**第1页**

Good afternoon, everyone. We are group 6.

**We propose to explore the differences between Li Bai's and Du Fu's Poems in "Three Hundred Tang Poems".**

**第2页**

**We will talk about our research topics from the following aspects.**

**第3页**

**Let me start with a brief introduction.**

**第4页**

We all know that the poetry culture has a long history, and the Tang Dynasty is the prosperous era of poetry rising. Here, poetry is popular and widely spread, and countless famous great poets have come out. Li Bai and Du Fu are the two bright pearls among the stars of the Tang poetry circle, attracting countless future generations to explore.

On this basis, we want to make full use of the statistical knowledge we have learned in this semester to study literature and poetry, and prove that statistics is useful in many aspects.

We applied descriptive statistics and statistical inference to complete our research. In order to verify the length difference between poems, we choose the t test of hypothesis test. Hypothesis test, also known as statistical hypothesis test, is a statistical inference method used to determine whether the differences between samples and samples, between samples and population are caused by sampling errors or essential differences. The method of goodness of fit test is used to judge image. Goodness of fit test is one of the important contents of statistical significance test with chi square statistics. It calculates the expected frequency of each category in the classification variable according to the overall distribution, compares it with the observation frequency of the distribution, and judges whether there is significant difference between the expected frequency and the observation frequency, so as to achieve the purpose of analyzing from the classification variable. **These scientific statistical methods make the results more reliable and convincing.**

**第5页**

Thank you, kong deran. Kong deran has just introduced abstract and introduction to us. Now I will introduce the specific contents of materials and method.

About the source of the material, in order to ensure the accuracy and consistency of the data, the poems we choose are mainly from three hundred tang poems.

**Let's take a look at the calculation we used.**

**第6页**

The first is about comparing the length of the poems written by the two great poets.

As for judging standard, we use the number of words in poems to determine whose poem is longer.

Li bai's 25 poems and du fu's 25 poems were selected as samples from “Three hundred poems of tang dynasty”. After recording and calculating the average number of words and the variance of the number of words in 25 poems by li bai and du fu, the double-sample hypothesis test method was used for data processing. **Under the condition of significance level α= 0.05.**

**第7页**

When counting the length of these poems and the variance of the word count, we use a program written by python to do the statistics. **The specific program is as follows.**

**第8页**

The above is the specific method to compare the length of poems.

The following is : "are the poetic images of li bai and du fu the same in the three hundred poems of tang dynasty?" This problem is specific to the steps we use.

In the selection of data materials: of course, in order to ensure the consistency of the data, we selected 25 poems of li bai from 300 poems of tang dynasty as samples and 35 poems of du fu as the population.

In the process of data processing, we first programmatically extract the images contained in the selected poems and calculate their word frequency, and then classify them. In our data processing, there are five types of images: pride and ambition, sadness and melancholy, nobility and loyalty, separation and pain, grace and leisure.

For the five types of images, we take the image word frequency of li bai after classification as a sample, and the image proportion of du fu as a whole. **Under the assumption of 0.05 significance level, Chi-square Goodness of Fit tests were used to explore whether there are differences in the use of imagery between the two.**

**第9页**

**Here, of course, are the python programs used to count word frequency.**

**第10页**

**Next, I will talk about the results and discussion of the group project.**

**第11页**

Our team members used Hypothesis testing about two Population and chi-square tests to **study the difference between the length and image of Li Bai's poems and Du Fu's poems in "Three Hundred Tang Poems".**

**第12页**

The first part is about Hypothesis testing about two Population.

In order to find out the length relationship between Li Bai's poems and Du Fu's poems, at a significance level of 0.05, we are going to conduct a hypothesis testing based on two population.

We set Null hypothesis: Libai’s poems don’t differ from Dufu’s poems in length. Through our calculation of sample variance, we use the unequal variance procedure. The calculation process is shown in the PPT, and finally we conclude that:

**we do not reject Null hypothesis in favor of Alternative hypothesis at a significance level of 0.05. That is, we don’t have enough evidence to say that Libai’s poems differ from Dufu’s poems in length.**

**第13页**

The second part is Chi-square Goodness of Fit Test.

At the 0.05 significance level, can we conclude that the image content distribution of Libai is different from Dufu in "Three Hundred Tang Poems"?

Null hypothesis is that The image distribution is the same of both poets.

**Then we draw histograms and pie charts based on the number of images in Libai’s poems and Dufu's poems. From the chart, we can clearly obtain the image distribution of the poems.**

**第14页**

The calculation process is shown in the PPT, and finally we conclude that: In conclusion, **it cannot reject the null hypothesis, the sample shows no difference between the image content distribution of two poets**.

**第15页**

Next is Discussion. Combined with project goals, From the previous results, we can see that our team members do not have enough evidence that **Li Bai's poem length is different from Du Fu's poem length and not a difference between the two poets' image distribution.**

**第16页**

Then we comment on the project and improve it: our group found that Libai had a poem with a length of 1114. This value is an obvious outlier. Pay attention to the occurrence of outliers and analyzing the causes of them often become an opportunity to find problems and improve decision-making.

**Then we saw from "The Box Display": There are two outliers in Libai's poems, which are 369 and 361, and one out of Dufu's poem is 1114.**

**第17页**

Then we summarized the characteristics of "The Box Display" and put suggestions. When studying the length of Li Bai and Du Fu, **our team can first calculate the outliers in the two sets of data, and then consider whether to delete the outliers.**

**第18页**

Chen ziqi has introduced the two parts of results and discussion to us. Now I would like to summarize the conclusion for you.

In our exploration of the similarities and differences between the poems of li bai and du fu in the three hundred poems of tang dynasty. **Mainly used two methods, two-sample hypothesis test and chi-square goodness of fit test.**

**第19页**

Double sample hypothesis test on the length of two great poet li bai and du fu's poetry played a key role on contrast, after collecting materials, numerical extraction, after calculation, statistics and a series of steps, getting the result: in the case of alpha = 0.05, there is not enough evidence to prove the started the first length of li bai and du fu's poems are different.

Chi-square test of goodness of fit is mainly in detecting image content distribution and li bai, du fu in the tang started for different use, after data collection, image classification, image extraction, word frequency statistics, **calculation and a series of steps, in the case of alpha = 0.05, concluding that the two poets the conclusion of the image content distribution of no difference.**

**第20页**

Du fu and li bai were both representative poets of the tang dynasty. They stood on two peaks of poetry in the tang dynasty, one was called realism, the other was called romanticism.

Which of them is the wiser or the better can’t be said, because both of them are geniuses. To understand their differences in poetry, big data statistics are needed.

This time, the writing habits of the two poets were statistically analyzed and calculated in terms of the length and image of the poems, so as to obtain the data of their "writing preference", so as to better understand the writing habits of the two poets and their era.

Li bai represents the youth of the tang dynasty, while du fu represents the maturity of the tang dynasty. In Yu qiuyu's Chinese context, it is mentioned that students at Peking University ranked their favorite poets, among which li bai was the first and du fu was the second.

In the eyes of most people, the length of poetry and the use of imagery are undoubtedly the two important factors affecting the choice, so this is the breakthrough of the two studies.

**Surprisingly, according to the calculated data, the two great poets were almost identical in both respects.**

**第21页**

**This strongly suggests that the factors that influence people's preference for these two poets are not the length of the poems or the use of imagery, but other factors.**