Research on the Influencing Factors of Consumer Seafood Consumption Willingness in the Bohai Rim Area Based on Random Forest and Structural Equation Model

Yucong Peng[1, *], Le Kang[1], Zhongxu Ma[3], Junchen Liu[1], Yizhuo Sun[1], Zhirui Zhang[1] and Yichen Xiang[1]

[1]School of Economics and Management, Northwest A&F University, Xian, China

[3]School of Landscape Arts, Northwest A&F University, Xian, China

* Corresponding Author Email: 269640133@qq.com

Abstract. With the increasing prominence of environmental issues, the impact of nuclear pollution events on the marine ecological environment has attracted widespread concern. This study aims to explore the influencing factors of consumers' willingness to consume seafood in the Bohai Rim area when facing nuclear pollution events. By combining the use of random forest and structural equation modeling (SEM), this research conducted an in-depth analysis of the collected survey data. Initially, the random forest model was applied to rank the importance of satisfaction scale indicators, factor analysis was used to reduce the scale indicators to obtain factors affecting consumer satisfaction, and finally, the SEM model was employed to study the impact of various factors and indicators on consumer satisfaction, identifying the key factors affecting the willingness to consume prepared dishes. The study results show that consumers pay high attention to the origin, quality, and nutritional value of seafood, and these factors are proven to have a significant impact on consumption willingness through the measurement model. The structural model analysis reveals the importance of business services and product features in enhancing consumers' willingness to purchase. Additionally, the model fit indices indicate that the proposed model structure fits the actual data well. This study provides targeted marketing strategy recommendations for seafood suppliers to adapt to changes in consumer behavior under environmental changes.

Keywords: Random Forest; Structural Equation Modeling; Factor Analysis; Consumer Willingness; Bohai Rim Area; Seafood.

1. Introduction

As globalization continues to advance, environmental issues have become a common challenge that transcends national borders. In recent years, the issue of nuclear-contaminated water management in Japan has particularly attracted widespread international attention. Reports indicate that Japan's storage capacity for nuclear-contaminated water is nearing its limit, and on August 24th of 2023, the Japanese government decided to release nuclear wastewater into the ocean, disregarding opposition from the international community. This decision has sparked global concern and controversy and poses a severe threat to marine ecological safety [1].

The radioactive substances contained in nuclear wastewater are undeniably destructive to marine ecosystems and also have profound implications for human societies that rely on ocean resources, especially for the consumption of seafood products [2]. The difficulty in treating radioactive contaminants is significant, and the self-purification capacity of seawater is insufficient to fully eliminate these pollutants, particularly radioactive materials. This has intensified public anxiety regarding the safety of seafood products.

Against this backdrop, the Bohai Rim area, as an important marine economic region in China, has seen a notable impact on its seafood consumption market. Consumer perceptions of the safety of seafood, willingness to purchase, and consumption behaviors may change significantly in the wake of nuclear pollution events. Therefore, this study aims to delve into the influencing factors of consumers' willingness to consume seafood in the Bohai Rim area when faced with nuclear pollution events, and how these factors affect the consumer purchasing decision process.

By combining the use of random forest and structural equation modeling, this research seeks to reveal consumers' attitudes and behavioral changes towards seafood consumption willingness, providing scientific decision-making evidence for seafood suppliers and policymakers [3, 4]. This is to better respond to the challenges brought about by environmental changes and promote the sustainable development of the seafood market [5].

2. Data Analysis

After completing the questionnaire design, we distributed questionnaires on the network questionnaire survey platform. A total of 211 questionnaires were collected and 174 valid questionnaires were obtained after screening. We conducted item analysis and reliability and validity tests on the data based on the recovered questionnaires to check the differentiation of the questionnaire questions, the reliability of the measurement results and the validity of the questionnaires.

2.1. Reliability Test

Reliability is the analysis of the accuracy of the measurement results of the questionnaire during reliability test, that is, the accuracy and reliability test of the data results obtained under the repeated use of the designed questionnaire. It mainly tests the consistency, consistency, reproducibility and stability of the results [6].

This paper adopts Cronbach'α (Alpha) reliability coefficient method to test the reliability of the questionnaire. In general, the internal consistency coefficient of the questionnaire items is mainly considered, and the value range of the reliability coefficient is [0,1]. The reliability coefficient is calculated as follows.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum S_i^2}{S_t^2}\right) = \frac{N}{N-1}\left[1 - \frac{N}{N+2r}\right] \quad (1)$$

Where, N is the total number of questions in the questionnaire, S_i^2 is the variance of the score in question i, and S_i^2 is the variance of the total score of all questions. There was inconsistency in the scores of each survey item in the Cronbach α time scale. The greater the reliability coefficient, the greater the reliability of the measurement.

We conducted KMO and Bartlett tests and reliability analysis on the survey data of scale questions in the questionnaire through SPSS, and obtained that the reliability coefficient score of consumers' basic information was 0.754, and the information reliability coefficient score of consumers' willingness was 0.785. According to the reliability test, Cronbach's Alpha coefficient all exceeded the minimum acceptable level of 0.7, which indicated that the questionnaire had good internal consistency and reliability

2.2. Validity Test

KMO and Bartlett tests were conducted on the survey data, and the results were shown in Table 1. As can be seen from the structural validity results in Table 1, the Barrlett test of willingness scale was significant (significance Sig=0.000), and the KMO of consumption willingness table was 0.701>0.5, indicating that the questionnaire had good willingness validity and effective results.

Table 1. KMO and Bartlett tests

Sample the Kaiser-Meyer-Olkin measure of adequacy		0.701
Bartlett tests	Approximate chi-square	623.198
	Degree of freedom	129.000
	Significance	0.000

2.3. Run Test

The statistic of the run test is R, which is the number of runs. $X_1, X_2, \ldots X_n$ is a sequence of 0 or 1, and the hypothesis testing problem is: H_0 is the data appear in random order, H_1 is the data appear in non-random order.

If H_0 is true, $X_i \sim b(i, p)$, p is the probability of 1, determined by n/m, and the distribution of the number of runs R is related to p. When the sample size is large, under the null hypothesis, it can be obtained according to the properties of the exact distribution.

$$E(R) = 2\frac{mn}{m+n} + 1. \tag{2}$$

$$Var(R) = \frac{2mn(2mn-m-n)}{(m+n)^2(m+n+1)} \tag{3}$$

Therefore, the P-values and test results can be obtained by using the normal distribution table. At this time, after the level α is given, the critical values c_1 and c_2 of the rejection domain can be obtained by an approximate formula, which satisfy P（R≤c_1）≤α and P（R≤c_2）≤α.

We took the gender of the respondent as the variable X to test whether the questionnaire data was random. It can be seen from the results that the gender-specific run test with the mean as the segmentation point has a significance of 0.893>0.05, so it passes the run test. The results show that the questionnaire used in the pre-survey can achieve the purpose of this survey.

Table 2. Run test results

	Sex
Test value (average)	1.150
Number of cases < test value	168
Total number of cases	174
Z	0.1245
Asymptotic significance (two-tailed)	0.893

3. Analysis of Influencing Factors

In this part, we analyze the consumption intention of consumers in various aspects of the process of purchasing seafood in the Bohai Rim area. We first use random forest to rank the importance of the six indicators, then use factor analysis to classify the six indicators into three main factors that affect the consumer's consumption intention, and finally use structural equation model to study the impact of indicators and factors on the consumption intention.

3.1. Ranking the importance of indicators based on random forest model

In order to study how different factors in the purchasing process affect consumers' willingness to consume, we first need to understand the importance of different factors affecting the willingness to consume. According to the results of the study on seafood consumption and consumption intention, we selected 6 possible influencing factors and ranked their importance by random forest model. A random forest is a (parallel) integrated algorithm made up of decision trees that can be used for data classification or index ranking. This paper uses the random forest model to rank the importance of satisfaction indicators, aiming to understand the importance of different satisfaction indicators and evaluate the business development of suppliers. The importance ranking results are shown in Fig. 1.

The top three in terms of importance are: origin, quality and nutritional value; The last three are: green food logo, purchase channels, purchase convenience. According to the results, origin and quality are relatively important satisfaction indicators, while the product itself is relatively important.

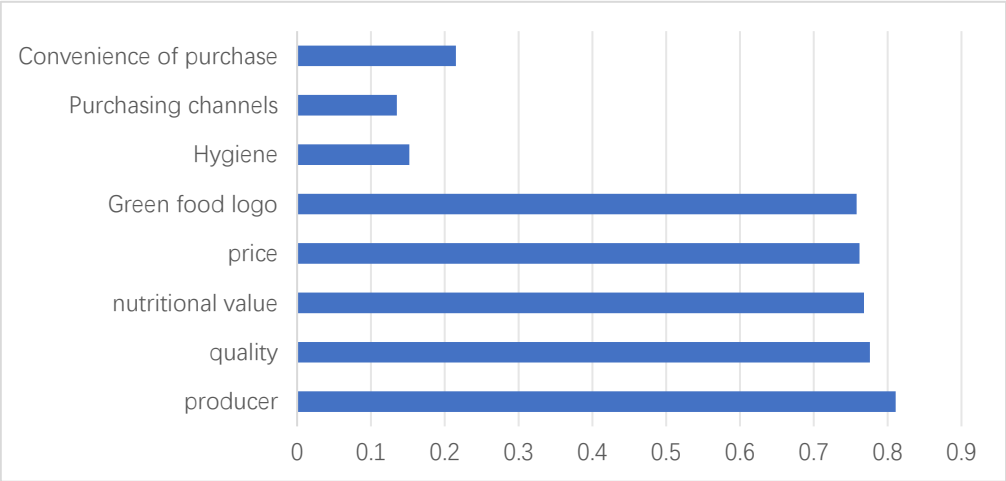


Fig. 1 Ranking the importance of indicators that measure the willingness of consumers to spend

3.2. Analysis of seafood consumption willingness factors

Structural equation model includes measurement model and structure model. The measurement model describes the relationship between each latent variable and the corresponding observed variable, while the structural model describes the relationship between each latent variable. In this paper, factor analysis method is used to construct measurement model.

The reliability and validity of the data of consumption intention scale are good, and factor analysis can be carried out. SPSS23.0 was used to select principal component analysis for factor dimensionality reduction, and 8 measurement indicators were converted into 2 factors. The factor loads of each index in different factors were shown in Table 3.

Table 3. Run test results

Index	Divisor		Name
	1	2	
Place of origin	0.811	0.198	
Quality	0.776	0.108	
Nutritional value	0.768	0.156	Product characteristics
Price	0.762	0.165	
Green food label	0.758	0.132	
Sanitary condition	0.152	0.786	
Purchase channel	0.135	0.814	Merchant service
Ease of purchase	0.215	0.756	

By observing the result of factor analysis, it is found that there are significant differences between the factor loads of the 8 indicators in different factors. Component factors of the measurement indicators with factor loads greater than 0.7 in each factor are extracted and named as "product features" and "business services" as latent variables of the model. Combined with the importance ranking results, we can see that among the two latent variables, the importance ranges from large to small: merchant service and product features, while the level of consumers' willingness to consume ranges from large to small: product features and merchant service, and the absolute level of merchant service is significantly different from the former.

3.3. Analysis of influencing factors of seafood consumers' consumption intention

Based on the measurement model, we put forward two hypotheses for the relationship between latent variables according to the current research status of seafood consumption and the research theory of influencing factors of consumers' consumption intention. First, the product characteristics of seafood have a significant positive impact on consumers' consumption intention. When purchasing

seafood, origin, quality, nutritional value and price are the core attributes of consumers to decide whether to buy or not, and are also fundamental to the formation of consumers' willingness to consume. The better the overall quality of the product, the higher the willingness to buy will be brought to consumers. Secondly, the merchant service of seafood has a significant positive impact on consumers' consumption intention. When purchasing seafood, merchant service determines the quality of consumers' purchasing experience, which has an important impact on consumers' consumption intention. The better the service of the merchant, the better the purchasing experience of the consumer, and the higher the purchase intention of the consumer.

Based on the index system shown in Table 3, we built the structural equation model shown in Fig. 2.

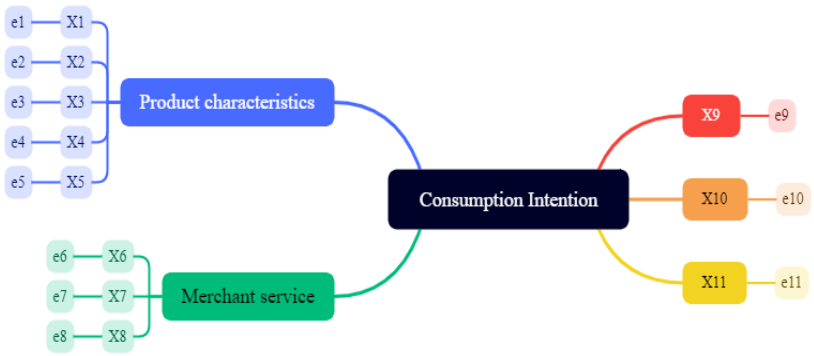


Fig. 2 Structural equation model diagram

SPSS23.0 and AMOS25.0 were used to construct structural equation models and carry out fitting analysis. The output results are shown in Table 4. In the table, *** means significant at 0.001 level and ** means significant at 0.01 levelThe factor loads of the three latent variables, product features, merchant services and consumption intention, are all greater than the minimum standard of 0.6. The values are all greater than 0.7, which accords with the reference value, indicating that the model has a good fitting result.

Table 4. Structural equation model index test results

Latent variable	ID	Observed variable	Estimate	S.E.	C.R.	p-value
Product characteristics	X1	Place of origin	0.821	—	—	—
	X2	Quality	0.754	0.054	17.985	***
	X3	Nutritional value	0.809	0.052	19.355	***
	X4	Price	0.732	0.054	17.185	***
	X5	Green food label	0.684	0.058	15.836	***
Merchant service	X6	Sanitary condition	0.739	—	—	—
	X7	Purchase channel	0.736	0.064	14.711	***
	X8	Ease of purchase	0.764	0.068	15.196	***
Consumption satisfaction	X9	Purchase intention	0.726	—	—	—
	X10	Repurchase intention	0.738	0.076	12.843	***
	X11	Recommendation intention	0.705	0.073	12.566	***

After fitting modification, the fit index value of the final model is shown in Table 5. Compared with the recommended value given by the fitting index, the fitting value of all the fitting indexes falls within the recommended value range. It can be seen that the setting of the theoretical model is acceptable and the effect is good.

Table 5. Model test result

Fit Indicator	Recommended Value	Fit Value
x^2	—	122.869
x^2/df	<3.0	1.502
GFI	>0.9	0.953
AGFI	>0.8	0.946
RMSEA	<0.08	0.026
NNFI	>0.9	0.975
IFI	>0.9	0.996
CFI	>0.9	0.976

The Estimate value in the table above represents the degree of influence of observed variables on latent variables. It can be seen that seafood suppliers do better in the aspects with greater impact, such as "health conditions", "purchase channels" and "convenience of purchase", and consumers have a higher overall purchase intention. Then AMOS25.0 was used to test the proposed hypothesis, and the results were shown in Table 6.

Table 6. Path coefficient and significance index value

Influencing factor index	Estimate	S.E.	C.R.	P	Hypothesis
Satisfaction (Product Characteristics)	0.203	0.060	3.136	**	Set up
Satisfaction (Merchant Service)	0.22	0.062	3.866	***	Set up

The results obtained from the above table are as follows: product characteristics have a significant impact on consumption intention (P<0.01), H1 hypothesis is valid. The standardized path coefficient of the influence of product features on consumption intention is 0.203, indicating that when consumers' evaluation of product features is increased by one unit, consumers' consumption intention will become 1.203 times of the original.

Merchant service has a very significant effect on consumption intention (P<0.001), and H2 hypothesis is valid. The standardized path coefficient of the influence of merchant services on consumption intention is 0.22, indicating that when consumers' evaluation of merchant services is increased by one unit, consumers' consumption intention will become 1.22 times of the original.

To sum up, all hypotheses passed the significance test and all hypotheses were valid, that is, product features and merchant services have different degrees of positive impact on the consumption intention of consumers who buy seafood. The degree of influence of these two factors on consumption willingness is ranked from large to small: publicity and sales, business services. It is consistent with the results of importance ranking. Suppliers should not only maintain the outstanding results in the existing business (such as product features), but also implement targeted measures to optimize the business with low consumption willingness as soon as possible, and improve the comprehensive service ability.

4. Conclusion

The present study set out to investigate the factors influencing consumers' willingness to consume seafood in the Bohai Rim area amidst concerns over nuclear pollution events. Utilizing a combination of random forest and structural equation modeling (SEM), we have identified and analyzed key determinants that shape consumer behavior in the context of environmental challenges. Our findings underscore the heightened awareness and sensitivity of consumers towards the origin, quality, and nutritional value of seafood products. The significant role of these factors in enhancing consumer willingness to purchase was corroborated by the measurement model. Furthermore, the structural model revealed that business services and product features play a pivotal role in fostering a positive attitude towards seafood consumption, despite the looming threat of nuclear pollution.

The robustness of our model, as indicated by the model fit indices, suggests that the proposed framework effectively captures the complex relationships between various factors and consumer willingness to consume seafood. This study not only contributes to the academic discourse on consumer behavior in the face of environmental risks but also provides actionable insights for seafood industry stakeholders.

References

[1] Li, Zichao, et al. "Research on calculation and verification of hazardous area for radioactive sewage entering into the ocean." Progress in Nuclear Energy 161 (2023): 104739.

[2] Kamei-Ishikawa, Nao, et al. "Fate of radiocesium in sewage treatment process released by the nuclear accident at Fukushima." Chemosphere 93.4 (2013): 689-694.

[3] Rigatti, Steven J. "Random forest." Journal of Insurance Medicine 47.1 (2017): 31-39.

[4] Bowen, Natasha K., and Shenyang Guo. Structural equation modeling. Oxford University Press, 2011.

[5] Horn, Katrin. "Consumer values and product perception." Consumer Perception of Product Risks and Benefits (2017): 283-299.

[6] Elsayed, Elsayed A. "Overview of reliability testing." IEEE Transactions on Reliability 61.2 (2012): 282-291.