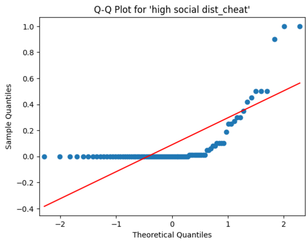
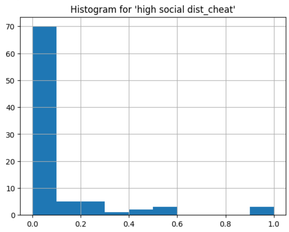
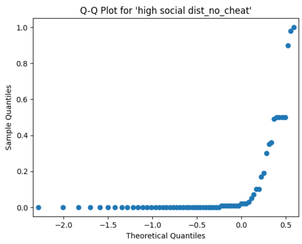
Checking Normality of our data:
Using Q-Q Plots, Histograms and Shapiro-Wilk Test.

In [2]:
```
import pandas as pd
import statsmodels.api as sm
from scipy.stats import shapiro
import matplotlib.pyplot as plt

# Load the data from Excel
df = pd.read_excel(r'C:\Users\adiit\OneDrive\Desktop\New Microsoft Excel Worksheet.xlsx')

# List of column names
columns = ['high social dist_cheat', 'high social dist_no_cheat', 'low social dist_cheat', 'low social dist_no_cheat']

# Loop through each column
for column_name in columns:
    data = df[column_name]

    # Q-Q Plot
    sm.qqplot(data, line='s')
    plt.title(f"Q-Q Plot for '{column_name}'")
    plt.show()

    # Shapiro-Wilk Test
    shapiro_test_stat, shapiro_p_value = shapiro(data)
    print(f"Shapiro-Wilk Test p-value for '{column_name}':", shapiro_p_value)

    # Histogram
    data.hist()
    plt.title(f"Histogram for '{column_name}'")
    plt.show()
```

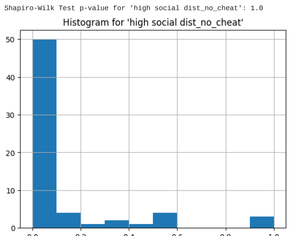


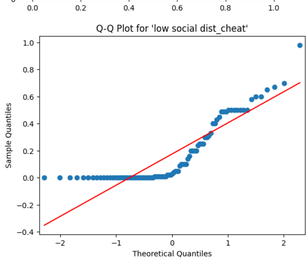

```
Shapiro-Wilk Test p-value for 'high social dist_cheat': 6.864573115532768e-16
```



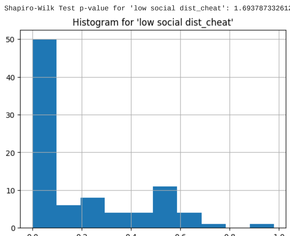


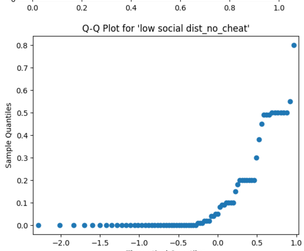

```
Shapiro-Wilk Test p-value for 'high social dist_no_cheat': 1.0
```



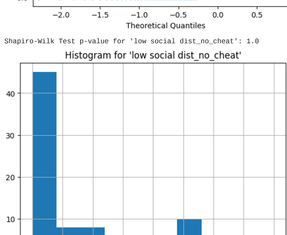



```
Shapiro-Wilk Test p-value for 'low social dist_cheat': 1.6937873326128283e-10
```






```
Shapiro-Wilk Test p-value for 'low social dist_no_cheat': 1.0
```



Analysis Results

1. Q-Q Plots

The Q-Q plots for each leg do not exhibit a straight line, indicating that the data in each leg does not follow a normal distribution. This observation aligns with the visual evidence from histograms, which show a right-skewed distribution.

2. Histograms

The histograms further support the visual evidence of non-normality. The right-skewed distribution, particularly with a majority of observations at 0%, contributes to the deviation from normality.

3. Shapiro-Wilk Test

The p-values from the Shapiro-Wilk test provide strong evidence against the null hypothesis of normality for 'high social dist_cheat' and 'low social dist_cheat.' The p-values for 'high social dist_no_cheat' and 'low social dist_no_cheat' are larger but still indicative of non-normality.

Summary

In summary, the analysis of Q-Q plots, histograms, and Shapiro-Wilk tests collectively suggests that normality assumptions are violated in each leg of the study.