# Week 8 Assignment: Predictive Modeling Using Generalized Linear Models

Jonathan Ibifubara Pollyn

College of Science, Engineering and Technology, Grand Canyon University

DSC-530: Predictive Modeling

Filippo Posta

October 27, 2021

## Introduction

Linton, I. (n.d.). *How does advertising affect product awareness & use?* Small Business - Chron.com. https://smallbusiness.chron.com/advertising-affect-product-awareness-use-36750.html

The author is making a point that the attention of potential customers can be drawn with the use of advertisements. Advertisement is used to establish a preference for that product over its rivals. If advertising accomplishes these two goals, customers are more likely to buy the advertised product the next time they shop. Because a single campaign's impact is temporary, it's critical to budget for advertising over time to keep people aware and to use the product. The author affirmed that two requirements apply when consumers choose between alternative items or brands. In other words, the author stated that the product must be an option that a customer may choose from, and that customer must prefer one option over all others in that available selection. The author referred to consumer awareness surveys; if they know the three competing items, they must be made aware of your product to become a viable option. The author stipulated that consumer could learn about your product's existence and availability through advertising. The first step in getting customers to buy and use your product is to raise their awareness of it. Your marketing materials must persuade them of the product's value. Advertisements that feature client testimonials or highlight the attainment of quality criteria do better generate interest and sales. The author stated that awareness and preference rise or diminish over time based on the amount and frequency of advertising. According to the author, frequent advertising can raise awareness to the point where a product's nature and benefits are fully understood, from simple recognition of some of the product's features. The author

affirmed that if you stop advertising or your competitors raise the amount of advertising, they undertake compared to yours, awareness and preference may decrease.

The author also stipulated that advertising on the Internet makes it easier to quantify the effects of advertising on product awareness and use. When customers click on your Internet ad to find out more about your product, you can see right away if the campaign was a success based on how many people responded. Also, the author stated that when customers buy the advertised product, you know that the campaign successfully got them to use it. Although advertising can enhance product awareness and usage, other means customers utilize to receive product information must also be considered.

**Information about the model**

As a result of using a logistic model, you can forecast the response variable's outcome. Using the advertising dataset to model a logistic regression model, I understood how Age and Area income influences the customer to click on the Ad. The predictor variables are the age and area income from the dataset, while its response variable is the clicked-on Ad. The dataset shows that the more time a customer spends on the site, the better the chance of clicking on the Ad. Figure 1 shows that 75 percent of the time a customer spent on the site, there is a chance that they will click on the Ad, while zero chance of them clicking on the Ad if they spent just 25 percent.

| | Daily Time Spent on Site | Age | Area Income | Daily Internet Usage | Male | Clicked on Ad |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.00000 |
| mean | 65.000200 | 36.009000 | 55000.000080 | 180.000100 | 0.481000 | 0.50000 |
| std | 15.853615 | 8.785562 | 13414.634022 | 43.902339 | 0.499889 | 0.50025 |
| min | 32.600000 | 19.000000 | 13996.500000 | 104.780000 | 0.000000 | 0.00000 |
| 25% | 51.360000 | 29.000000 | 47031.802500 | 138.830000 | 0.000000 | 0.00000 |
| 50% | 68.215000 | 35.000000 | 57012.300000 | 183.130000 | 0.000000 | 0.50000 |
| 75% | 78.547500 | 42.000000 | 65470.635000 | 218.792500 | 1.000000 | 1.00000 |
| max | 91.430000 | 61.000000 | 79484.800000 | 269.960000 | 1.000000 | 1.00000 |

Figure 1: Summary of the dataset

The data also shows exciting information; the Ad performs better with younger customers than the more elderly ones. Figure 2 shows that the younger a customer, the more frequently they are likely to click on an Ad.

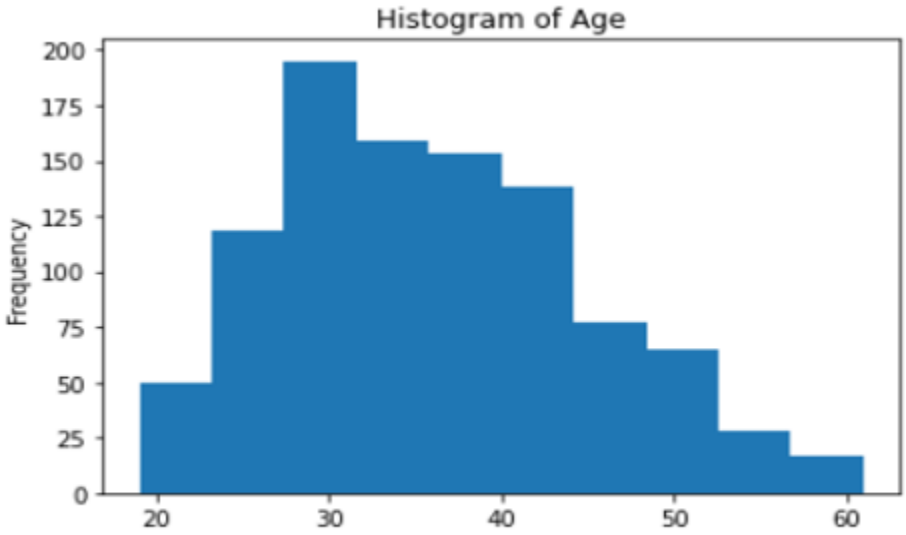


Figure 2: Histogram of Frequency against Age

The model shows that both predictors (Age and Area Income) are statistically significant; figure 3 shows that the p-values for both Age and Area Income are < 0.05, which is the cutoff value.

| | | | |
|---|---|---|---|
| Model: | Logit | Pseudo R-squared: | 0.365 |
| Dependent Variable: | Clicked on Ad | AIC: | 886.0035 |
| Date: | 2021-10-25 14:52 | BIC: | 900.7268 |
| No. Observations: | 1000 | Log-Likelihood: | -440.00 |
| Df Model: | 2 | LL-Null: | -693.15 |
| Df Residuals: | 997 | LLR p-value: | 1.1490e-110 |
| Converged: | 1.0000 | Scale: | 1.0000 |
| No. Iterations: | 7.0000 | | |

| | Coef. | Std.Err. | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 0.0916 | 0.5399 | 0.1697 | 0.8653 | -0.9666 | 1.1498 |
| **Age** | 0.1626 | 0.0126 | 12.9008 | 0.0000 | 0.1379 | 0.1874 |
| **Area Income** | -0.0001 | 0.0000 | -12.5857 | 0.0000 | -0.0001 | -0.0001 |

Figure 3: Summary of Model

Through the summarized model from figure 3, we can obtain the descriptive form of the final regression model as

$$\text{yhat(income)}=\frac{\exp\left(exp(-0.00916+0.1626(age)-0.0001(Area\ Income)\right)}{1+exp(-0.0916+0.1626(age)-0.0001(Area\ Income))}$$

The age coefficient reveals that a customer is increasingly likely to click on an Ad with an age of approximately 117% or 1.17. At the same time, the probability of having a higher click on Ads every five years after the customer has clicked it before is likely to increase by 235% or 2.35. It means that the Ads, once clicked, are likely to capture the customer's attention for a long time. The coefficient of the Area Income demonstrates that a customer clicking on an Ad is expected to increase approximately 99.9% or 0.99; also, if a customer Area Income increased by 5,000, then the clicked-on Ads is likely to increase approximately 60.7% or 0.61. 60.7% is a drop from the collected data, which indicates that customers' interest is likely to drop as their area income increases. The predicted value is obtained from the prediction of the predictors, while the target value is the actual response variable that did not tell any difference in information. The Poisson regression with the same advertising data shows that all variables belong to the model, and we do not have to remove anyone.

The summary below shows that the p-values for all two variables are below the cutoff value of 00.05.

Generalized Linear Model Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | Clicked on Ad | **No. Observations:** | 1000 |
| **Model:** | GLM | **Df Residuals:** | 997 |
| **Model Family:** | Poisson | **Df Model:** | 2 |
| **Link Function:** | log | **Scale:** | 1.0000 |
| **Method:** | IRLS | **Log-Likelihood:** | -758.24 |
| **Date:** | Tue, 26 Oct 2021 | **Deviance:** | 516.47 |
| **Time:** | 07:32:37 | **Pearson chi2:** | 452. |
| **No. Iterations:** | 5 | | |
| **Covariance Type:** | nonrobust | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | -0.9793 | 0.280 | -3.496 | 0.000 | -1.528 | -0.430 |
| **Age** | 0.0415 | 0.005 | 8.556 | 0.000 | 0.032 | 0.051 |
| **Area Income** | -2.495e-05 | 3.1e-06 | -8.050 | 0.000 | -3.1e-05 | -1.89e-05 |

Figure 4: GLM Regression Result

The descriptive form of the Poisson regression from can be derived from figure 4 as

$$yhat = e^{(-0.9793+0.0415(age)-2.495e-05(Area\ Income)}$$

## References List

Larose, C. D., Larose, D. T., & Larose, Chantal D., Author. (2019). *Data science using python and r*. John Wiley & Sons,inc,.

Linton, I. (n.d.). *How does advertising affect product awareness & use?* Small Business – Chron.com. https://smallbusiness.chron.com/advertising-affect-product-awareness-use-36750.html