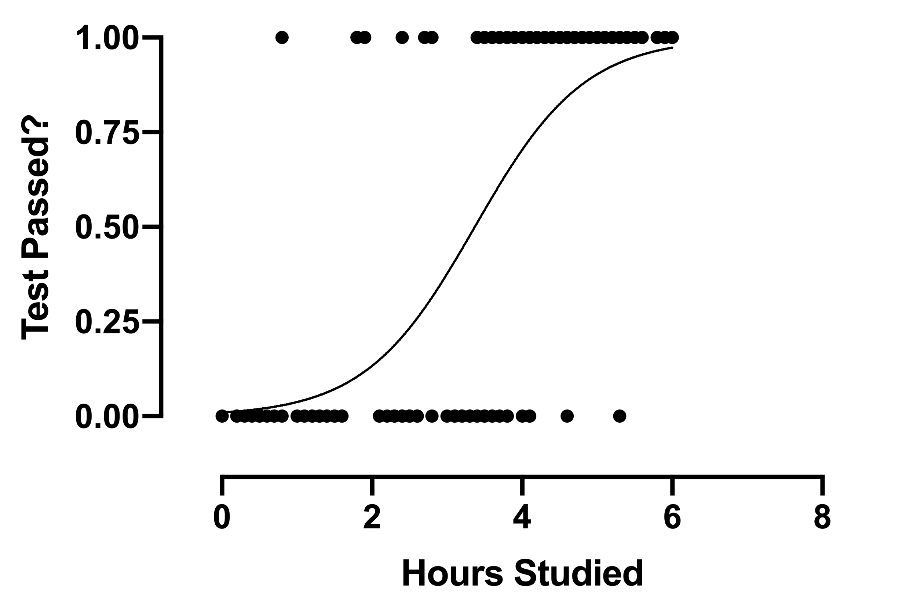
Last week I discussed the inner workings of linear regression, or the prediction of a variable’s value by using the equation of the least-squares fit line. The equation of the line was calculated to be that which minimized the sum of the squared residuals (the difference between the data and the equation output). While linear regression is a commonly used form of regression, it is not the only one. What if we wanted to predict whether someone will or won’t get a certain disease? What if we wanted to predict the number of days a certain event will occur? These questions can be answered by *logistic* and *Poisson* regression, respectively. I will discuss each type of regression and some of their use cases.

Logistic regression is used when the predicted (dependent) variable is a Boolean; that is, it is one of two values (true/false, pass/fail, etc). Its equation takes the form of the sigmoid function y = 1/(1+exp(Ax1 + Bx2 +…)), where each x represents an input variable and each letter represents the variable’s coefficient. Below is a visualization of a logistic regression function predicting if a student is likely to pass a test based on their studying. The data points correspond to a training data set of students’ hours studied and whether they passed the test. A logistic regression function outputs the probability of an event occurring (in this case whether a student will pass a test), which is classified as a 1 or 0 depending on a cutoff probability defined based on the application; usually around 50% is the cutoff. Below we can see that the more hours students study, the greater their probability is of passing.



Perhaps one of the most famous examples of logistic regression would be credit card fraud detection. Banks form a logistic regression function for each customer that takes input variables such as transaction date, place, amount, and type. The function then outputs the probability of the transaction being fraud which is classified as 1 (fraud) or 0 (not fraud). Another example would be predicting whether a tumor is benign or malignant. Variables such as tumor volume and blood flow to the area are input to the logistic function which outputs probabilities that are then classified as binary outcomes.

Poisson regression is similar to logistic regression; however, it is designed for predicting a wide range of discrete numbers. Unlike the logistic function, the dependent variable for Poisson regression is a range of integers rather than just 0 or 1, making it ideal for predicting variables such as the number of days an event will take or the number of children a woman might have. Because the output can only be a positive integer, the Poisson regression function is often called the log-linear function. Its function takes the form y = exp(Ax1 + Bx2 + …), where each x and the letters are the same as they were above. Rearranging the equation, we can see that log(y) is linearly correlated to the input variables, which makes sense since y must always be a positive integer.

Poisson regression appear in many places, not only in predicting time or number of children, but also in predicting the mortality of a diseased patient. For example, a patient’s age, sex, and time since diagnosis can be used to build a Poisson regression model that outputs an integer value for the patient’s mortality percentage. I could even use Poisson regression in my own line of work to predict the number of bacteria in my samples, a number that is very large but nonetheless still an integer. By building a Poisson regression model from data I’ve collected on temperature, pH, sugar, and fermentation time, I can find out what number of bacteria per unit volume I can expect for each of my experiments.

References:

https://ademos.people.uic.edu/Chapter19.html

https://magoosh.com/data-science/4-logistic-regressions-examples/#:~:text=Logistic%20Regression%20Example%3A%20Credit%20Card%20Fraud&text=When%20a%20credit%20card%20transaction,the%20transaction%20is%20a%20fraud.

https://www.researchoptimus.com/article/regression-analysis.php

http://www.pauldickman.com/software/strs/modelling\_poisson/