

Construction and application of the multidimensional quantitative model for game evaluation

Jinkun Hong

Aquinas International Academy, 3200 Guasti Rd, Suite 100, Ontario, CA 91761, USA

Jinkunhong42@gmail.com

Abstract. This article is a modeling study on the self-rating of board games. This article has reviewed and understood board games and various indicators of board games and stated the background of the topic and the questions raised. It has determined the dimensions of game characteristics that need to be evaluated. Through keyword retrieval, the text is transformed into the required numerical information, and through public calculation, the required data is obtained to fit the geek ratings and average player ratings of the 25 games in the target. Next, analyze the specific issues of each data through descriptive statistics. Then, through correlation analysis, which indicators are related were obtained, and a series of conclusions were drawn. Next, a second linear regression analysis will be conducted on the Geek rating and average rating, and a series of conclusions will be drawn. Summarize the conclusion, consider the advantages of the model, and reflect on the shortcomings of the model. Use sensitivity analysis to check if the conclusion is complete. This article randomly selected a game and analyzed it through a model. And draw relevant conclusions. Finally, the analysis and summary improved the modeling.

Keywords: probability distribution, quadratic linear regression, descriptive analysis, correlation analysis.

1. Introduction

The earliest board game can be traced back to the ancient Egyptian game *Seni* in the 3rd century Before Christ. Remains the most well-known board game, originating from ancient Indian chess in the 6th century Anno Domini. In 1974, with the release of the first version of *Dungeons and Dragons*, tabletop games also ushered in a new era. Since the millennium, the form and content of desktop games have experienced unprecedented development. While some video games try to simulate the experience of a tabletop game as closely as possible, a notable feature of the tabletop game is the role of the *Dungeon Lord*, who not only adds a unique personal touch to the game but also provides the basis for flexibility in rule interpretation. This characteristic results in significant differences in consideration between tabletop game design and video game design. This creates some profound differences between the design of desktop games and the aspects considered in electronic game design. With the diversification of desktop game types, there has gradually been a column for evaluating the game itself.

Petri presented a systematic evaluation of educational games focusing on the evaluation process [1]. The study results are based on 11 relevant articles describing 7 approaches to systematically evaluate educational games. The study confirmed that only a few approaches are available to systematically

evaluate educational games. However, the research results are based on only 7 encountered approaches where no clear pattern emerged on which factors are essential to evaluate educational games [1].

However, current evaluations often focus on subjective experiences, making it difficult to achieve horizontal quantitative comparisons between games. From the perspective of board game developers, to better understand a board game, this paper needs an objective measurement. The evaluation given by the player's comprehensive rating model is often too vague, so mathematical modeling can be well combined with it. A stable mathematical model helps people choose a good board game while considering game balance, stability, and other game metrics.

The linear regression model provides a powerful device for organizing data analysis. Models are specified, variables are measured, and equations are estimated with ordinary least squares (OLS). All goes well if the classical linear regression assumptions are met. However, several assumptions are likely to be unmet if the dependent variable has only two or three response categories. In particular, with a dichotomous dependent variable, assumptions of homoskedasticity, linearity, and normality are violated, and OLS estimates are inefficient at best. The logistic regression model takes the natural logarithm of the odds as a regression function of the predictors [2]. The maximum likelihood estimation of a logistic regression overcomes this inefficiency [3].

For the numerous games that have already been released, it is hoped to purposefully select some reference dimensions to anticipate a comprehensive evaluation and response of a game within a controllable range. This article attempts to reconstruct this problem using mathematical models using the following methods: 1. Determine the dimensions for evaluating game characteristics: adversarial, balance, strategic, difficulty, and interactivity. 2. Clean existing comment text and convert it into comparable numerical factors. 3. Select factors to construct a regression model to fit the player ratings and expert ratings of 25 games in the target. 4. Analyze the causal relationship between significant factors in the model and their correlations. 5. Consider games outside of the sample to test the model's universality and discuss the phenomenon of insignificant factor reactions in the model. 6. Based on the model results, provide an experimental report and scoring guidelines for game designers.

2. Model assumptions

The data set for directed content analysis comprised data extracted from 58 articles on Game-Based Learning (GBL) evaluation literature from our previous systematic literature review. The selected articles comprised GBL evaluation frameworks, evaluation studies, and reviews [4].

The corpus is entirely focused on GBL evaluation literature, rather than the integration of gaming and learning fields, to ensure alignment with the study's objectives. The data items extracted from the selected papers include dimensions, factors, sub-factors, metrics, relationships among these dimensions, factors, and sub-factors, the nature or description of these relationships, as well as definitions of the dimensions, factors, and sub-factors [5].

This article will consider extracting a linear regression model from the text. However, considering the natural ambiguity/ambiguity of natural language and the incomplete nature of player comments, the following idealized assumptions need to be made. Artificial intelligence using machine learning (ML) is an ensemble of techniques that automatically learn patterns from data and that require no assumptions regarding the structure of the data [6]. The reflection on the rationality of assumptions will be presented in the sixth part of this article:

1. As this article obtained factors through text analysis of game reviews, these factors can be classified as emotional factors. Firstly, it is assumed that comments always comprehensively and truthfully reflect the most memorable features of the game for players.

2. Assuming that there are no interaction features between these factors, this assumption is achieved through the selection of factors: this study is confident that the selected dimensions are independent of each other, so the conditional probability will not change due to changes in other dimensions.

3. As the purpose of this article is to guide designers, it is assumed that the player ratings and expert ratings referred to in the following text are stable (stationarily) for the rating time, that is, there is always

a constant square difference between these two ratings in any given period. Only in this way can the model conclusions be used to predict the performance of new games.

3. Model establishment

3.1. Acquisition of emotional indicators

In the process of model building, due to the lack of existing analysis data, the most representative rules and mechanisms of the top 25 games were selected based on the game descriptions provided on the Board Game Geek website. Based on the keywords that appear in the hot reviews, specific reference values for gameplay, strategy, difficulty in getting started, randomness, player interaction intensity, and player ranking balance have been summarized for each game. For example, by analyzing all the comments of these 50 people and drawing a conclusion. If there are words such as "high difficulty coefficient" or "not recommended for beginners to try" in the evaluation, one point will be added to the difficulty of getting started, resulting in a total score of 50. If the difficulty of getting started is 27 points, and 5 people in the comments think that the difficulty of getting started is high but it is relatively easy to master proficiently, they are considered neutral and included in the calculation range. If they calculate them in words, then the difficulty of getting started accounts for 5%, and neutral accounts for 5%. If the highest score is set at 10 points, which means 10 is completely satisfied, then the difficulty of getting started with this game is 5.9. However, due to some games being too obscure, it is difficult to find 50 samples for analysis, so many two-decimal places appear.

The formula can be expressed as:

$$S_{ij} = \frac{P_{ij} + \frac{1}{2}M_{ij}}{P_{ij} + N_{ij} + M_{ij}} \quad (1)$$

3.2. Linear regression model

Due to $S_{ij} \in [0,1]$, as mentioned earlier, for each factor, on average, players have a preference range, and the difficulty and strategy of the game should be kept neither too low nor too difficult. To describe this nonlinear property, a quadratic term S_{ij}^2 is introduced.

The preference for indicators should be symmetrical. Therefore, the regression model is:

$$\begin{cases} \ln G_i = \sum_j g_j^{[1]} l_{ij} + g_j^{[2]} l_{ij}^2 + C_g \\ \ln A_i = \sum_j a_j^{[1]} l_{ij} + a_j^{[2]} l_{ij}^2 + C_a \end{cases} \quad (2)$$

Among $g_j^{[i]}, a_j^{[i]}$, g and a represent the impact of the j th indicator on geek/average ratings, while C_g, C_a are the basic ratings for the two ratings. The above form can also be organized as:

$$\begin{cases} \ln G_i = \sum_j g_j^{[2]} (l_{ij} - \gamma_j)^2 + C_g^* \\ \ln A_i = \sum_j a_j^{[2]} (l_{ij} - \alpha_j)^2 + C_a^* \end{cases} \quad (3)$$

$$\gamma_j = \frac{g_j^{[2]}}{2g_j^{[1]}}, \alpha_j = \frac{a_j^{[2]}}{2a_j^{[1]}}$$

So $S_j^a = e^{\alpha_j}$, $S_j^g = e^{\gamma_j}$ are the best choice. Symbol description as Table 1.

Table 1. Symbol description

Symbol	Symbol description
G_i	i th Geek rating of the game
A_i	i th average rating of the game
P_{ij}	Number of positive reviews for i game j factor

Table 1. (continued)

N_{ij}	Number of negative reviews for i game j factor
M_{ij}	The number of neutral reviews for i game j factor
S_{ij}	i the indicator of the j factor of the game
l_{ij}	$l_{ij} = \ln S_{ij}$

4. Model results

The following table presents descriptive statistics of various data collected from 25 games.

$$\begin{cases} \ln G_i = \sum_j g_j^{[2]} (l_{ij} - \gamma_j)^2 + C_g^* \\ \ln A_i = \sum_j a_j^{[2]} (l_{ij} - \alpha_j)^2 + C_a^* \end{cases} \quad (4)$$

$$\gamma_j = \frac{g_j^{[2]}}{2g_j^{[1]}}, \alpha_j = \frac{a_j^{[2]}}{2a_j^{[1]}}$$

So $S_j^a = e^{\alpha_j}$, $S_j^g = e^{\gamma_j}$ are the best choice.

4.1. Descriptive statistics of raw data

Table 2 presents descriptive statistics of various data collected from 25 games.

Table 2. Descriptive statistics of various data

Variable Name	Maximum Values	Minimum Values	Average Value	Standard Deviation	Kurtosis	Skewness	Coefficient of Variation (CV)
Gameplay log	0.971	0.669	0.855	0.077	-0.25	-0.66	0.09
Strategic log	0.951	0.659	0.86	0.074	1.115	-1.168	0.086
Difficulty log	0.937	0.736	0.848	0.056	-0.901	-0.48	0.066
Randomness log	0.952	0.734	0.866	0.065	-0.905	-0.513	0.075
Player's initial	0.948	0.551	0.829	0.103	1.046	-1.086	0.125
Strength log							
Sequential							
Balance log	0.992	0.659	0.875	0.075	1.461	-0.873	0.086

It is noteworthy that all indicators exhibit high kurtosis and negative bias. That is to say, these variables all exhibit extremely low scores and are densely concentrated around the mean. Through logarithmic transformation, so adjusted the factors to $\ln S$ and $\ln G_i$, $\ln A_i$. The descriptive statistics of each factor after logarithmic transformation are as follows:

4.2. Correlation between two indicators

After solving the distribution problem of the indicators themselves, interested in whether the transformed indicators are linearly correlated. If there is a high correlation, the lecture speculates that there are some endogenous variables between certain factors. The Pearson correlation analysis in Table 3, found that the indicators it collected effectively avoided collinearity:

Table 3. Effectively avoided collinearity

	Gameplay	Strategic	Difficulty in Getting started	Randomness	Hands on intensity	Sequential balance
Gameplay	1(0.00***)	0.19(0.38)	0.01(0.96)	-0.1(0.64)	- 0.03(0.90)	0.25(0.24)
Strategic	0.19(0.38)	1(0.00***)	-0.26(0.21)	-0.10(0.64)	0.2(0.32)	0.13(0.55)
Difficulty in Getting started	0.01(0.96)	- 0.26(0.21)	1(0.00***)	-0.05(0.80)	0.21(0.32)	0.42(0.04**)
Randomness	-0.1(0.64)	- 0.10(0.64)	-0.05(0.80)	1(0.00***)	0.05(0.83)	-0.09(0.68)
Hands on intensity	- 0.03(0.90)	0.21(0.32)	0.21(0.32)	0.05(0.83)	1(0.00***)	0.22(0.30)
Sequential balance	0.25(0.24)	0.13(0.55)	0.42(0.04**)	-0.098(0.68)	0.22(0.30)	1(0.00***)

* p < 0.1, ** p < 0.05, *** p < 0.01

4.3. Linear regression analysis

The Lasso regression algorithm identified measurability, difficulty, and hands-on strength as the most effective factors. A quadratic model was developed based on these factors, as shown in Table 4.

Table 4. Result of the quadratic model

Non standardized Coefficient						
	B	Standard error	t	P	VIF	R2 adjustment F
constant	-68.013	31.96	-2.128	0.048**	-	
Strategic logging approach	-43.997	17.574	-.2504	0.023**	413.05	F=3.479
Strategic log	74.153	28.688	2.585	0.019**	408.06	
Difficulty in getting started with log method	-68.427	50.875	-1.345	0.196	2035.026	0.393
Difficulty log	109.806	86.204	1.274	0.220	2090.652	
Player's initial Strength log	5.687	20.707	0.275	0.787	410.623	P=0.020**
Dependent variable: Geek rating/Green eating						

* p < 0.1, ** p < 0.05, *** p < 0.01

F statistics show that the explanatory power of the model is significant, especially in terms of strategy, but geek players are not very significant in terms of difficulty and intensity in getting started. This also aligns with an understanding of geek players. By considering:

$$\frac{\partial G}{\partial \ln s_1} = 0 \quad (5)$$

Can get:

$$S_1^* + 10^{\left(\frac{74.513}{2 \times 43.997}\right)} = 7.146 \quad (6)$$

The probability has a sigmoidal relationship with the independent variable, and the estimated probabilities are now appropriately constrained between 0 and 1 [7].

Easy to find strategic rating at 7.146 Geek rating. Table 5 shows the linear regression analysis between various game metrics and average player ratings. Simple linear regression lives up to its name: it is a very straightforward simple linear regression approach for predicting a quantitative response Y based on a single predictor variable X. It assumes that there is approximately a linear relationship between X and Y. Mathematically, it can write this linear relationship as $Y \approx \beta_0 + \beta_1 X$ [8]. For ordinary players, strategy remains an important indicator for evaluating games

Table 5. Result of linear regression analysis

Non standardized Coefficient							
B		Standard error	t	P	VIF	R2 adjustment	F
Constant	-24.771	16.664	-1.486	0.155	-		
Strategic logging approach	-31.637	9.163	-3.453	0.003***	413.05		F=4.118
Strategic log	54.085	14.958	3.616	0.002***	408.06		
Difficulty in getting started with log method	-16.765	26.527	-0.632	0.536	2035.026	0.449	
Difficulty log	27.697	44.947	0.616	0.546	2090.652		
Player's initial Strength log	-1.877	10.797	-0.174	0.864	410.623		P=0.010***
Dependent variable: Average rating/Average rating							

5. Advantages and disadvantages of the model

5.1. Model advantages

The model is simple and easy to understand, which aligns well with the general focus of board game players. In the future, a questionnaire survey can be designed to obtain clearer data.

The simplicity of the model has revealed an effective factor for guiding game development: strategy. Scalability is also a strength of this model. Given the ease of iterating multiple factors in linear models, future research could generate new factors to add to the model, such as topicality, peripheral factors, and more.

With the rise of large language models like ChatGPT, sentiment analysis algorithms have become more sophisticated, allowing researchers to extract positive and negative emotions from large volumes of text. This advancement will help overcome the challenges posed by small sample sizes.

5.2. Model disadvantages

Table 5 shows that the presence of three outliers in the small sample prevents the model from accurately estimating these data points. Additionally, the analysis reveals that even after logarithmic transformation, the metrics do not follow a normal distribution. This issue may stem from the fact that the ratings are based on subjective evaluations, with only users who have a certain emotional attachment to the game participating. This introduces bias into the sample set.

Deviation in natural language analysis is also inevitable. In this model, data was processed manually using keyword retrieval and reading. However, this evaluation system, particularly for neutral comments, is neither objective nor stable. These factors contribute to the lack of reliability in the metrics themselves [9, 10].

6. Conclusion

The passage randomly selected a game called 'Earth', which is a board game just released in 2023. The paper used the model to analyze it and obtained a series of data. Consider first the case with a single preassigned independent variable. Summarize the evaluations of this game on the market by picking up keywords and converting them into numerical information. Then calculated that the metrics provided by the model did not match the actual player ratings. This confirms the answer.

Ordinary players tend to place greater emphasis on whether the game provides solutions within a limited time frame when purchasing, but if the game's strategic requirements are too high, ordinary players' evaluation of the game will also decrease. The paper sample size is too small and it has not found suitable indicators, so there is no significant correlation between indicators and geek ratings for desktop games. Secondly, the evaluations on the Board Game Geek website are too subjective and only rely on regression error analysis

Searching for keywords with less than or equal to 50 hot reviews is currently unable to draw relatively accurate conclusions. Further algorithm research will be conducted on the evaluation criteria for games, and a relatively objective and unified desktop game evaluation standard will be developed to enhance the reference value of the evaluation criteria for players.

References

- [1] Tahir, R., & Wang, A. I. (2017, October). State of the art in game-based learning: Dimensions for evaluating educational games. In European Conference on Games Based Learning (pp. 641-650). Academic Conferences International Limited.
- [2] Menard, S. (2020). Applied logistic regression analysis (Vol. 106). Sage.<https://scholar.google.com.hk/>
- [3] Nick, T. G., & Campbell, K. M. (2007). Logistic regression. Topics in biostatistics, 273-301.
- [4] Djelil, F., Sanchez, E., Albouy-Kissi, B., Lavest, J. M., & Albouy-Kissi, A. (2014, October). Towards a learning game evaluation methodology in a training context: A literature review. In 8th European Conference on Games-Based Learning.
- [5] LaValley, M. P. (2018). Logistic regression. Circulation, 117(18), 2395-2399.
- [6] BoardGameGeek <https://boardgamegeek.com/browse/boardgame>
- [7] Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society Series B: Statistical Methodology, 20(2), 215-232.
- [8] Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. International journal of medical informatics, 151, 104484.
- [9] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear regression. In An introduction to statistical learning: With applications in python (pp. 69-134). Cham: Springer International Publishing.
- [10] Mawardi, V. C., & Darmaja, E. (2023). Logistic Regression Method for Sentiment Analysis Application on Google Playstore. International Journal of Application on Sciences, Technology and Engineering, 1(1), 241-247.