A Logistic Function for Hiring University Professors

Dominador Gila Dawa
Roberto B. Corcino
Cebu Normal University
Cebu City, Philippines

Abstract

The main objective of this study was the derivation of a probability model for hiring university professors. This was done by considering the logistic curve as the primary function and by letting $\bar{P}$ as the number of actual research output of the professor with $k$ as his expected research output in the university. This will give yield us $\bar{P}/k$ as the probability that a professor will do and publish research for a given number $Y_i$ of research produced and published. Factors for research productions and publications are integrated using the independent variables $x_1, x_2, \ldots, x_n$. Then the identified predictors are then become the basis for the probability model, the derived statistics logistic function from the odds-ratio of doing to not-doing research. More precisely, the resulting model is

$$P(Y_i) = \frac{1}{1 + e^{(-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}}$$

where the arbitrary constants $\beta_0, \beta_1, \ldots, \beta_n$ can be obtained by means of the usual multiple regression analysis.

Keywords: Hiring; Logistic function; Odds-ratio; Regressions

Introduction

In universities, research is generally conducted to support government priorities. The professor should bring to the class their own discoveries so that the students should have firsthand experience on these discoveries. If professors are active in doing and publishing research, they can influence their students and so research culture is developed in the university. Research adds value to both undergraduate education and to the society. Educational programs would be of poor quality without it. In a globalized economy where knowledge and information are increasingly becoming important, diminishing the role of research in universities will be a disservice to all those they educate (Puri, 2015). However, it was observed that for many years, that the Philippines had the lowest research publication growth rate of all ASEAN countries. The Philippines had been ahead of Indonesia and Vietnam in the early 80s, but by the mid-2000’s Vietnam has overtaken the country; and the gap is alarmingly increasing. Being ranked as the second to the lowest increase rate among ASEAN countries, the Philippines may well soon end up at the bottom, unless concrete actions will be done to rectify it. With the highest rate of increase in the number of publications, Cambodia could overtake the Philippines by 2020 (Nguyen & Pham, 2011). Of all the Philippine universities, only the University of the Philippines (UP) landed a spot in the Times Higher Education (THE) World University Rankings. UP was in the 800 bracket in 2017 and in the 601-800
bracket in 2018. The Times Higher Education World University Ranking lists the top 1,000 universities in the world. It bases its rankings on 13 performance indicators grouped into the following areas: teaching, research, citations, industry income and international outlook (Dablo, 2017). This ranking is consistent to the findings of Nguyen and Pham in 2011 who reported that the University of the Philippines (UP) had the highest number of publications in 2011 with 343 articles (up from 302 articles in 2010). UP was followed by the International Rice Research Institute (IRRI) with 139 articles (109 in 2010). For 2011, De La Salle University (in 4th place previously) overtakes the University of Santo Thomas (3rd place previously); the former with 48 papers and the latter having 29 papers (Dablo, 2017). Many factors are attributed to research production and application but are not usually integrated in the hiring process. The rare use of objective indicators, and the prevalent use of personal judgment by nonscientists (who lack valid publications), are the two major causes of the poor state of Philippine higher education research production (Lacanilao, 2009). Most of the formulas related to hiring are broad and not specific to the job description. For example, in order to determine the applicant’s integrity, a personality questionnaire should be included in the selection process (Hughes, 2017). In connection with this, a hiring formula for university professors must consist of all identified factors or predictors for research production and publication to capture the applicant’s true capabilities. It has been observed that recruiters and candidates are typically talking about the wrong things during the job interview. The hiring officer should first approach a candidate about a role they would be good for. This entails the kinds of questions that are embedded on the feasible region of concerns (Adler, 2017). The university should consider a hiring technology for better results (e.g. research and publication ethics). Such technology has to be plugged with sufficient data - candidates. Hiring tools need to offer a comprehensive view of candidates, incorporating their qualitative and quantitative aspects. Cognitive and behavioral data and qualities, which do not often emerge until faced in one-on-one interviews, must become part of the data collection further upstream, and as such, form an integral part of the selection process (Sackett, 2012). A logistic curve is a popular model for describing the evolution of systems in technology (Kucharavy & De Guio, 2008). In the application of curve model as a fitting methods for the settlement of prediction, it was found that the logistic curve is more accurate and reasonable because it has the value of popularization and application in engineering (Wu & Niu, 2014). It was also proven that logistic regression is a good model in the prediction of population (Ravichandran, 2013) and in the prediction of traffic crash on the curve (Xi & Liu, 2014). In order to use the logistic curve, the set of factors or predictors for research productions and publications are considered as primary sources of data. A study regarding the research productivity and its policy Implications to higher education institutions revealed some factors that motivates a professor to do and publish research. These factors include; educational attainment, research benefits and incentive system, the need to have a strong faculty development program, enhanced research collaboration, and improved research productivity (Quimbo & Sulabo, 2011). A good incentive system is a good support for achievement in doing research (Edgar & Gear, 2011). This means that the support from the administration is a predictor of research productivity (Fawzi & Al-Hattani, 2017). The research organization plays an important role in order to promote and enhance the research culture in the university (Bay & Clerigo, 2013). Higher education institutions aiming to foster productivity and stability among their academic staff should adapt their incentive systems to the preferences and concerns of this particular type of worker and promote their intrinsic motivation (Albert and Davia, 2016). A study on factors affecting faculty research
productivity using a systematic analysis of existing literature enumerated several factors such as; self-efficacy, affiliation, motivation, commitment, orientation, basic and advance research skills, sense of achievements, contribution to society, sense of responsibility, scholarly pursuit, autonomy and flexibility, satisfying interest and curiosity, institutional factors (have fewer course preparations, staff support, advising and mentoring, resources, rewards, sufficient work time, culture, research emphasis, tenure and promotion, financial rewards, satisfying performance standards, peer and social recognition), leadership factors (highly regarded able scholar, research oriented, work for departments with a similar priority placed on research), instructive factors (gender, age of a faculty member at a given point in time, intelligence, and personality of the individual (Mantikayan & Abdulgani, 2017). Moreover a study on research productivity and job satisfaction among Spanish professors in English pointed out the impact of research productivity and other satisfaction drivers, such as other research outcomes – research stays abroad and cooperation with teams abroad – and teaching load, marital status and young children (Albert & Davia, 2016). In addition, research collaboration and past research experience of the professor and his mentor are found to be very significant (Diaz, McCall, & Geesy, 2008). The variables or factors for research publications should be integrated in the probability model for selecting a faculty who can do and publish research. This strategy is the basic idea of association and regression that links causality (Freedman, 2009). The probability model can be derived from the logistic curve. A logistic curve is used for forecasting and induces the correct measurement of the growth process that in turn can be applied to quantitatively identify the law of natural growth quantitatively and to reveal the value of the ceiling (upper limits of growth/higher probability) and steepness of the growth (slope of curve). Obviously, the more precise the data and the bigger the section of the logistic curve they cover lead to a lower level of uncertainties. In other words, one can identify a more accurate ceiling and steepness with a larger data set (Kucharavy & De Guio, 2008). Thus, the researchers would like to design a probability model that can help the university in selecting the right person, specifically, the research-doer who implements research in the university and publishes in reputable academic journals. With this initiative, the university maximizes its research productions and publications and delivers its purpose towards excellence. Faculty members, on the other hand, are able to elevate their ranks and at the same time upgrade their professional skills. Furthermore, students are exposed to new knowledge and are able to conduct publishable research of their own.

**Methodology**

The researchers transformed the logistic curve to logistic regression, a type of probabilistic statistic model using the logistic function in predicting the probability of the professor to do and publish researches.

\[
Q(x) = \frac{e^t}{1 + e^t} = \frac{1}{1 + e^{-t}} \tag{1}
\]

The logistic function is useful because it can take an input with any value from negative to positive infinity, whereas the output always takes values between zero and one and hence is interpretable as a probability. Figure 1 shows the graph of the logistic function as shown below.

The logistic function \(Q(t)\) is an element \([0,1]\) for all \(t\). If \(t\) is viewed as a linear function of an explanatory variable (or of a linear combination of explanatory variables), then we express \(t\) as follows:

\[
t = \beta_0 + \beta_1 x \tag{2}
\]

And the logistic function can now be written as:
Note that $F(x)$ is interpreted as the probability of the dependent variable equaling a "success" or "case" rather than a failure or non-case. It is clear that the response variables $Y_i$ are not identically distributed: This means that every dependent variable $y_i$, in symbol we say; $P(y_i=1|x)$ differs from one data point $x_i$ to another, though these $Y_i$ are independent, they still share common factors or parameters called (Colosimo, 2008). Logistic regression can be binomial or multinomial. In this study binary logistic regression was used which deals with situations where the observed outcome for a dependent variable can have only two possible types (for example, success or failure). Logistic regression is used to predict the odds of being a case based on the values of the independent variables (predictors). The odds are defined as the probability that a particular outcome is a case divided by the probability that it is a non-case (Zagreb, 2014). The researchers also used the concept of response surface methodology (RSM) in statistics that explores the relationships between several explanatory variables and one or more response variables. The main idea of RSM is to use a sequence of designed experiments to obtain an optimal response. Box and Wilson suggest using a second-degree polynomial model to do this. They acknowledge that this model is only an approximation, but they use it because such a model is easy to estimate and apply, even when little is known about the process (Hill and Hunter, 2012). Statistical approaches such as RSM was employed to maximize the production of a special substance by optimization of operational factors. In contrast to conventional methods, the interaction among process variables can be determined by statistical techniques (Hill and Hunter, 2012). In this study, the researchers identified the common variables for research productions and publications in the private and public universities in the Philippines. Standardized research instruments were used to extract data or variables from the 10 private and public universities in the Philippines. Checklists, aptitude test scales and personal data sets.
of question are given to the respondents and data are obtained in terms of the following variables and measurement.

The 30 identified variables for research productions and publications were subjected to the Anderson Darling normality test. Only normal variables were selected for the multiple linear regression analysis. Significant factors or variables in the multiple linear regression model was used for another multiple linear regression analysis that will be used as the exponent of e in the probability equation.

Results and Discussion

The model obtained in this study was based on the general concept of the logistic curve, which is the model for the change of population size with respect to time expressed in terms of growth rate and carrying capacity. For this study, the researchers took the number of actual research output of the faculty to be the population size and the expected research output of the faculty to be the carrying capacity.

More precisely, the resulting model is given by

\[ P(Y_i) = \frac{1}{1 + e^{(-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \]  

(4)

where \( x_1, x_2, \ldots, X_n \) are the independent variables that determines the number \( Y_i \) of research produced and published by the \( i \)th faculty. In other words, the model can be used to determine the probability that a faculty will do and publish research. The arbitrary constants \( \beta_0, \beta_1, \ldots, \beta_n \) can be obtained by means of the usual multiple regression analysis.

Note that if \( P(Y_i) \) is the probability that a faculty will do and publish research, then we say that, 1-\( P(Y_i) \) is the probability that a faculty will not do and publish research. This means that if \( P(Y_i) \) is the probability of success, then 1-\( P(Y_i) \) is the probability of failure. For example, if the probability of success is 60% or 0.60 then the probability of failure is 1-0.60 which is equal to 0.40 or 40%. Now if \( A \) is the event of doing and publishing research, then the odds ratio of \( A \) is the ratio of the probability of success in doing and publishing research over the probability of failure of doing and publishing research. In symbol, we say

\[ \text{odds}(A) = \frac{P(A)}{1 - P(A)} \]  

(5)

Hence, by substitution, the odds of doing research can be expressed as

\[ \text{odds}(Y_i) = \frac{P(Y_i)}{1 - P(Y_i)} \]  

(6)

Now, from equation 4, it is set that,

\[ P(Y_i) = \frac{1}{1 + e^{(-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \]  

(7)

Through substitution,

\[ 1 - P(Y_i) = 1 - \frac{1}{1 + e^{(-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \]  

(8)

Then the odds ratio of doing and publishing research \( y_i \) is equal to

\[ \text{odds}(Y_i) = \frac{1 + e^{(-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}}{1 + e^{(-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \]  

(9)

\[ \text{odds}(Y_i) = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n} \]  

(10)

Taking the natural logarithm (ln) of both sides gives

\[ Z_i = \ln(\text{odds}(Y_i)) \]  

(11)

\[ Z_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \]  

(12)

which is the usual multiple regression model. On the other hand, the odds of not doing research is given by
Table 1. Model summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R-Square</th>
<th>Std Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.974a</td>
<td>.948</td>
<td>.943</td>
<td>2.13562</td>
</tr>
</tbody>
</table>

Table 2. ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>Df</th>
<th>Mean Squares</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Regression</td>
<td>10452.346</td>
<td>4</td>
<td>7273.514</td>
<td>782.301</td>
<td>0.000a</td>
</tr>
<tr>
<td>Residual</td>
<td>76.415</td>
<td>48</td>
<td>6.380</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10520.749</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>St. Error</td>
<td>Beta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. constant</td>
<td>-1.9562</td>
<td>0.4387</td>
<td>-2.167</td>
<td>-4.46</td>
</tr>
<tr>
<td>x3</td>
<td>0.3295</td>
<td>0.1197</td>
<td>0.124</td>
<td>2.75</td>
</tr>
<tr>
<td>x9</td>
<td>0.3200</td>
<td>0.0735</td>
<td>-7.67</td>
<td>4.36</td>
</tr>
<tr>
<td>x13</td>
<td>0.0368</td>
<td>0.0103</td>
<td>0.654</td>
<td>3.56</td>
</tr>
<tr>
<td>x15</td>
<td>0.1572</td>
<td>0.1490</td>
<td>1.432</td>
<td>7.76</td>
</tr>
</tbody>
</table>

\[
\text{odds}(Y_l) = \frac{1 - P(Y_l)}{P(Y_l)} \quad (13)
\]

\[
\text{odds}(\tilde{Y}_l) = \frac{e^{-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n}}{1 + e^{-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n}} \quad (14)
\]

\[
\text{odds}(\tilde{Y}_l) = e^{-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n} \quad (15)
\]

This implies that the odds ratio OR is equal to

\[
OR = e^{2(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)} \quad (16)
\]

Illustrative Example

Supposed that in university A, out of 30 identified factors for research publication and production, 10 were found to be normally distributed by Anderson Darling test and 4 of which i.e., advance degree of specialized discipline (x3), number of research collaboration (x9), research experience of the mentor (x13), and number of times a researcher’s work is cited by other researchers were found to be significant by the multiple linear regression. These four independent variables were re-subjected to multiple linear regressions with number of research publications and productions as (y), the dependent variable, to find out the beta coefficients of the corresponding variables. For instance, a regression analysis of this scenario are presented on the tables below.

A multiple linear regression was calculated to predict the number of research productions and publications (y) based on their educational qualification (x3), number of research collaboration (x9), research experience of the mentor (x13), and the number of times the researchers work is cited by the other researchers (x15). A significant regression equation was found $F(4,48) = 782.301$, $p < 0.000$ with an $R^2$ of 0.948. Participants predicted $y$ is equal to $-1.9562 + 0.3295 \times (x3) + 0.3200 \times (x9) + 0.0368 \times (x13) + 0.1572 \times (x15)$, where x3, x9, x13, and x15 are coded as continuum. Participant’s research production experience of the mentor (x13), and number of times a researcher’s work is cited by other researchers were found to be significant by the multiple linear regression. These four independent variables were re-subjected to multiple linear regressions with number of research publications and productions as (y), the dependent variable, to find out the beta coefficients of the corresponding variables. For instance, a regression analysis of this scenario are presented on the tables below.

A multiple linear regression was calculated to predict the number of research productions and publications (y) based on their educational qualification (x3), number of research collaboration (x9), research experience of the mentor (x13), and the number of times the researchers work is cited by the other researchers (x15). A significant regression equation was found $F(4,48) = 782.301$, $p < 0.000$ with an $R^2$ of 0.948. Participants predicted $y$ is equal to $-1.9562 + 0.3295 \times (x3) + 0.3200 \times (x9) + 0.0368 \times (x13) + 0.1572 \times (x15)$, where x3, x9, x13, and x15 are coded as continuum. Participant’s research production
and publication increases in 32.95% for each increase of educational qualification, 32% for each number of research collaboration, 3.68% for each research experience of the mentor, and 15.72% for each number of times other researchers cite the researcher’s work. The 94.8% variance of y is attributed to x3, x9, x13, and x15, which means that x3, x9, x13, and x15 are significant predictors of y. From the resulting multiple linear regression equation y = -1.96 + 0.330 x3 + 0.320 x9 + 0.0368 x13 + 1.16 x15, the probability model for hunting a research-doer in university A can be defined by substituting the value of y or zi to the logistic function, that is:

\[ P(z_i) = \frac{1}{1 + e^{(-1.96 + 0.330x_3 + 0.320x_9 + 0.0368x_{13} + 1.16x_{15})}} \]

where:

- \( P(Z_i) \) = is the probability of the ith faculty to do and publish researches in university A.
- \( X_9 \) = number of research collaboration
- \( X_{13} \) = research experience of the mentor
- \( X_{15} \) = number of times the researchers work is cited by the other researchers

Hence, if the applicant got the following scores in the four predictors respectively, i.e., 4, 3, 2, and, 1 then the probability of hiring that applicant is 82.54%. The applicant with the highest probability will be considered. Moreover, the odds of this faculty to do research are

\[ e^{-1.96 + 0.330x_3 + 0.320x_9 + 0.0368x_{13} + 1.16x_{15}} = e^{3.1072} = 22.36 \]

This formula is very easy to use by a non-mathematics expert. The best way of using this formula is the construction of spreadsheet in Microsoft Excel. In this manner, the hiring committee will just input the scores of the applicant in the different columns and the last 2 columns will automatically give the corresponding probability result and rank.

**Conclusion**

The logistic curve is a useful tool in determining the probability of a professor to do and publish research in the university. Finding the different predictors for research production and publication will result to the maximum number of research done and published in the university. Using a calculator or Microsoft Excel spreadsheet, the hiring committee can easily determine the probability whether an aspiring academic applicant is a research-doer and publisher or not. This method saves time and effort and defies the question, answer, and demo method of hiring university professors. It is more relevant and accurate because it is substantiated by documentary proof and analyses.

**Recommendations**

To establish reliability and validity, this formula must be tested to the different private and public universities in region 7. Findings of the study will enable each university to polish or redirect research policies most especially in the hiring process of qualified academics. To come up with a more conclusive result, a survey using a more comprehensive instrument must be conducted to the universities in Region 7. This will enable the researcher to capture all the necessary variables that affect research productivity and publication rate. In addition, more number of respondents delimits the possibility of rejecting the normality of data which is the basis for the multiple linear regressions. A study on the non-normal
distributions of data must be conducted to find out other intervening factors that cause the heterogeneity of the responses of the respondents. This may entail a discovery of another probability model for hiring a university faculty that follows the stochastic probability path.

## References Cited


