EC50の計算方法について

Tatsuo YAMAGUCHI

2021-12-06

## 読み込みパッケージ

library(tidyverse)  
library(drc)  
library(plotly)  
library(rgl)  
library(minpack.lm)

## はじめに

用量反応曲線はシグモイド曲線（S字曲線）の形を取る。 シグモイド曲線を取る曲線として以下のものが上げられる。

* シグモイド曲線を取る曲線
  + 正弦曲線
  + ロジスティック曲線
  + プロビット曲線

プロビット曲線

このうちよく使われているものはロジスティク曲線とプロビット曲線である。

* ロジスティック曲線

説明変数と出現率の関係がロジスティック曲線で表される曲線

以上の数式をに関する指数式に変換すると、

と表すことが出来る。

* プロビット曲線

説明変数と出現率の関係が累積正規分布曲線で表される曲線

## Model選択の際の統計的手法

モデル選択の際、尤度比検定もしくはAIC（赤池情報量規準）が用いられる。

### 尤度とは

確率密度関数に変数に観測値を代入したときに、その確率密度を観測値で評価したときの度合い。つまり。観測値の確率変数に近い「もっともらしい度合い」である。

尤度を目的変数とする関数のことを尤度関数という。

尤度関数の自然対数をとったものを対数尤度関数といい、対数尤度関数あるいは尤度が最も大きくなる最尤推定量、あるいは最尤推定量となり得るパラメータ（最尤推定値）を探索する。

### Wald統計量

推定値を標準誤差の推定値で除した値。平均が、Wald統計量の絶対値、標準偏差が1の正規分布におけるから0までの値をとる確率の2倍が0に近ければ推定値がゼロに近いと見なすことが出来る。

### 尤度比検定

尤度比検定について説明する前に、逸脱度についてせつめいする。

逸脱度（）とは「あてはまりの悪さ」の指標で、最大対数尤度を用いて、

と表される。

ここで最小逸脱度（Full model時の逸脱度）との残差逸脱度と定義する。

注：Full modelとは目的変数の個数だけ観測変数のパラメータをあてはめたものをイメージすればよい。

そして、尤度比検定とは他のモデルの残差逸脱度との差

を検定統計量として用いる検定のことである。

この検定統計量を評価できる方法はパラメトリックブートストラップ法という乱数発生によりシミュレーションする方法とparameterの数の差を自由度とする検定を用いた近似計算方法が用いられる。

「正規分布を仮定出来るなら」、その結果は最小二乗法の推定結果は正規線形モデルに従う最尤法の結果と一致する。そのため逸脱度を最小二乗法での指標である残差平方和を置き換えて、F検定で検定できる。

この場合、残差平方和の差の増分が自由度数（データ数ーパラメータ数)の差の増分より大きいかどうかの判断する。つまり残差平方和の差を自由度数（データ数ーパラメータ数)の差で除したF比

DF:自由度

が1.0より大きくなるかどうかをF検定により判断する。

#### AIC（Akaike Information Criterion）

AICとはモデルの複雑さと、データとの適合度とのバランスの良いモデルを選択するための基準である。

k:パラメータの数(ｋには定数項もパラメータ数として含む)

と表される。このAICの値が小さいほど「あてはまりが良い」モデルであると言える。

そこでモデルのAICの差は

「正規分布を仮定出来るなら」、その結果は最小二乗法の推定結果は正規線形モデルに従う最尤法の結果と一致する。そのため逸脱度を最小二乗法での指標である残差平方和を置き換えることが出来るため、

と表すことが出来る。

# EC50計算の具体的方法

## 手計算で求められる簡便な方法

### Reed and Muench法

比例配分法によってを求める。

ただこの方法は用量反応関係を純粋に直線になるようにplotする方法であり、plotした0と100の部分がやでない可能性がある。そのため、次に説明する方法で求めることが多い。

## ロジスティック曲線を用いた近似

基本的に近似式で使われるモデルは以下のものが上げられる。

#### 3パラメータモデル

Topは最高値。Hill係数はロジスティク曲線の傾きの度合いを表し、は右辺の値が最高値と最低値のr％になったときのxである。

#### 4パラメータモデル

Baseは右辺の最低値、Topは最高値。Hill係数はロジスティク曲線の傾きの度合いを表し、は右辺の値が最高値と最低値のr％になったときのxである。

を算出する場合は

となる。 これを非線形最小二乗法という方法で残差平方和が最も小さくなるパラメータの値を探索する。GraphPadPrismでは探索の方法として、Marquardt法を採用している。

### 例

以下のロジスティクモデルを仮定して、パラメータ推定を行う。

具体的なやり方を以下で説明する。

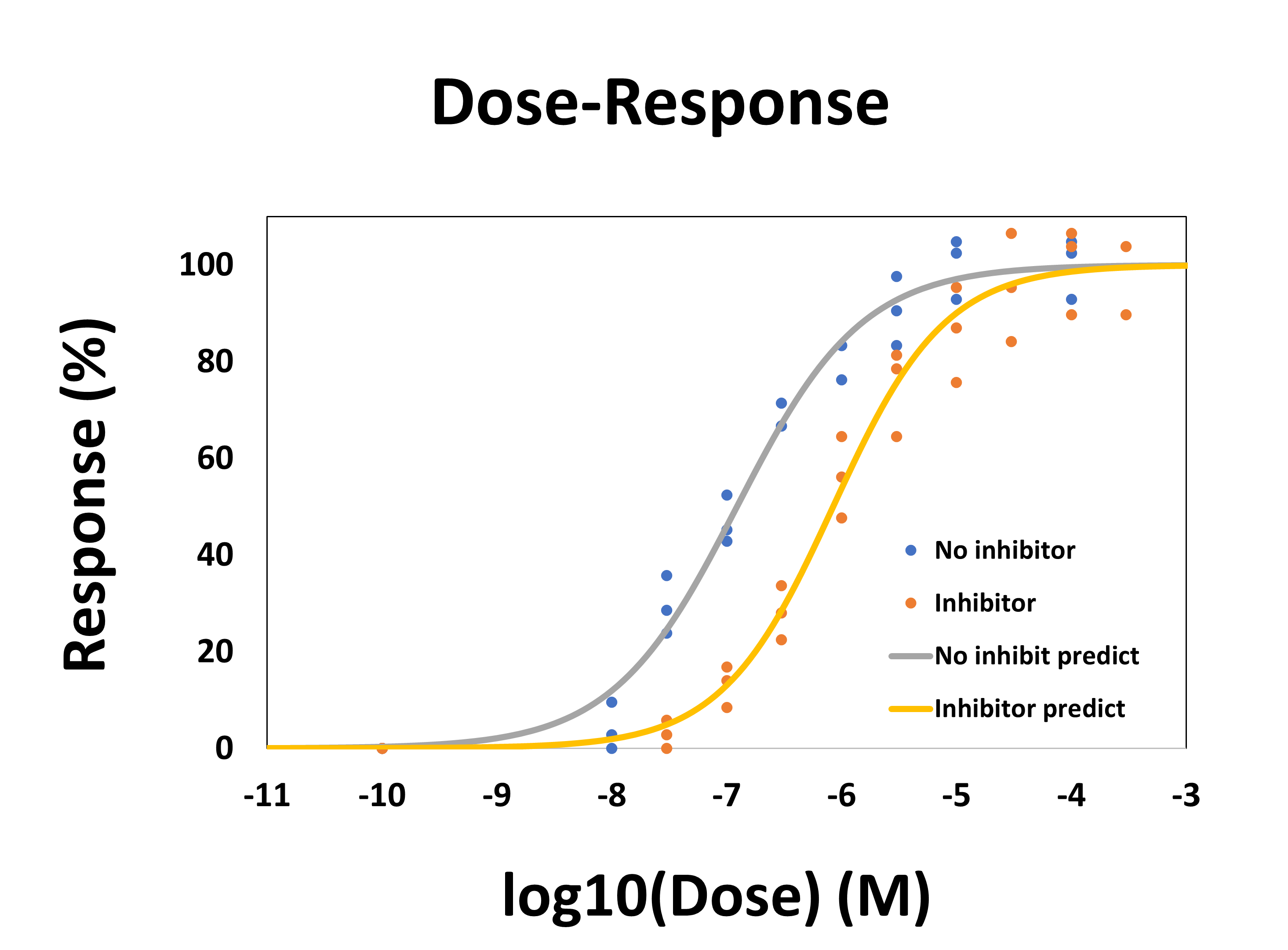
データセットは以下の通りである。

dat<-read.csv("D:/VBA/Nov\_25th\_EC50\_training.csv",header=TRUE)  
dat<-dat[,-1]  
dat

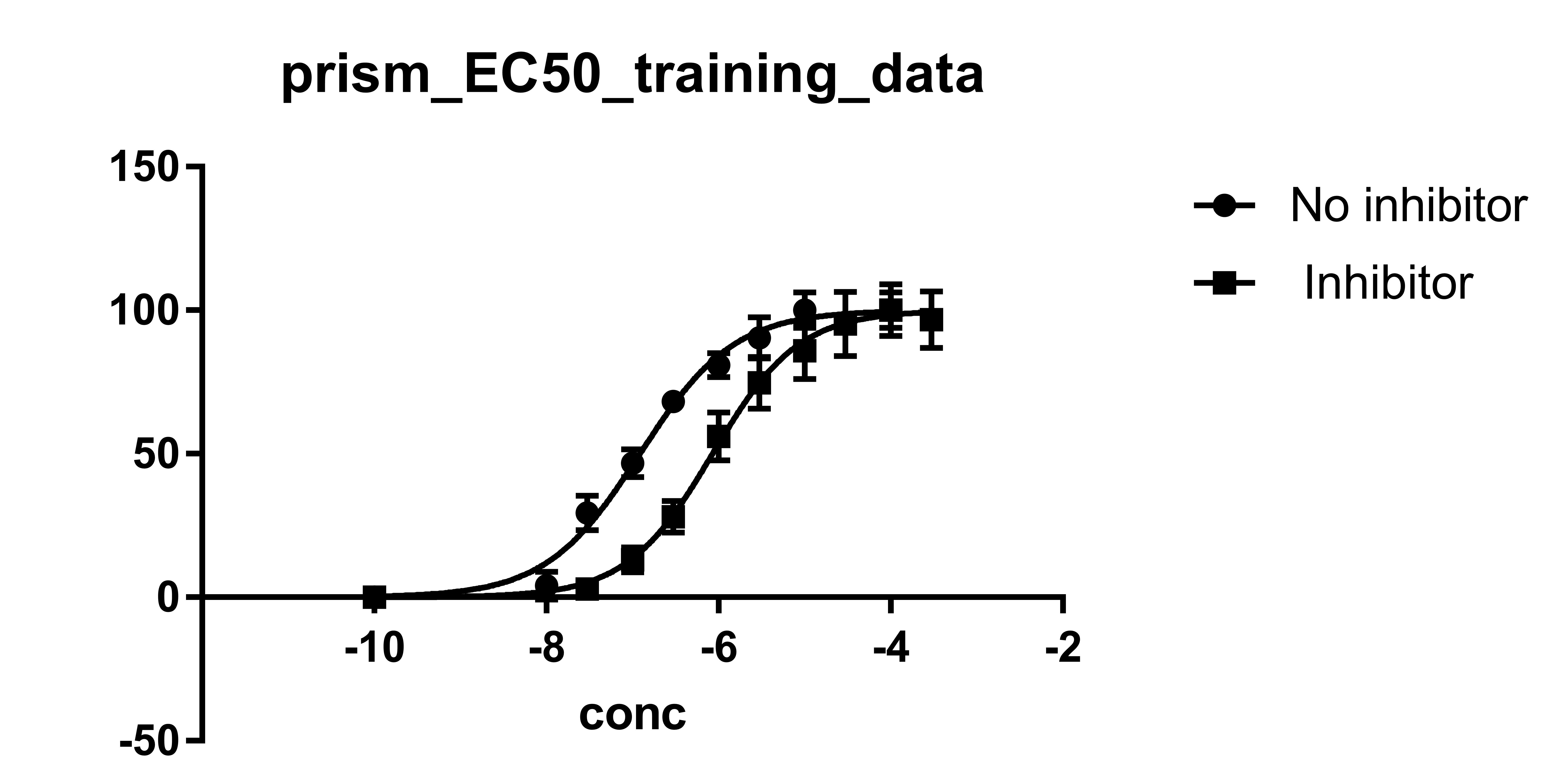
## conc No.inhibitor No.inhibitor.1 No.inhibitor.2 Inhibitor Inhibitor.1  
## 1 -10.000 0.000 0.000 0.000 0.000 0.000  
## 2 -8.000 2.831 0.000 9.524 NA NA  
## 3 -7.523 28.571 23.810 35.714 2.804 0.000  
## 4 -7.000 45.238 42.857 52.381 14.019 8.411  
## 5 -6.523 66.667 71.429 66.667 22.430 28.037  
## 6 -6.000 76.190 83.333 83.333 47.664 56.075  
## 7 -5.523 83.333 97.619 90.476 64.486 78.505  
## 8 -5.000 92.857 102.381 104.762 75.701 86.916  
## 9 -4.523 NA NA NA 84.112 95.327  
## 10 -4.000 92.857 102.381 104.762 89.720 103.738  
## 11 -3.523 NA NA NA 89.720 103.738  
## Inhibitor.2  
## 1 0.000  
## 2 NA  
## 3 5.807  
## 4 16.822  
## 5 33.645  
## 6 64.486  
## 7 81.308  
## 8 95.327  
## 9 106.542  
## 10 106.542  
## 11 NA

ExcelやRなどで解析する際は、予め薬物濃度を常用対数に変換して解析することをお勧めする。

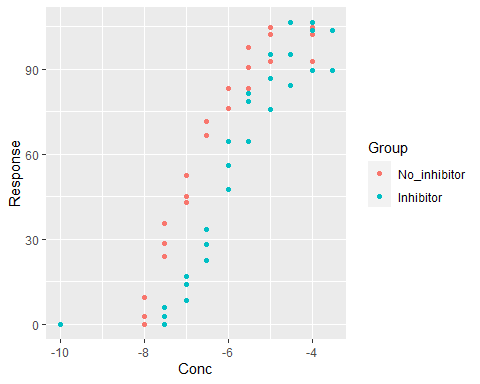
まずはグラフをplotしてみて、大方のhill係数やEC50 の幅を推定する。 というのも、出力されたパラメータが、本来最も最適なパラメータではない、いわゆる局所最適解を示す可能性があるためである

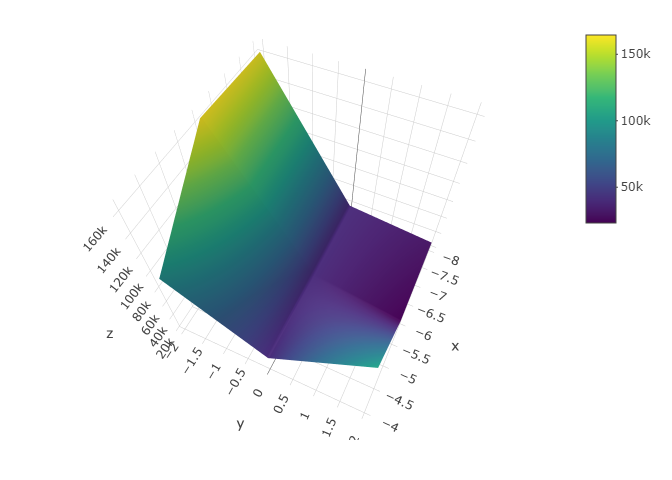


ExcelによるPlot



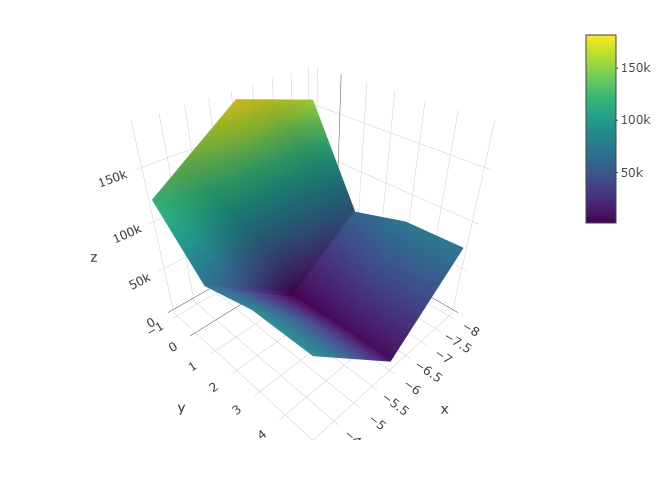
Rは以下の通りである。





No inhibitor

この上の図はNo inhibitorの残差平方和の大きさを表した図である。X軸がEC50,Y軸がHill係数、Z軸が残差平方和を示している。EC50が6、Hill係数が0付近に解があることが分かる。

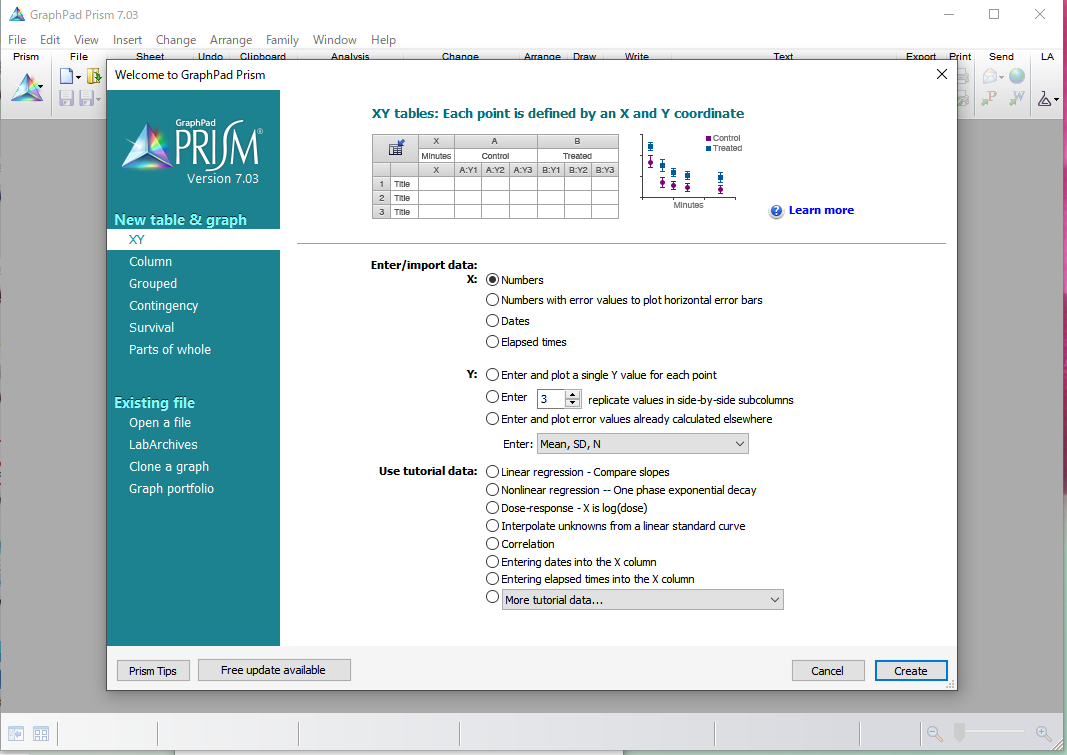


Inhibitor

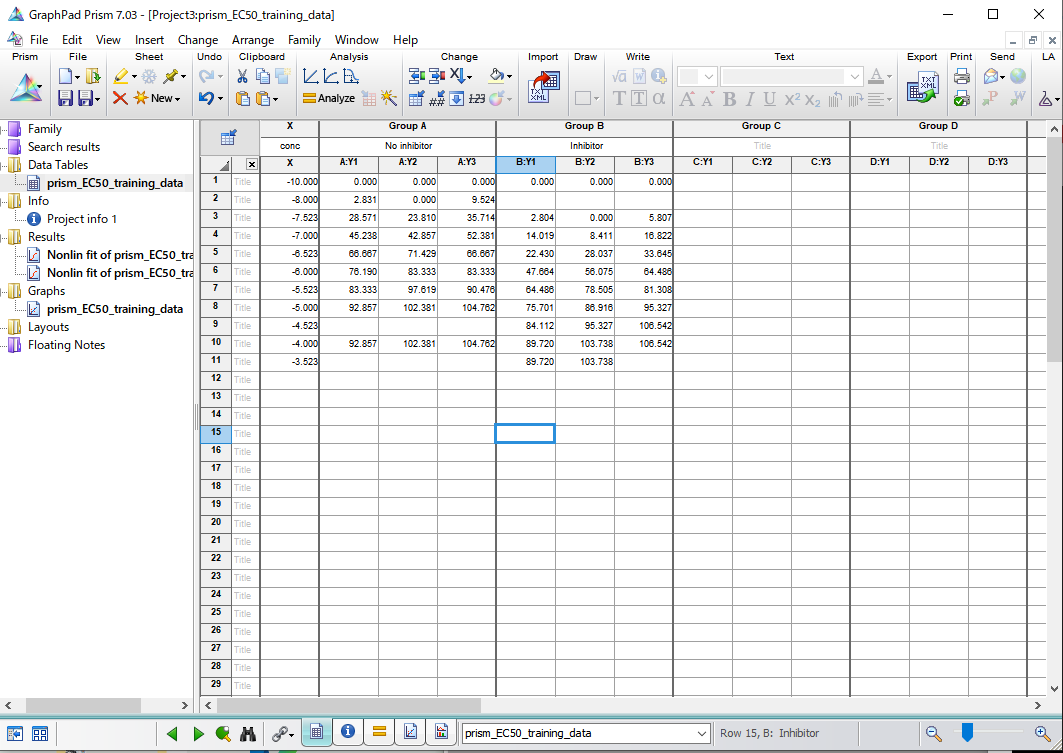
上図はInhibitorの残差平方和の大きさを表した図である。X軸がEC50,Y軸がHill係数、Z軸が残差平方和を示している。EC50が6、Hill係数が0付近に解があることが分かる。

### Graphpad Prismを用いる方法

GraphPad prismで非線形回帰を使うためには、データ読み込みの際に「XY data」を選択する。

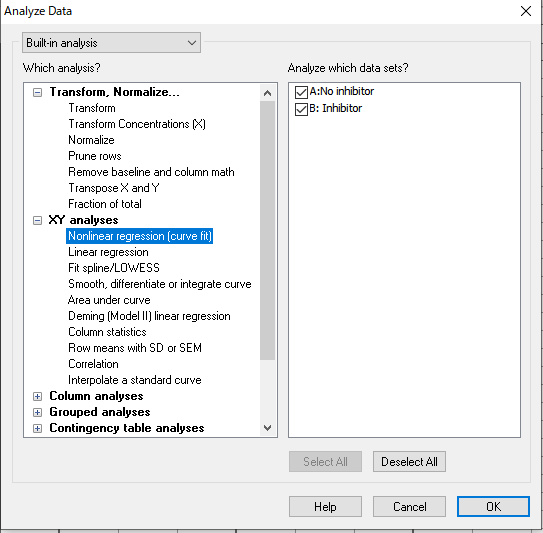


prism\_datainput



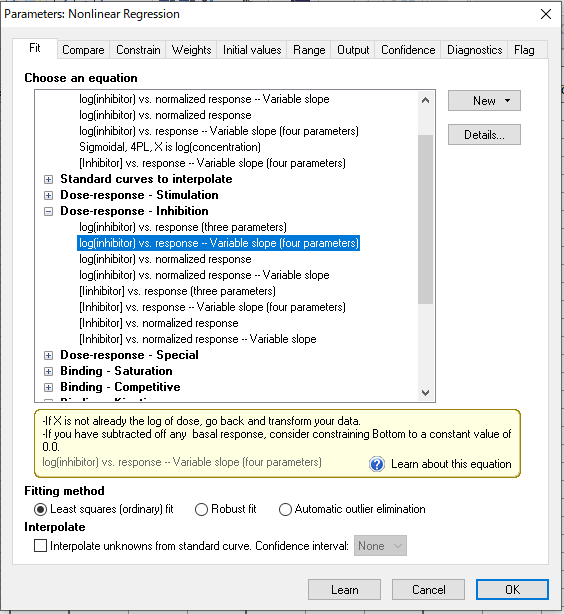
EC50\_data

「Analyzing Data」タブ

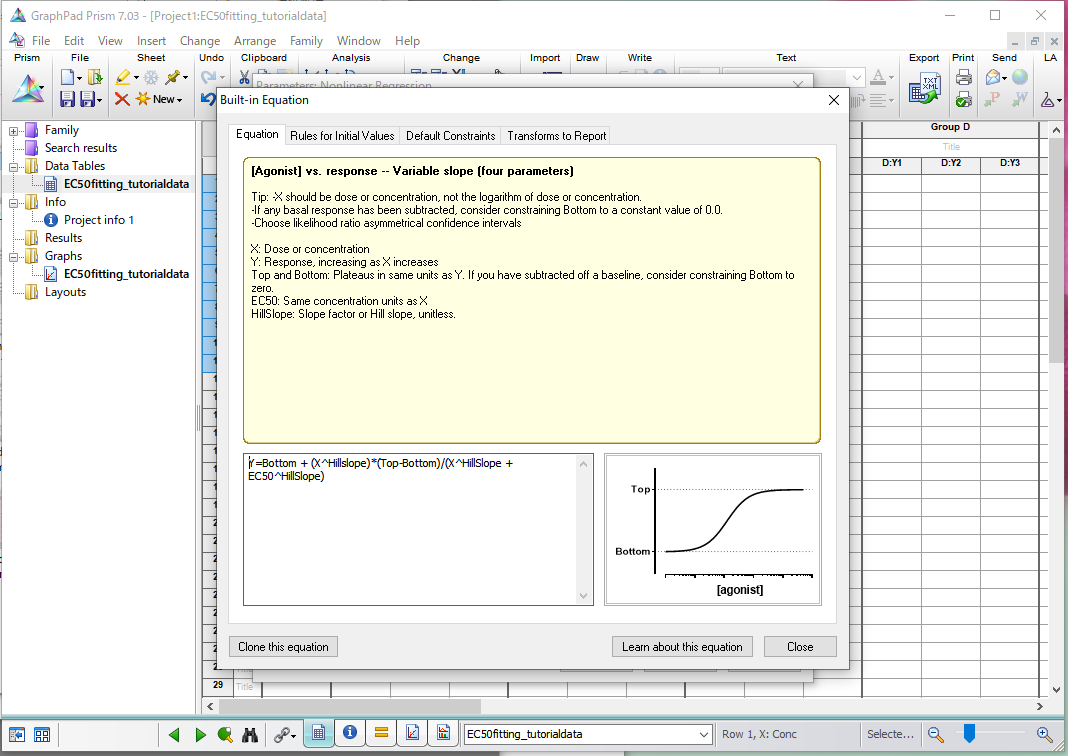


modelselect

そして、このうち最適なequationを選択する（今回は vs responsed -Variable slope (log parameters：これは4パラメータモデルで使うequationである。)

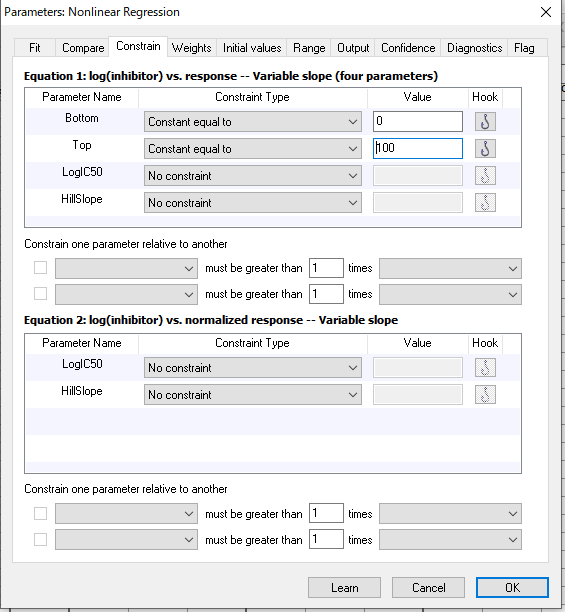


細かいデータは数式などはDetailをクリックして確認する。



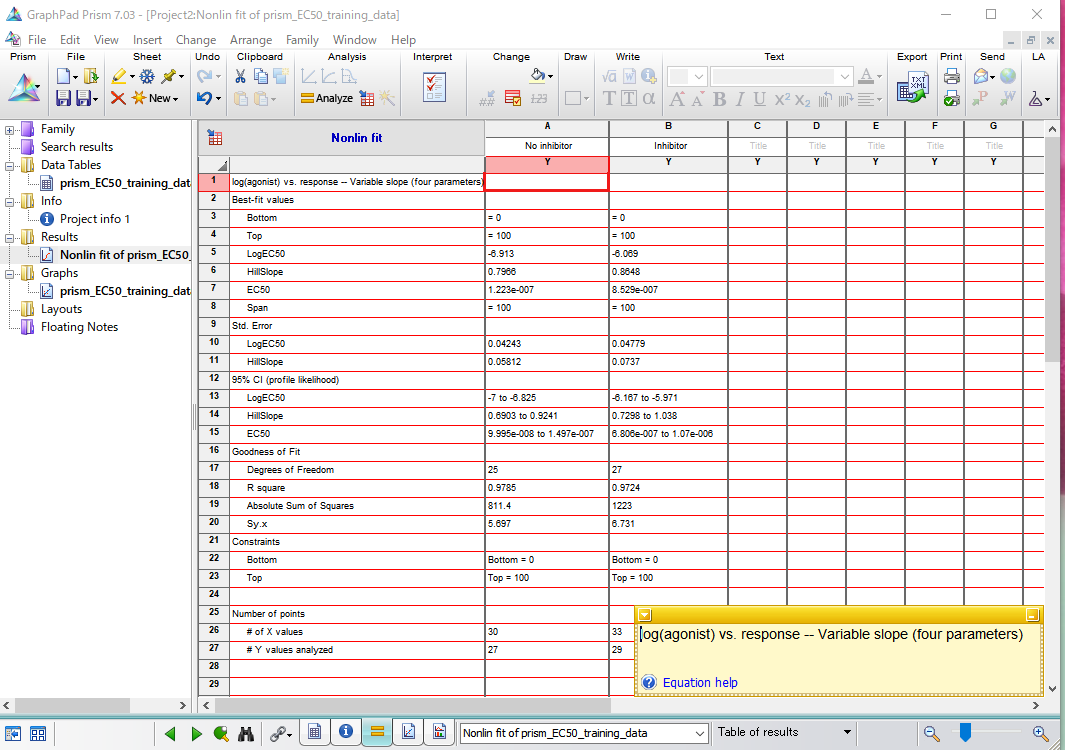
Model select2

TopとBaseの差が100なので、Bottomに0、Topに100を入力。



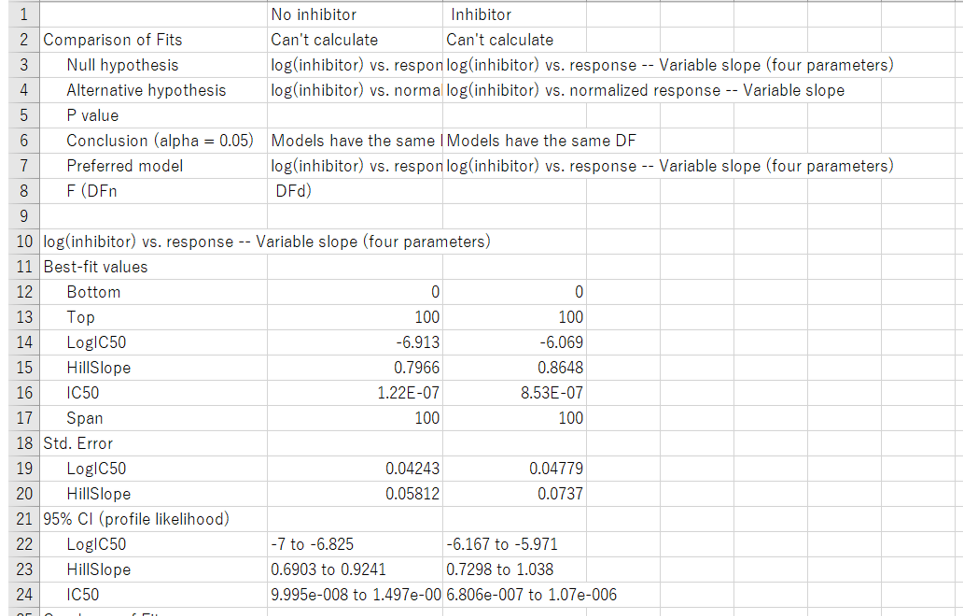
定数値の設定

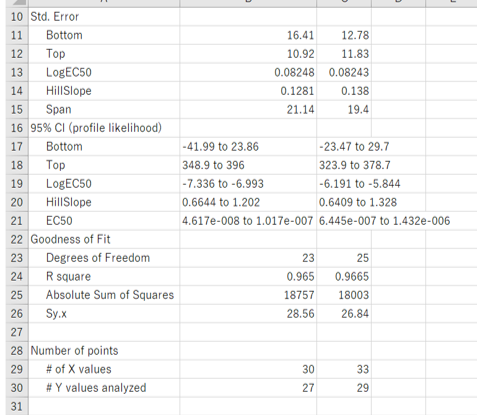
すべての設定を終えるとOKをクリック。すると、



Prism結果

Prismによる推定結果は以下のようになる

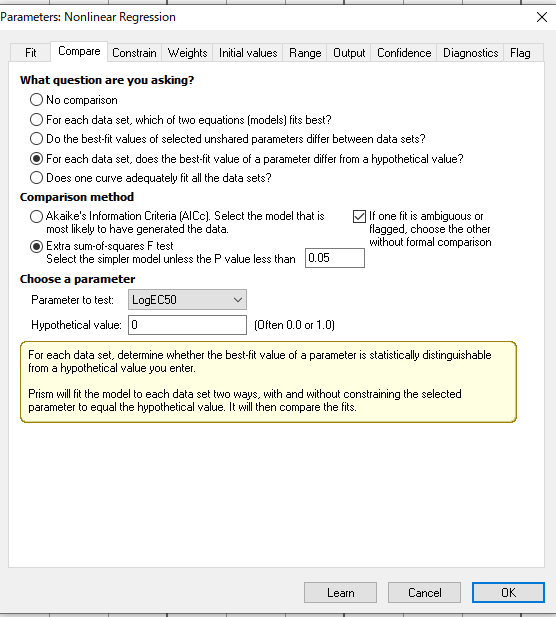




そして、結果よりNo No inhibitorとInhibitorのの95%信頼区間が重なっていないため、Inhibitorでは抑制効果があることが分かった。 またNO inhibitorとInhibitorの間でhill係数の95%信頼区間がで重なっているため、用量反応関係が相似であり左右の平行移動のみであることが読み取れる。

では、Top、Bottom、Hill係数が二群間で同一の場合のモデリングを 上記と同じ手順で行う。

上記と一点違うのは定数値の設定で、Hill係数の設定を「All data shared」にすれば良い。

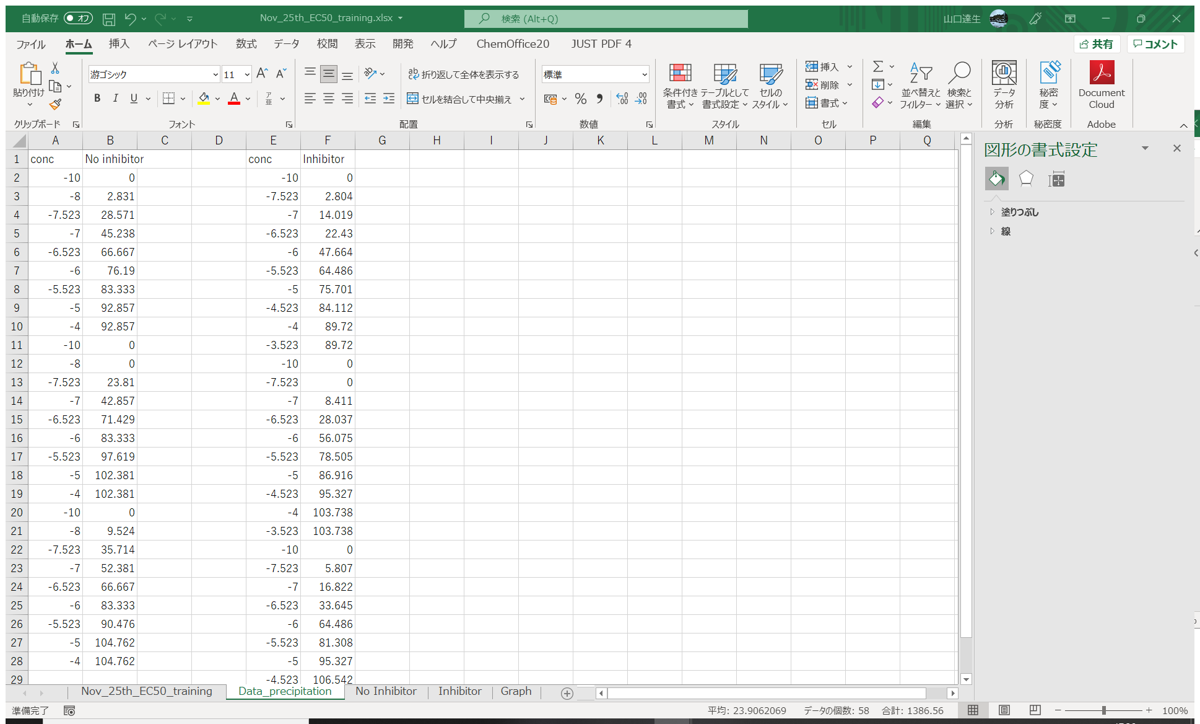


ちなみにTop、Bottom、Hill係数が二群間で同一の場合を仮定した時は、から効力比を推定出来る。

ただ効力比を推定する場合には、だけでなくやなども計算した方が良い。

### Excelを用いる方法

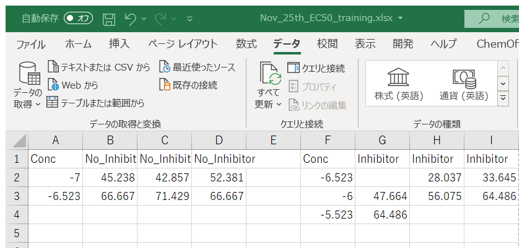
データの読み込みが終わった後は、欠損値を除いていく。



欠損値処理後Data

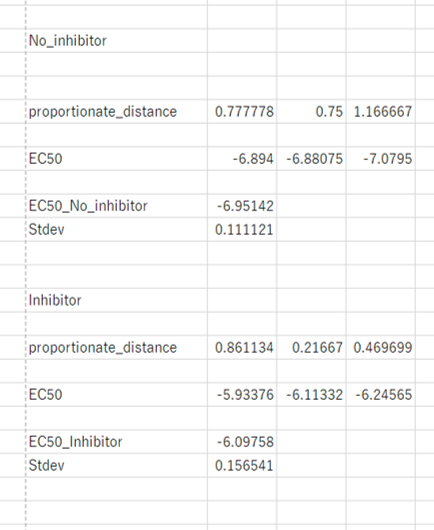
そして、No inhibitor、Inhibitorのパラメータをそれぞれ推定に入る。 まずはReed-Muench法で算定する。

まずは50%近傍のデータを抽出してきて、



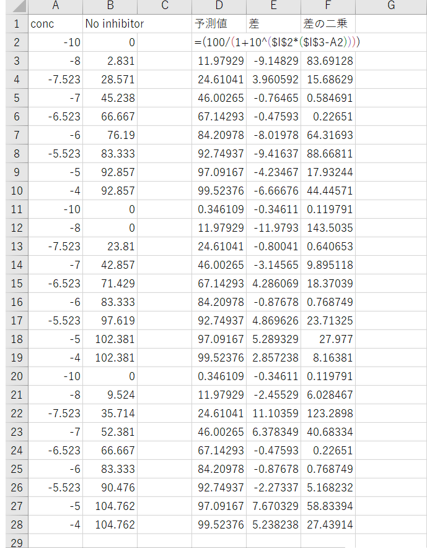
No inhibitor

PD;proportionate, distance



次にロジスティク曲線を用いた近似計算方法について説明する。 まずI2をhill係数、I3をEC50のセルとして、それぞれ初期値を、と入力する。 次にDの列に

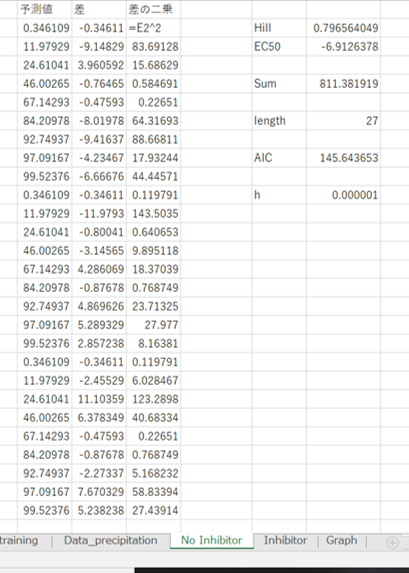
となるように数式を打ち込み、



Excel\_pred

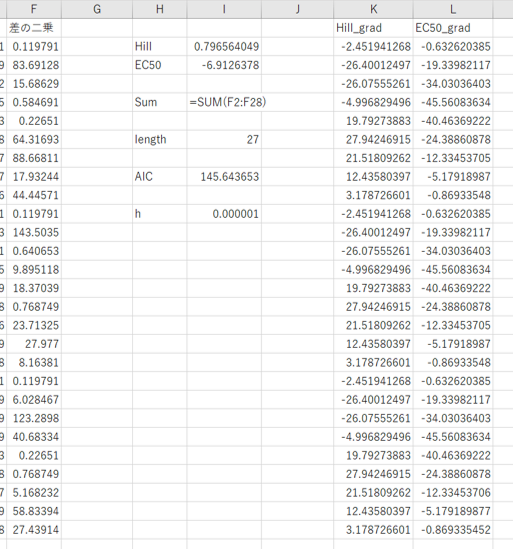
隣のE列にResponseと各予測値との差(Residue)を求める。

そしてその横の列に各Residueの2乗を求める。



Excel\_Residue

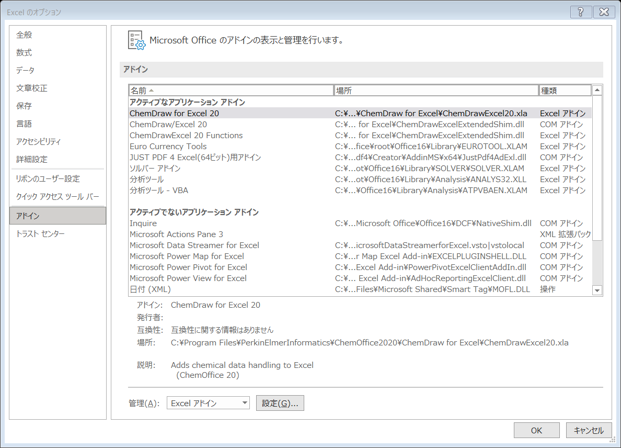
そして、その残差平方和を足し合わせる。



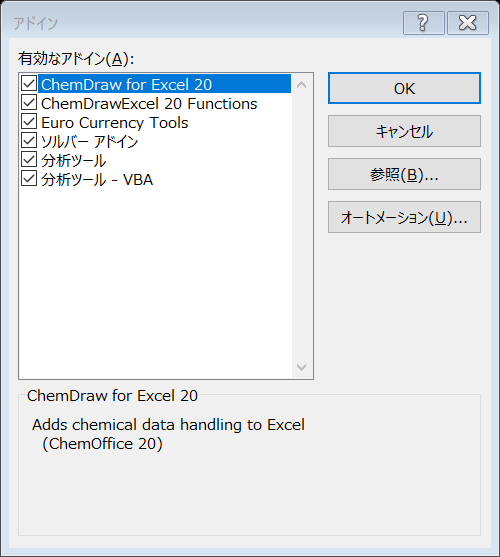
Excel\_Residue\_sum

この残差平方和が最も小さくなるパラメータを見つけるというものが非線形最小二乗法の原理である。 Excelではソルバーを起動して、この残差平方和が最小になるような条件を探索する。

まずソルバーの追加はファイルのタブのオプションから行う。

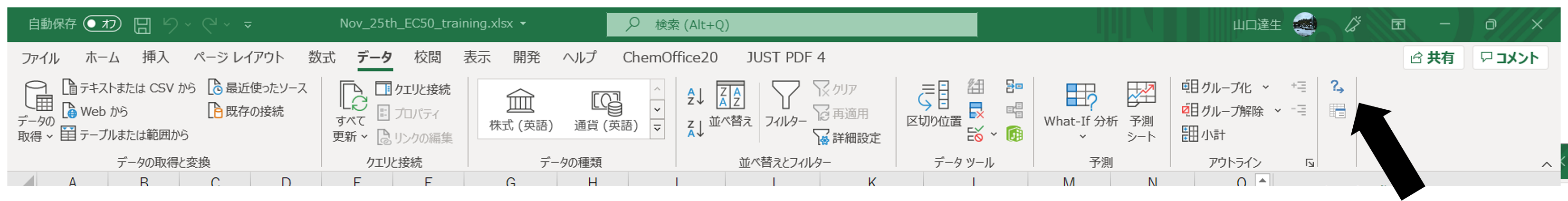


Excel\_addin\_1



Excel\_addin\_2.

ソルバーはデータのタブの右端に存在するところから起動する。

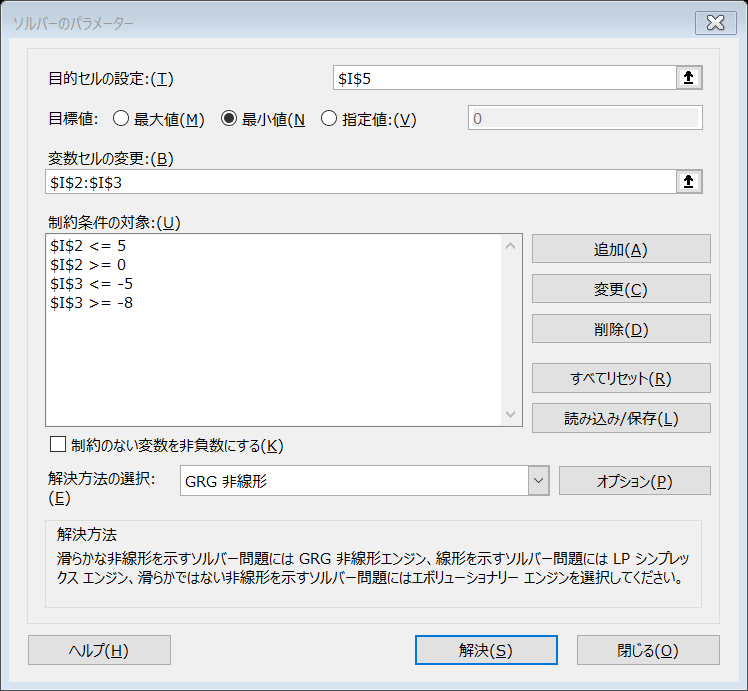


Excel\_solver起動

このソルバーによる探索の際、予め制約条件を入力するとより正確なパラメータを推定しやすくなる。

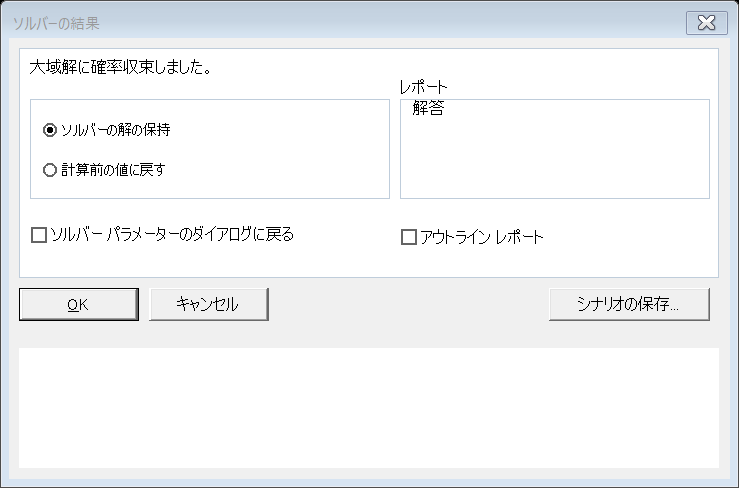
今回は

の条件、解決方法「GRG非線形」で探索を行う。



Excel\_solver\_param

するとI5の値が最小化され、I2、I3に入力したhill係数、EC50の値が求まる。（注：このとき、大域解でなく局所最適解が出力され、グラフから予測されるパラメータと比較しおかしいパラメータが出力される場合がある。このときは制約条件変更などを行うこと。）



Excel\_Solver結果

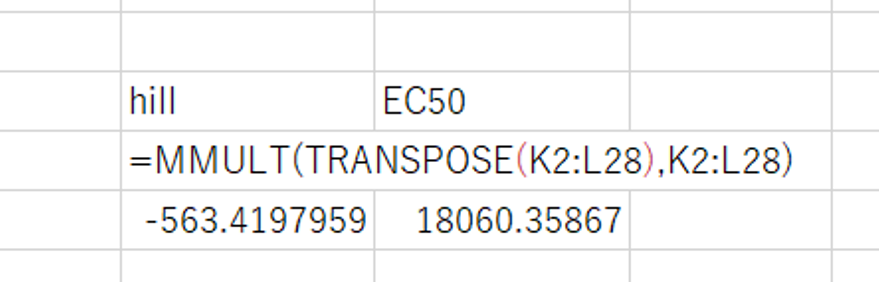
そして、その横のK5、L5に今回の関数をhill係数、EC50のパラメータで微分した勾配()を入力する。 入力する計算式は、

Hill係数の勾配関数

の勾配関数

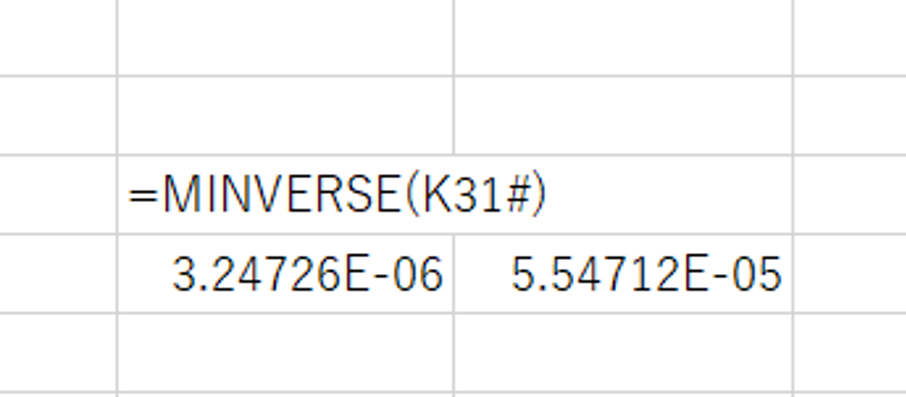
である。（ちなみに勾配は中心差分で計算している。）

実際には I11にと打ち込む。 そして、各xについて勾配を計算する。 その後、計算した各勾配を転置させた行列と転置させていない行列を掛け合わせる。



Excel\_MMULT

そして計算した行列の逆行列を求め、



Excel\_MINVERSE

そしてソルバーで最小化した残差平方和（Residue）を自由度（今回はデータの大きさからパラメータ2を引いた値の25）で割り、先ほどの行列と掛け合わせる。

そして各行列の対角項の平方根

をとると、hill係数とEC50の標準誤差が求められる。 また標準誤差に両側97.5%に入るt統計量を掛けると、信頼区間が求められる。

自由度は

また決定係数は、

で求められる。

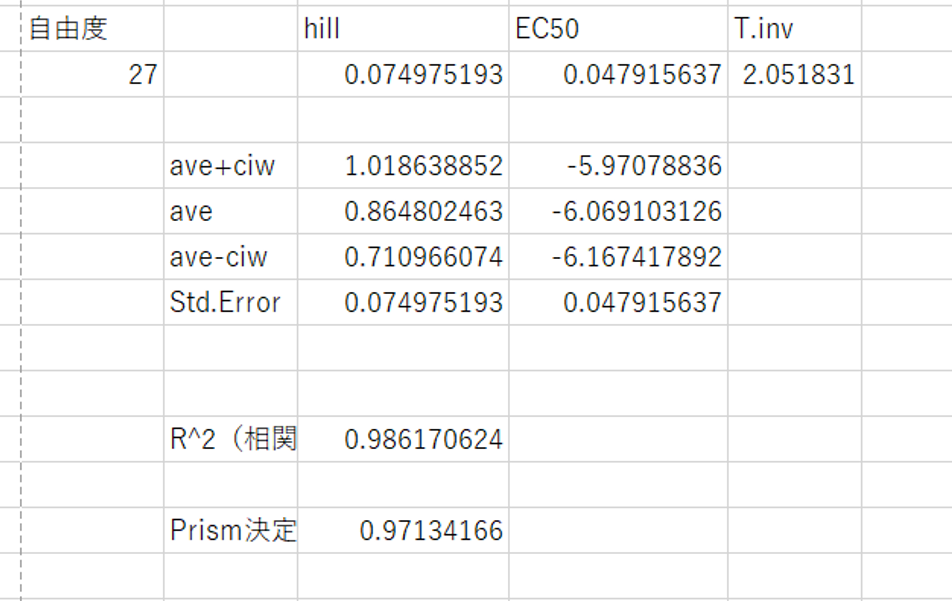
以上の手順よりNo　inhibitorの結果は下記のようになる。



No\_inhibitor結果

Inhibitorの項も同様にして計算する。

Inhibitorの値は以下の通りである。



Inhibitor結果

Hill係数の値が二群間で共通の場合は以下のように行う。

まずは、No inhibitorの最終行の一個下の行にInhibitorのデータを貼り付ける。 次に、No inhibitorを0、Inhibitorを1と各処理群に0、1の因子を割り付ける。 そして、以下の数式を入力する。

後の手順は、これまで説明したように残差平方和を求め、Solverで各係数の値を変数にして、残差平方和が小さくなるようにすればよい。

結果は以下のようになる。

