# Research on the Z-Generation Health Market and Its Buying Behavior under the New Economy of Punk Health

Danni Liu

Guangdong Pharmaceutical University, Guangzhou, 510006, China

ldn0523@126.com

**Abstract.** With the young people of Generation Z joining the ranks of health care, the new concept of "punk health care" has emerged, which refers to the formal health care to make up for the loss of health and comfort the indulgent psychology. In this study, a variety of sampling methods were used to collect and analyze the use of health care products and customer satisfaction through online surveys. The questionnaire recovery rate was high and the data quality was reliable. The study found that the market potential of health care products is huge and is in a period of rapid development. By analyzing online public opinion and comments on e-commerce platforms, using Python to crawl data and carry out sentiment analysis, it reveals consumers' praise, concerns and dissatisfaction with health care products. Further, through cross-correlation analysis and PCA-systematic clustering, the influencing factors of Gen Z's cognition and satisfaction with health care products were explored, and the characteristics of customers and their continuous purchase intention were analyzed. The conclusion is that the usage rate and recognition rate of punk health care products are improved, and the characteristics of potential customer groups are obvious. Based on Logistic regression and systematic cluster analysis, potential customers are classified. Finally, from the three aspects of government supervision, enterprise R & D marketing and consumers' own behavior, the paper puts forward suggestions to promote the healthy development of punk health market.

**Keywords:** health products; Generation Z; User portrait; Buying behavior; Sentiment analysis.

## 1 Introduction

Punk health is represented by Generation Z (after 95), which has a persistent unhealthy lifestyle but still has a sense of health care. It reflects young people's concern about their own health.

"Punk health" sounds like a joke, but behind it is tens of billions or even billions of levels of health subdivision track, with great market development potential. According to the "2017-2022 China Health and wellness industry market Development status and investment prospects forecast Report" shows that the current size of China's health and wellness market has exceeded one trillion yuan, and the average urban


Q. Wu et al. (eds.), *Proceedings of the 2024 3rd International Conference on Public Service, Economic Management and Sustainable Development (PESD 2024)*, Advances in Economics, Business and Management Research 309,
https://doi.org/10.2991/978-94-6463-598-0_78

resident spends more than 1,000 yuan per year on health and wellness, while the proportion of young people aged 18-35 is as high as 83.7%[1].

External factors: Due to the intense pace of work, heavy economic pressure, and poor lifestyle, various health problems have entangled young people who should have unlimited energy. At the same time, they also clearly know their health needs, and young people generally start to keep healthy[2].

Internal factors: Generation Z has been instilled with the concept of "health" by their parents since they were very young, but they do not want to be as disciplined as their elders and do not want to lose the youthful vitality of young people, but their poor physical fitness and gradually improving health awareness have prompted them to pick up a thermos cup for health.

The innovation point of this paper is that on the basis of the Expectation confirmation model (ECM), a theoretical model of users' continuous purchase intention is constructed. According to the characteristics of punk health care products, five basic potential variables of the model are constructed, which are subdivided into perceived value, reference information, product trust, use feedback and emotional feedback, so as to study the factors affecting the continuous purchase intention of the buyers of punk health care products[3].

## 2 Market Survey Scheme and Sampling Scheme Design

### 2.1 Survey Scheme

1. Survey respondents

The main body of "punk health" is generation Z young people, so our survey objects are mainly young people born after 1995, supplemented by other age groups.

2. Questionnaire design

The first part is personal basic information, including the gender, age, education, occupation, monthly income level of the survey object.

The second part is the scale part, including perceived value, product trust, reference information, emotional feedback four dimensions.

The third part is the non-scale, including personal rest and physical status, usual exercise frequency, average sleep time and average sleep duration in a week.

The fourth part is the capacity of behavior.

The fifth part is health awareness, and the sixth part is the use and choice of health products.

### 2.2 Sampling Scheme Design

The cross-validation method is adopted in this study. The total sample is randomly divided into two equal parts. Exploratory Factor Analysis (EFA) and Confirmatory Cactor Analysis (CFA) are carried out respectively. Therefore, the sample size requirements of both EFA and CFA methods should be considered. Most researchers believe that the sample size of EFA is at least 50. According to previous survey experience, we believe that it is more reasonable to collect a total sample size of >200 for

pre-survey. For models with 3-4 indicators per factor, researchers believe the sample size should be at least N = 100 (Boomsma,1985; Marsh and Hau,1999).

In addition, we used the pwr package in R language to analyze the required sample size in the study, and assumed that there was a medium effect size d=0.50(Cohen,1988), a statistical test force 1-β=0.8, and a significance level α=0.05. The results showed that at least 36 subjects were required in each group. Researchers argue that SEM requires a larger sample size, e.g., N = 200 (Hoogland and Boomsma 1998; Boomsma and Hoogland, 2001; Kline, 2005). Simulation studies have shown that a reasonable sample size for a simple CFA model is about N = 150 for normally distributed indicator variables with no missing data (Muthen and Muthen, 2002). According to the previous survey experience, we believe that it is more appropriate to collect a total sample size of >400 for the survey[4].

# 3 Survey Implementation and Questionnaire Quality Control

## 3.1 Quality Control

1. Data sorting:

The questionnaire was distributed through the network, EpiData3.1 was used to establish a database, double-blind parallel data entry, and SPSS26.0 and AMOS26.0 were used for data analysis. The criteria for determining the validity of a questionnaire are: missing less than 10% of the items, answering less than half of the average time, no two or more questionnaires have the same answers, and there is no situation in which all items of the questionnaire are the same. At the same time, we set up a group of questions with the same meaning and different expressions interspersed in the questionnaire. If the answers to these two questions are contradictory in a questionnaire, the data of the questionnaire is considered to be wrong and will be deleted. After strict quality control and cleaning and screening, the data obtained will be scientifically and accurately analyzed to provide powerful data for the conclusion of this report.

## 3.2 Pre-Survey Data Processing and Inspection

1. Questionnaire issuance and recovery:

Before the formal distribution of the questionnaire, we now select a small sample in the sampling box for pre-investigation, so as to find unreasonable places in the questionnaire and make timely modifications. In the pre-survey implementation stage, 250 questionnaires were distributed according to the random principle, and 224 were valid, with an effective recovery rate of 89.6%, which was statistically significant. According to the data of the pre-survey sample, the reliability test and validity test of the scale in the questionnaire were carried out, and then the questionnaire was adjusted to improve the effectiveness of the questionnaire.

2. The pre-survey passed the reliability and validity test

In this study, the influencing factors of continuous purchase intention were measured in the form of a scale, so testing the data quality of the measurement results is an

important prerequisite to ensure the significance of subsequent analysis. Generally speaking, the correlation coefficient between the scores of each item and the total score of the scale is greater than 0.3 and has statistical significance, indicating that the discrimination of this item is up to the standard. The greater the correlation coefficient, the higher the discrimination, but if it is too high (above 0.85), there may be multiple collinearity problems, which will lead to unreliable factor analysis results. Since the scores of each item were grade variables, Spearman rank correlation was used for analysis. As shown in the appendix, all the correlation coefficients were between 0.3 and 0.85, and were statistically significant (P>0.0001), so no item needed to be deleted.

We used SPSS26.0 to conduct exploratory factor analysis and reliability test on the pre-survey data to ensure the reliability and validity of the final questionnaire. The analysis results showed that five common factors were extracted from the 33 measurement items, and the Cronbach's Alpha coefficients of the reliability test of the five factors were all higher than 0.7, as shown in the appendix. However, the Cronbach's Alpha coefficients of the two dimensions of health care product experience and emotional appeal were all lower than 0.7. Therefore, eliminate Q9. Q10. Q11. Q12. Q13. Q14 during April 12-15. Q15, keep 26 item. The pre-survey Bartlett sphericity test KMO>0.6, and the significance <0.05, as shown in Table 1.

**Table 1.** KMO and Bartlett sphericity tests for the pre-survey

| KMO sample appropriateness measure | | 0.875 |
|---|---|---|
| Bartlett sphericity test | Approximate chi-square | 2373.673 |
| | Degree of freedom | 528 |
| | significance | 0.000 |

# 4 Consumer Clustering and Review Sentiment Analysis Based on PCA- Systematic Clustering

## 4.1 Research Method

Principal component analysis (PCA) can not only reduce the dimensionality of high-dimensional data, but more importantly, eliminate data redundancy and noise. If the dimensionality of the final cluster is very high, the clustering effect and running speed of spectral clustering are not good because of insufficient dimensionality reduction. Therefore, cluster analysis based on a few comprehensive indicators or dimensions after PCA is used for dimensionality reduction will produce a more concise and clear result[5].

Text sentiment analysis, also known as opinion mining, refers to the process of extracting, analyzing and processing emotionally colored text by using natural language processing, text mining and computer technology, so as to help users obtain effective information. Through emotion analysis, we can understand the views of netizens on a certain event and identify the emotional trend of the content posted by netizens, such

as support, opposition, happiness or sadness. Through emotion analysis, we can further predict the evolution law of emotions over time. There are two main methods of text sentiment analysis: rule-based (sentiment dictionary) and machine learning. The dictionary-based method is to match the emotion words in the pre-processed text by making an emotion dictionary. This method does not require manual annotation, but the same word may express different emotion tendencies in different contexts.

Based on the event corpus, this study uses Dalian Institute of Science and Technology emotional Vocabulary Ontology Database and ROSTCM6 word segmentation system as the basic word database. At the same time, through observation and collection, the self-defined word database including the evaluation terms of punk health groups and network popular words is sorted out, and the self-defined word database is imported into the word segmentation dictionary of ROSTCM6 text analysis tool. Through the emotion analysis of Weibo and Zhihu comment data text, the emotional tendency distribution of punk health care crowd is obtained, and countermeasures and suggestions are provided for the expansion and subdivision of punk health care market[6].

### 4.2 Data Processing and Preprocessing

1. Weibo and Zhihu comment data capture

In this paper, "punk health" as the search keyword, from Sina Weibo, Zhihu according to the relevance of 3112 comment text data.

Using the manual screening method, after the manual cleaning of 1062 texts, the remaining 2050 valid text data are finally divided into words, and will be filtered again, only adjectives, nouns, verbs are retained.

1) Preparation: Enter the developer mode through the browser to view and analyze Zhihu and Weibo pages.

2) Get the page: Send a request to the URL to return the data of the entire page. This is similar to when you click the URL in your browser and press Enter, then you can see the entire page.

3) Parse the web page: Extract the desired data from the data of the entire web page. Similar to seeing the entire page of the website in a browser, the data needed for the research is mainly related to punk health comments and microblog posts. What needs to be parsed is HTML structure data, which is parsed using regular expressions.

4) Data storage: The data is stored, and a total of 3112 data are obtained.

Text word segmentation refers to the process of splitting each clause in a document into a single word according to certain rules, which is a key step in semantic understanding. In this article, use the Jieba word segmentation tool kit in python to perform Chinese word segmentation. After word segmentation, the jieba word in the data set refers to removing the words or words that have no practical significance and have no obvious topic for subsequent work, such as modal words and special symbols. The text preprocessing process is shown in Figure 1.

Use Python crawler data to capture users' concerns about punk health. Social platforms (such as Sina Weibo and Zhihu) can directly capture the attention Angle of the majority of netizens on the current hot topics. This paper analyzes the word frequency

of all texts after preprocessing[7-8]. The Top29 high-frequency words are shown in the table, and the word cloud graph generated by high-frequency words is shown in Figure 1.

From the word cloud map can be card, in the top 500 words in the text word frequency, the higher the frequency of occurrence, the larger the font, the word frequency according to the number of "health", "punk", "young people", "health", "stay up", "body", "wolfberry", "health products"[9].

2. Modification of the model by the elbow rule

The elbow method is a rough estimate of the optimal number of clusters by the graph[10]. The coefficients in the SPSS output document were applied, and after sorting them, Excel was used to generate images, as shown below:

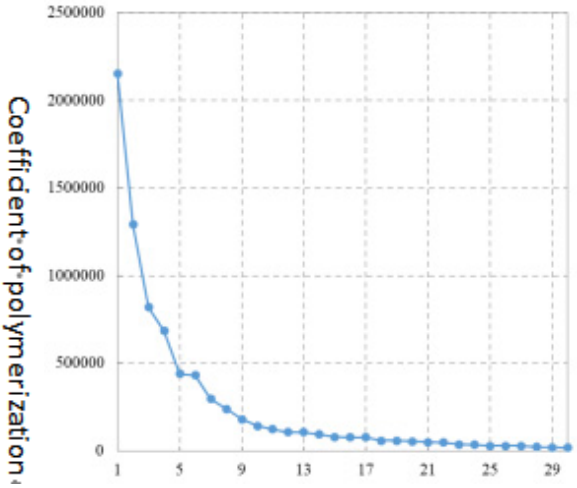


**Fig. 1.** Distortion degree

When K=5, the contour coefficient is the largest, which means that the clustering effect is relatively reasonable at this time.

## 5 User Cognition Status of Punk Health Products

### 5.1 Correlation, Significance and Difference Analysis

1. There is a strong positive correlation between impact factors and satisfaction

Correlation analysis is a statistical method to study the strength of the relationship between variables in the model. Correlation coefficient is usually used to describe the direction and degree of the linear relationship between variables. In this study, Pearosn correlation coefficient is used to describe the relationship between variables.

**Table 2.** Correlation analysis between impact factors and satisfaction

| correlation | | | | | | |
|---|---|---|---|---|---|---|
| | Perceived value | Product trust | Reference information | Emotional feedback | Use feedback | satisfaction |
| Perceived value | 1 | | | | | |
| Product trust | .744** | 1 | | | | |
| Reference information | .737** | .733** | 1 | | | |
| Emotional feedback | .660** | .665** | .668** | 1 | | |
| Use feedback | .624** | .704** | .536** | .603** | 1 | |
| satisfaction | .692** | .697** | .781** | .536** | .766** | 1 |

Note: ** means P<0.01, * means P<0.05

There was a strong positive correlation between perceived value, product trust, reference information, emotional feedback, use feedback and satisfaction at 0.01 significance level.

2. Five-dimension variables significantly affect user satisfaction

**Table 3.** Five-dimension significance analysis

| | Nonnormalized coefficient | | Standard coefficient | | | |
|---|---|---|---|---|---|---|
| model | B | Standard error | β | t | significance | VIF |
| (constant) | -0.014 | 0.136 | | -0.105 | 0.917 | |
| Perceived value | 0.015 | 0.01 | 0.083 | 1.615 | 0.108 | 2.938 |
| Product trust | -0.022 | 0.019 | -0.062 | -1.131 | 0.259 | 3.395 |
| Reference information | 0.189 | 0.015 | 0.615 | 12.217 | 0 | 2.841 |
| Emotional feedback | -0.1 | 0.02 | -0.23 | -5.118 | 0 | 2.261 |
| Use feedback | 0.315 | 0.024 | 0.568 | 12.86 | 0 | 2.185 |
| | R2 | | | | 0.806 | |
| | F | | | | 180.678 | |
| | P | | | | <0.001 | |
| Dependent variable: user satisfaction | | | | | | |

R2=0.806>0.6, indicating that the calculation results can very truly and reliably reflect the impact of perceived value, product trust, reference information, emotional feedback, and use feedback on user satisfaction. At the same time, there is no multicollinearity among the five independent variables, and VIF is all less than 5. In addition, the regression equation is significant, F value is 146.583, P<0.001, which means that at least one of the five independent variables can significantly affect the dependent variable.

Perceived value had no significant positive effect on user satisfaction (β=0.083>0, P=0.108>0.05). Product trust had no significant negative effect on user satisfaction (β=-0.062<0, P=0.259>0.05). The reference information had significant positive effect on user satisfaction (β=0.615>0, P=0<0.05). Emotional experience had a significant reverse effect on user satisfaction (β=-0.23<0, P=0<0.05). The use of feedback can significantly positively affect user satisfaction (β=0.568>0, P=0<0.05). Therefore, the regression equation between the variables is as follows:

User satisfaction =-0.014+0.615* Reference information -0.23* Emotional feedback +0.568* Use feedback

## 6 Conclusion

This study reveals that the emerging concept of "punk health care" among Generation Z has significant market potential and is in a period of rapid development. Through comprehensive analysis, we find that reference information, emotional feedback, and use feedback are key factors affecting user satisfaction with punk health care products. Specifically, reference information and use feedback have significant positive effects on user satisfaction, while emotional feedback has a significant reverse effect. Perceived value and product trust, however, do not significantly impact user satisfaction. Based on these findings, we suggest that enterprises and policymakers focus on improving reference information, enhancing user experience, and addressing emotional feedback to promote the healthy development of the punk health care market. Furthermore, targeted marketing strategies and government supervision are essential to meet the unique needs and preferences of Generation Z consumers.

## References

1. Li Yazheng. A study on the influencing factors of online education platform users' continuous usage intention and course payment willingness [D]. University of Science and Technology of China,2016.
2. Exploring childhood and retaining youthful memories -- An investigation report on influencing factors of college students' purchase intention of nostalgic products [D]. China Society of Business Statistics.2022.
3. Smart Bracelet use status and product development analysis report -- Text mining based on e-commerce platform consumer purchase evaluation and sampling survey in Wuhan [D].2016.
4. 2017-2022 China Health and Wellness Industry Report [R].2022.
5. "Healthy China 2030" Plan Outline [R].2016.
6. National Nutrition Plan (2017-2030) [R].2017.
7. "Nutrition Survey Questionnaire 2021" iResearch.[R].2021.
8. Hao Miaomiao. Analysis on the mentality and causes behind the culture of "Punk health preservation" [A]. Jinguwenchuang, 2022, (15).
9. Ren Baolong. Typical characteristics, formation mechanism and resolution Strategy of Youth "punk health" phenomenon [A].2022.

10. Hu Liangyi. Liu Yongning. Wu Chunmei. Analysis of the Youth subculture Phenomenon of "Punk Health Preservation" [A].2019.1006-1789(2019)01-0075-05.

## Appendix

**Table 4.** KMO and Bartlett sphericity tests for the pre-survey

| KMO sample appropriateness measure | | 0.875 |
|---|---|---|
| Bartlett sphericity test | Approximate chi-square | 2373.673 |
| | Degree of freedom | 528 |
| | significance | 0.000 |

Table 2: Correlation coefficient and significance of each measurement item and total score

| Measurement item | Spearman rank correlation coeffi-cient | Significance (P-value) |
|---|---|---|
| Q1 | 0.456 | <0.0001 |
| Q2 | 0.521 | <0.0001 |
| ... | ... | ... |
| Q9 | 0.287 | <0.0001 |
| Q10 | 0.293 | <0.0001 |
| Q11 | 0.301 | <0.0001 |
| Q12 | 0.312 | <0.0001 |
| Q13 | 0.324 | <0.0001 |
| Q14 | 0.335 | <0.0001 |
| Q15 | 0.467 | <0.0001 |
| ... | ... | ... |
| Q33 | 0.765 | <0.0001 |