Combating Cancer:
The Critical Role of Computer Science in Cancer Biology

Landon Tally, Andra Johnson, Nykara Brown*, Keeley Cleghorn, and Catherine E. Steding
Department of Biology, The Center for Genomic Advocacy, Indiana State University, Terre Haute, IN.

Abstract

Breast cancer kills thousands of women yearly, and effectively combatting this disease requires collaboration from more than one discipline. This project emphasizes the essential role of computer science in the evaluation of cancer therapeutics. Evaluation on cellular consequences of a given treatment requires the utilization of sub-lethal doses. To find the appropriate treatment parameters, knowledge of both statistics and computer science is essential. This project utilized R programming and excel to create a logistic regression model for data analytics. This work provides the foundation for future analysis of the consequences of individual therapeutic interventions allowing for highly reproducible results across multiple experiments.

Background

.

- The data comes in the form of an excel file which contains the absorbance ratio and the drug concentration.
- A logistic regression was used to accurately calculate the sublethal dose for the phytol derivative drug.
- The model compares the absorbance of the drug to the concentration of the drug to generate a dose-response curve.
- The methods performed were the logit and probit regression models and the five-parameter logistic model.

Regression

Regression analysis is used to find trends in data. The relationship between a dependent variable and independent variable(s), or predictors, is often the focus of this statistical model.

Logit and Probit Regression
They estimate that the probability of an observation with particular characteristics will fall into a specific category, and moreover use a binary classification model.

N-Parameter Logistic Regression
[1]N-Parameter Logistic Regression (NPLR) fits dose-response data to easily accommodate asymmetry in data, such as assays, in its predictive model.

Methods

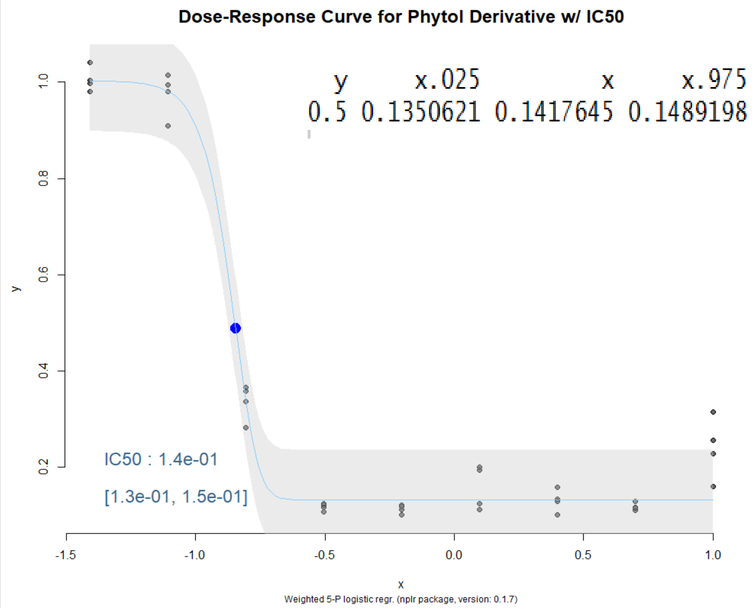
```
#Sets working directory for personal use
setwd("C:/Users/ltally/Documents/R/Datasets")

#Reads in the excel worksheets
drug <- read_excel("MBPADrug2-48Hours.xlsx",
                   sheet = "Drug Treated Plate 48-Hour", na = "NA
#Reshapes the information provided into something more usable; rename
drugMelt <- melt(drug)
colnames(drugMelt) <- c("Column", "Absorbance")

#Adds a new column for the Concentration
drugMelt <- add_column(drugMelt, Concentration = c(0, 0, 0, 0, 0.039,

#Creates a new dataset excluding the values at 0 concentration since
drugMelt2 <- drugMelt[5:40, ]

#Fitting the model w/o replicates
np1 <- nplr(x = drugMelt2$Concentration, y = drugMelt2$Absorbance)

np1

#Visualizing the model
plot(np1, cex.main = 1.2,
     main = "Dose-Response Curve for Phytol Derivative")

plot(np1, pcol="grey40", lcol="skyblue1", xlab = "Concentration", ylab = "Absorbance",
       main="Dose-Response Curve for Phytol Derivative w/ IC50", cex.main=1.5)
```

Figure 1. Melded Data in the NPLR Package. This was the code that was used in order to calculate the sublethal dose of the phytol derivative that was produced here at ISU. The data is read into the program, melded into two columns, the addition of the concentrations column is created, and then plugged into the NPLR function to produce a plot visualizing the location of the sublethal dose in terms of concentration by absorbance.

Conc, C	log(C)	Alive	Dead	Prop, P	Corr, P	Logit(P)	Probit(P)
0	#NUM!	0.991	0.000	0.000	0.000	#NUM!	#NUM!
0.039	-1.409	1.005	-0.015	-0.015	-0.015	#NUM!	#NUM!
0.078	-1.108	0.975	0.016	0.016	0.030	-3.479	3.124
0.156	-0.807	0.336	0.655	0.661	0.666	0.675	5.429
0.3125	-0.505	0.118	0.873	0.881	0.883	2.007	6.191
0.625	-0.204	0.113	0.877	0.886	0.887	2.049	6.213
1.25	0.097	0.158	0.833	0.841	0.843	1.666	6.007
2.5	0.398	0.131	0.860	0.868	0.870	1.886	6.126
5	0.699	0.118	0.873	0.881	0.883	2.003	6.188
10	1	0.240	0.751	0.758	0.762	1.146	5.711

Figure 2. Logit and Probit Model. This is the data that was derived from using logit and probit prediction models to figure out the sublethal dose of the phytol derivative. The absorption rate indicates the amount of cells alive in each well. The control was assumed to be 100% of the cells alive. A proportion of dead over total was created for each concentration, a corrected proportion was created to assume that any external factors altered the absorption for the control seeing as how the smallest dose of the drug contained a higher value, and the natural log of the corrected proportion over one minus the proportion was used to find the logit value. The probit value was created using five plus the standard normal distribution of the corrected proportion. The logit and probit values were compared to a test value of 0 and 5 respectively. The three numbers surrounding these test values were used to create a predictive slope and intercept to discover the sublethal dose of the phytol derivative.

Results



	Logit	Probit
Slope	9.100902	5.086494
Intercept	7.075724	9.017584
Test Value	0	5
Log(C%)	-0.77747	-0.78985
LC50	0.166926	0.162236

Figure 3. Sublethal Dose Response for Phytol Derivative Drugs at 50%. The value of the sublethal dose using the logit and probit models were a little high when comparing the results to the proportion at 0.156 concentration. However, the results of the NPLR seem much more concise at 0.14 concentration.

Future Direction

- To provide a foundation for reproducibility and complete accuracy.
- The goal is to create a web application using the R package Shiny that will produce all of this work with no R programming knowledge required. This will allow researchers to input files and get immediate results.

Acknowledgements

The Center for Genomic Advocacy
Big Data 4 ISU
SURE 2018
Department of Biology, Indiana State University
National Institutes of Health
Indiana State University Center for Student Research and Creativity

Dr. Inlow
Dr. Kinne
Dr. Gonser
Dr. Latimer

*Nykara Brown provided the phytol derivative data using a methylene blue assay.

[1]Commo, Frederic, and Briant M Bot. *R Package Nplr n-Parameter Logistic Regressions*. 25 Dec. 2016, cran.r-project.org/web/packages/nplr/vignettes/nplr.pdf.
[2]Currell, Graham. *Scientific Data Analysis*. Oxford University Press, 2015.