Predicting Flu Shot Incidence

Alexis Arthur B. Garcia

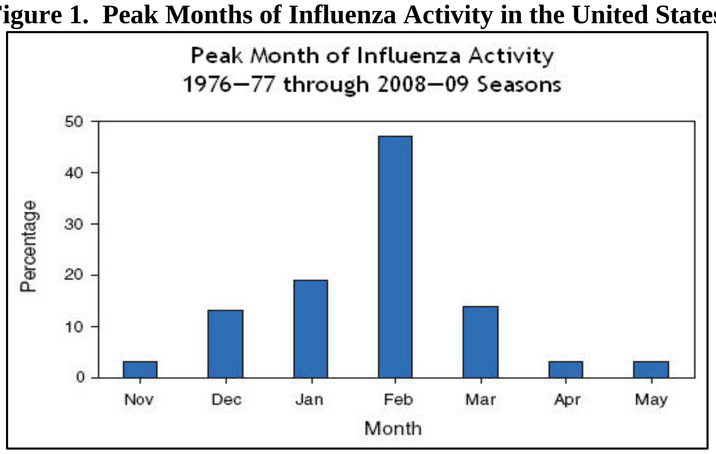
Abstract

An important issue among public health officials when flu epidemics are imminent, is the encouraging of the taking of flu shots, especially among older people who are at higher risk of developing complications. This study was conducted to determine whether age and health awareness were important predictors of the incidence of flu shots in a certain population. A total of 50 people were randomly selected for this study. Results of the binary logistic regression shows that with a health awareness index of 60, the odds of getting a flu shot is increased by 71.7% as a person gets older by a year. On the other hand, the predicted probability of getting a flu shot given a person is 50 years old with health awareness index of 60 is 0.988 which indicates that older people with high health awareness would probably get a flu shot.

Introduction

Influenza (Flu) is a contagious respiratory illness caused by influenza viruses and can cause mild to severe illness, and at times can lead to death. It has long been considered as a serious contagious disease by the Centers for Disease Control and Prevention (CDC) here in the United States. According to the CDC, each year on average, 5% to 20% of the U.S. population gets the flu and more than 200,000 people are hospitalized from flu-related complications and about 36,000 people die from flu-related causes. The gravity and pervasiveness of this disease made it a very important issue among public health officials in encouraging taking of flu shots especially when flu epidemics are imminent in the months of winter (Figure 1).

Figure 1. Peak Months of Influenza Activity in the United States.



Source: www.cdc.gov

The CDC had profiled people that are at increased risk for serious complications from seasonal flu illness. Among these are older people, young children, pregnant women and people with certain conditions such as asthma, diabetes, or heart disease. This study was conducted to determine whether age and health awareness were important predictors of the incidence of flu shots in a certain population. Results of this study will aid in enabling public health officials to effectively target specific population that are at high risk and with high probability of taking flu shots vaccination.

The remainder of this article proceeds as follows: Data description is discussed in the next section. The statistical methods are then discussed followed by the presentation and discussion of the results. Conclusions are presented in the final section.

Data

In determining the predictive capability of age and health awareness on the incidence of flu shots, 50 people were randomly selected toward the end of the flu season. They were asked whether they had actually received a flu shot that season as well as their age along with the information relating to their awareness on general health issues which was combined into a health awareness index. These variables are described in Table 1 along with its summary statistics. The flu shot is the response variable in this study and is coded as yes (1) if one had the flu shot at the season, otherwise it is coded as no (0). About 42% (21 respondents) responded as having (yes) taken a flu shot. The average age of the randomly selected sample is 46 years while their health awareness index is 51.24, on the average, with 27 and 71 recorded as the lowest and the highest health awareness index, respectively

Table 1. Summary Statistics of the Logistic Regression Variables.

Variable	Description	N	Mean	Std. Dev	Min	Max
Flu Shot	Whether the person had a flu shot (yes = 1, no=0)	50	0.42	0.4985694	0	1
Age	Age in Years	50	46.42	8.7972398	28	64
Awareness	Health Awareness Index (Higher values indicate greater awareness)	50	51.24	11.7726074	27	71

Statistical Methods

Chi-Square Test of Independence

The Chi-Square Test of Independence between the factors Flu_Shot (yes or no) and Age (young or old where young is <50 years old, otherwise old) was employed to determine whether the incidence of flu shots is independent of the age. In the context of this study wherein we compare two nominal variables, this test compares the frequency of one nominal variable with the frequency of the second nominal variable. The setup of the data is best described in the Row-Column Form of Two-way Table. The null hypothesis under this test assumes that there is no association between the two variables while the alternative hypothesis assumes otherwise that is there is an association between the two variables. The hypothesis testing under Chi-Square Test of Independence is similar to the ANOVA test. If the calculated value of the Chi-Square test is greater than the table value, we will reject the null hypothesis in favor of the alternative hypothesis.

Binary Logistic Regression

A logistic regression model (McCullagh and Nelder, 1998) would best model the relationship between a binary response variable Y and the predictors $X = [X_1, \ldots, X_s]$ and takes the form of

$$logit\ (p) = ln\left(\frac{p}{(1-p)}\right) = X\beta = \beta_0 + \beta_1 X_1 + \cdots + \beta_s X_s \qquad \textbf{Eq (1)}$$

where $p = \Pr(Y = 1)$ is the probability of having flu shots. $\beta_j (j = 1, \ldots, s)$ is interpreted as a unit increase in X_j with other predictors held fixed will increase $ln\left(\frac{p}{(1-p)}\right)$ by β_j where $ln\left(\frac{p}{(1-p)}\right)$ is called as the log-odds of $Y = 1$. The estimated of the regression coefficients, $[\hat{\beta}_0, \ldots, \hat{\beta}_s]$, are obtained by Maximum Likelihood Estimation (MLE) based on a dataset D=$[X_1^T, \ldots, X_N^T]^T$, where $X_i = [X_{i1}, \ldots, X_{is}]$ is a vector of measurements on the predictors.

For a given observation $X_0 = [X_{01}, \ldots, X_{0s}]$, $logit\ (\hat{p}_0) = \hat{\beta}_0 + \hat{\beta}_1 X_{01} + \cdots + \hat{\beta}_s X_{0s}$ and thus the predicted probability $\hat{p}_0$, algebraically, is

$$\hat{p}_0 = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{01} + \cdots + \hat{\beta}_s X_{0s}} \Big/ (1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{01} + \cdots + \hat{\beta}_s X_{0s}}) \qquad \textbf{Eq (2)}$$

Similar to the linear regression model, the coefficient of the parameters were tested for significance using Wald tests. The overall fitness of the model was also evaluated using several goodness of fit measures such as Pearson χ^2, Deviance and Hosmer-Lemeshow Tests. These tests assume good model fit under null hypothesis. Influential observations was also evaluated using delta χ^2 which indicates influential or poorly fit for observations with values larger than 3.8. The predictive ability of the model was also assessed using the percentage of concordants which indicates higher predictive ability for larger percentages.

Results

The two-way table for chi-square test for independence is shown in Table 2. Only 42% (21) of the 50 people randomly selected responded as having flu shots for the season. Eleven (52.38%) of it were classified as old respondents while the rest (11 respondents) have responded as having not taken flu shots at the season.

Table 2. Two-Way R-C Table for Chi-Square Test for Independence.

Age	Legend	Flu Shot		
		Yes	No	Total
Old	Frequency	11	5	16
	Percent	22	10	32
	Row Pct	68.75	31.25	
	Col Pct	52.38	17.24	
Young		10	24	34
		20	48	68
		29.41	70.59	
		47.62	82.76	
Total		21	29	50
		42	58	100

Results of the chi-square test for independence show that there is an association between age and flu shots variable. The chi-square test statistic has a value of 6.9116 with 1 degree of freedom. Its probability value is 0.0086 which is less than 1% experimentwise error rate, thus, we reject the null hypothesis of no association in favor of the alternative hypothesis which is there is dependence between flu shots incidence and the age.

The final specification of the binary logistic regression model ($P(Y = 1 or\ Yes)$) includes the interaction between age (in years) and the awareness index. Results of the estimation are shown in Table 3. The final model is highly statistically significant compared to an intercept only model as indicated by the likelihood ratio statistic with probability value of <0.0001. The estimation results also show that the main effects by the predictors age and awareness were not statistically significant at 5% experimentwise error rate although the predictor awareness is statistically significant at 10% experimentwise error rate. However, even if the main effects were significant, we cannot have a straightforward interpretation of its effects since their interaction is statistically significant at 5% experimentwise error rate. Thus, we have to consider the parameter estimate of the interaction term in determining the odds ratio corresponding to either age or awareness since this odds ratio depend also on awareness. Given that awareness = 60, then the odds ratio corresponding the age is computed as

$$\boldsymbol{\Pi_{Age}} = \boldsymbol{e}^{(\hat{\beta}_1 + \hat{\beta}_3 Age*Awareness)} = \boldsymbol{e}^{(-0.8814+0.0237*60)} = \mathbf{1.717}.$$

An odds ratio of 1.717 corresponding age and given awareness=60 means that an increase in age by one year increases the odds of having flu shots by 71.7%. On the other hand, the predicted probability that a person aged 50 with a health awareness index of 60 gets a flu shot can be computed based on the fitted model. From Equation 2, we have

$$\hat{p}_0 = \frac{e^{26.7559-0.8814*50-0.8223*60+0.0237*50*60}}{1+e^{26.7559-0.8814*50-0.8223*60+0.0237*50*60}} = 0.\mathbf{988}.$$

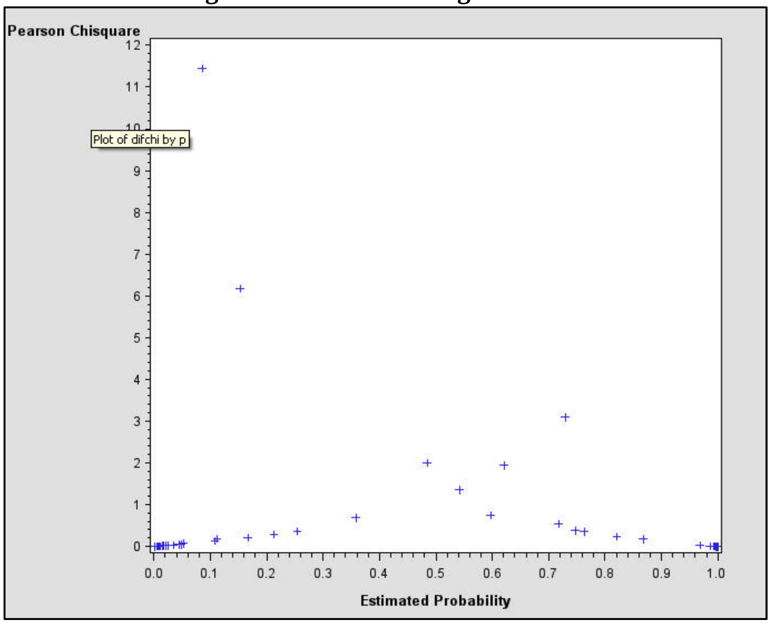
Table 3. Estimation Results of the Binary Logistic Regression Model.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	26.7559	23.436	1.3034	0.2536
Age	1	-0.8814	0.5447	2.6181	0.1056
Awareness	1	-0.8223	0.4993	2.7128	0.0995
Age*Awareness	1	0.0237	0.0119	3.961	0.0466

Likelihood Ratio χ^2 (df=3)= 43.7461 with Pr > ChiSq <0.0001
Hosmer and Lemeshow Goodness-of-Fit Test χ^2 (df=8)= 2.4527 with Pr > ChiSq 0.9639
% Concordant = 96.1%

Based on the fitted logistic regression model, a 50 year old person with health awareness of 60 is predicted to get a flu shot with a probability of 0.988 which is highly probable. The fitted model has also high predictive ability as indicated by 96.1% concordant. On the other hand, the final model has also good model fit as indicated by an insignificant chi-square statistic. Influential observations were also checked using the delta chi-square. Figure 2 shows the diagnostic plot of delta chi-square vs. estimated probability. Two observations were highly above the threshold of 3.8. A model was fitted less these two observations but all the parameter estimates were found to be statistically insignificant.

Figure 2. Influential Diagnostic Plot



Conclusion

The results of this study show that there is dependence between age and the incidence of flu shots based on the chi-square test for independence. Based on the 50 randomly selected persons, most of the older persons (> 50 years old) were shown to have flu shots vaccine during the season. Estimation of the final model revealed that there is a positive interaction effect between age and awareness in predicting the probability of a person getting a flu shot. Given an awareness of 60, the odds of getting a flu shot is increased by 71.7% as a person gets older by a year. The predicted probability of getting a flu shot given a person is 50 years old with health awareness index of 60 is 0.988 which indicates that older people with high health awareness would probably get a flu shot.

References

McCullagh, P. and J.A. Nelder. 1998. *Generalized Linear Models,* Chapman & Hall/CRC.

Ott, R.L., M. Longnecker. *An Introduction to Statistical Methods and Data Analysis*, 5th ed. Duxbury-Thomson-Brooks/Cole, Belmont, 2001.

www.cdc.gov.