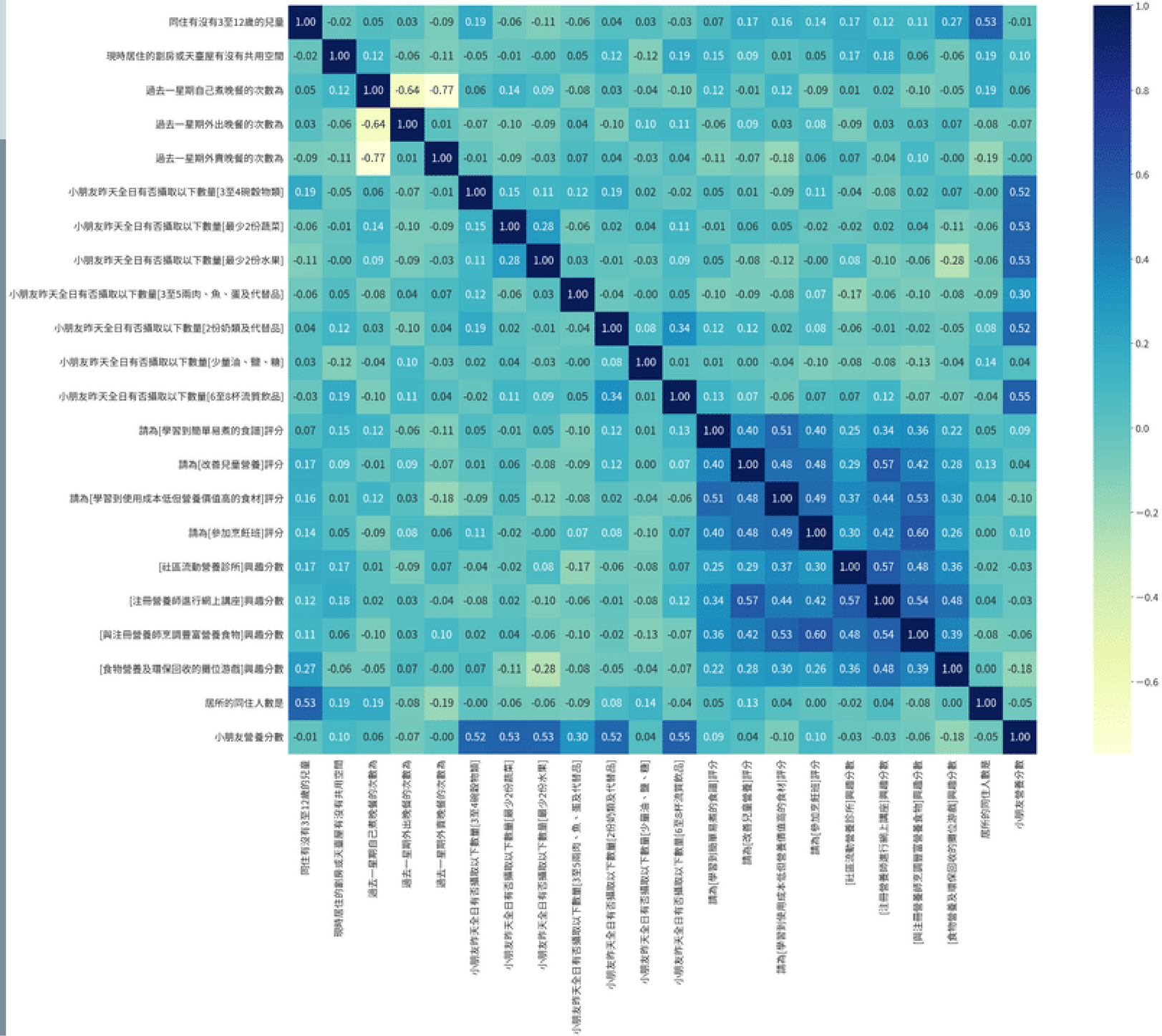
WILLINGNESS ON ATTENDING ACTIVITIES DATA



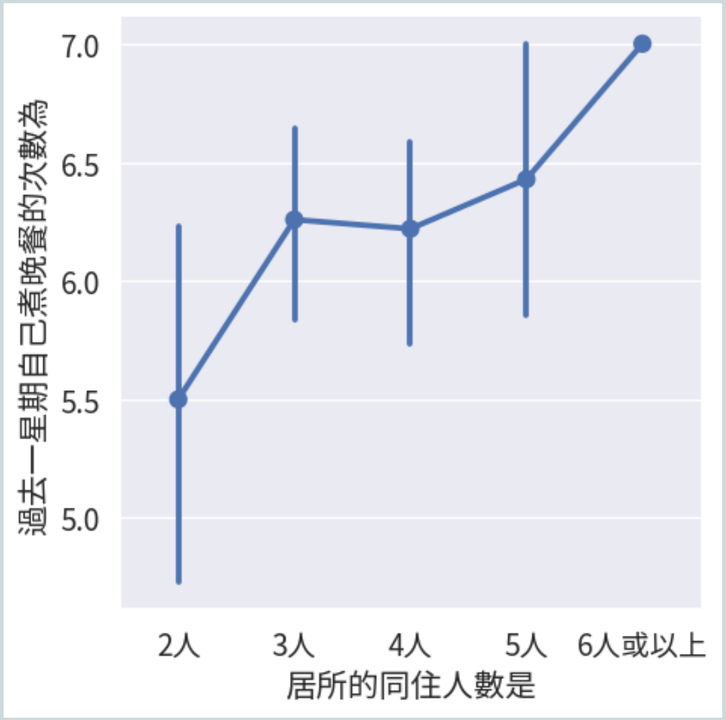
More children and families are willing to visit the game booth.

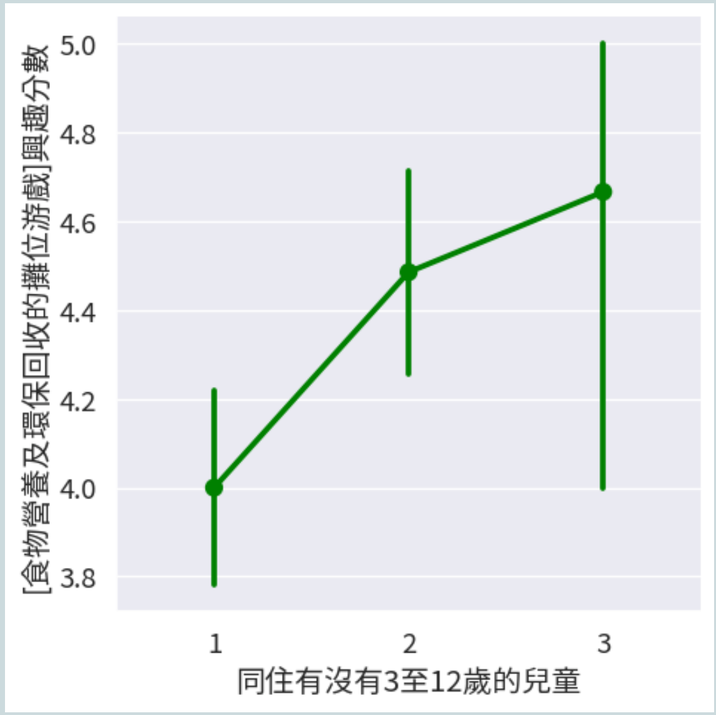
HEAT MAP

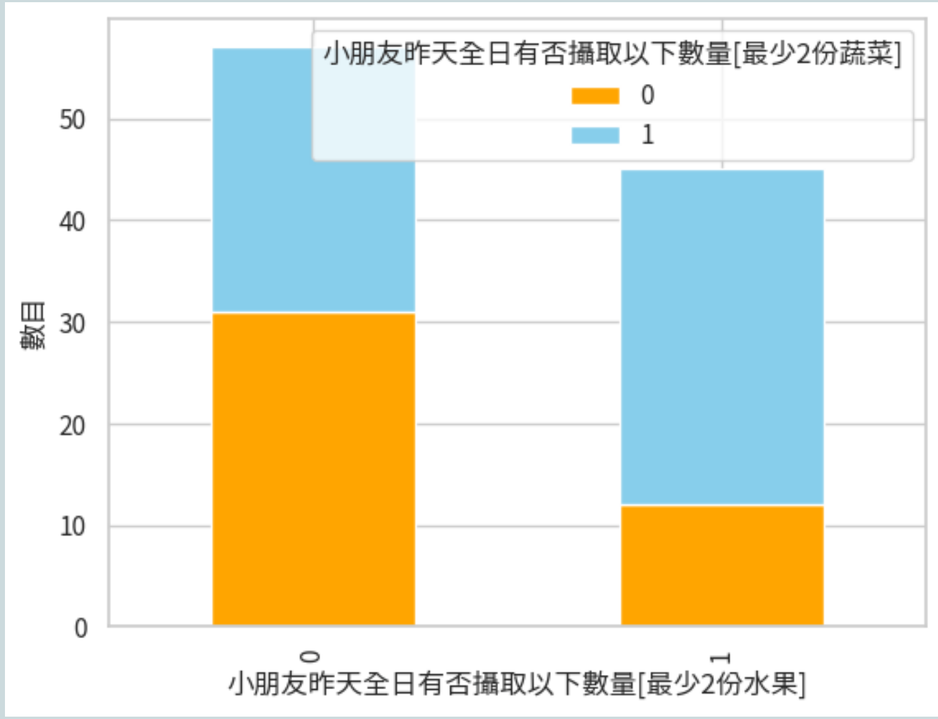
The plots and further analyses of the data from two questions with magnitudes greater than 0.2 are provided below.



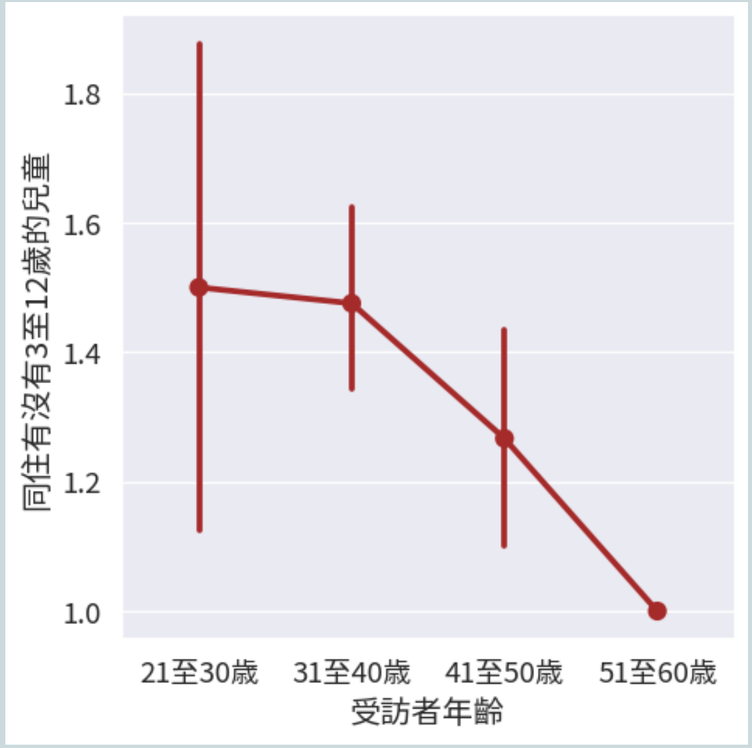
FURTHER INVESTIGATIONS ON THE RELATIONSHIP OF TWO ASSIGNED DATA

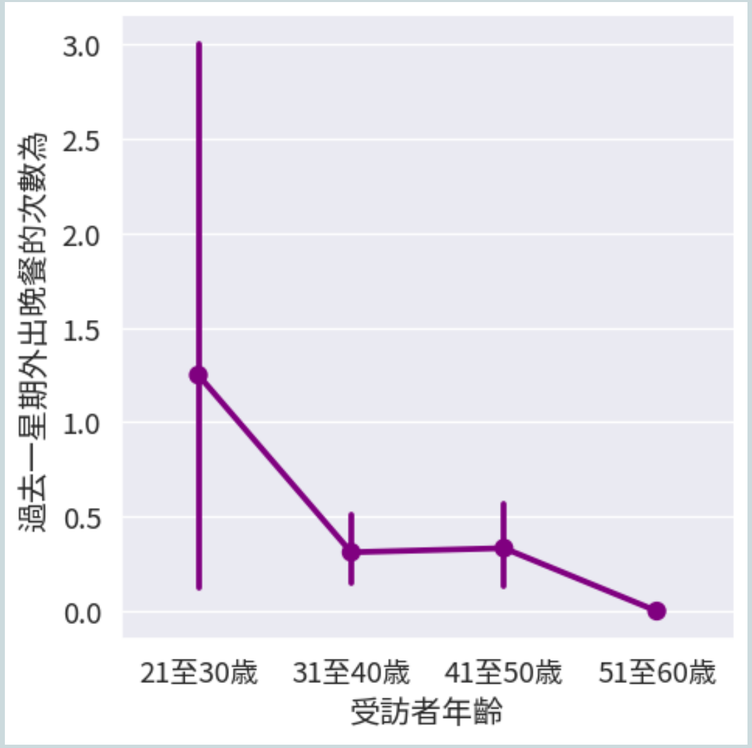




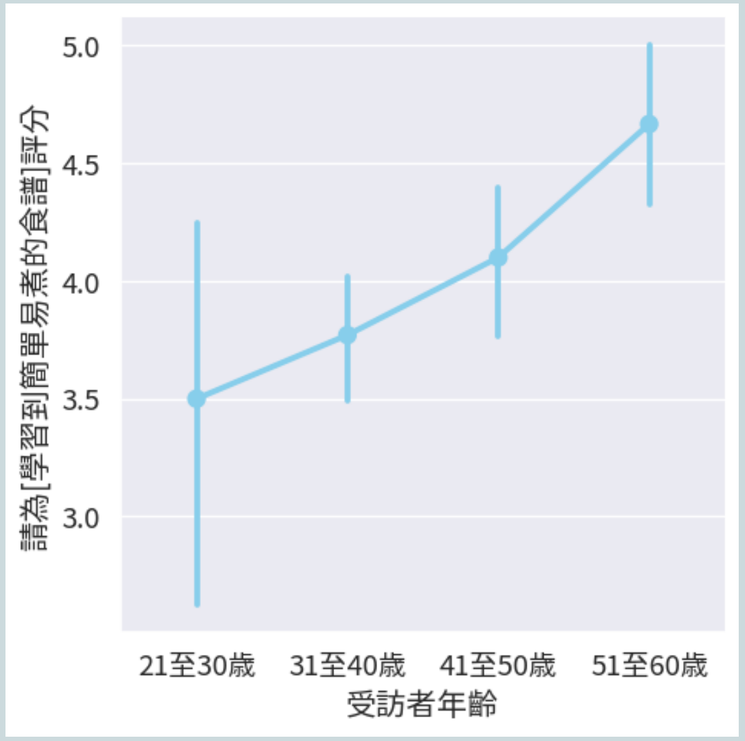


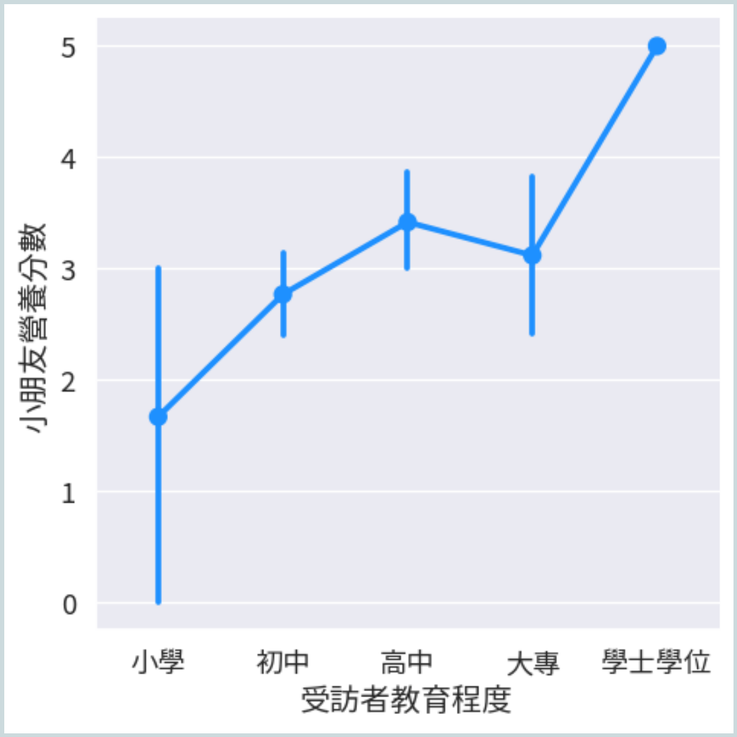
FURTHER INVESTIGATIONS ON THE RELATIONSHIP OF TWO ASSIGNED DATA

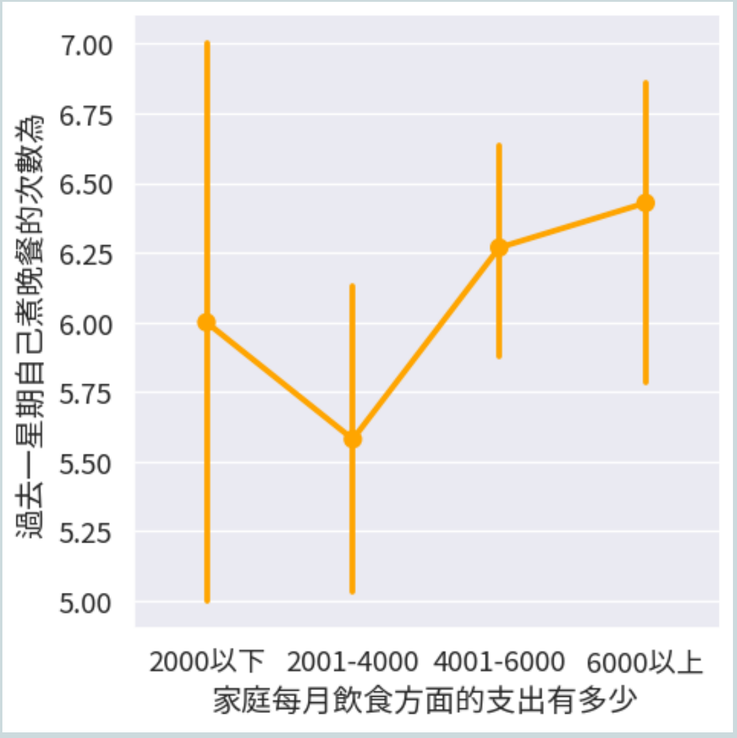




FURTHER INVESTIGATIONS ON THE RELATIONSHIP OF TWO ASSIGNED DATA



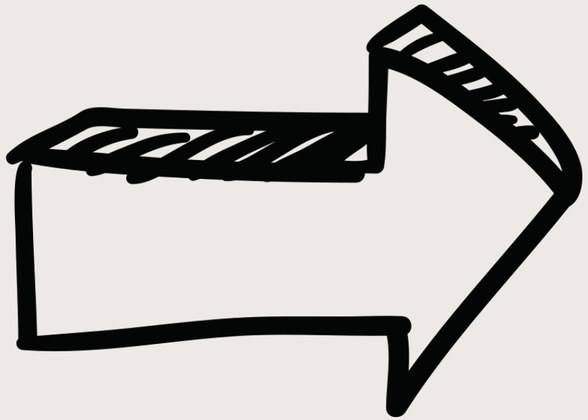




FINDINGS

- A partial eclipse and an unbalanced diet
- Low milk intake, less fruit consumption
- Insufficient drinking water, educational level related to nutrition score
- Simplified daily dining places, cooking at home accounts for a large proportion
- Children's nutrition score has no obvious relationship with all the factors involved in the survey

HYPOTHESIS TEST



	Categorical	Numerical
Categorical	Chi-Square Test	T-test ANOVA
Continuous	Regression	Correlation Test

CHI-SQUARE TEST

Test of Independence of our Data

Null Hypothesis H_0: The two categorical variables have no relationship
Alternative hypothesis H_1: There is a relationship between two categorical variables
$\alpha = 0.05$

Assumed independent:

$$\mathbf{E_{ij}} = (R_i \cap C_j) = P(R_i) \times P(C_j) \times \text{Grand Total}$$

$$= \frac{R_i \text{ Total}}{\text{Grand Total}} \times \frac{C_j \text{ Total}}{\text{Grand Total}} \times \text{Grand Total}$$

$$= \frac{\mathbf{R_i \text{ Total} \times C_j \text{ Total}}}{\mathbf{\text{Grand Total}}}$$

$$\chi^2 = \sum \frac{(O - E)^2}{E} \sim \chi^2_{(r-1)(c-1)}$$

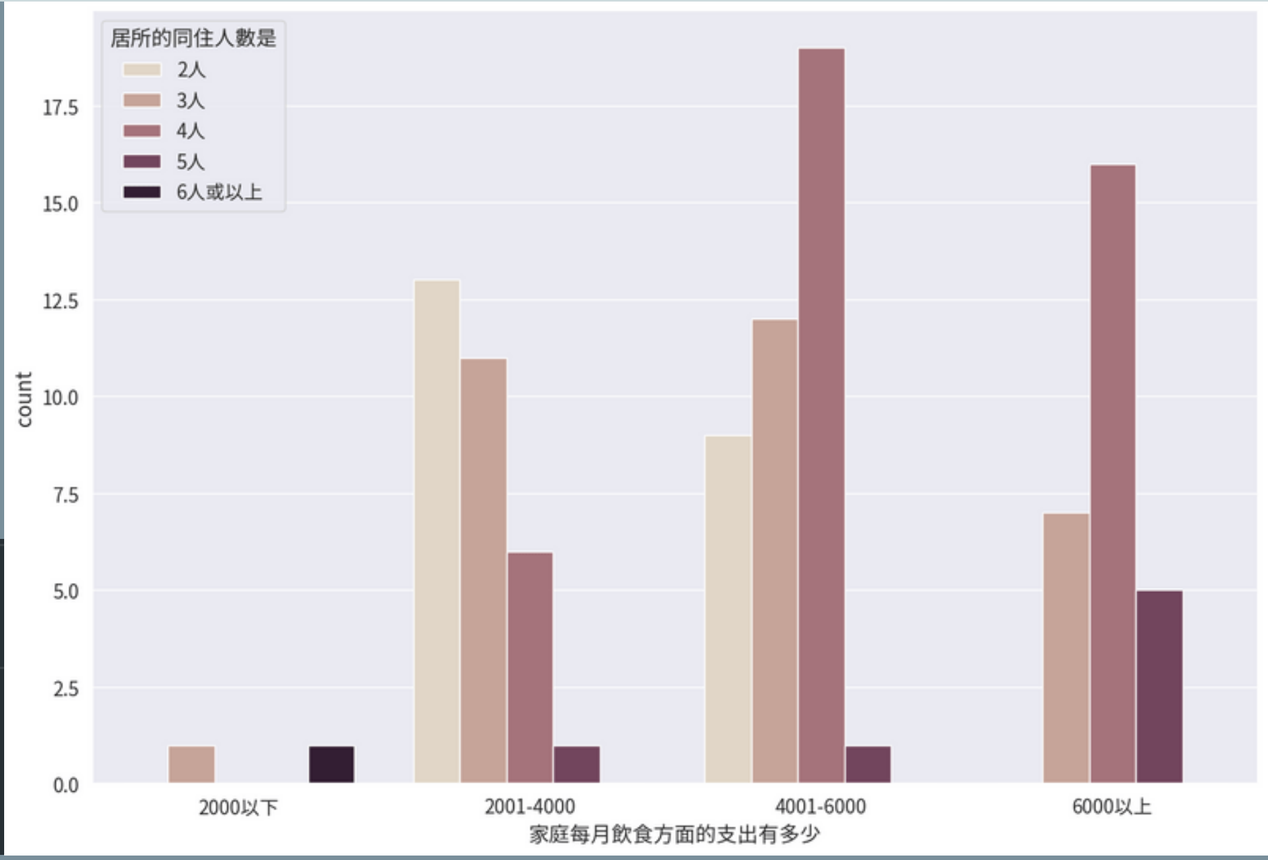
EXAMPLE SET

居所的同住人數是	2人	3人	4人	5人	6人或以上	Total
家庭每月飲食方面的支出有多少						
2000以下	0	1	0	0	1	2
2001-4000	13	11	6	1	0	31
4001-6000	9	12	19	1	0	41
6000以上	0	7	16	5	0	28
Total	22	31	41	7	1	102

```
chiTest(data['家庭每月飲食方面的支出有多少'],data['居所的同住人數是'])
```

✓ 0.8s

```
The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 76.61283553950395 and p value is: 1.8189560968551177e-11
Null Hypothesis is rejected.
```



H_0: The two categorical variables have no relationship

H_1: There is a relationship between two categorical variables

FOOD EXPENSES VS THE NUMBER OF FAMILY MEMBERS HAVE A RELATIONSHIP

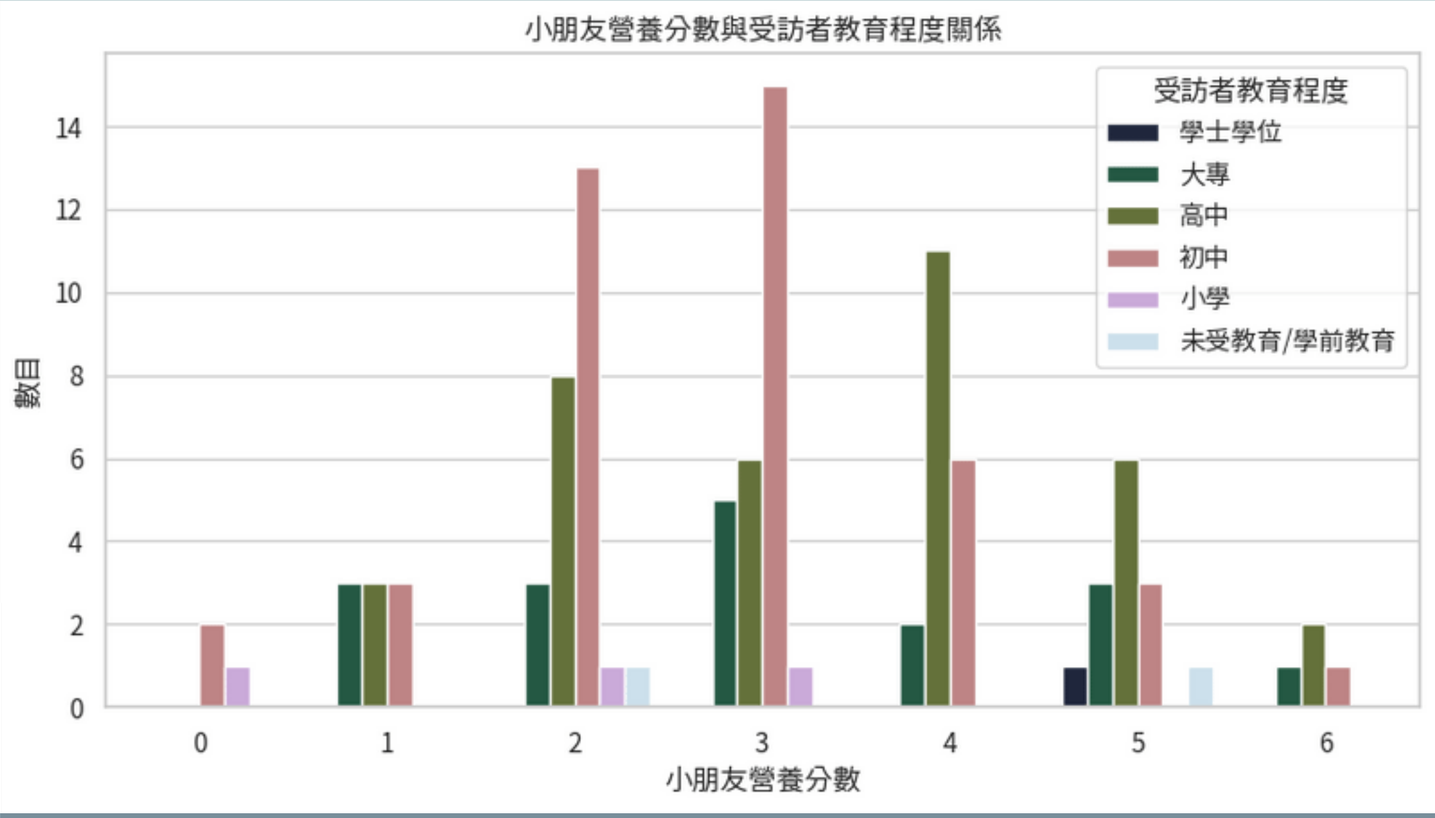
CASE 1 NUTRITION SCORE VS RESPONDENT'S EDU.LEVEL

小朋友營養分數	0	1	2	3	4	5	6	Total
受訪者教育程度								
初中	2	3	13	15	6	3	1	43
大專	0	3	3	5	2	3	1	17
學士學位	0	0	0	0	0	1	0	1
小學	1	0	1	1	0	0	0	3
未受教育/學前教育	0	0	1	0	0	1	0	2
高中	0	3	8	6	11	6	2	36
Total	3	9	26	27	19	14	4	102

```
chiTest(data['受訪者教育程度'],data['小朋友營養分數'])
```

✓ 0.2s

```
The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 34.12491400435096 and p value is: 0.2758663725165147
Failed to reject the null hypothesis.
```



H₀: The two categorical variables have no relationship

H₁: There is a relationship between two categorical variables

NUTRITION SCORE OF THE CHILD VERSUS THE RESPONDENT'S EDUCATION LEVEL HAVE NO RELATIONSHIP

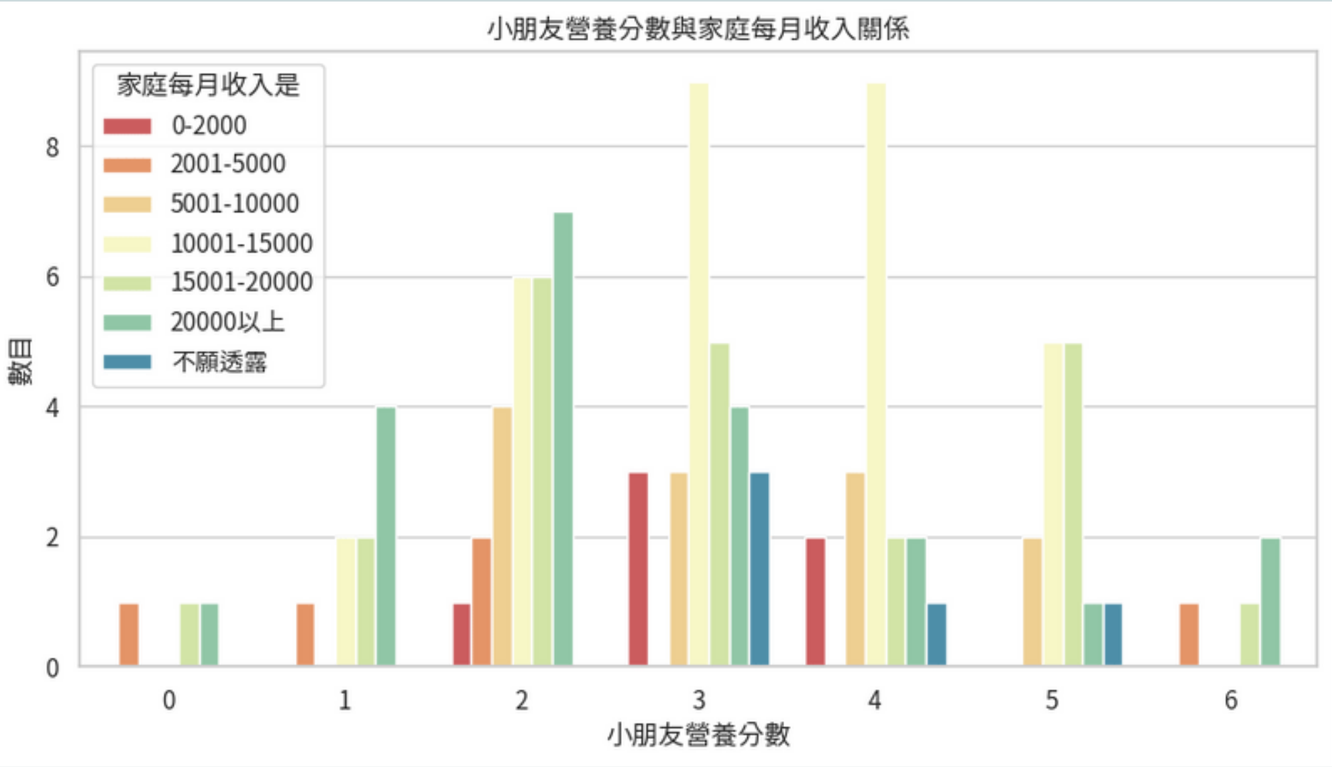
CASE 2 NUTRITION SCORE VS FAMILY'S MONTHLY INCOME

小朋友營養分數	0	1	2	3	4	5	6	Total
家庭每月收入是								
0-2000	0	0	1	3	2	0	0	6
10001-15000	0	2	6	9	9	5	0	31
15001-20000	1	2	6	5	2	5	1	22
20000以上	1	4	7	4	2	1	2	21
2001-5000	1	1	2	0	0	0	1	5
5001-10000	0	0	4	3	3	2	0	12
不願透露	0	0	0	3	1	1	0	5
Total	3	9	26	27	19	14	4	102

```
chiTest(data['家庭每月收入是'],data['小朋友營養分數'])
```

✓ 0.8s

```
The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 38.092028437153346 and p value is: 0.37439851705479565
Failed to reject the null hypothesis.
```



H₀: The two categorical variables have no relationship

H₁: There is a relationship between two categorical variables

NUTRITION SCORE OF THE CHILD VS THE FAMILY'S MONTHLY INCOME HAVE NO RELATIONSHIP

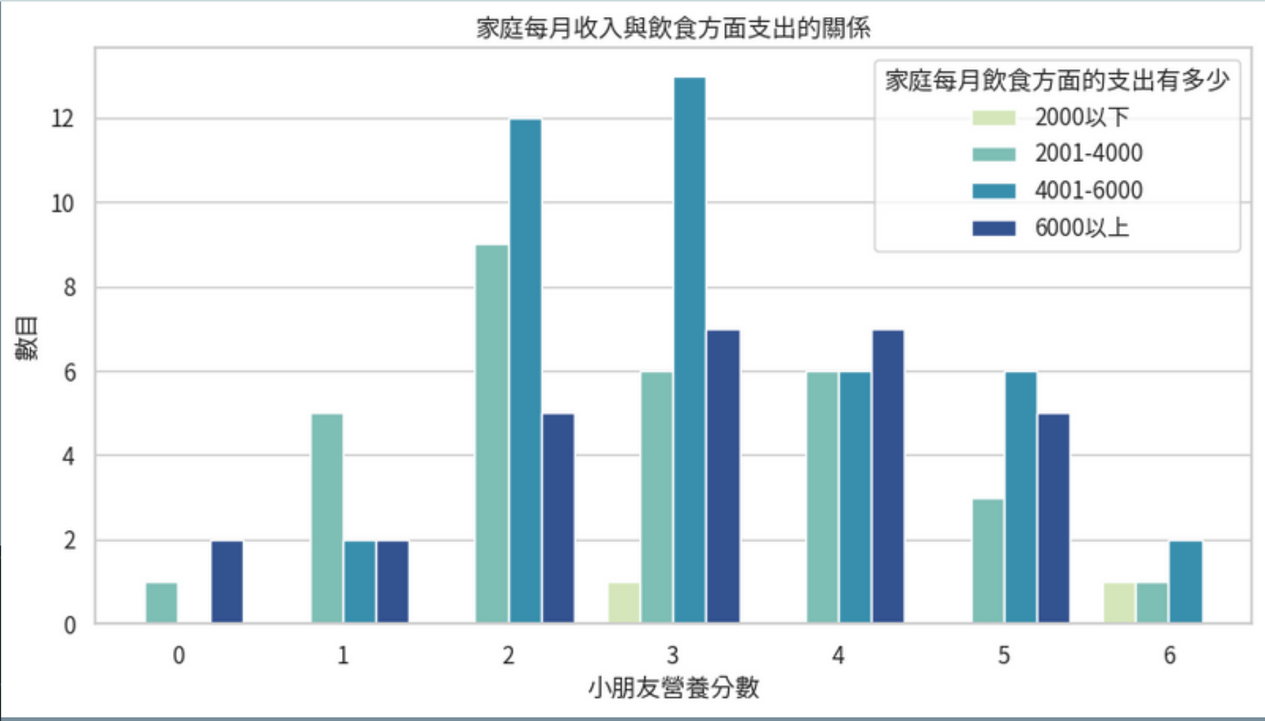
CASE 3 NUTRITION SCORE VS FAMILY FOOD MONTHLY EXPENSES

小朋友營養分數	0	1	2	3	4	5	6	Total
家庭每月飲食方面的支出有多少								
2000以下	0	0	0	1	0	0	1	2
2001-4000	1	5	9	6	6	3	1	31
4001-6000	0	2	12	13	6	6	2	41
6000以上	2	2	5	7	7	5	0	28
Total	3	9	26	27	19	14	4	102

```
chiTest(data['家庭每月飲食方面的支出有多少'],data['小朋友營養分數'])
```

✓ 0.6s

```
The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 23.230007694973708 and p value is: 0.1819054730527253
Failed to reject the null hypothesis.
```



H_0: The two categorical variables have no relationship

H_1: There is a relationship between two categorical variables

NUTRITION SCORE OF THE CHILD VERSUS THE FAMILY FOOD MONTHLY EXPENSES HAVE NO RELATIONSHIP

CONCLUSION ON CHI-SQUARE TEST

```
chiTest(data['受訪者教育程度'],data['小朋友營養分數'])
```
✓ 0.2s

```
The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 34.12491400435096 and p value is: 0.2758663725165147
Failed to reject the null hypothesis.
```

```
chiTest(data['家庭每月收入是'],data['小朋友營養分數'])
```
✓ 0.8s

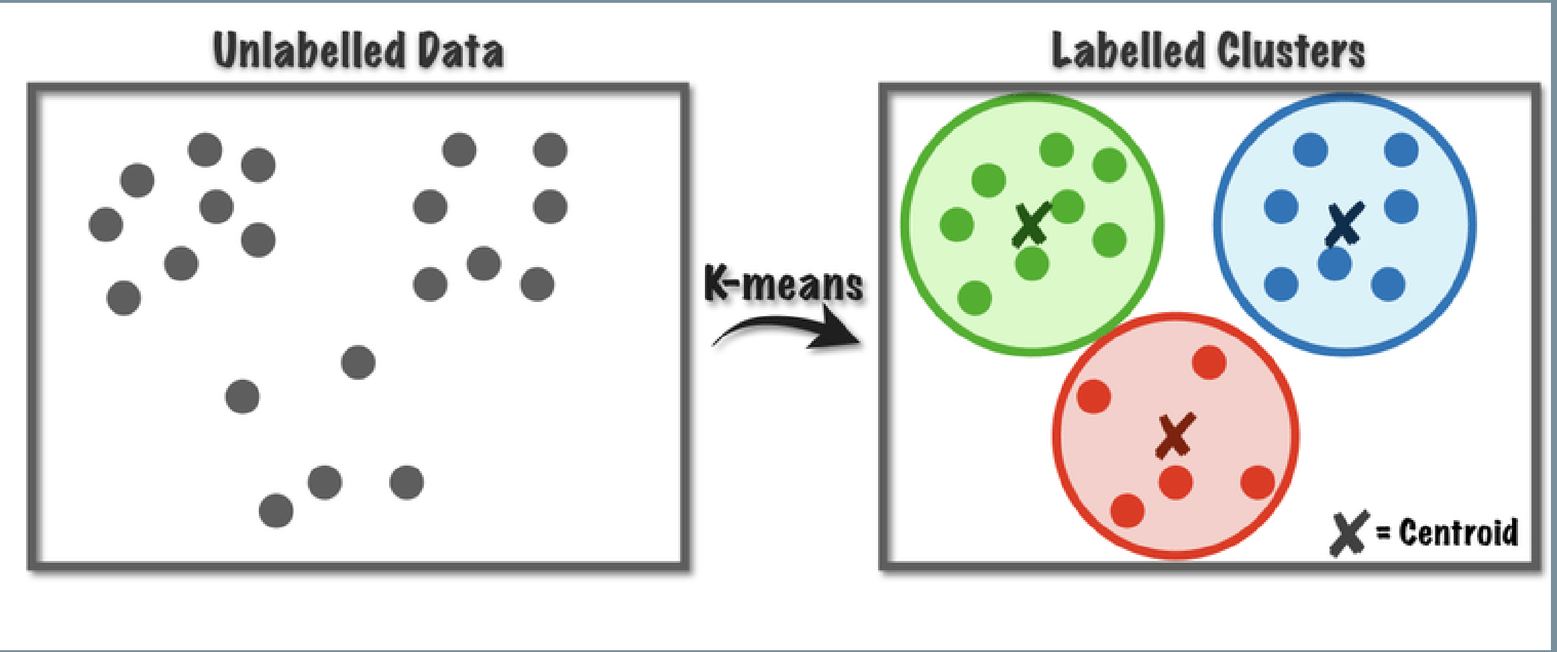
```
The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 38.092028437153346 and p value is: 0.37439851705479565
Failed to reject the null hypothesis.
```

```
chiTest(data['家庭每月飲食方面的支出有多少'],data['小朋友營養分數'])
```
✓ 0.6s

```
The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 23.230007694973708 and p value is: 0.1819054730527253
Failed to reject the null hypothesis.
```

the Respondent's education level
/
Family's monthly income
/
Family food monthly expenses
with no relationship with
Nutrition score

Clustering for our data



Aims:

->Classifier different classes of family

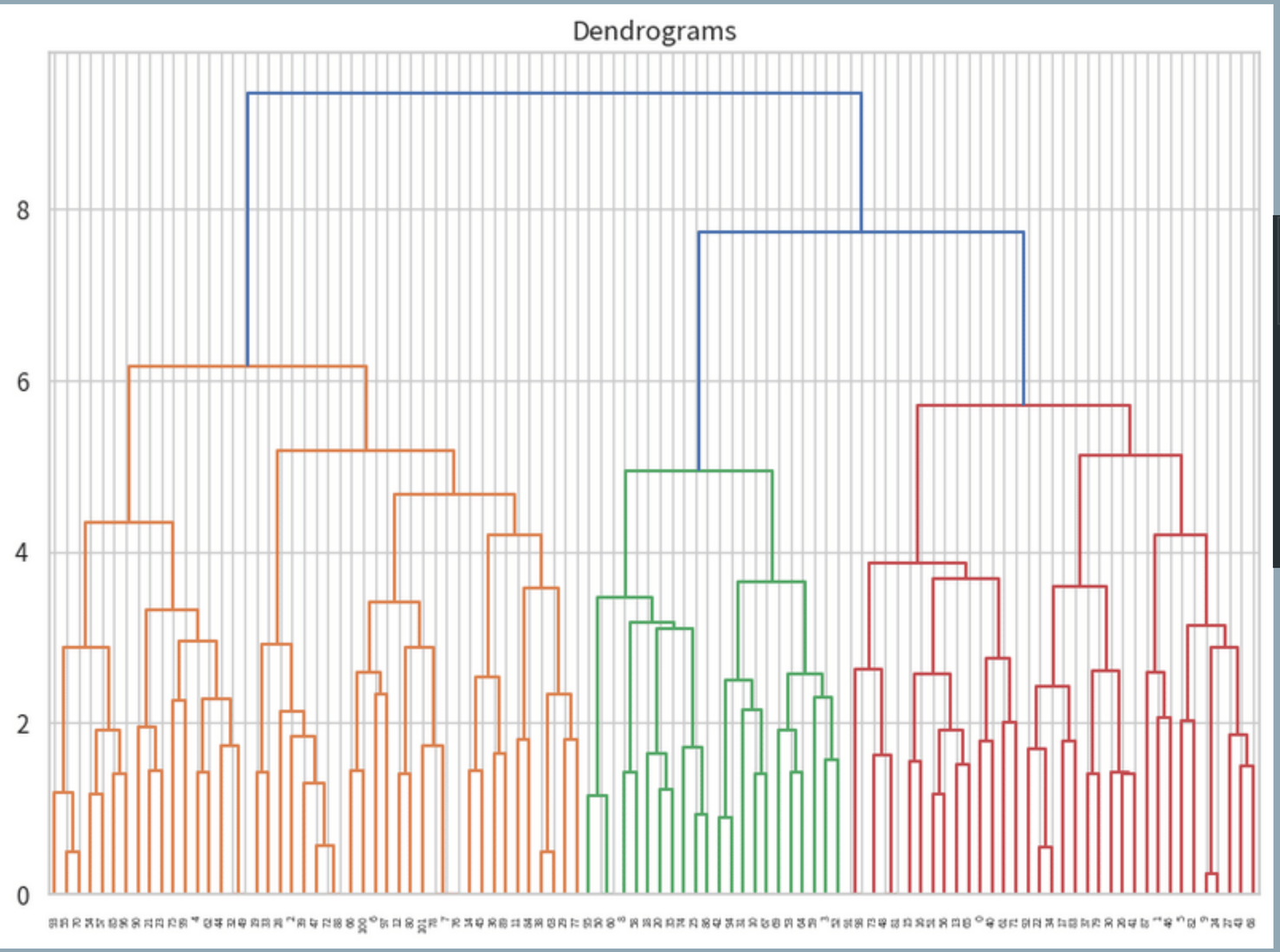
->Discover the prefence of the events of different classes

DEMOGRAPHIC DATA -> DATA PREPROCESSING

	家庭每月飲食方面的支出有多少_2000以下	家庭每月飲食方面的支出有多少_2001-4000	家庭每月飲食方面的支出有多少_4001-6000	家庭每月飲食方面的支出有多少_6000以上	受訪者性別_女	受訪者性別_男	受訪者年齡_21至30歲	受訪者年齡_31至40歲	受訪者年齡_41至50歲	受訪者年齡_51至60歲	...	家庭每月收入是_15001-20000	家庭每月收入是_20000以上	家庭每月收入是_2001-5000	家庭每月收入是_5001-10000
0	0	0	1	0	1	0	0	0	1	0	...	0	0	0	0
1	0	0	0	1	1	0	0	0	1	0	...	1	0	0	0
2	0	0	1	0	1	0	0	1	0	0	...	0	0	0	0
3	0	1	0	0	1	0	0	0	1	0	...	0	0	0	0
4	0	1	0	0	1	0	0	0	1	0	...	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
97	0	0	0	1	1	0	1	0	0	0	...	0	1	0	0
98	0	0	0	1	1	0	0	1	0	0	...	0	0	0	1
99	0	1	0	0	1	0	0	1	0	0	...	0	1	0	0
100	0	0	0	1	1	0	0	0	1	0	...	0	1	0	0
101	0	0	0	1	1	0	0	1	0	0	...	1	0	0	0

- One Hot encode
- MinMax Scaler

HIERARCHICAL CLUSTERING



-> We can defifne our data into three groups

```
Cluster_data.groupby(['HierCluster'])['HierCluster'].count()
```

✓ 0.3s

```
HierCluster
1    45
2    35
3    22
Name: HierCluster, dtype: int64
```

CLASSES

HierCluster	受訪者性別	受訪者年齡	受訪者教育程度	居所的同住人數是	家庭每月收入是	同住有沒有3至12歲的兒童	家庭每月飲食方面的支出有多少
1	女	31至40歲	初中	4人	20000以上	2	4001-6000
2	女	31至40歲	初中	3人	10001-15000	1	4001-6000
3	女	31至40歲	初中	2人	10001-15000	1	2001-4000

First Group <-Families have 4
Second Group <- Families have 3
Third Group <- Families have 2

HierCluster	[社區流動營養診所]興趣分數	[注冊營養師進行網上講座]興趣分數	[與注冊營養師烹調豐富營養食物]興趣分數	[食物營養及環保回收的攤位游戲]興趣分數
1	3.777778	3.733333	4.000000	4.333333
2	3.885714	3.371429	4.142857	4.028571
3	3.772727	3.727273	4.227273	4.136364

HierCluster	請為[學習到簡單易煮的食譜]評分	請為[改善兒童營養]評分	請為[學習到使用成本低但營養價值高的食材]評分	請為[參加烹飪班]評分
1	3.933333	4.288889	4.444444	3.666667
2	3.828571	4.142857	4.285714	3.628571
3	3.818182	4.000000	4.272727	3.727273

SUGGESTION (PERSONAL)

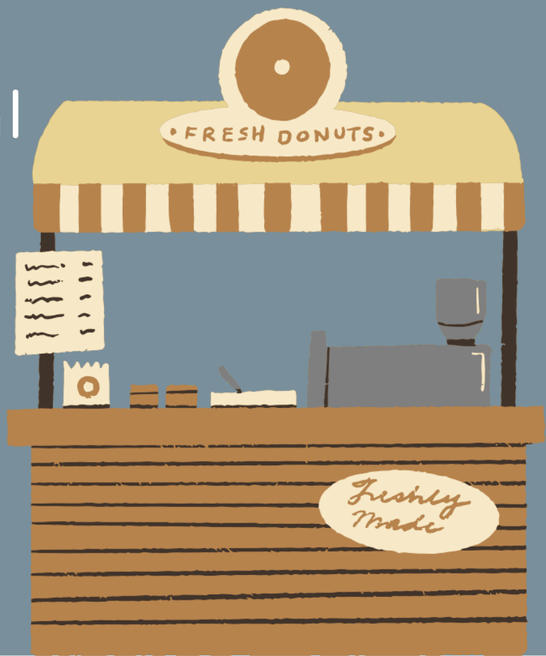
- Optimize food consumption structure

- Established correct eating habits

- Learn more about healthy eating

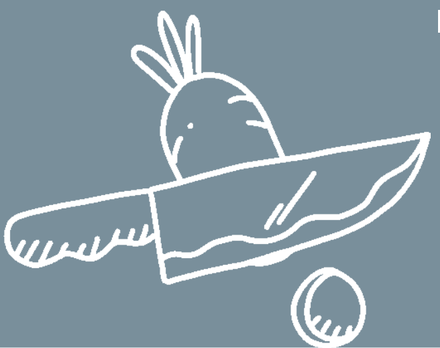
SUGGESTION (ORGANIZATION)

- Host booth game day
  - For the first group (a family of four members),
    - as participants
      - Taught the benefits of a balanced diet
    - as organizers
      - volunteer work opportunity
      - learn from the process of preparing material
    - spend their family day
    - learn through play



SUGGESTION (ORGANIZATION)

- Host cooking classes
  - For the majority of groups (groups 2 and 3)
    - Believe that the dishes made by themselves are especially tasty
      - Encourage their children to try new foods
      - Improve their eating habits
    - Gain a better understanding of food and nutrition
    - Spend family time together
    - Learn to cherish
    - Be grateful



SUGGESTION (SOCIETY)

- Create a supportive environment
  - i.e. Schools 
    - Prohibits the sale and promotion of unhealthy snacks
    - Encourage tuckshops to provide fruits and other healthy snacks at discounted prices
- Promote a balanced diet 
  - i.e. Department of Health
    - Actively promote a balanced diet
    - Take the lead in encouraging fruit and vegetables

LIMITATION

- Limited Responses
  - Fails to capture busy, lazy, or indifferent respondents
    - Who make up a significant portion of the population
  - Limitations of the questions set
    - Small amount of questions
    - Not specific enough
  - As a result
    - Relationships aren't obvious or even contradictions
    - Data may not be adequately explained by the clustering

LIMITATION

- Unreliability
  - Misinterprets a question
  - Gives an incomplete or indefinite response
  - Human errors
    - Mark answers incorrectly
    - Manipulate entries by asking leading questions

Appendix


```
def chiTest(df1,df2):
    data_crosstab = pd.crosstab(df1,
                                df2,
                                margins=True, margins_name="Total")
    alpha = 0.05
    chi_square = 0
    rows = df1.unique()
    columns = df2.unique()
    for i in columns:
        for j in rows:
            O = data_crosstab[i][j]
            E = data_crosstab[i]['Total'] * data_crosstab['Total'][j] / data_crosstab['Total']['Total']
            chi_square += (O-E)**2/E

    print("The p-value approach to hypothesis testing in the decision rule")
    p_value = 1 - stats.chi2.cdf(chi_square, (len(rows)-1)*(len(columns)-1))
    conclusion = "Failed to reject the null hypothesis."
    if p_value <= alpha:
        conclusion = "Null Hypothesis is rejected."

    print("chisquare-score is:", chi_square, "and p value is:", p_value)
    print(conclusion)
```

✓ 0.4s

Appendix


```
import scipy.cluster.hierarchy as shc
from sklearn.preprocessing import MinMaxScaler
X=data[['家庭每月飲食方面的支出有多少','受訪者性別', '受訪者年齡', '受訪者教育程度', '居所的同住人數是', '家庭每月收入是']]
X=pd.get_dummies(X)
Y =data[['同住有沒有3至12歲的兒童', '現時居住的劏房或天臺屋有沒有共用空間',
        '過去一星期自己煮晚餐的次數為', '過去一星期外出晚餐的次數為', '過去一星期外賣晚餐的次數為',]]
G = pd.concat([X,Y],axis=1)
MinMax = MinMaxScaler()
G_std = MinMax.fit_transform(G)
plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = shc.dendrogram(shc.linkage(G_std, method='ward'))
```

```
from sklearn.cluster import AgglomerativeClustering
cluster = AgglomerativeClustering(n_clusters=3, affinity='euclidean', linkage='ward')
data['HierCluster'] = cluster.fit_predict(G_std)
data['HierCluster'] = data['HierCluster'] + 1
```