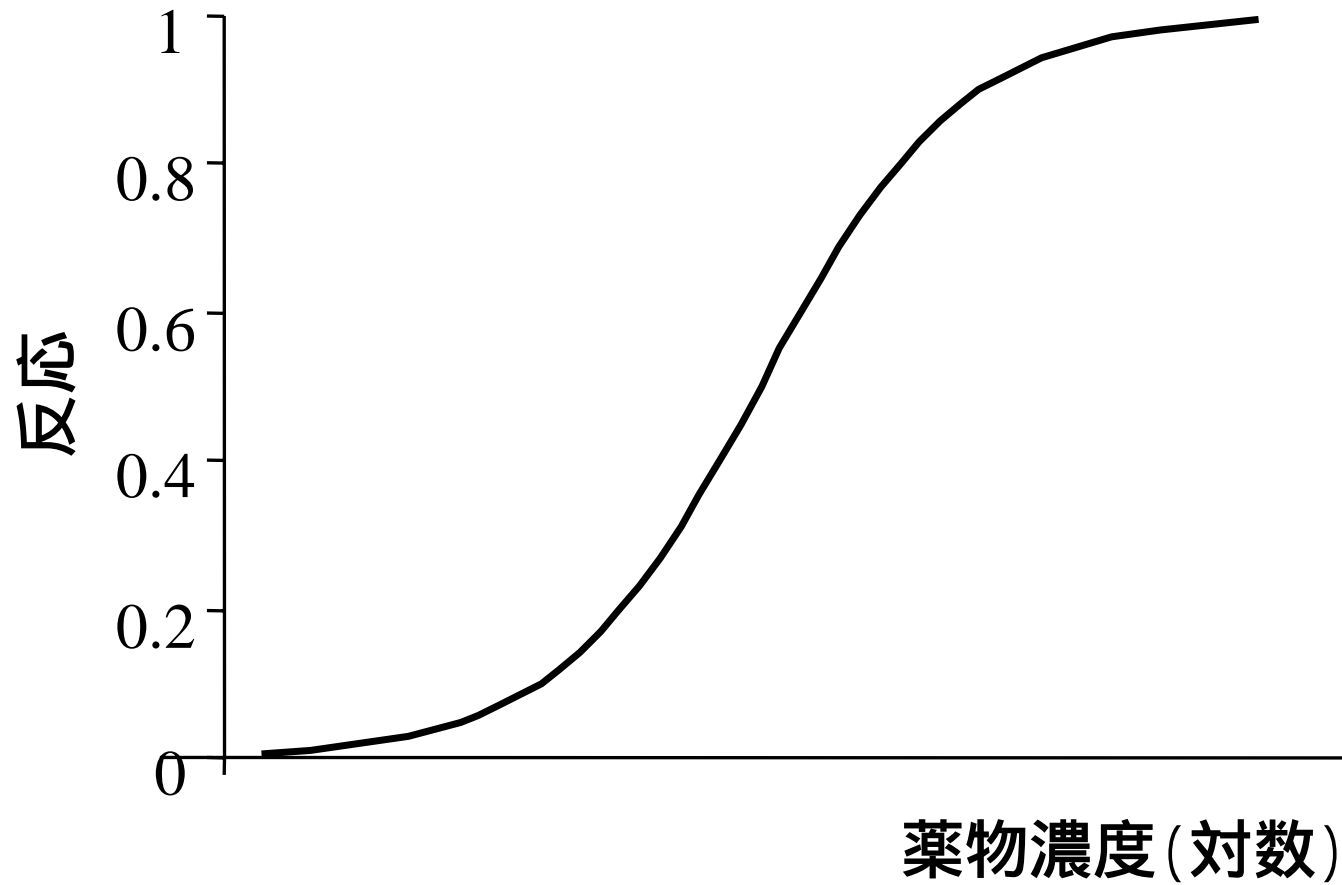


シグモイド曲線に回帰される反応の 信頼区間算出法

大日本製薬 生物統計室
杉本 忠則

第59回関西SASユーザー会 (2005年5月13日)

シグモイド曲線に回帰される反応



シグモイド曲線に回帰される反応

マグヌス実験

in vitroにおける生化学反応

酵素活性測定

binding実験

など

解析方法

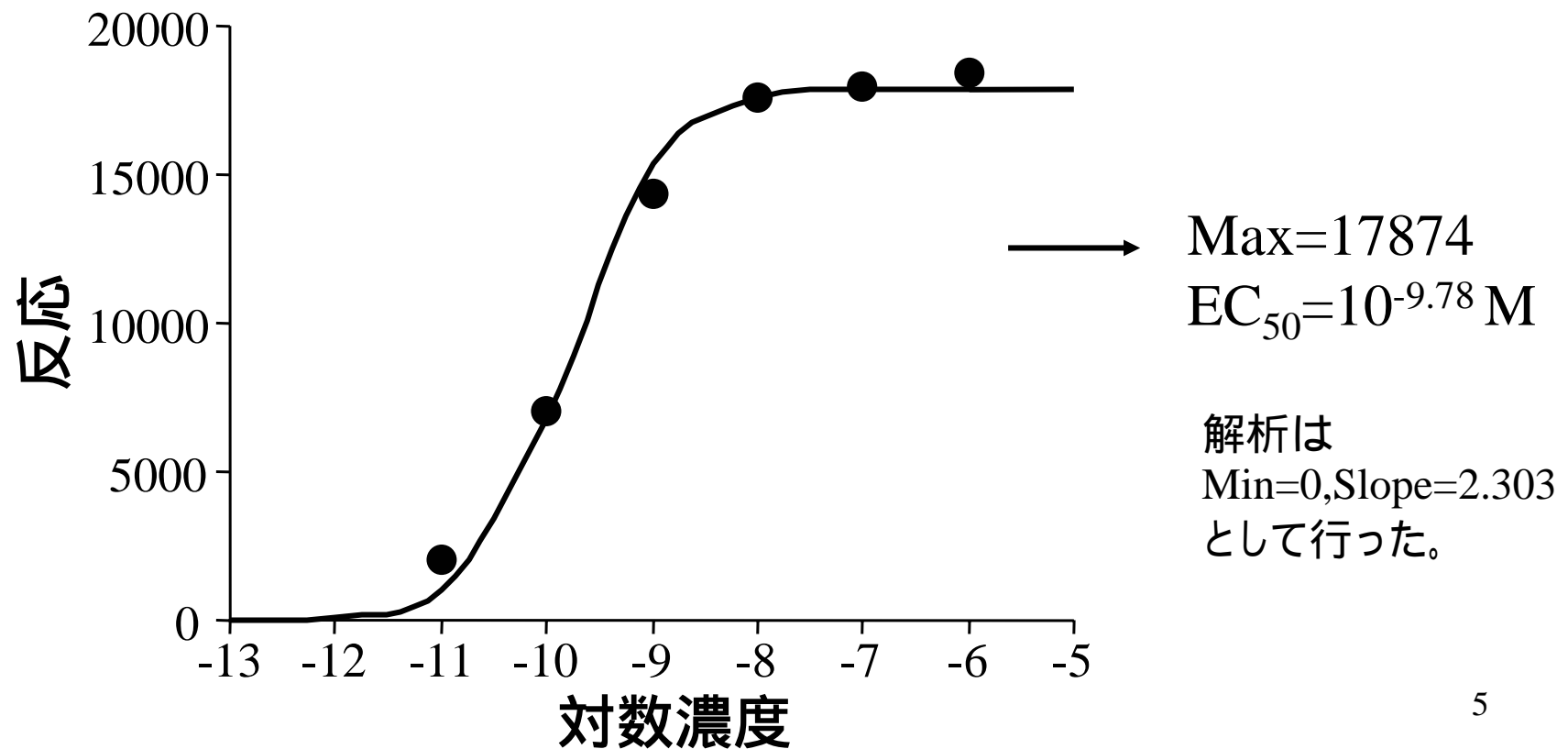
解析方法として、ロジット曲線に回帰することにより、EC₅₀ 値を推定する。

ロジット曲線:
$$F(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$$

生物反応:
$$response = (Max - Min) \times \frac{1}{1+e^{-Slope(\log C - \log EC_{50})}} + Min$$

CHO細胞におけるシグナル上昇作用

特定の受容体を発現させた、CHO細胞にアゴニストを作用させると、細胞内のシグナル濃度が上昇する。この反応はシグモイド曲線を描く。



シグモイド曲線への回帰の解析法

解析方法としては、非線形最小2乗法を適用し、
解析ソフトウェアとしては

SAS

SAS前臨床パッケージ

JMP

Excelのソルバー

を用いることにより各パラメータを推定できる。

本日は第59回関西SASユーザー会なので SASを使った解析を考えてみる

SASの非線形最小2乗法を行うには、proc nlinを用いる。
proc nlinの主な文法は次の通りである。

```
proc nlin   data = SAS-data-set;  
  model    dependent = expression;  
  parameters parameter = values;
```

data : 解析したいデータセット
model : 回帰式
parameters : 初期値

SASプログラム例

CHO細胞のデータをSASで解析してみる

```
data chocell;
  input response x;
  cards;
2024.8 1e-11
7041.6 1e-10
14312.8 1e-9
17604.6 1e-8
18004.6 1e-7
18404.6 1e-6
;

proc nlin data = chocell;
  model response = max / ( 1 + exp( -2.303*( log10(x)-log10(ec) ) ) );
  parameters max = 18000 ec = 1e-9;
run;
```


SASの出力結果(一部)

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	2	1.2289E9	6.1444E8	995.66	<.0001
Residual	4	2468497	617124		
Uncorrected Total	6	1.2314E9			
Corrected Total	5	2.3308E8			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
max	17874.2	437.6	16659.3	19089.1
ec	1.65E-10	2.93E-11	8.37E-11	2.46E-10

Approximate Correlation Matrix

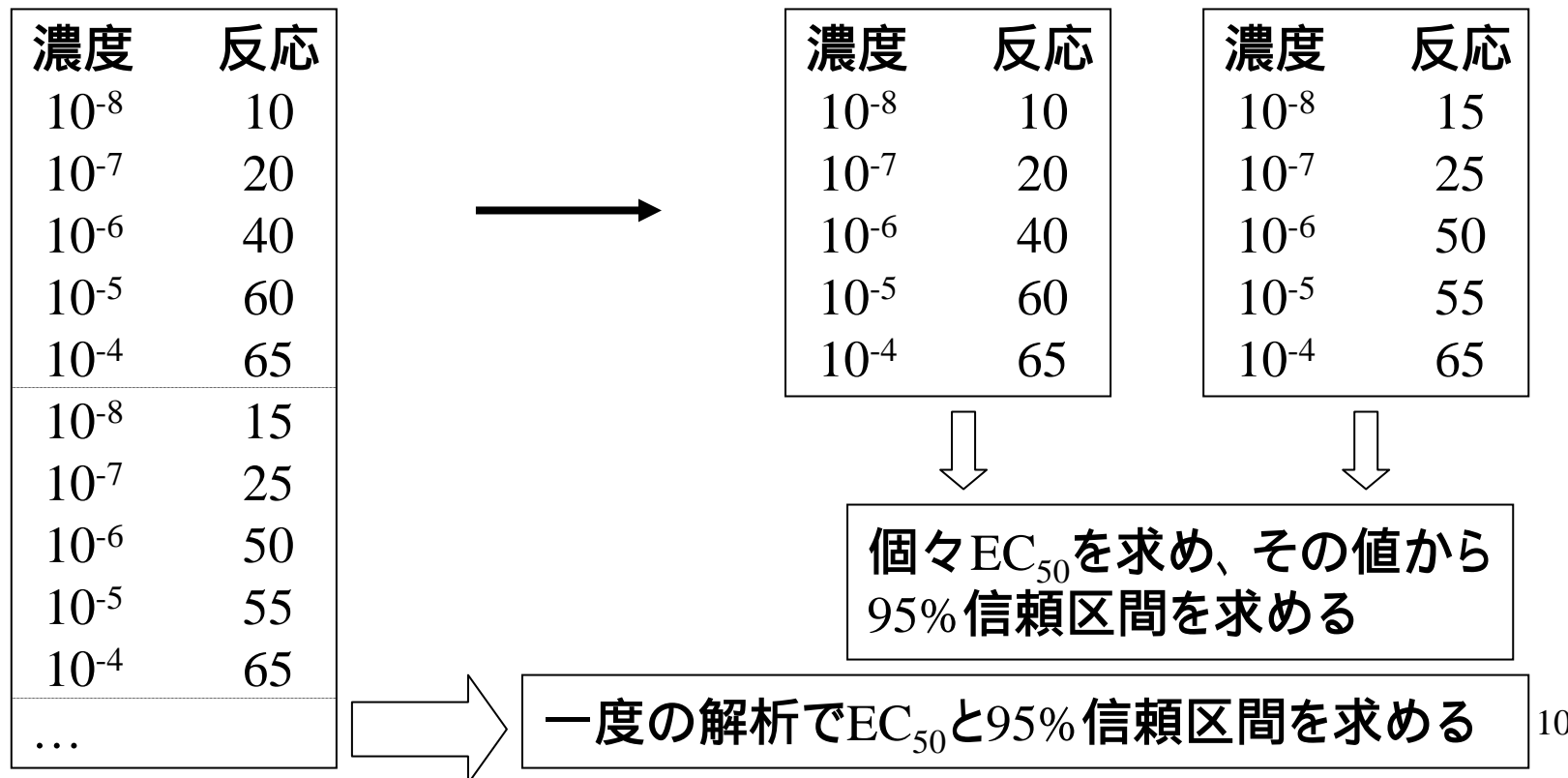
	max	ec
max	1.0000000	0.4026834
ec	0.4026834	1.0000000

ここに推定値が出力される。

解析時の疑問

解析ソフトウェアによっては、95%信頼区間も算出されるが、これはいったい何を意味しているのか？

複数の実験を行った場合にはどのように解析するのがよいのか？



シミュレーションによる検討

コンピュータ内でデータを発生させ、解析法をいろいろ検討する。

シグモイド曲線に回帰する反応式(1)

$$response = (Max - Min) \times \frac{1}{1 + e^{-Slope(\log C - \log EC_{50})}} + Min \quad (1)$$

今回は、Max=100、Min=0、Slope=2.303、 $EC_{50}=10^{-9.75}$ とした式(2)に誤差を加えたデータを作成する。

$$response = \frac{100}{1 + e^{-2.303(\log C + 9.75)}} \quad (2)$$

誤差について

今回誤差の発生が反応の大きさに依存せず一定の大きさの正規分布をすると仮定しデータを発生させた。

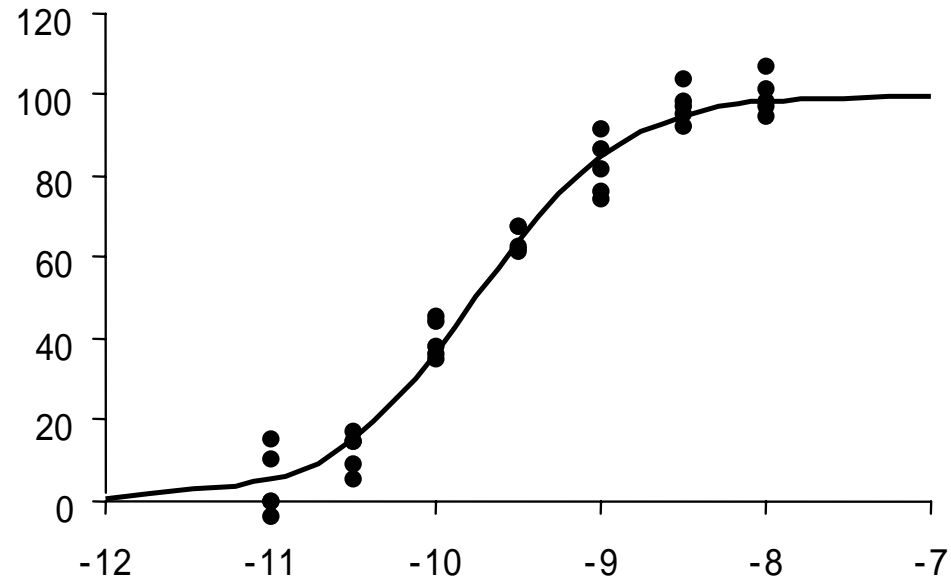
$$response = \frac{100}{1 + e^{-2.303(\log C + 9.75)}} + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$$

解析: Min=0、Slope=2.303が既知であるが、Maxと EC₅₀を推定する目的でSASを使用する。

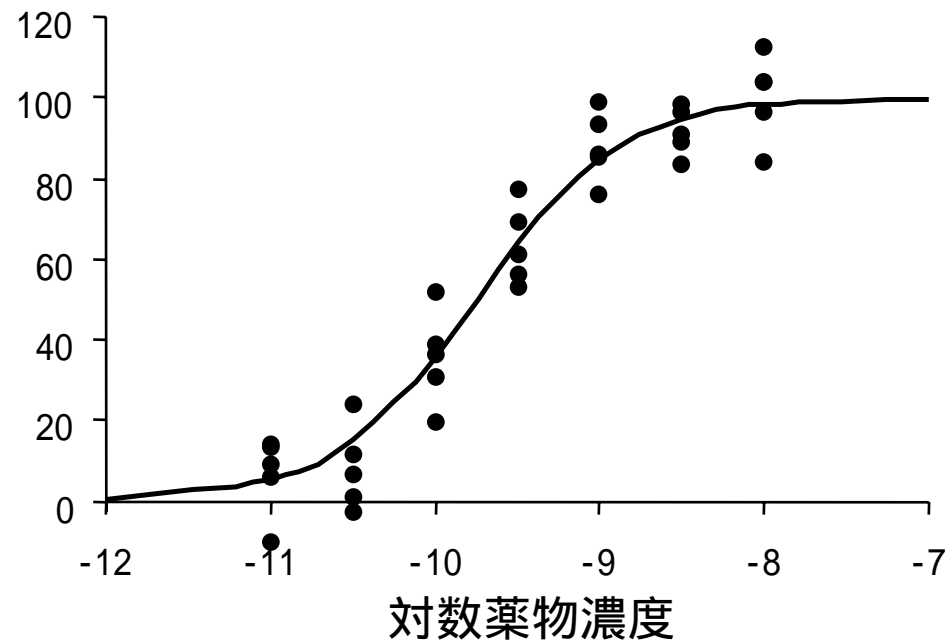
つまり、推定される Maxと EC₅₀が真の値 Max=100, EC₅₀=10^{-9.75}からどれくらい離れているかを調べる。

誤差分散 (σ^2) とデータの発生

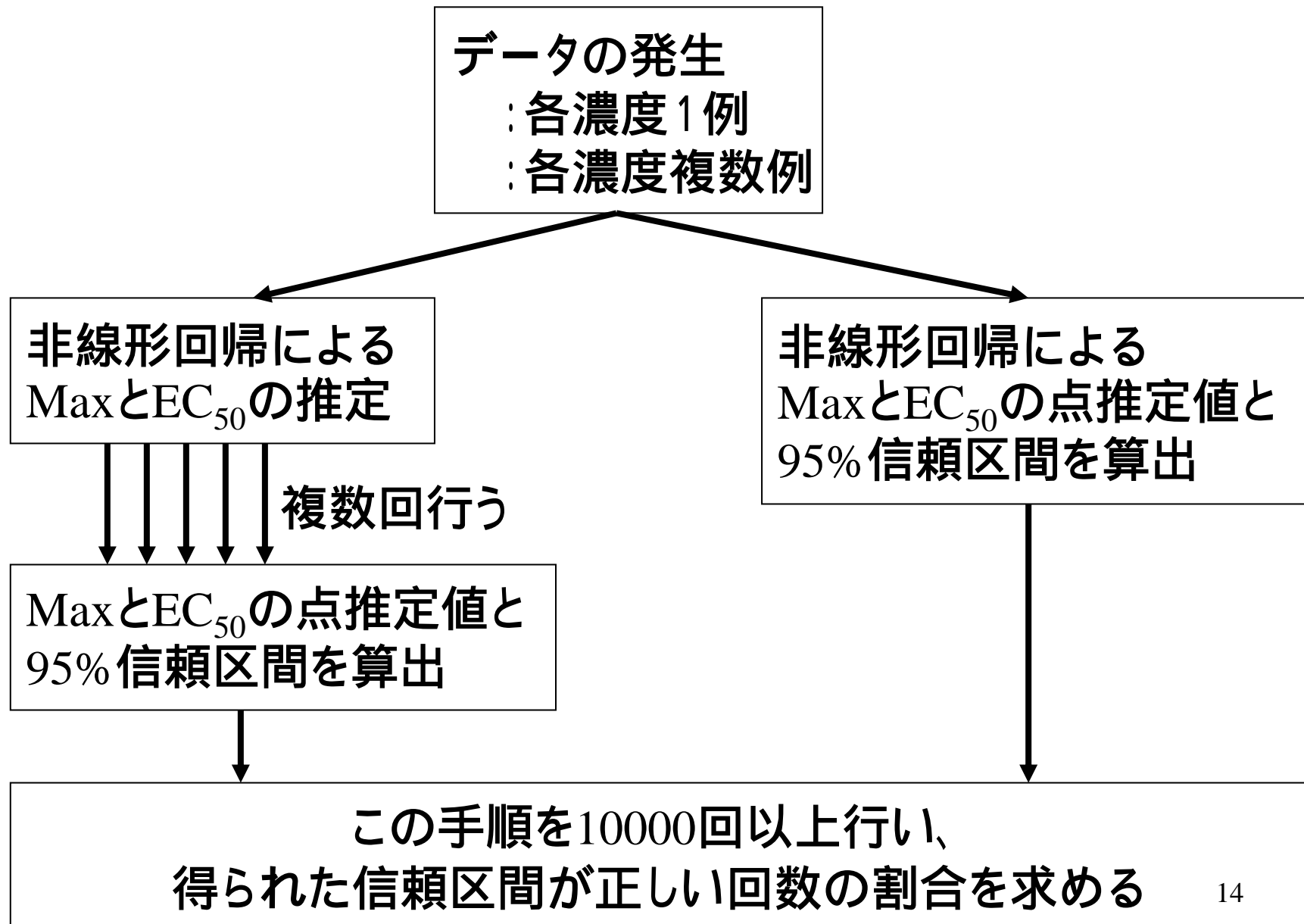
$\sigma^2=5^2$ の場合



$\sigma^2=10^2$ の場合



シミュレーション実験の手順



シミュレーション結果

データ例と解析結果

本データは $\sigma^2=5^2$ のときの仮想データである

		濃度					Max	EC ₅₀
10 ⁻¹¹	10 ^{-10.5}	10 ⁻¹⁰	10 ^{-9.5}	10 ⁻⁹	10 ^{-8.5}	10 ⁻⁸		
1	14	39	60	90	101	100	104.8	1.98E-10
7	11	38	72	89	96	98	101.6	1.59E-10
4	15	27	58	81	104	98	105.2	2.61E-10
9	15	30	59	83	86	105	100.5	2.23E-10
-6	25	41	59	84	100	101	102.0	1.77E-10
						mean	102.3	2.04E-10
						SE	0.9	1.79E-11
						95%信頼区間	100.3~	1.54E-10~
							105.4	2.53E-10

10000回行ったところEC50の95%信頼区間が
10^{-9.75}(=1.78E-10)を含んでいた割合は

シミュレーション結果

データ例と解析結果

濃度						
10^{-11}	$10^{-10.5}$	10^{-10}	$10^{-9.5}$	10^{-9}	$10^{-8.5}$	10^{-8}
1	14	39	60	90	101	100
7	11	38	72	89	96	98
4	15	27	58	81	104	98
9	15	30	59	83	86	105
-6	25	41	59	84	100	101

	Max	EC ₅₀
estimate	102.6	1.99E-10
SE	1.7	1.54E-11
95%信頼区間	99.1 ~	1.68E-10 ~
	106.1	2.31E-10

10000回行ったところEC₅₀の95%信頼区間が
 $10^{-9.75}$ (=1.78E-10)を含んでいた割合は

シミュレーション結果

各解析におけるEC₅₀の95%信頼区間の成功率
(%)

σ^2	解析 繰り返し				解析 各濃度				
	2反応	3反応	4反応	5反応	1反応	2反応	3反応	4反応	5反応
1 ²	0.950	0.950	0.948	0.948	0.951	0.948	0.949	0.943	0.944
3 ²	0.948	0.950	0.947	0.948	0.949	0.947	0.948	0.943	0.944
5 ²	0.949	0.951	0.947	0.949	0.944	0.946	0.947	0.942	0.944
10 ²	0.949	0.950	0.947	0.949	0.936	0.939	0.941	0.938	0.943
30 ²	0.926	0.939	0.944	0.951	0.858	0.886	0.901	0.910	0.912
50 ²	0.843	0.917	0.939	0.952	0.788	0.828	0.848	0.864	0.876
100 ²	0.693	0.868	0.924	0.949	0.678	0.728	0.763	0.782	0.795

薬理・生化学研究者がよく行なう解析例

各濃度で数点の反応を得る実験 (duplicate、triplicate) を行い、その平均値を解析に使用する。

このような解析には、問題はないのであろうか？

シミュレーション結果

データ例と解析結果

濃度								
10^{-11}	$10^{-10.5}$	10^{-10}	$10^{-9.5}$	10^{-9}	$10^{-8.5}$	10^{-8}		
1	14	39	60	90	101	100		
7	11	38	72	89	96	98		
4	15	27	58	81	104	98		
9	15	30	59	83	86	105		
-6	25	41	59	84	100	101		
3	16	35	62	85	97	100	Max	EC ₅₀
					estimate		102.6	1.99E-10
					SE		1.1	9.67E-12
					95%信頼区間		99.9~	1.74E-10~
							105.4	2.24E-10

10000回行ったところEC₅₀の95%信頼区間が
 $10^{-9.75}$ (=1.78E-10)を含んでいた割合は

シミュレーション結果

各解析におけるEC₅₀の95%信頼区間の成功率
(%)

σ^2	解析(:全反応を用い解析、 :各濃度の反応の平均値を用い解析)										
	各濃度		1反応		2反応		3反応		4反応		5反応
1 ²	0.951	0.948	0.949	0.949	0.950	0.943	0.950	0.944	0.950	0.944	0.950
3 ²	0.949	0.947	0.948	0.948	0.949	0.943	0.949	0.944	0.951	0.944	0.951
5 ²	0.944	0.946	0.947	0.947	0.948	0.942	0.949	0.944	0.949	0.944	0.949
10 ²	0.936	0.939	0.942	0.941	0.945	0.938	0.945	0.943	0.947	0.943	0.947
30 ²	0.858	0.886	0.892	0.901	0.908	0.910	0.918	0.912	0.924	0.912	0.924
50 ²	0.788	0.828	0.835	0.848	0.856	0.864	0.877	0.876	0.888	0.876	0.888
100 ²	0.678	0.728	0.733	0.763	0.767	0.782	0.790	0.795	0.808	0.795	0.808

まとめ

- シグモイド曲線を描く反応を示す実験において、 EC_{50} 値の95%信頼区間を求める方法の解析を検討した。
- シグモイド曲線を描く反応の理論反応値に誤差を加え仮想データを発生させ、これを解析に用いた。
- 各濃度での反応値が複数ある場合の EC_{50} 値の算出方法として、3つの方法を比較した。
 - ：複数のシグモイド曲線を描き、得られた EC_{50} 値から95%信頼区間を算出する。
 - ：全ての反応値を解析に使用する。
 - ：各濃度毎に反応値の平均値を算出し、これを解析に使用する
- いずれの、解析方法で算出された95%信頼区間も、95%の確率で真の EC_{50} 値を含んでいた。このことより、いずれの方法を用いても EC_{50} 値の95%信頼区間が算出できることがわかった。

実際のデータの誤差

薬理・生化学の研究者は、反応が小さいところでのデータのバラツキは小さく、反応が大きいたところでのデータのバラツキは大きいことを経験している。

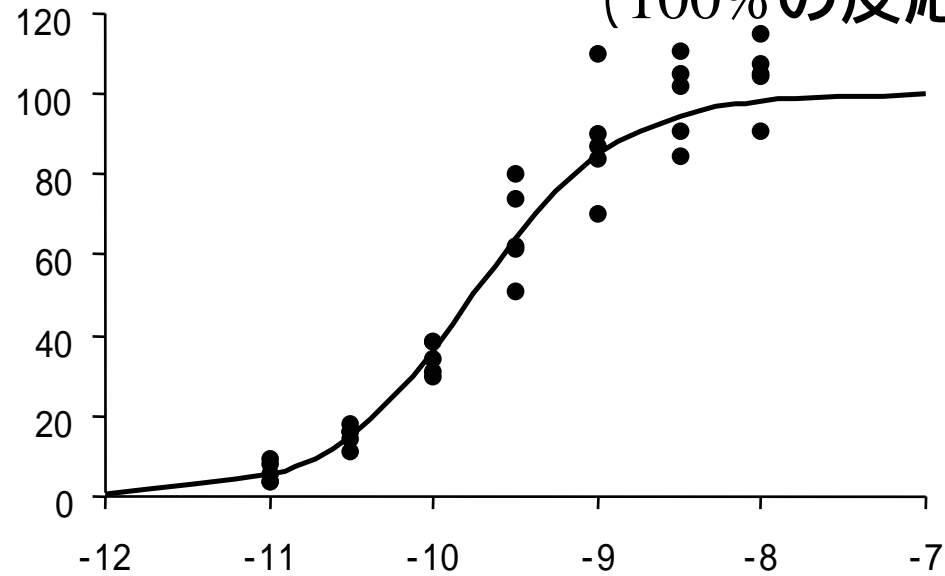
しかし、今回のシミュレーションでは、反応のどの部分でも、同じ誤差を発生させ解析した。

本来は、反応の大きさに応じて誤差を発生させるべきでは？

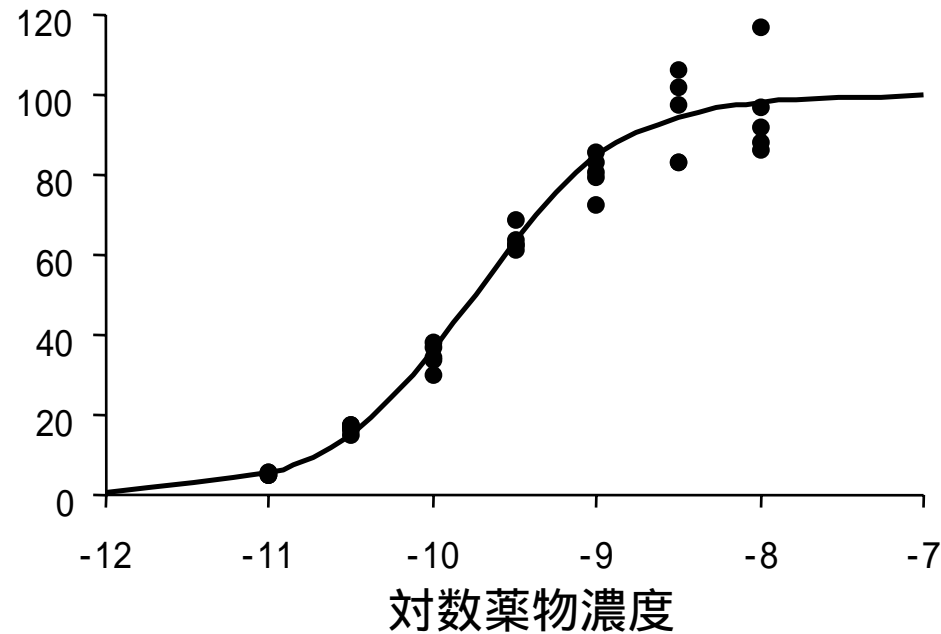
誤差分散 (σ^2) が反応とともに大きくなる場合

(100%の反応 $\sigma^2=10^2$)

$\sigma^2 \propto$ 反応



$\sigma^2 \propto$ 反応²



今後の課題

反応の誤差分散が一定でない場合は、分散の逆数を重みとして解析を行なうことが一般的である。

次のスライドに示すようにSASのproc nlinの場合は、_weight_オプションにより設定できる。

誤差分散(σ^2)が反応とともに大きくなる場合は、_weight_オプションで対応可能であるが、これにより出力される信頼区間が正しいかは今後の検討課題としたい。

proc nlinの主な文法

proc nlinでは以下のように重みを付けて、解析を行うことができる。

```
proc nlin    data = SAS-data-set;  
  model      dependent = expression;  
  parameters parameter = values;  
  _weight_ = expression;
```

data: 解析したいデータセット

model: 回帰式

parameters: 初期値

weight: 重み