Sage Research Methods: Business

Using Linear Modelling and Predictive Analytics Make Future Decisions Based on Large Employee HR Datasets

For the most optimal reading experience we recommend using our website.

A free-to-view version of this content, which is easy to navigate and search, and includes interactive questions, is available by clicking on this link

Student Guide

Introduction

The following exercises are designed to help HR personnel and other aspiring researchers develop basic data analytic skills using Microsoft Excel spreadsheets and statistical software. Included are three scenarios that each present different statistical methods for answering common HR questions using employee data. HR personnel within large organizations are responsible for collecting, analyzing, and reporting on various data metrics. One area that has gained attention in the workplace is diversity analytics. To keep up with new regulations concerning equality, diversity, and inclusion in the workplace, HR personnel are tasked with collecting, analyzing, and reporting to their superiors on various diversity metrics. For example, organizations may be interested in knowing whether female representation is equal across different job roles, or whether certain ethnicities are paid less than other ethnicities, or even if turnover rates are higher amongst one demographic more than another. Properly coding and organizing various employee demographic data can be vital for organizations seeking to answer such diversity questions. However, to do so, one must understand how to statistically analyze data, interpret output tables, and report findings to superiors in laymen terms. For simplicity, the same excel dataset is used for all exercises in this document. Each exercise progresses in difficulty by including more variables and deeper analysis than the previous exercise. The present exercise uses prescriptive data analytics to 'predict' future outcomes.

Background – HR Predictive Analytics

Very few HR practitioners have the relevant competencies and skills to apply statistical techniques that enable them to make *predictive* inferences using datasets. Bersin et al. (2013) outline the importance of using predictive analytics to help organizations predict and understand the performance of a person (or indeed a group of people) based on available historical data. When we produce statistical reports without applying them for predictive or evaluative purposes, we simply get a 'snapshot' of what is happening at a current point in time. The current trend (and need) for HR practitioners is to develop predictive analytic skills. Predictive HR analytics can be defined as 'the systematic application of predictive modelling using inferential statistics to existing HR

people-related data to inform judgements about possible causal factors driving key HR-related performance indicators. The results of this modelling can be used (where appropriate) to make tangible predictions about results or people outcomes' (Edwards & Edwards, 2019). Research shows that HR analytics are important for business performance but that very few practitioners possess the skills necessary to conduct the relevant analyses (Kryscynski et al., 2018). HR analytics skills are increasing in demand due to their potential to create analytical models that help steer, adjust, and even drive business strategy. HR Analytics can assist with strategic organisational decision making.

Multiple Linear Regression

An MLR is used to collect data about a dependent variable (continuous) and independent variables (either continuous or categorical). MLR models the relationship between two or more independent (or 'explanatory') variables and a dependent (or 'response') variable by analysing many observations. The model can then be used to predict and inform future behaviours with a degree of statistical confidence. The equation and descriptions for a typical MLR can be found below:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i$
Y=Dependent variable

β_0: Intercept

β_i: Slope for X_i

X= Independent variable

Theoretically, think of plotting a number of data points on a graph to see if they form into some sort of straight line. The MLR is the line of best fit for those data points.

Illustrative Example: Predictive HR Analytics

Example 3: Using MLR to Model and Predict Intention to Quit and Job Satisfaction for ABC Insurance Co.

As mentioned above, ABC Insurance co. conducted an employee survey last year to better understand its employee's job satisfaction and their intention to quit. This was measured using a scale of 1–5. Imagine we are interested in understanding what other variables in our dataset could potentially influence our employee's job satisfaction and/or intention to quit. To do so, we could develop a predictive model.

Initial analysis and running the data through a statistical software may tell us that the correlation between job satisfaction and/or intention to quit is significantly related to some of our remaining variables. For example, imagine if we were interested in predicting whether Salary grade and Job evaluation impacted **Job satisfaction**. The correlation table (developed in Dataset II of this series) found a significant correlation of .355 (Salary grade) and .515 (Evaluation). Then a preliminary rough regression formula might look like:

Job satisfaction = .515 * × (Evaluation) + .355 * × (Salary grade)

Therefore, if Salary and Evaluation were the main predictors of Job satisfaction and we had the option to influence one or both of these variables (i.e., pay increase across the organization; or boosting evaluations), then what would we expect for job satisfaction?

Mathematically, we could predict this. For example, if we increased Salary grades by 1 level across the organization, we would expect and increase of job satisfaction of .355. The same for Evaluations, an increase in 1 point would increase job satisfaction by .515.

What this means is that the level of job satisfaction that someone feels increases by .355 or .515 for every one-point increase in Salary grade or Evaluation, respectively.

Predictive Analytics Exercises

Next, how can we informatively use this predictive model to increase job satisfaction at ABC Insurance co.? Well, our model suggests that for every 1-point increase in Salary Grade (on a 5-point scale), Job satisfaction increases by .263. And for every 1-point increase in Evaluation (on a 5-point scale), Job satisfaction should

increase by .462.

Look at these 3 illustrative examples:

a. An employee with a Salary grade of 1 (i.e., $20–40,000) and an Evaluation of 3 (i.e., *met expectations*) we get 1.098 + .263(1) + .462(3) = 2.747. This result is closest to '3' or 'neither satisfied nor dissatisfied'.
b. If our organizational target is to have Job satisfaction be a minimum of '4' (i.e., *somewhat satisfied*), by how many points would we have to increase Salary grade? Or, how much would we need to see an increase in Job evaluations?
c. Imagine ABC Insurance co. are concerned about employees with intention to quit of level 4 or above. It is your responsibility to make suggestions as to which strategies would be best to reach this goal. First, create an MLR model using the following variables: Intention to quit, Job Satisfaction, Age, Job evaluations, and Salary grade. Be sure to place your Independent/Dependent variables in the correct boxes using statistical software. Run the model (similar to above) and write out the formula using the beta coefficients and intercept. Set the 'Y' to a 4 and experiment with different values for the remaining variables. What would your suggestions to senior management be?

Review

Linear multiple regression is a useful tool for 'modelling' variables in a dataset. The model can then help to predict how changes in one variable, can affect a variable of interest. HR managers may be interested in understanding and predicting what factors contribute most to intention to quit, job satisfaction, etc. With large datasets at their disposal, HR managers (and others) can move beyond simply summarizing datasets (i.e., descriptive statistics) and move towards using them for predictive analytics.

Your Turn

You can download this sample dataset along with a how-to guide demonstrating how to create linear models to predict future outcomes. The how-to guide uses SPSS statistical software. The sample dataset includes 9 variables that you can explore with linear modelling. See whether you can reproduce the results presented here.