**I. Research Background and Motivation**

Customer personality analysis represents a comprehensive examination of a company's ideal clients. It aids businesses in better understanding their customers, enabling them to tailor their products more effectively based on the specific needs, behaviors, and concerns of various customer types. Customer personality analysis assists companies in modifying their offerings for target customer segments drawn from different types of customer groups. For instance, instead of spending money to promote new products to every customer in the company's database, a firm can analyze which customer group is most likely to purchase the product and then market the product exclusively to that particular group.

To identify target customers for wine, meat products, and gold, it is proposed to first investigate the factors that influence customer purchasing intentions

**II. Research Objectives**

The study aims to analyze whether customers' spending on wine, meat products, and gold is related to their educational level, emotional status, and household annual income.

**III. Variable Introduction**

1. **ID (Customer ID)**: A unique identifier assigned to each customer.
2. **Year\_Birth (Customer's Year of Birth)**: The year in which the customer was born.
3. **Education (Customer's Education Level)**:

* Basic: Elementary or Middle School
* Graduation: High School
* 2n Cycle: Master's Degree
* Master: Bachelor's Degree
* PhD: Doctorate

1. **Marital\_Status (Emotional Status)**:

* Divorced: The customer is divorced.
* Married: The customer is married.
* Single: The customer is single.
* Together: The customer is in a relationship but not married.
* Widow: The customer is a widow/widower.

1. **Income (Household Annual Income)**: The total yearly income of the customer's household.
2. **Recency**: The number of days since the customer's last purchase.
3. **MntWines (Spending on Wines in the Last Two Years)**: The amount of money spent on wine in the past two years.
4. **MntMeatProducts (Spending on Meat Products in the Last Two Years)**: The amount of money spent on meat products in the past two years.
5. **MntGoldProds (Spending on Gold in the Last Two Years)**: The amount of money spent on gold in the past two years.
6. **Wine**: Categorization of the MntWines variable into five groups based on spending amount:

* Less than 300
* 300 to 600
* 600 to 900
* 900 to 1200
* More than 1200

1. **Meat**: Categorization of the MntMeatProducts variable into five groups based on spending amount:

* Less than 150
* 150 to 300
* 300 to 450
* 450 to 700
* More than 700

1. **Gold**: Categorization of the MntGoldProds variable into five groups based on spending amount:

* Less than 50
* 50 to 100
* 100 to 150
* 150 to 200
* More than 200

1. **Incomegp (Income Groups)**: Categorization of the Income variable into four groups based on amount:

* Less than 30,000
* 30,000 to 50,000
* 50,000 to 70,000
* More than 70,000

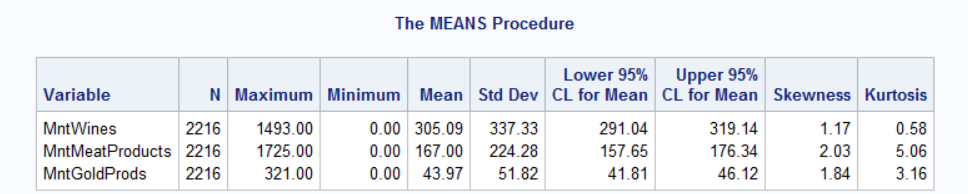
**IV. Data Source**

The dataset for this study was obtained from Kaggle, an online community of data scientists and machine learning practitioners. The specific dataset, titled "Customer Personality Analysis," was used to analyze customer spending behaviors on wine, meat products, and gold. This dataset includes detailed information on customer demographics, purchasing history, and socio-economic indicators. The dataset can be accessed through the following URL: <https://www.kaggle.com/code/armanmostafazadeh/customer-personality-analysis-k-means/input>, provided by the user Arman Mostafazadeh.

**V. Descriptive Statistics**

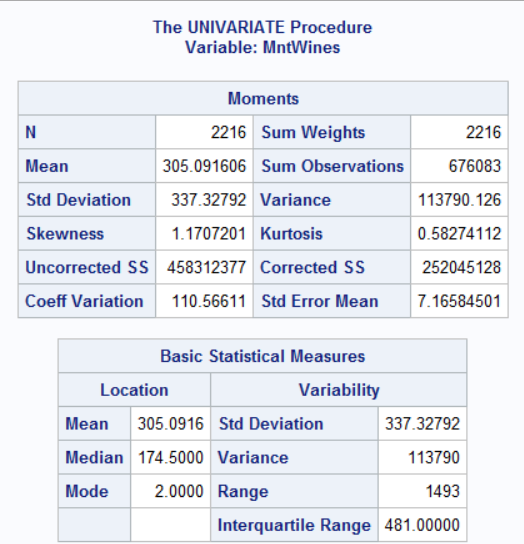
Descriptive statistics provide a concise summary of data through measures of central tendency (mean, median) and measures of variability (standard deviation, variance, range, quartiles).

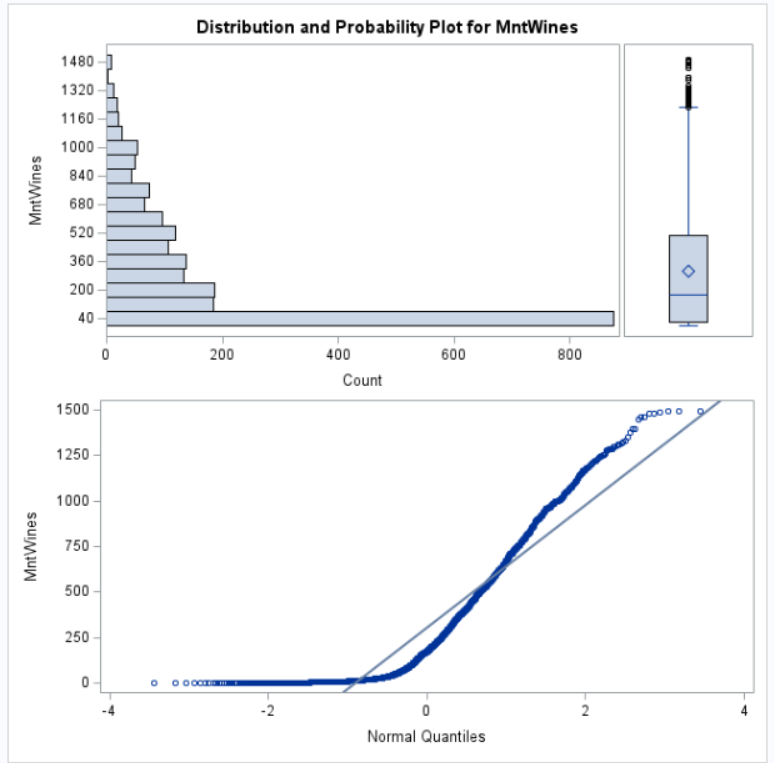
The descriptive statistics for continuous variables such as "MntWines" (amount spent on wines), "MntMeatProducts" (amount spent on meat products), and "MntGoldProds" (amount spent on gold) over the past two years :

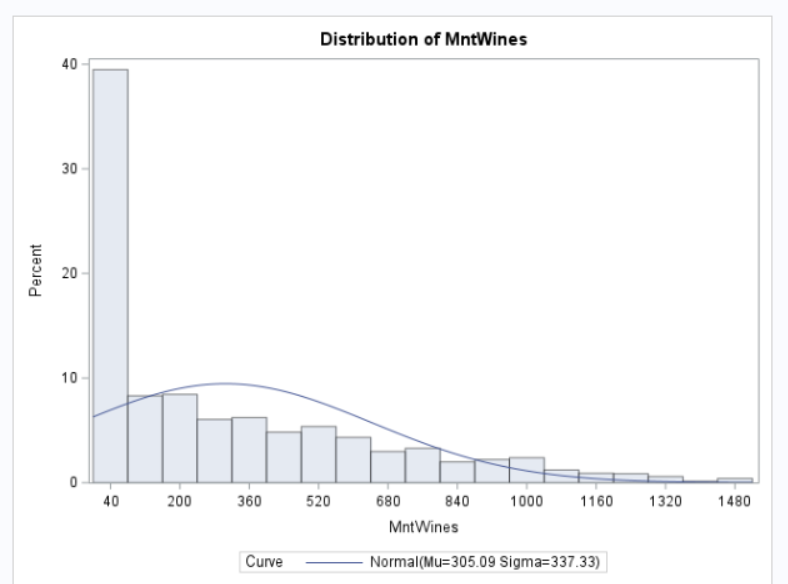


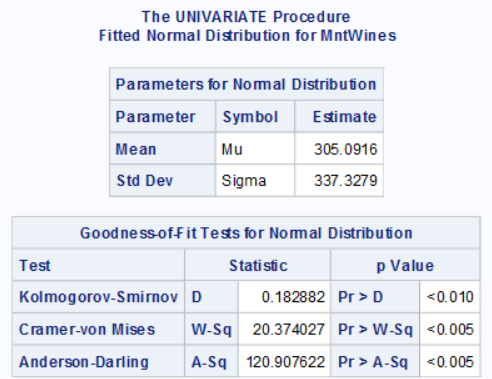
* **For the continuous data of "MntWines"**

as indicated by the part circled in red, the goodness-of-fit test for normal distribution indicates that the p-value is less than 0.05, indicating that this variable does not follow a normal distribution.

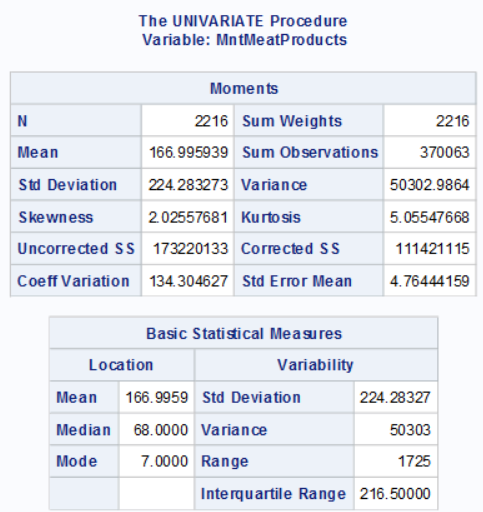


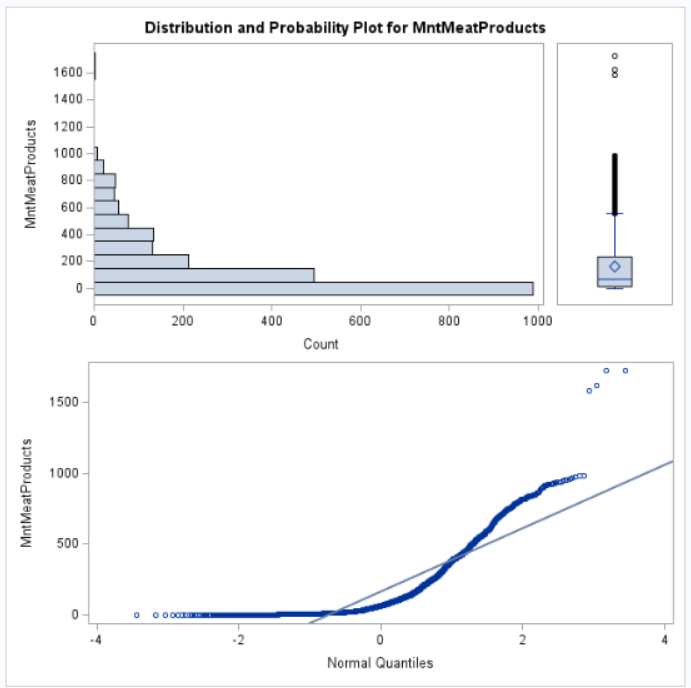


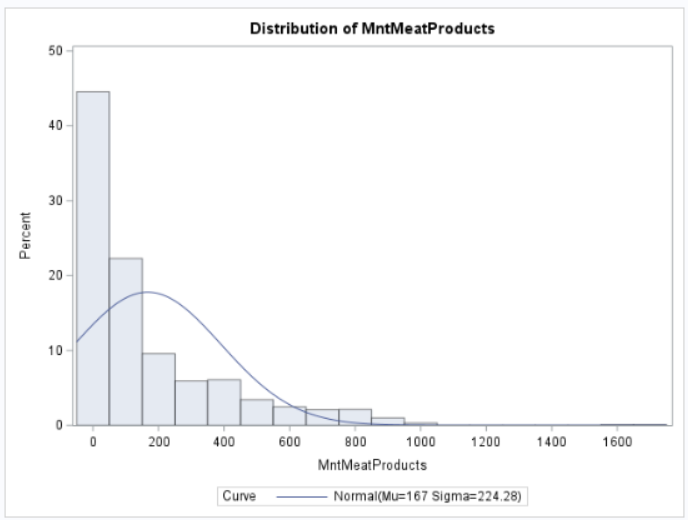


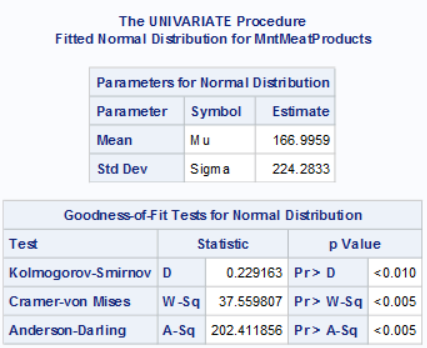


* **For the continuous data of "** **MntMeatProducts"**

/

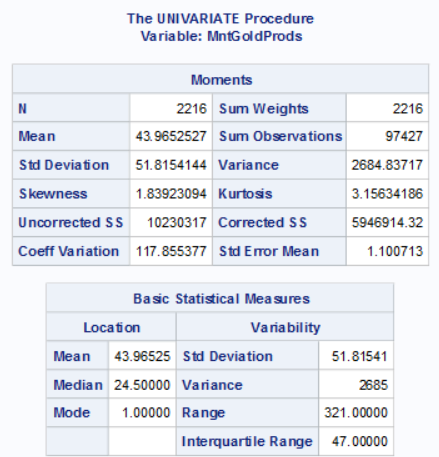


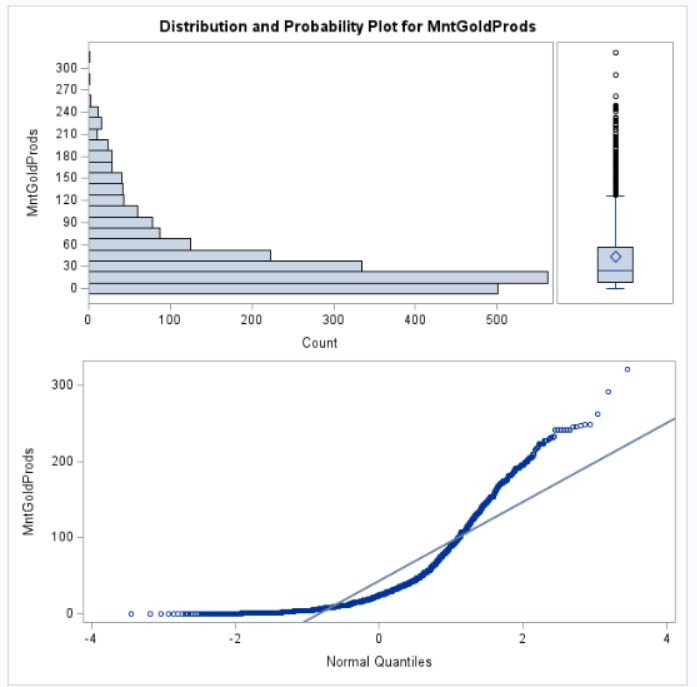


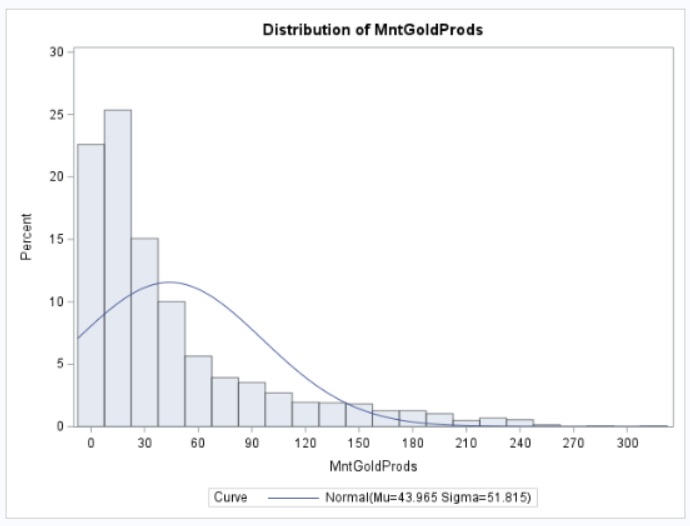


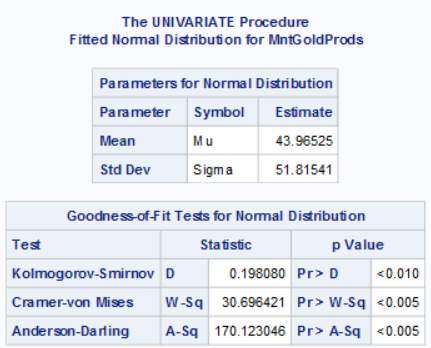
* **For the continuous data of "MntGoldProds"**

as indicated by the portion encircled in red, the goodness-of-fit test for normal distribution reveals that the p-value is less than 0.05. Consequently, this suggests that the variable does not adhere to a normal distribution.









**VI. Independence Test**

* Using independence test to test whether there is a significant relationship between customers' education level, marital status, household annual income, and their spending on wine over the past two years

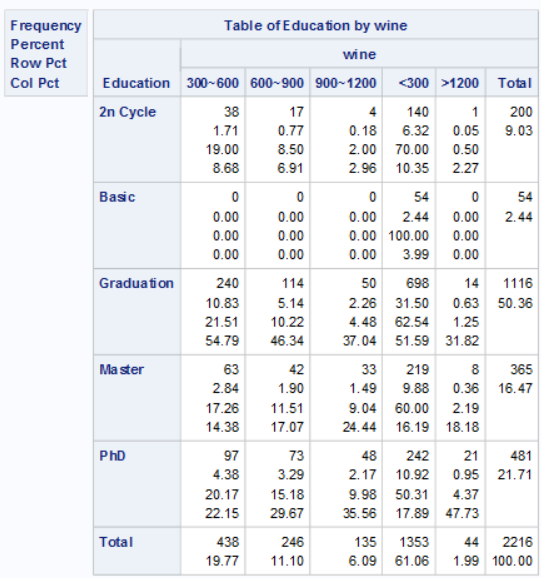


Figure 1: Contingency Table of Education Level and Wine Spending Over the Past Two Years

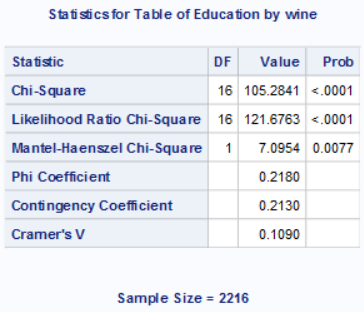


Figure 2: Chi-square Test for Independence Between Education Level and Wine Spending Over the Past Two Years

From Figure 2, it can be observed that the Chi-square p-value is less than 0.05, allowing us to reject the null hypothesis (H0). This means that education level and wine spending over the past two years are not independent of each other; rather, they are interrelated and influence each other.

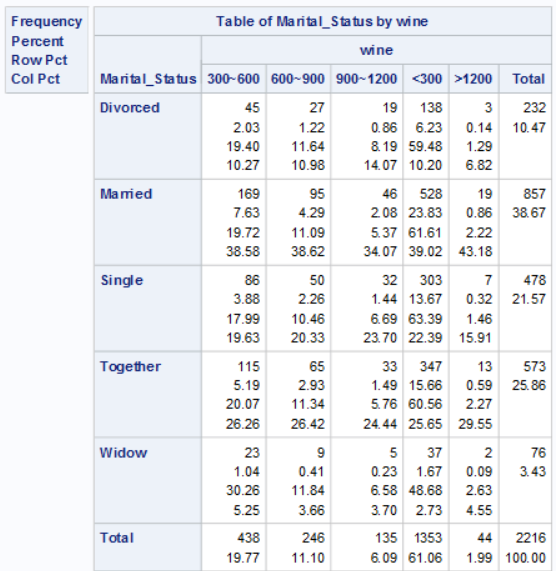


Figure 3: Contingency Table of Marital Status and Wine Spending Over the Past Two Years

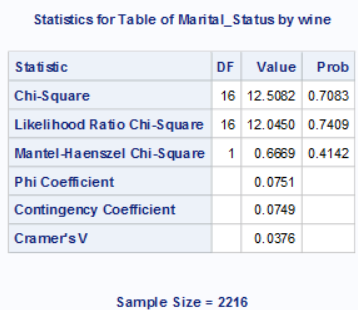


Figure 4: Chi-square Test for Independence Between Marital Status and Wine Spending Over the Past Two Years

From Figure 4, it is evident that the Chi-square p-value is greater than 0.05, therefore we cannot reject the null hypothesis (H0). This indicates that marital status and wine spending over the past two years are independent of each other, meaning they do not influence each other.

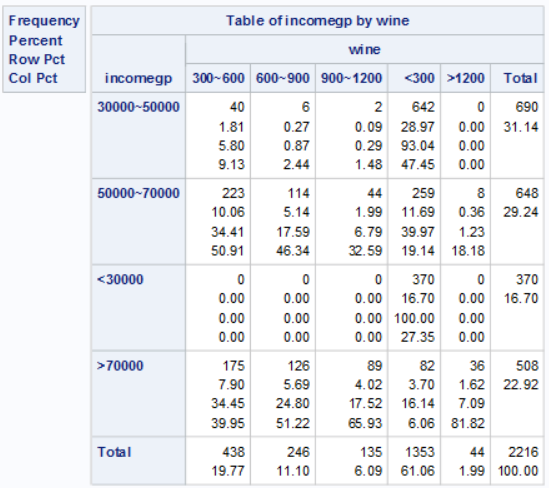


Figure 5: Contingency Table of Household Annual Income and Wine Spending Over the Past Two Years

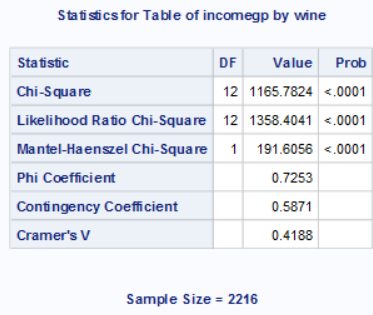


Figure 6: Chi-square Test for Independence Between Household Annual Income and Wine Spending Over the Past Two Years

From Figure 6, it is apparent that the Chi-square p-value is less than 0.05, allowing for the rejection of the null hypothesis (H0). This implies that household annual income and wine spending over the past two years are not independent of each other; rather, they influence each other.

* Testing whether education level, marital status, and household annual income are related to customers' spending on meat products over the past two years.

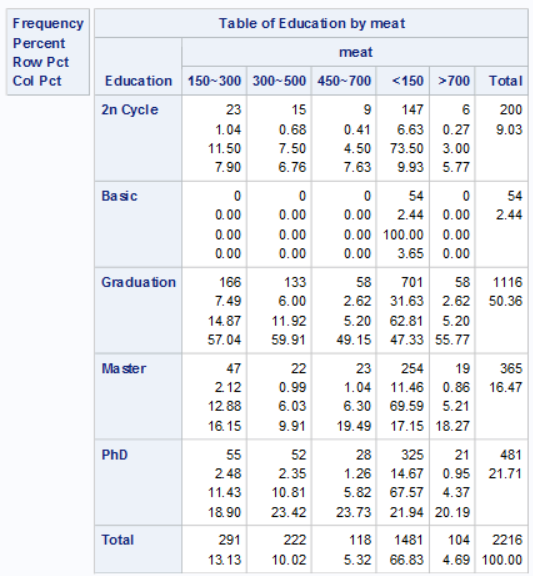


Figure 7: Contingency Table of Education Level and Meat Product Spending Over the Past Two Years

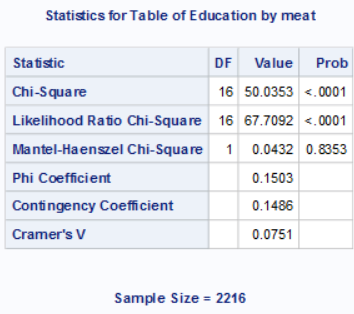


Figure 8: Chi-square Test for Independence Between Education Level and Meat Product Spending Over the Past Two Years

From Figure 8, it is evident that the Chi-square p-value is less than 0.05, thus we can reject the null hypothesis (H0). This means that education level and spending on meat products over the past two years are not independent of each other; instead, they influence each other.

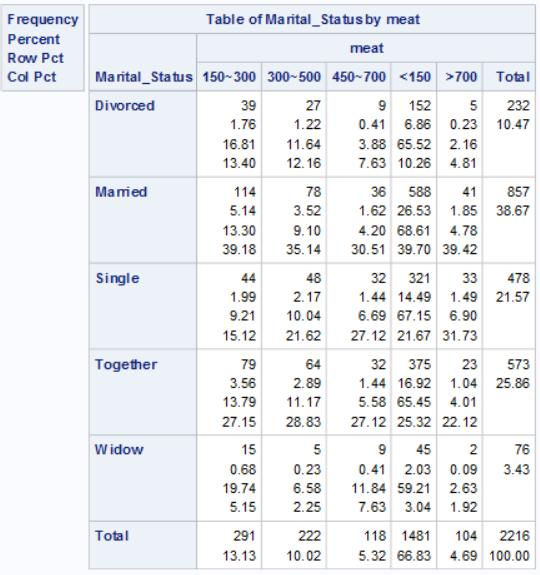


Figure 9: Contingency Table of Marital Status and Meat Product Spending Over the Past Two Years

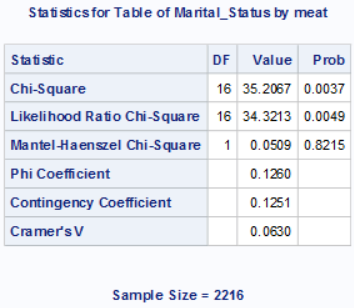


Figure 10: Chi-square Test for Independence Between Marital Status and Meat Product Spending Over the Past Two Years

From Figure 10, it is evident that the Chi-square p-value is less than 0.05, therefore we can reject the null hypothesis (H0). This indicates that marital status and spending on meat products over the past two years are not independent of each other; rather, they influence each other.

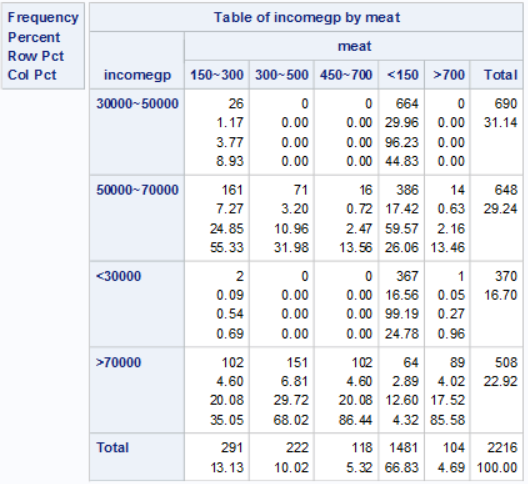


Figure 11: Contingency Table of Household Annual Income and Meat Product Spending Over the Past Two Years

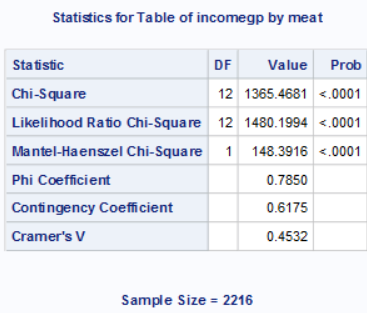


Figure 12: Chi-square Test for Independence Between Household Annual Income and Meat Product Spending Over the Past Two Years

From Figure 12, it is apparent that the Chi-square p-value is less than 0.05, therefore we can reject the null hypothesis (H0). This means that household annual income and spending on meat products over the past two years are not independent of each other; instead, they influence each other.

* Testing whether education level, marital status, and household annual income are related to customers' spending on gold over the past two years.

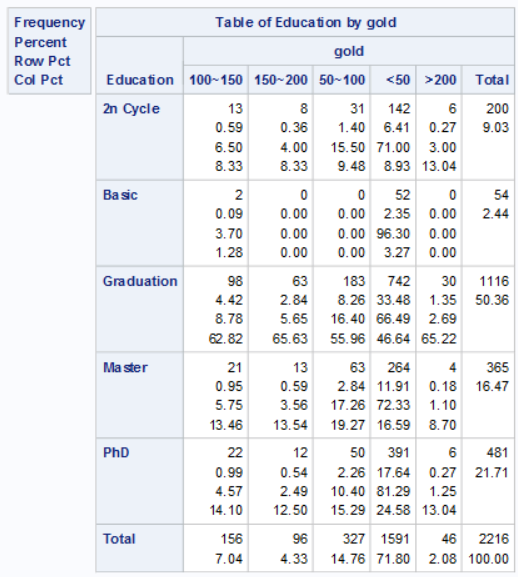
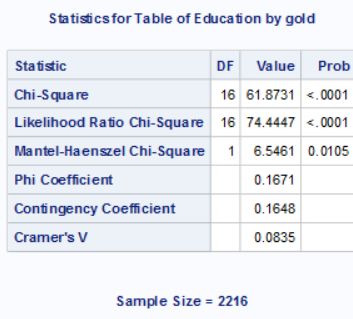


Figure 13: Contingency Table of Education Level and Gold Spending Over the Past Two Years



圖Figure 14: Chi-square Test for Independence Between Education Level and Gold Spending Over the Past Two Years

From Figure 14, it is evident that the Chi-square p-value is less than 0.05, thus we can reject the null hypothesis (H0). This indicates that education level and spending on gold over the past two years are not independent of each other; rather, they influence each other.

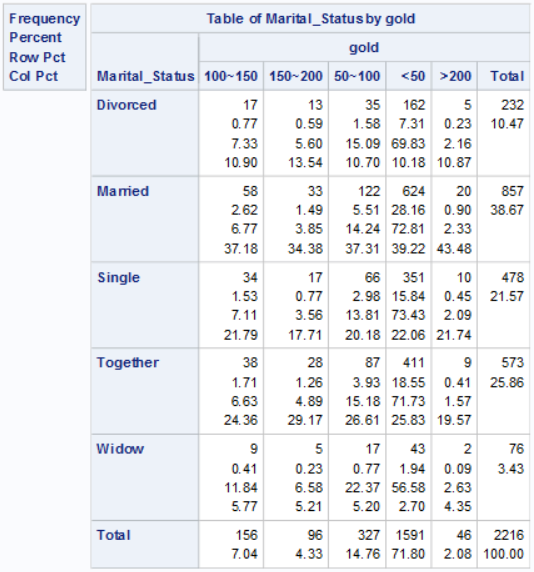


Figure 15: Contingency Table of Marital Status and Gold Spending Over the Past Two Years

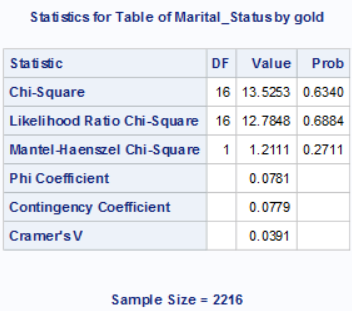


Figure 16: Chi-square Test for Independence Between Marital Status and Gold Spending Over the Past Two Years

From Figure 16, it is apparent that the Chi-square p-value is greater than 0.05, therefore we cannot reject the null hypothesis (H0). This means that marital status and spending on gold over the past two years are independent of each other, indicating they do not influence each other.

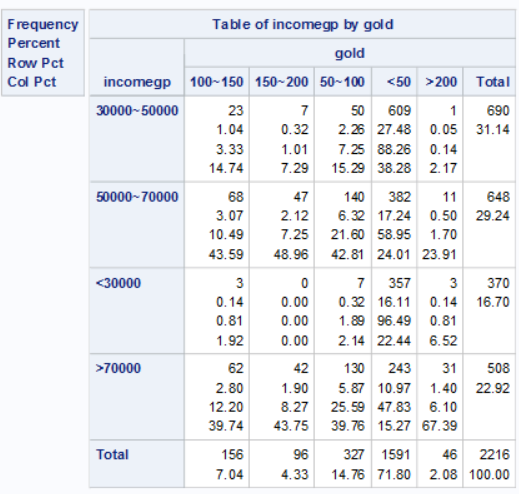


Figure 17: Contingency Table of Household Annual Income and Gold Spending Over the Past Two Years

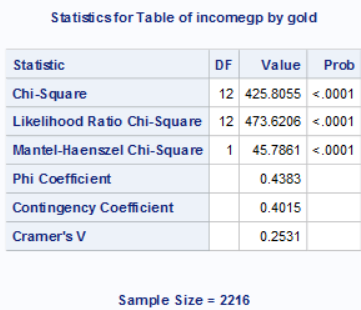


Figure 18: Chi-square Test for Independence Between Household Annual Income and Gold Spending Over the Past Two Years

From Figure 18, it is evident that the Chi-square p-value is less than 0.05, thus we can reject the null hypothesis (H0). This indicates that household annual income and spending on gold over the past two years are not independent of each other; rather, they influence each other.

**VII. Conclusion**

Synthesizing Figures 1 to 18, the following conclusions can be drawn:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Spending on Wine Over the Past Two Years** | **Spending on Meat Products Over the Past Two Years** | **Spending on Gold Over the Past Two Years** |
| Customer Education Level | Related | Related | Related |
| Customer Marital Status | Not Related | Related | Not Related |
| Customer Household Annual Income | Related | Related | Related |

Spending on wine, meat products, or gold is related to a customer's level of education and annual household income; however, a customer's relationship status only correlates with spending on meat products and is unrelated to expenditures on wine or gold. This means if a manufacturer wants to sell wine, meat products, or gold items, they can tailor their pricing and promotional activities for customers with different levels of education and household incomes. However, if a manufacturer wishes to set sales prices based on customers' relationship status for holidays like Valentine's Day, this approach would be more suitable only for meat product vendors.