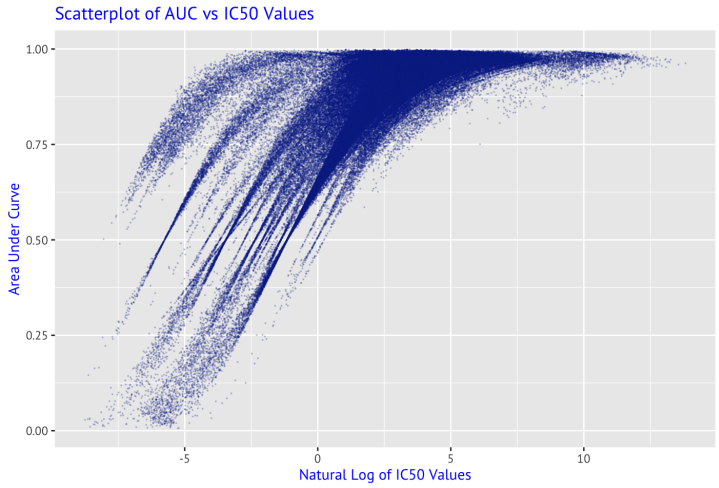
IC50 vs AUC Exploratory Data Analysis
Tianyi Chen

Introduction:

IC50– the half-maximal inhibitory concentration– is an important measurement in pharmacology, representing the concentration of a drug required to reduce cell viability by 50%. In cancer research, IC50 measures how much drug is needed for the biological/biochemical process of a cancer cell to be inhibited by 50%. As such, IC50 is a useful measure of assessing cell response to a particular drug. The lower the IC50 of a cell is in relation to a drug, the more effective the drug is on the cell; i.e. the more responsive a cell is to the drug.

AUC, or Area Under the Curve, is another pharmacokinetic measure of drug effectiveness. Often visualized as drug concentration over time, AUC helps to quantify how much drug is present in the cells after a dosage is given. A higher AUC indicates that the drug has stayed present in the cells for a higher concentration or longer period of time. An extremely high AUC may lead to toxicity, so the optimal AUC must be found specific to each type of cancer cell.



Visualization: Utilizing the Genomics of Drug Sensitivity in Cancer (GDSC) dataset on Kaggle, let's analyze a potential relationship between IC50 and AUC. It should be noted that the data takes the natural logs of the IC50 values to compress the scale and normalize the data. Utilizing R Studio, we have extracted the data pertaining to IC50 and AUC and created a scatter plot of AUC vs IC50. Based on the initial graph, it looks like there is a general trend of AUC vs IC50 in which AUC increases as IC50 increases. *This suggests an inverse relationship between drug efficacy and cell response. As cell response to a drug decreases (higher IC50 values), the more time the drug has stayed within the cells (higher AUC values) without being effectively inhibiting cell viability.*

Summary Statistics:

Based on summary statistics, the IC50 values are slightly left skewed as the mean (2.817) is lower than the median (3.237). 50% of the IC50 values fall between 1.508 and 4.700. AUC values range from 0.006 to 0.998, indicating that some drugs are fully retained while some have negligible exposure to the cancer cells. With a mean lower than the median, the AUC values are slightly left skewed as well. This is consistent with the trend we see in the graphical representation; we see a large portion of values between (-5, 0.50) and (10, 1.00); with a trailing "tail" of values below, representing drugs with lower effectiveness and retention times.

```
     LN_IC50              AUC        
 Min.   :-8.748   Min.   :0.006282  
 1st Qu.: 1.508   1st Qu.:0.849452  
 Median : 3.237   Median :0.944197  
 Mean   : 2.817   Mean   :0.882593  
 3rd Qu.: 4.700   3rd Qu.:0.974934  
 Max.   :13.820   Max.   :0.998904  
```

Next Steps:

Moving forward, it would be interesting to try and fit first a linear model to the data; although, based on the general trend of the data observed in the visualization, a linear model would likely not yield the best fit. The linear model could still be experimented with, though, and the R-squared value evaluated. Because the data is biological in nature, it is possible that a supervised exponential or logarithmic regression model would better capture underlying relationships. All three models– linear, exponential, and logarithmic– can be implemented in R Studio. We could also potentially experiment with taking the log of AUC values to create more normalized data.