Chapter 4

Data Analysis Results

Independent Study title was: The Impact of Online Word- of- Mouth from Multi-Platform on Product Sale Research objectives were: (1) To research the comprehensive impact of various dimensions of online word-of-mouth on a single e-commerce platform; (2) To research on the impact of integrating social media platforms' online word-of-mouth on product sales; (3) To research and identify the most important online word-of-mouth factor that affects product sales. The main research method of this thesis is quantitative method. This research uses crawler technology to obtain one month's comment data for 14 skincare products from JD e-commerce platform and third-party social media platform Xiaohongshu. The number of total effective comments is 20578. By gathering statistics on daily comments for every product, we can get 420 valid sample data. This chapter mainly carries out statistical analyses and hypothesis tests on the obtained effective samples. The statistical analysis is mainly divided into three parts, including descriptive statistical analysis, pearson correlation analysis and regression analysis. Finally, the results of regression analysis are used to conduct hypothesis testing.

Descriptive Statistical Analysis Results

Descriptive statistical analysis is a method used to describe and analyze data, which can help us understand the distribution patterns and trends in data, and better understand the relationships and impacts between data. Descriptive statistical analysis mainly includes the following aspects: first, collect data, and then perform preliminary processing, such as conducting normality tests, analysis of variance, etc. Secondly, classify the data, for example, divide the data into different types, such as normal test, non-normal distribution, variance analysis, etc. Finally, conduct statistical analysis on the data, such as descriptive statistical analysis, descriptive statistical

inference, etc. Through these methods, we can better understand the distribution patterns and trends in data.

The research sample in this chapter is the data of experiential products (facial essence in skincare products) from JD Mall and Xiaohongshu. Research variables include product sales, star rating, number of pictures, and number of social media platform comments. This chapter first carries out descriptive statistical analysis on these research variables, sorts out and counts the scattered data captured on the website through descriptive statistical analysis, and finally presents the distribution laws of statistical dispersion and central tendency of each variable. The contents of descriptive statistical analysis mainly included the maximum and minimum values, mean values, and standard deviations of the data. The specific results of descriptive statistical analysis are shown in table 4.1 and table 4.2.

Table 4.1 Descriptive Statistics Analysis on JD

	N	Minimum	Maximum	Mean	Std. Deviation
Num of JD's Comments	420	1.00	149.00	26.2959	23.08164
Num of JD's Image	420	0.00	116.00	21.0136	22.01581
Star rating of JD	420	3.67	5.00	4.8946	0.17118
Sales volume of products	420	25.00	330.00	138.5986	80.00129
Valid N (listwise)	420				

By observing the data in **Table 4.1**, we can conclude that the minimum number of product comments on the JD platform is 1, and the maximum number is 149. Each product receives about 26.29 comments on average, with a standard deviation of 23.08. This indicates that there is a certain degree of difference in the number of product comments for different products on the JD platform, some products receive relatively more user discussions while others have fewer comments.

The minimum number of product images is 0, and the maximum number is 116. Each product has about 21.01 images on average, with a standard deviation of 22.02. Similar to the number of comments, there is also a certain difference in the number of images of the product. Some products may have more images, which may be related to the product's characteristics and sales strategy. The star rating data range is between 3.67 and 5.00, with an average star rating of approximately 4.89 and a standard deviation of 0.17. This indicates that most products have received high star ratings on the JD platform, and most consumers may tend to give positive comments which is in line with consumers' rating habits. The product sales data in the dataset ranges from 25 to 330, with an average sales volume of approximately 138.60 per product and a standard deviation of 80.00. The standard deviation of sales is relatively large, indicating a significant difference in product sales. Some products may have higher sales, while others may have lower sales.

Table 4.2 Descriptive Statistics Analysis Integrating XHS

	N	Minimum	Maximum	Mean	Std. Deviation
Num of JD's Comments	420	1.00	149.00	26.2959	23.08164
Num of JD's Image	420	0.00	116.00	21.0136	22.01581
Star rating of JD	420	3.67	5.00	4.8946	0.17118
Num of XHS's Comments	420	1.00	107.00	23.5034	17.63427
Sales volume of products	420	25.00	330.00	138.5986	80.00129
Valid N (listwise)	420				

By looking at the data in **Table4.2**, we can draw the following conclusions: on the JD e-commerce platform, the number of reviews, the number of pictures and the star ratings are consistent with the data from the fusion third-party platform. However, the minimum value of the number of product comments on the third-party social media platform Xiaohongshu is 1 and the maximum value is 107, which indicates a wide variation in the degree of product discussion involved on this platform. This wide range of distribution may reflect differences in the popularity of different products among users as well as differences in the level of attention. Each product receives about 23.5 reviews on average in Xiaohongshu. This value can be regarded as a reference indicator of how much attention a product receives. A high average value indicates that most products are discussed by a considerable number of users, which is significant for brand communication, marketing and consumer. The standard deviation is 17.6, which indicates the degree of dispersion of the number of reviews relative to the mean. A large standard deviation may indicate that the distribution of the number of reviews is more dispersed, with a large difference in the level of discussion between products. This may be due to the fact that some products attract more attention in a particular period or under a particular event, while others receive relatively less attention. The maximum value of the number of comments is 107, some products have attracted considerable discussion in Xiaohongshu. This shows that Xiaohongshu as a social media platform have influence in brand communication and product promotion.

Correlation Analysis Results

Correlation analysis, as an important means of data analysis, can be used to explore the interrelationships and interactions between two variables. It can be used to determine the correlation between different variables and their impact on the results. When conducting relevant analysis, a series of steps are required, including data collection, data organization and analysis, and establishing a model. When collecting data, it is necessary to ensure its reliability and accuracy, and perform necessary screening and cleaning to ensure the reliability of the collected results.

Before regression analysis of variables, Pearson correlation analysis can be used to preliminarily investigate whether there is correlation between their respective variables, dependent and independent variables. The Pearson correlation analysis results can only determine whether the variables are closely correlated, but not reflect the causal relationship between the two variables. In general, the larger the absolute value of the correlation coefficient, the stronger the correlation. The correlation is stronger when the correlation coefficient is around 1 or -1, and weaker when it is around 0. Regarding the significance P-value, if P>0.05, it means that there is no significant correlation between the two variables. At the end of the analysis, there is no need to analyze the Pearson correlation coefficient again. If P<0.05, it means that there is a significant correlation between the two variables, and further investigation is needed to determine the magnitude of the correlation. The correlation coefficient can be used to determine the correlation strength of variables through the following range of values. The correlation coefficient is highly correlated from 0.8 to 1.0; Strong correlation is between 0.6 and 0.8; Moderate correlation is between 0.4 and 0.6: weak correlation is between 0.2 and 0.4; Very weak or no correlation is between 0.0 and 0.2.

The Pearson correlation coefficient calculation formula is as follows:

$$r = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum(X - \overline{X})^2 \sum(Y - \overline{Y})^2}}$$

Where X and Y are the values of two variables, $\overline{X}$ and $\overline{Y}$ are the average values of the two variables, respectively.

Correlation analysis results of a single e-commerce platform JD as shown in table 4.3.

Table 4.3 correlation analysis on JD

		Num of JD's Comments	Num of JD's Image	Star rating of JD	Sales volume of products
Num of JD's Comments	Pearson Correlation	1	.891**	-0.065	.831**
	Sig. (2-tailed)		0.000	0.266	0.000
	N	420	420	420	420
Num of JD's Image	Pearson Correlation	.891**	1	0.022	.790**
	Sig. (2-tailed)	0.000		0.712	0.000
	N	420	420	420	420
Star rating of JD	Pearson Correlation	-0.065	0.022	1	-0.007
	Sig. (2-tailed)	0.266	0.712		0.910
	N	294	294	294	294
Sales volume of products	Pearson Correlation	.831**	.790**	-0.007	1
	Sig. (2-tailed)	0.000	0.000	0.910	
	N	420	420	420	420
**. Correlation is significant at the 0.01 level (2-tailed).					

From **Table 4.3**, it can be seen that after performing Pearson correlation analysis, a significant correlation ($p < 0.05$) is found between the number of reviews, the number of picture reviews and sales volume of JD Mall.

Further analysis shows that the correlation coefficient between the number of comments on JD Mall and sales is 0.83, which is a significant positive correlation. This means that as the number of comments increases, sales also increase, and the two variables show a certain degree of positive correlation. In addition, the value of the correlation coefficient is 0.83, indicating a strong correlation between the number of comments and sales.

The correlation coefficient between image comments and sales is 0.79, which is a significant positive correlation. This means that as the number of image

comments increases, sales also increase, and the two variables show a certain degree of positive correlation. In addition, the value of the correlation coefficient is 0.79, indicating a strong correlation between the number of comments and sales.

There is no significant correlation between star rating and sales volume due to P>0.05

Correlation analysis results of integrated social media platform Xiaohongshu as shown in table 4.4.

Table 4.4 correlation analysis Integrating XHS

		Num of JD's Comments	Num of JD's Image	Star rating of JD	Num of XHS's Comments	Sales volume of products
Num of JD's Comments	Pearson Correlation	1	.891**	-0.065	-0.085	.831**
	Sig. (2-tailed)		0.000	0.266	0.145	0.000
	N	420	420	420	420	420
Num of JD's Image	Pearson Correlation	.891**	1	0.022	0.021	.790**
	Sig. (2-tailed)	0.000		0.712	0.723	0.000
	N	420	420	420	420	420
Star rating of JD	Pearson Correlation	-0.065	0.022	1	0.007	-0.007
	Sig. (2-tailed)	0.266	0.712		0.908	0.910
	N	420	420	420	420	420
Num of XHS's Comments	Pearson Correlation	-0.085	0.021	0.007	1	.129*
	Sig. (2-tailed)	0.145	0.723	0.908		0.026
	N	420	420	420	420	420
Sales volume of products	Pearson Correlation	.831**	.790**	-0.007	.129*	1

	Sig. (2-tailed)	0.000	0.000	0.910	0.026	
	N	420	420	420	420	420
**. Correlation is significant at the 0.01 level (2-tailed).						
*. Correlation is significant at the 0.05 level (2-tailed).						

As can be seen in **Table 4.4**, after Pearson correlation analysis, a significant correlation is found between the number of JD Mall reviews, the number of picture reviews, Xiaohongshu reviews and sales (P < 0.05), while no significant correlation exists between star ratings and sales (P > 0.05). The correlation coefficients of the total number of JD Mall reviews and picture reviews are consistent with the results of the single e-commerce platform analysis.

The correlation coefficient between Xiaohongshu and sales is 0.13, and this correlation coefficient is a weak correlation. This means that as the number of Xiaohongshu comments increases, the sales volume also increases, and the two variables show a certain degree of positive correlation. In addition, the value of the correlation coefficient is 0.13, which indicates that the degree of positive correlation between the number of comments and sales is weak correlation. Even though they present a positive relationship, the strength of the correlation is low and needs to be further verified.

Regressive Analysis Results

Regression analysis is a statistical analysis method that can be used to determine the relationship between two variables and determine their linear relationship. By using this method, we can establish a model that can perform numerical predictions on other variables. When conducting multiple linear regression, appropriate independent variables should be selected, and relevant mathematical models should be constructed to provide a reasonable description. In linear regression analysis, testing the results is a very important step, which mainly includes the following indicators: 1. goodness of fit test (R-square test): test the model's goodness of fit through multiple determinable coefficients and modified determinable coefficients, verify the explanatory degree of the independent variable

70

to the dependent variable. 2. Significance test of regression equation (F value and its significance level Sig.): Directly test the overall significance of the regression equation from the regression effect. The F-test is used to judge whether the regression model is significant as a whole, that is, whether the combined influence of independent variables on dependent variables is significant. The Sig. value (significance level) represents the critical probability of rejecting the original hypothesis. 3. Significance test of regression parameters (T-test): it mainly verifies the individual effects of independent variables on dependent variables. The T-test is used to test whether the regression coefficient is significantly not equal to zero. If the P-value of the T-test is less than the set significance level, it can be considered that the independent variable has a significant impact on the dependent variable. 4. VIF collinearity test: VIF (variance expansion factor) is used to test whether there is Multicollinearity between independent variables. The larger the VIF value, the more severe the collinearity between independent variables. 5. Residual diagnosis: used to evaluate the suitability of the regression model and check whether the model meets its assumptions.

1. Goodness of fit test (R-square test)

In regression analysis, the R-squared test is a method used to measure the fit between the regression model and the dependent variable. The R square reflects the proportion of the regression model in explaining the variation of the dependent variable, is the degree to regression model which explains the variation of the dependent variable. The R value ranges between 0 and 1. When the R square is equal to 0, it indicates that the regression model has no explanatory power for the variation of the dependent variable, that is, the model has no fitting effect; When the R square is equal to 1, it indicates that the regression model fully explains the variation of the dependent variable, the model fitting effect is very good. Generally speaking, the more closer to 1 the R-square is, the better the regression model fits the data, that can better explain the variation of the dependent variable, and its predictive ability is also stronger. However, a very high R-squared value does not necessarily mean that the model is necessarily the best, because overfitting may lead to the model being too complex and not suitable for future predictions. In

summary, the goodness of fit test (R-squared test) is an important evaluation indicator in regression analysis, used to measure the degree of fit of the regression model to the dependent variable, which means that the model can explain the proportion of variation in the dependent variable, provides a basis for evaluating the effectiveness and predictive ability of the model.

        1.1 Goodness-of-fit test (R-square test): result of a single e-commerce platform (JD)

Table 4.5 Model goodness of fit on JD

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.838[a]	0.703	0.700	43.84348
a. Predictors: (Constant), Star rating of JD, Num of JD's Image, Num of JD's Comments				
b. Dependent Variable: Sales volume of products				

From the results in **Table4.5,** the following conclusions can be drawn: R^2 (Coefficient of determination) = 0.7, i.e., the fit of the model is 70%,. It means that the independent variables of the model (number of comments, number of picture reviews, star ratings) explain the changes in product sales relatively well. In this model, the number of comments, the number of picture comments, and the star ratings are the key factors in explaining the product sales. They play an important role in the model and have a significant impact on consumers' willingness to buy again. Based on the fitting results of the model, we can infer that these factors play an important role for the product sale. Since R^2 is not equal to 100%, this means that there is still 30% of the variation in product sales that is not explained by the model. This part of unexplained variation may be affected by other factors such as comment depth, comment validity, preference differences, price, etc. The standard estimation error is 43.8, it indicates the accuracy of the prediction model, and a

smaller standard estimation error usually means that the model's prediction is more accurate. In summary, these results are an important guide for developing marketing strategies, improving service quality and increasing customer loyalty. It is also necessary to continue to study other potential influencing factors to improve the explanation and prediction of product sales.

     1.2 Goodness-of-fit test (R-square test) result that integrated social media platform (Xiaohongshu)

Table 4.6 Model goodness of fit Integrating XHS

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.858[a]	0.736	0.733	41.37400
a. Predictors: (Constant), Num of XHS's Comments, Star rating of JD, Num of JD's Image, Num of JD's Comments				
b. Dependent Variable: Sales volume of products				

From the results in **Table 4.6**, the following conclusions can be drawn: when the data from the social media platform Xiaohongshu are integrated, the R-squared increases, and R^2 value ((Coefficient of determination) is 0.736, it indicates that the model's independent variables (number of comments, number of picture comments, star ratings, and number of comments on Xiaohongshu) explain the variations in the product sales relatively well. The increase in the R-squared implies that the model is more capable of explaining the product sales. Specifically, the newly added Xiaohongshu data in the model can further explain the variation in product sales, make the model more accurate in predicting actual sales. The standard estimation error is 41.4, smaller than before. The standard estimation error measures the average error between the model's predicted value and the actual observed value, and a smaller standard estimation error indicates that the model's prediction is more accurate. Therefore, the prediction accuracy of the model is improved after fusing the Xiaohongshu data.

2.   Significance test of regression equation

The significance test of regression equations is a statistical test used to test the global significance of regression models. It is used to determine whether the joint effect of the independent variable on the dependent variable is significant, that is, whether the regression model has statistical significance in interpreting the data. In the testing of regression equations, a regression model containing multiple independent variables was first established, and then the model was fitted using methods such as least squares. Then, compare the fitting effect of the regression model with the simple mean to determine the significance of the overall regression model.

2.1   Significance test of regression equation result of a single e-commerce platform (JD

Table 4.7 Analysis of variance on JD

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1317807.863	3	439269.288	228.518	.000[b]
	Residual	557452.776	419	1922.251		
	Total	1875260.639	416			
a. Dependent Variable: Sales volume of products						
b. Predictors: (Constant), Star rating of JD, Num of JD's Image, Num of JD's Comments						

As can be seen in **Table 4.7,** the F-value of the significance test is 228.518 and the level of significance (Sig.) P<0.05, which means that the whole regression model is highly significant in general. That indicates star rating, number of images comments and number of comments have a statistically significant effect on explaining the product sales. This means that these independent variables together play a very important role in the variation of product sales and we can rely on this model for prediction and analysis with great confidence.

2.2 Significance test of regression equation result of integrated social media platform (Xiaohongshu)

Table 4.8 Analysis of variance Integrating XHS

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1380548.242	4	345137.060	201.621	.000[b]
	Residual	494712.397	415	1711.808		
	Total	1875260.639	419			
a. Dependent Variable: Sales volume of products						
b. Predictors: (Constant), Num of XHS's Comments, Star rating of JD, Num of JD's Image, Num of JD's Comments						

As can be seen in **Table 4.8**, in the regression model after incorporating the number of reviews on Xiaohongshu, the overall significance test has an F value of 201.621, with a significance level (Sig.) of P<0.05, which means that the whole regression model is highly significant in general and all of them are able to explain the product sales effectively. We can rely on this model with great confidence for forecasting and analysis. However, in order to get a more comprehensive understanding of the validity and robustness of the model, further analysis of the regression parameter significance, regression coefficients, etc., as well as in-depth interpretation and application in the context of the actual business situation are still required.

3. Significance test of regression parameters (T-test)

The significance test of regression parameters is a statistical method used to determine whether the relationship between independent and dependent variables in a regression model has statistical significance. In linear regression models, we attempt to predict the value of the dependent variable through independent variables (or explanatory variables). Each independent variable is accompanied by a regression coefficient (parameter), which represents the relationship between the unit change corresponding to the dependent variable and the unit change of the

independent variable. T-test is a commonly used statistical method for evaluating the significance of regression coefficients in regression models. In regression analysis, we hope to identify that independent variables have a significant impact on the dependent variable, that is, whether the relationship between their regression coefficients and the dependent variable is significantly non zero in the population.

3.1 Significance test of regression parameters (T-test) result of a single e-commerce platform (JD)

Table 4.9 Regression coefficients on JD

Model		Unstandardized	Coefficients	Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-2.913	74.908		-0.039	0.969		
	Num of JD's Comments	2.172	0.249	0.627	8.739	0.000	0.199	5.018
	Num of JD's Image	0.839	0.260	0.231	3.226	0.001	0.200	4.999
	Star rating of JD	13.639	15.232	0.029	0.895	0.371	0.965	1.036
a. Dependent Variable: Sales volume of products								

Based on the results in **Table 4.9**, the following conclusions can be drawn: (1) The number of comments has a significant impact on sales (P=0.00 < 0.05). Furthermore, the regression coefficient is 2.17 and is positive (>0), indicating that the number of comments has a significant positive impact on sales. Specifically, when the number of comments increases by 1, corresponding sales will also increase by 2.17. This quantitative relationship tells us that increasing the number of comments is very helpful in promoting product sales. Because the regression coefficient is

positive, it indicates that the increase in comments is positively correlated with the increase in product sales. That is to say, the more comments, the more product sales may also increase. (2) The number of image comments has a significant impact on sales (P=0.00<0.05). Furthermore, the regression coefficient is 0.84 and is positive, means that the number of comments has a significant positive impact on sales. Specifically, when the number of comments increases by 1, corresponding sales will also increase by 0.84. This quantitative relationship tells us that increasing the number of comments is very helpful in promoting product sales. Because the regression coefficient is positive, it indicates that the increase in comments is positively correlated with the increase in product sales. That is to say, the more comments, the more product sales may also increase. (3) The star rating has no significant impact on sales (p=0.37>0.05).

3.2 Significance test of regression parameters (T-test) result integrated social media platform (Xiaohongshu)

Table 4.10 Regression coefficients Integrating XHS

Model		Unstandardized	Coefficients	Standardized Coefficients	t	Sig.	Collinearity	Statistics
1		B	Std. Error	Beta			Tolerance	VIF
	(Constant)	-40.975	70.968		-0.577	0.564		
	Num of JD's Comments	2.510	0.241	0.724	10.410	0.000	0.189	5.301
	Num of JD's Image	0.509	0.251	0.140	2.025	0.044	0.191	5.245
	Star rating of JD	16.922	14.384	0.036	1.176	0.240	0.964	1.038
	Num of XHS's Comments	0.853	0.141	0.188	6.054	0.000	0.946	1.057
	a. Dependent Variable: Sales volume of products							

From the results in **Table 4.10**, in the regression model after integrating the number of Xiaohongshu comments, the number of Xiaohongshu comments has a significant effect on sales (P = 0.00 < 0.05). Further, the regression coefficient is 0.85 and positive, which means that the number of reviews has a significant positive effect on sales. Specifically, when the number of reviews increases by 1, and sales increase by 0.85. The number of JD comments and image comments have a significant effect on sales (p1 < 0.05), and star ratings still do not have a significant effect on sales (p > 0.05).

   4. VIF collinearity test

   If there is multicollinearity between independent variables, the significance test of variables in regression analysis will not be meaningful when conducting regression analysis, and the estimation results of the entire regression model is inaccurate. Therefore, before regression analysis, it is necessary to test whether there is multicollinearity between variables. Generally, when the value of variance expansion factor (VIF) is greater than 0 and less than 10, it means that there is no multicollinearity between variables; When the value of variance expansion factor (VIF) is greater than or equal to 10 and less than 100, it indicates that there is a strong Multicollinearity among variables; When the value of variance expansion factor (VIF) exceeds 100, it indicates that there is a very strong multicollinearity among variables.

   4.1 VIF collinearity test: result of a single e-commerce platform (JD)

   From the results in **Table 4.9,** it is clear that the value of VIF for each variable is maximum 5.0, much smaller than 10, which indicates that there is no problem of covariance between the independent variables. Therefore, the results of the regression model are stable and reliable. We can interpret the regression coefficients with relative confidence and use the model for forecasting and decision making. The result is very important for the validity and interpretability of the regression analysis. When there is multicollinearity between the independent variables, the explanatory power of the regression model may be affected and it is difficult to determine the independent contribution of the respective variables to the dependent variable. In this case, due to the smaller VIF value and the smaller impact

of covariance between independent variables, we can more accurately assess the impact of each independent variable on the dependent variable and more reliably make relevant analyses and predictions.

4.2 VIF colinearity test: result integrated Xiaohongshu social media platform

From the results in **Table 4.10**, in the regression model after incorporating the number of reviews of Xiaohonghsu, the value of variance inflation factor (VIF) of each variable is the maximum of 5.3, which indicates that there is no covariance problem among the independent variables. Therefore, the results of the regression model are stable and reliable. We can interpret the regression coefficients with relative confidence and use the model for forecasting and decision making. However, the VIF value has slightly increased, which it may have some correlation with the existing independent variables or co-influence the dependent variable. When new independent variables are incorporated, it may have changed the relationship between the respective variables in the model, resulting in an increase in the VIF value. The new independent variables may have introduced new explanatory power and at the same time may have increased the correlation between the independent variables and affected the stability of the model.

5. **Residual diagnosis**

In regression analysis, residual refers to the difference between observed and predicted values. The purpose of regression models is to find the optimal fitting line or curve to minimize residual error. If the residual error is normal distribution, the model can better capture the random error rather than the systematic error. This improves the reliability of the model. If the residual value in the regression model is close to the normal distribution, we can be more confident in the prediction and interpretation of the model. The reason is that the residual of normal distribution indicates that the model has a good fitting effect, so the prediction results are more reliable. On the contrary, if the residual error deviates from the normal distribution, its rationality should be reconsidered, and the model should be optimized or other regression methods should be used. To sum up, an important condition in regression analysis is that the residuals of the regression model basically follow normal

distribution. If this condition is met, it means that the calculation results of the regression model are stable and reliable, the fitting degree of the data is good, and it is suitable for forecasting and explaining the changes of dependent variables.

5.1 Residual diagnosis result of a single e-commerce platform (JD)
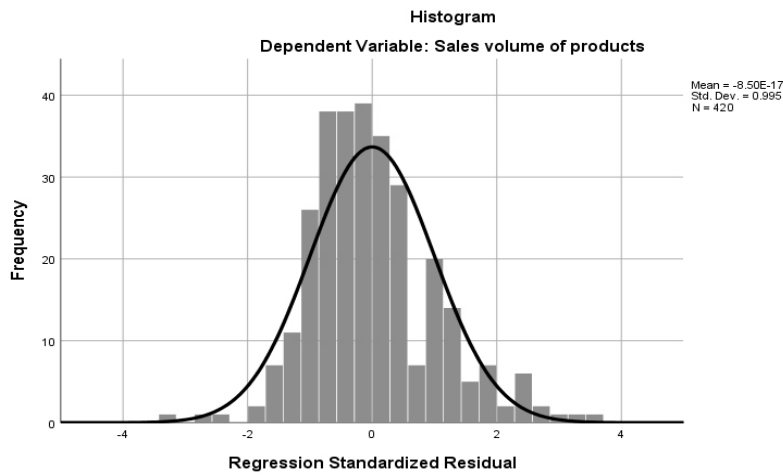

Figure 4.1 Histogram of residuals On JD platform

It can be seen from the results in **Figure 4.1** that the residuals of the regression model basically obey the normal distribution, that is, the operation results of the regression model are stable and reliable.

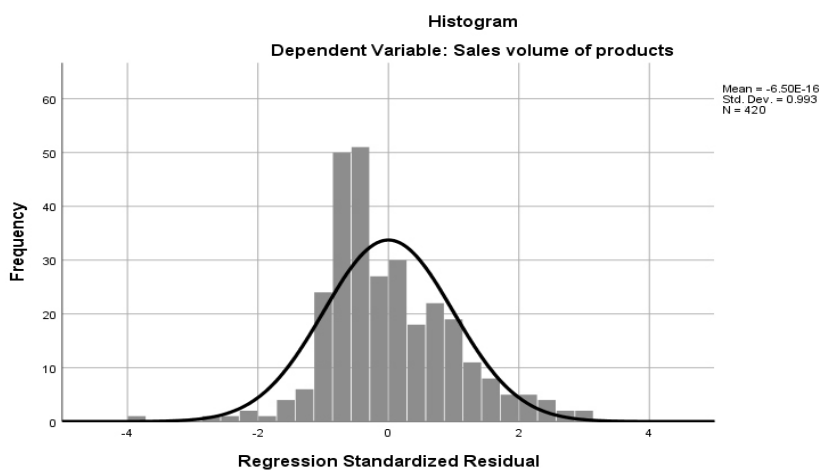### 5.2 Residual diagnosis result integrated social media platform (Xiaohongshu)

Figure 4.2 Histogram of residuals Integrating XHS social media platform

From the results in **Figure 4.2**, it can be concluded that the residuals of the regression model basically obey the normal distribution, i.e. the results of the regression model are stable and reliable. Incorporating the number of Xiaohongshu comments as a new independent variable may have provided the model with more information about product sales, thus enhance the explanatory power of the model. When the model explains the variability of the dependent variable (product sales) more accurately, the residual frequency will naturally decrease.

Hypothesis test results

After organizing and analyzing the data, the assumptions proposed in the previous text were analyzed and validated. The hypothesis test results obtained through regression analysis of the product are shown in **Table 4.11**.

Table 4.11 Hypothesis test results

Serial Number	Research Hypotheses	Conclusions
H1a	The more comments in the e-commerce platform reviews, the greater the impact on product sales, and it is a positive relationship	valid
H1b	The more image comments in the e-commerce platform reviews, the greater the impact on product sales, and it is a positive relationship	valid
H1c	The more star ratings in e-commerce platform reviews, the greater the impact on product sales, and it is a positive relationship	invalid
H2a	The higher the number of comments in social media platform, the greater the impact on product sales, and it is a positive relationship	valid

1. Influence of single e-commerce platform online word of mouth on product sales

  1.1 The more comments in the e-commerce platform reviews, the greater the impact on product sales, and it is a positive relationship.

  Research has shown that in comments on e-commerce platforms, it is a positive relationship between the number of comments and product sales, meaning that the more comments there are, the greater the impact on product sales. This conclusion is consistent with the research findings of previous scholars (Feng Jiao, Yao Zhong, 2016, pp.106-114) and (Shan Chunling, Zhao Hanyu, 2017,p,108-111). In general, the positive correlation between the number of comments and product sales is reasonable and understandable. Firstly, a high number of comments means that more users have purchased the product leading to increase in product sales. Secondly, users tend to write comments after purchasing, meanwhile products with comments are more likely to attract the purchasing interest of potential consumers. However, it is important to note that there may be other factors that affect this positive correlation, such as advertising, product awareness, etc. If the product has a good reputation, the likelihood of being purchased by other users will also increase.

  1.2 The more image in reviews on e-commerce platforms, the greater the impact on product sales, and it is a positive relationship.

  This conclusion is consistent with the research of (You Jun, Zhang Xiaoyu et al,2019, pp. 140-144). As a more intuitive display method than text, images can effectively present the appearance, characteristics, and practical application effects of products, thereby let consumers to have a deeper understanding of the product. When the product webpage contains more images, it is likely to attract more consumers to click and enter the webpage, thereby promoting the growth of product sales. However, it is worth noting that the factors that affect product sales are not limited to the number of images, but also include the quality of the images and the attractiveness of the product itself. It is crucial to fully utilize the advantages of images when online marketing strategies are implemented. High quality images can stimulate consumers' purchasing desire and give products a more realistic and eye-catching image. In addition, the uniqueness and attractiveness of the product

itself are also crucial for consumers' purchasing decisions. Therefore, taking into account the quantity, quality, and product characteristics of images can bring more positive effects to online marketing and enhance product sales performance.

1.3 Relationship between star rating and product sales is not significant

This conclusion is consistent with the studies of (Liu hua, 2020, pp. 83-91) and (Liu Yilin, Guan Hongbo, 2021, pp. 106-116). The star rating does not have a significant impact on the online sales of the product, which may be related to the product type. The evaluation of experiential products is usually more subjective, and consumers' experiences and feelings towards the product may vary depending on individual differences. Therefore, a single rating cannot fully reflect the quality and attractiveness of the product. In contrast, consumers tend to refer to the experiences and feelings of other consumers rather than relying solely on digital ratings. Moreover, customers' shopping decisions do not entirely depend on their scores, and they are also influenced by many factors. If these other factors have a more decisive impact, then the impact of high rated products on sales will be relatively small.

In summary: When using a single e-commerce platform, the more total comments and image comments, the greater the positive impact on product sales. This is because the number of comments and image comments usually reflects the popularity and consumers engagement of the product. When a product has more comments and image comments, it means that the product has attracted more attention and users' interaction, increases its visibility and reputation in the minds of consumers. These contribute to stimulate the interest of other potential buyers, increase product exposure, and promote product sales ultimately. On the other hand, products with high ratings may have high expectations in the market, leading to consumers having high expectations for the products, and to some extent, their sales growth may be relatively slow.

2. Influence of Integrating social media platform online word of mouth on Product Sales.

The more comments on social media platforms, the greater the impact on product sales, and it is a positive relationship. From the perspective of research hypotheses, factors related to online word-of-mouth on social media platforms can

have a positive impact. The phenomenon of positive impact of online word-of-mouth on product sales from social media platforms is known as the "Social Media Effect". The emergence of this phenomenon stems from the popularity of social media, which makes people easier to share their viewpoints and experiences. This conclusion is consistent with the research of (Zhao Zhirong, Fang Jiaming, 2015, pp. 668-673). The more comments there are, the more users are discussing and commenting on the product. This will increase the awareness and exposure of the product on social media platforms, attract more attention and attention to the product, thereby increase the number of potential consumers. Many consumers not only read reviews from e-commerce platforms, but also from third-party platforms before making online purchases-decision. They may feel that third-party platforms have higher authenticity, especially for experiential products. The more third-party platform users evaluate the product, the more they understand the product, and the more favorable the product selection is, thereby having a positive impact on sales.

In summary: When integrating into a third-party social media platform comments, the number of comments from the third-party platform has a more significant positive impact on product sales. This is because third-party social media platforms have a broader user base and a stronger social communication effect. When a product receives more reviews on third-party platforms, these reviews will further expand the product's exposure, attract more users' attention, and spread across more social circles, thus further increase the product's chances of sales.