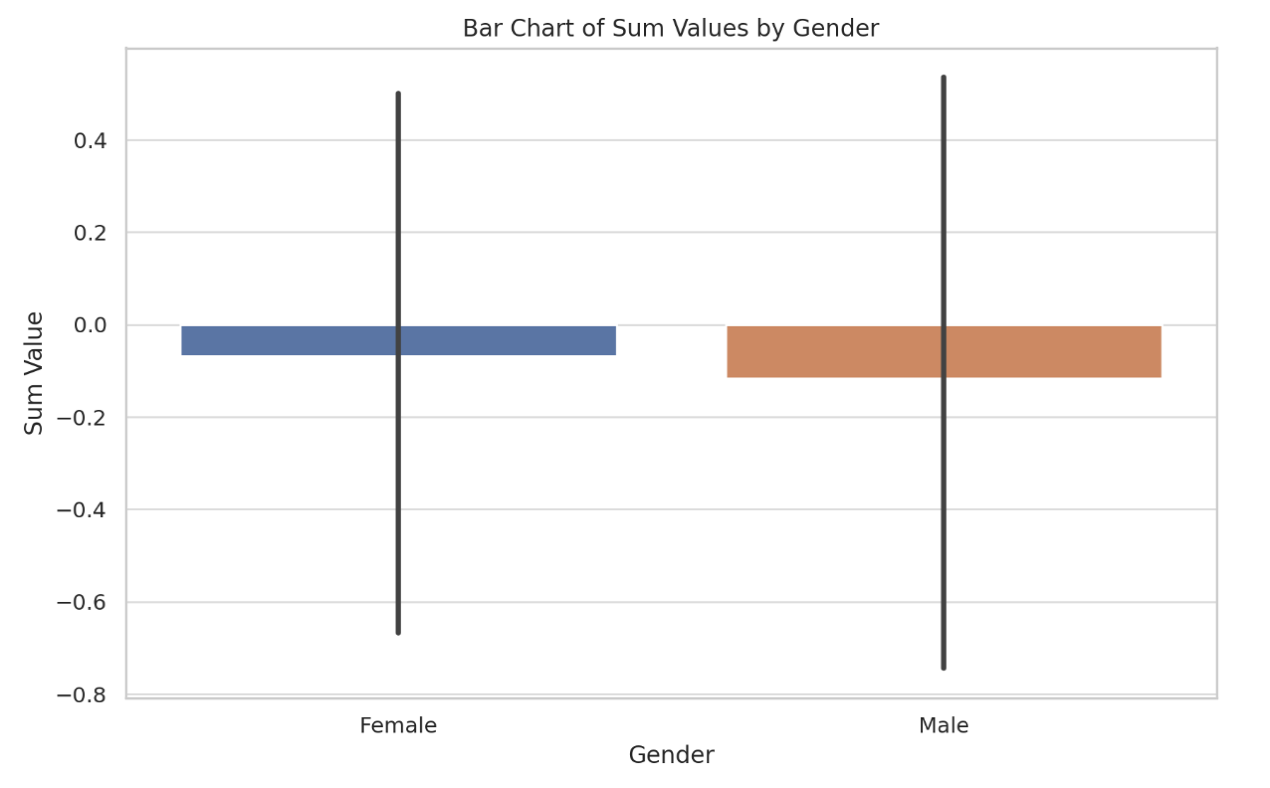
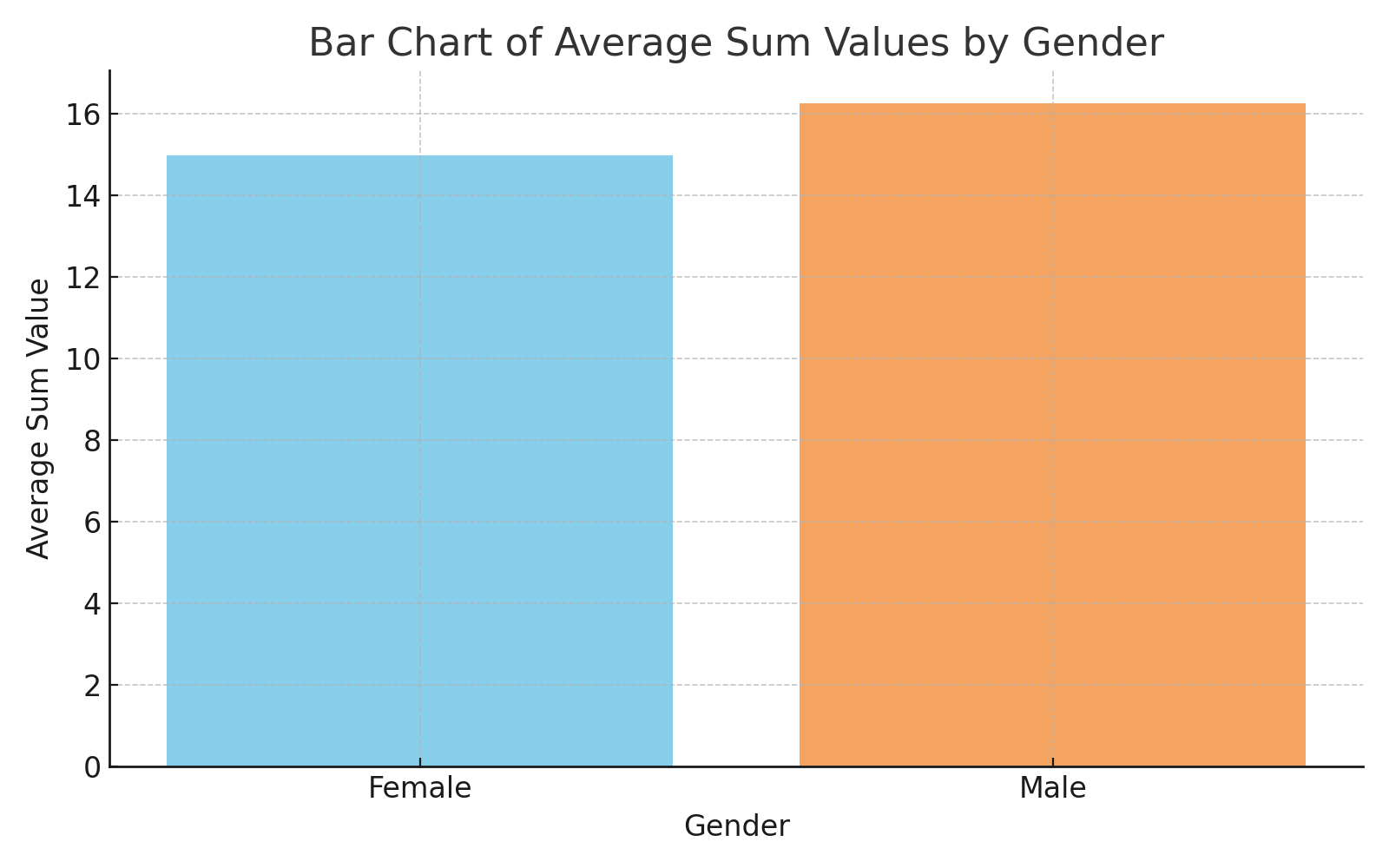
**1数据描述**

Output1:



This bar chart presents the distribution of sum values segmented by gender for a dataset containing complete entries (name, sum, and gender). The data is categorized into two groups: Male and Female. From the chart, it is observable that the sum values for both genders are closely distributed around similar ranges. The mean (average) sum value for the complete dataset is approximately -0.0959, indicating a slight negative average across the dataset. The median sum value is 0.0, suggesting that the central tendency of sum values lies at zero. This visualization effectively highlights the sum value distribution across different genders, providing a clear and concise representation of the data.

Output3



The bar chart displays the average 'sum' values categorized by gender. There are two bars representing the average 'sum' for each gender: one for females and one for males. The height of each bar corresponds to the mean value of the 'sum' data for that particular gender group within the dataset provided. The 'Female' bar is colored sky blue, while the 'Male' bar is in sandy brown, allowing for a clear visual distinction between the two categories. The x-axis of the chart is labeled 'Gender,' and the y-axis is labeled 'Average Sum Value,' indicating that the values we are observing are averages rather than individual data points. The title 'Bar Chart of Average Sum Values by Gender' succinctly conveys the content of the graph. The chart is designed for easy interpretation, with a clean layout and a straightforward color scheme that differentiates between the two groups.

Output4:

图表, 条形图

描述已自动生成

The bar chart illustrates the average 'sum' values, segregated by gender. It features two bars, one representing the average 'sum' for females and the other for males. The female category is indicated by a blue bar, while the male category is represented by an orange bar. The x-axis is labeled "Gender" to differentiate between the two groups, and the y-axis is labeled "Average Sum Value," denoting the metric being averaged. The title "Bar Chart of Average Sum Values by Gender" effectively communicates the chart's purpose. The visual is designed to be immediately understandable, with distinct colors for each gender and an uncluttered presentation that focuses on the average values.

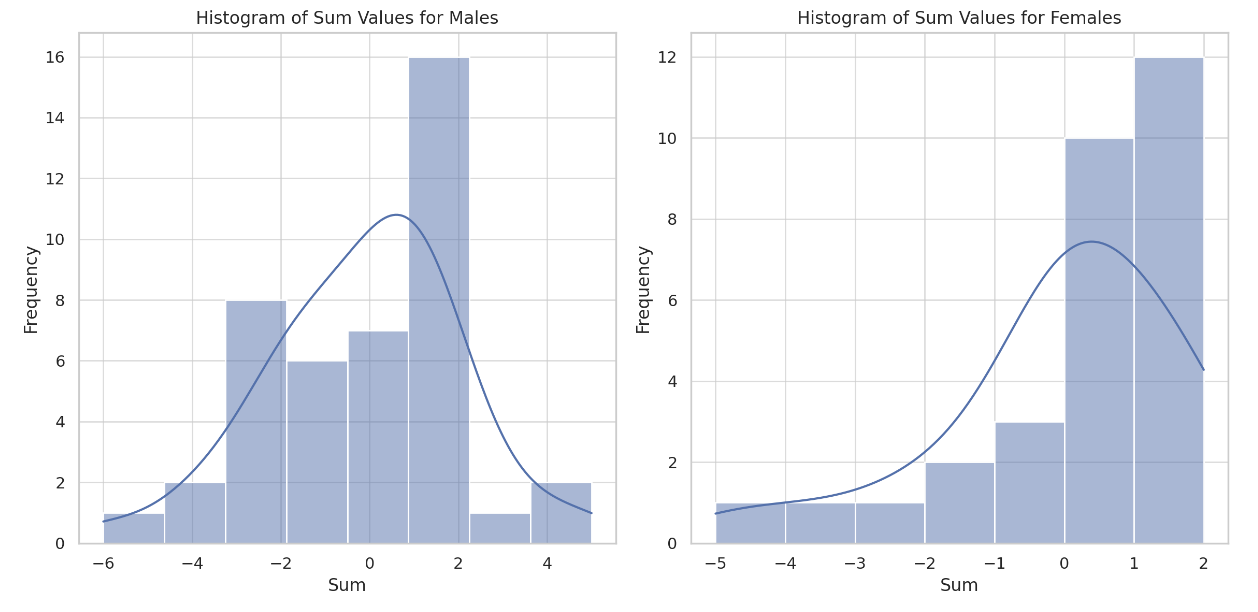
**2检验是否与性别相关过程-方法**

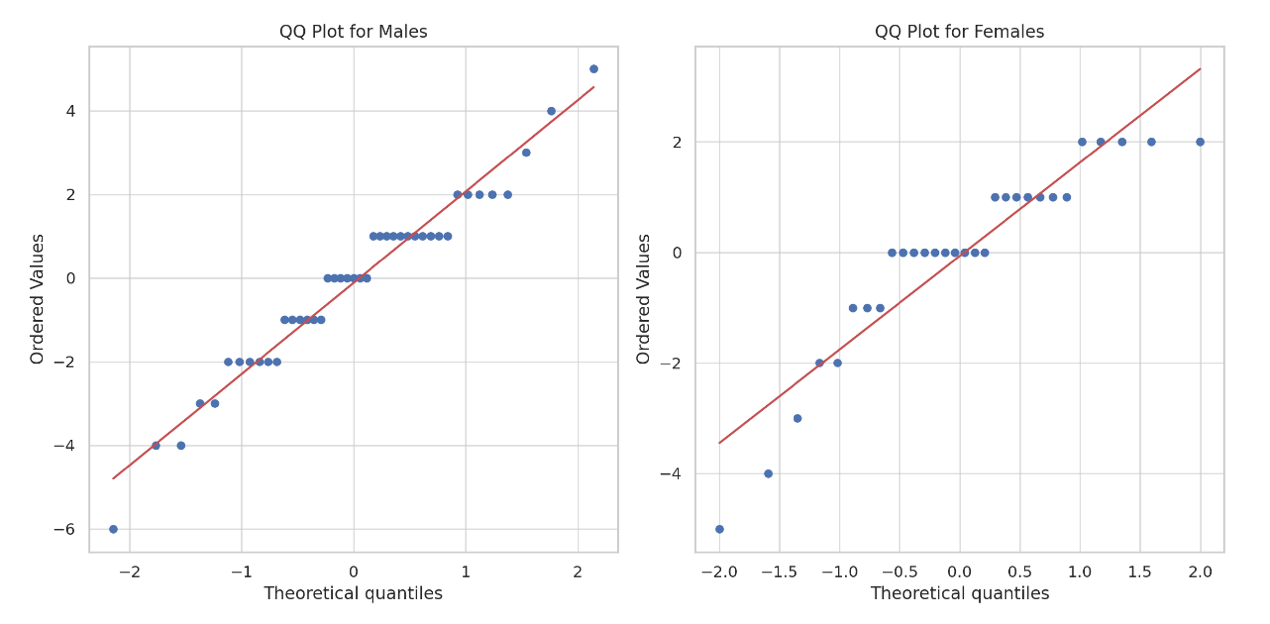
To analyze the relationship between gender (a binary categorical variable) and "sum" values (a continuous variable), an initial consideration would be to employ the Independent Samples t-test. This method is appropriate when comparing the means of two independent groups (in this case, males and females) on a continuous outcome. However, before applying the t-test, check for normality in the distribution of "sum" values within each gender group. The t-test assumes that the data are approximately normally distributed. Finally, if the normality assumption holds, perform the Independent Samples t-test to compare the means of the "sum" values between male and female groups. This test will determine if there is a statistically significant difference in the mean "sum" values across the two genders.

However, if it's found that the "sum" values significantly deviate from a normal distribution, which violates an assumption of the t-test, a non-parametric alternative becomes necessary. In such cases, It should employ a Non-Parametric Test. The Mann-Whitney U test is the non-parametric equivalent of the Independent Samples t-test. It does not assume normality in the data and is used to compare the distributions of "sum" values between the two gender groups. This test is based on the ranks of the data rather than the actual values and is suitable for data that are not normally distributed.

**3.结果**

**Output1的结果**：





The histograms and QQ plots above provide a visual assessment of the normality of the "sum" values for each gender group:

1. \*\*Histograms\*\*:

- The histogram for males shows the frequency distribution of their "sum" values. The shape of the distribution can be compared against a normal distribution (overlayed as a kernel density estimate).

- Similarly, the histogram for females displays their "sum" values distribution.

2. \*\*QQ Plots\*\*:

- The QQ plot for males compares the quantiles of their "sum" values against the quantiles of a normal distribution. If the points fall approximately along the straight line, it suggests normality.

- The QQ plot for females serves the same purpose, providing a visual indication of how closely the "sum" values follow a normal distribution.

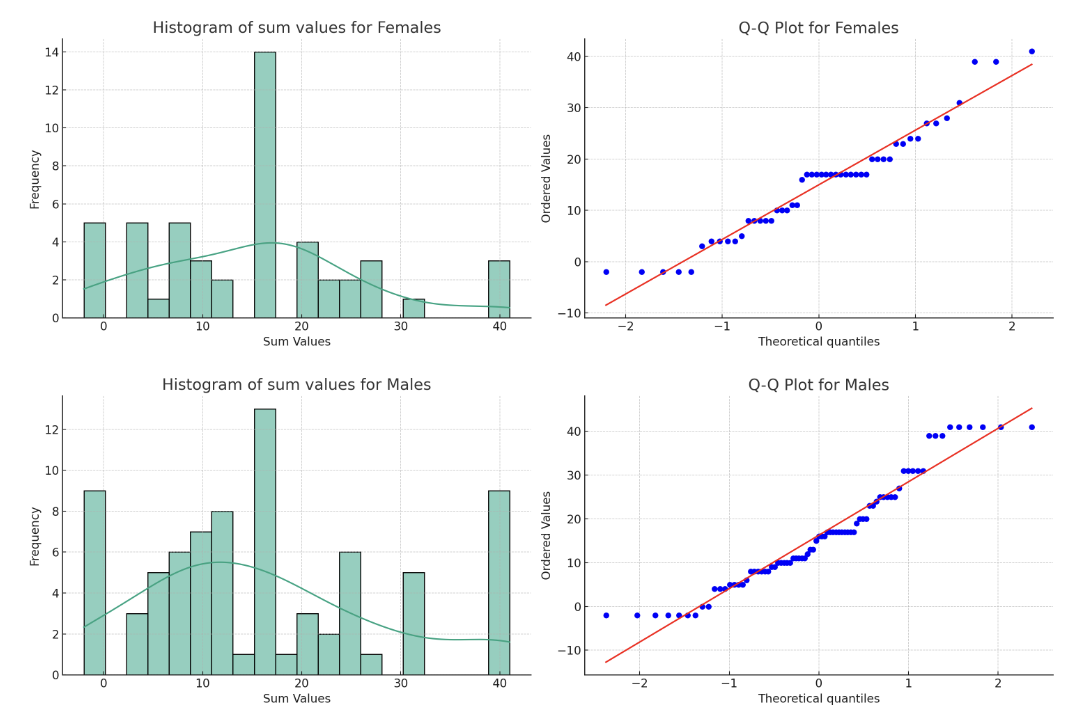
These visual methods are complemented by statistical tests for normality, such as the Kolmogorov-Smirnov test, to rigorously determine if the "sum" values for each gender conform to a normal distribution. If the data deviate significantly from normality, non-parametric tests like the Mann-Whitney U test are recommended for further analysis.

**最终采用Mann-Whitney U test方法判断没有相关性**

The Mann-Whitney U test was conducted to examine the relationship between gender and "sum" values. The result of the test yielded a U statistic of 626.0 and a p-value of approximately 0.833.

The p-value is significantly greater than the typical alpha level of 0.05, which suggests that there is no statistically significant difference in the distribution of "sum" values between male and female groups. In other words, based on this test, gender does not appear to be significantly related to the "sum" values in your dataset.

Output3:



The visual and statistical assessments provide insight into the normality of the 'sum' values distribution for both genders:

\*\*Visual Analysis:\*\*

1. The histograms for both females and males show the distribution of 'sum' values with a superimposed kernel density estimation (KDE) to estimate the probability density function.

2. The Q-Q (quantile-quantile) plots compare the quantiles of the 'sum' values with the quantiles of a normal distribution. A linear pattern in the points suggests that the data may follow a normal distribution.

\*\*Statistical Test:\*\*

The Kolmogorov-Smirnov test was performed for each gender group, comparing their 'sum' values to a normal distribution with the same mean and standard deviation as the observed data.

- For females, the K-S statistic is approximately 0.136 with a p-value of about 0.290.

- For males, the K-S statistic is approximately 0.134 with a p-value of about 0.105.

The histograms and Q-Q plots suggest that neither group perfectly follows a normal distribution; however, they do not show significant deviation from normality. The Q-Q plots do not display a perfectly straight line which would indicate normality, but they are not highly deviated either.

The Kolmogorov-Smirnov test results show p-values higher than the typical alpha level of 0.05, meaning that we do not reject the null hypothesis of normality for either group based on this test.

In conclusion, while the visual indicators suggest that the data is not perfectly normally distributed, the statistical test does not provide sufficient evidence to reject the normality of the 'sum' values for either gender. Therefore, the distribution of 'sum' values for both males and females could be considered approximately normal based on the K-S test results.

**最终采用Mann-Whitney U test方法判断没有相关性**

The Mann-Whitney U test yields a U statistic of approximately 1911 and a p-value of about 0.758. The p-value is well above the conventional alpha level of 0.05, indicating that there is no statistically significant difference in the 'sum' values distribution between genders based on this test. Therefore, we do not have evidence to suggest that 'sum' and 'gender' are correlated.

Output4:

图表

描述已自动生成

For the new dataset, the visual and statistical analysis of the 'sum' values by gender provides the following insights:

\*\*Visual Analysis:\*\*

- The histograms for both females and males display the frequency of 'sum' values, with the kernel density estimation (KDE) suggesting the shape of the distribution.

- The Q-Q plots show how the 'sum' values for each gender compare to a theoretical normal distribution. Deviations from the line in the Q-Q plots indicate departures from normality.

\*\*Statistical Test:\*\*

The Kolmogorov-Smirnov (K-S) test compares the 'sum' values to a normal distribution with the same mean and standard deviation.

- For females, the K-S statistic is approximately 0.171 with a p-value of about 0.038.

- For males, the K-S statistic is approximately 0.187 with a p-value of about 0.00064.

The histograms suggest that the distributions of 'sum' values for both genders have deviations from a normal distribution, as indicated by the shape of the KDE curves. The Q-Q plots for both genders also show some deviation from the expected straight line for a normal distribution, particularly in the tails.

The K-S test results for both genders yield p-values below the standard alpha level of 0.05, suggesting that the distributions of 'sum' values for both females and males are not normally distributed.

In summary, based on the K-S test, there is evidence to reject the null hypothesis of normality for the 'sum' values for both genders in the new dataset. The distributions are likely not normal for either group.

用Mann-Whitney U test检测是否sum和gender是否有相关性

The Mann-Whitney U test for the new dataset yields a U statistic of approximately 3464 with a p-value of about 0.478. This p-value is above the conventional alpha level of 0.05, suggesting that there is no statistically significant difference in the distribution of 'sum' values between the female and male groups. Therefore, the test does not provide evidence of a correlation between 'sum' and 'gender' for the given data.

27-Group1: shiqinruixu; haoli; youzhou

8-Group2: guiyuchen; jianqili; xiaohanlv; zetenglin; jiahuishi; ziyangwu

17-Group 3: hanyusi; yiningpeng; zuyuanwang; zihaozhu; xiangli

10-Group 4:xinyuma; chengyangli; haoqizhang; ziruzhang; xulingzhang; yijiexu; guangzhizhu; jiarongma

13-Group 5:yuchaozhuo; yuyangzhang; mingdashi; chenyusun

5-Group 6: jiayanghuang; jurongsong; yulinshen; shihanfu; yuxiwang

11-Group 9: xinshu; mengsun; pinxizhu; zhenwei; gahe; lanluo; chaozhezhang

19-Group 11: jieweili; haiyuli

25-Group 17: yanbodai; junzheqin; mengqichen; jinliu; mingzheliu; chenqingzhu

4-Group 23: runlinchen; yaocheng; tianxiangwang; yishuli; ruoshuipeng; rongzuo;yanfeizhou

10-Group 27: jingyu; yihancai; xiaolinchen; zeyuxiong; qiyingzhao; jianxingshi; jiaxinwu

-2-Group 28: danyunxiao; lingjin; jiachenlei; wenjingfang; xingruiliu; yiyuliu; yuchengliu; zhenlinzhong

8-Group 30: minjieliu; yuchengchen; zhaoyingchen; zipengwang; jingyipan; guangyuyan; guoqiangliang

16-Group 36: mzheng; wzhang; ygao; czhu; hqu; jli; wxu

17-Group 40: kaiwenliu; hongqiuwang; jianchen; xinxu

17-Group 44: nanzhu; zeyinggong

39-Group 47: yuzhuozhang; mengfanli; cehaoyang; hongtaowu; jinghuiye; lekangren

-2-Group 50: changhuang; yuruhuang; zhizhuokou; ziangzhao; mengyideng; yannanzhou; guoyangzhao

图表, 条形图

描述已自动生成

The bar chart has been created based on the groups specified. It shows the count of females (in pink) and males (in blue) for each group. The total 'sum' values are also indicated above the corresponding bars for each group.

The Mann-Whitney U test was performed to check if there is a statistically significant difference between the 'sum' values multiplied by the count of females and males across the groups. The test statistic is approximately 21.0 and the p-value is approximately 0.710.

A p-value higher than the common alpha level of 0.05 suggests that there is no statistically significant difference between the 'sum' values for the different gender counts within the groups. This implies that there is no evidence of a correlation between 'sum' values and gender quantity based on the groups provided.