In [1]:

```
import pandas as pd
import numpy as np
from scipy.stats import chi2_contingency

# Reading the data which was saved to an excel sheet after pre-processing
df_cleaned = pd.read_excel('finally_clean_data_for_further_processing.xlsx')
```

As for the ANOVA, the same dataset will be used again. For details, see week 1, 2 and 3 of the first module. I also included the automated assessment of the result, to judge whether the p-value is smaller than alpha = 0.05.

H1: "Drinking of wine increases with increasing income"

H2: "The amount of consumed wine is independent of age"

In [ ]:

```
# H1: Chi-squared test for income_category vs wine_frequency
contingency_table_h1 = pd.crosstab(df_cleaned['income_category'], df_cleaned['wine_fr
chi2_h1, p_h1, dof_h1, expected_h1 = chi2_contingency(contingency_table_h1)

print("H1: Drinking of wine increases with increasing income")
print(f"Chi-squared Statistic: {chi2_h1}")
print(f"P-value: {p_h1}")
print(f"Degrees of Freedom: {dof_h1}")

# Interpret the results for H1
alpha = 0.05
if p_h1 < alpha:
    print("Reject the null hypothesis: There is a significant association between inc
else:
    print("Fail to reject the null hypothesis: There is no significant association be
```

Output:

```
H1: Drinking of wine increases with increasing income
Chi-squared Statistic: 740.6435851068533
P-value: 4.2184192613836364e-63
Degrees of Freedom: 200
```

Reject the null hypothesis: There is a significant association between income and wine frequency.

Assessment: This means, that we can reject the null hypothesis, because the p-value is smaller than alpha = 0.05. Again, the result of the ANOVA is confirmed, that there is not enough evidence to conclude that wine consumption significantly varies with different income categories.

Let's continue with the second hypothesis:

In [ ]:

```
# H2: Chi-squared test for age_group vs wine_amount
contingency_table_h2 = pd.crosstab(df_cleaned['age_group'], df_cleaned['wine_amount'
chi2_h2, p_h2, dof_h2, expected_h2 = chi2_contingency(contingency_table_h2)

print("\nH2: The amount of consumed wine is independent of age")
print(f"Chi-squared Statistic: {chi2_h2}")
print(f"P-value: {p_h2}")
print(f"Degrees of Freedom: {dof_h2}")

# Interpret the results for H2
if p_h2 < alpha:
    print("Reject the null hypothesis: There is a significant association between age
else:
    print("Fail to reject the null hypothesis: There is no significant association be
```

Output:

```
H2: The amount of consumed wine is independent of age
Chi-squared Statistic: 550.0303994413212
P-value: 5.58604727899372e-68
Degrees of Freedom: 88
```

Reject the null hypothesis: There is a significant association between age and wine amount.

Assessment: This means, that we can reject the null hypothesis, because the p-value is smaller than alpha = 0.05. This means, that there is no significant association between age and wine amount.

In [ ]: