Nonlinear regression analysis

Peter Dalgaard (orig. Lene Theil Skovgaard)

Department of Biostatistics
University of Copenhagen

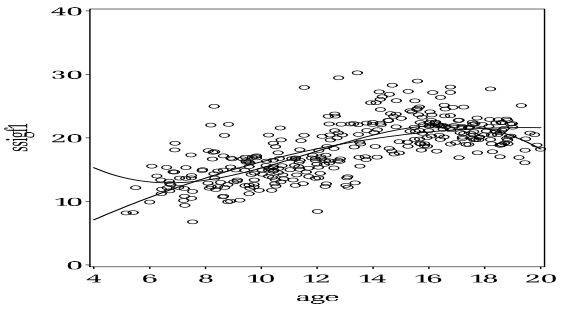
Variance & Regression, May 2008

Nonlinear regression

- Simple kinetic model
- Compartment models
- Michaelis Menten reaction
- Dose-response relationships

How can we model non-linear effects?

- Polynomials tend to be very wiggly
  (use '`i=rq`' or '`i=rc`' in the `symbol`-statement)



- Splines
  piecewise interpolations, often linear or cubic

Polynomial regression

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p$$

With the new covariates

$$Z_1 = X, Z_2 = X^2, \cdots, Z_p = X^p$$

this is just a **linear multiple regression**

$$Y_i = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \cdots + \beta_p z_{pi}$$

The model is **linear in the parameters**!

How can this work?

The covariates $Z_1, \cdots, Z_p$ are of course **correlated**, but they are not **linearly** dependent.

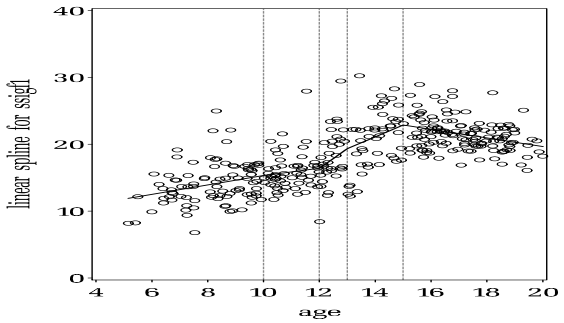
What do we use polynomial regression for?

- as model checking for the linear model
- as a smoothing method
- rarely as a 'final model' for publications

Linear splines

- Subdivide age into groups, using appropriate thresholds
- Fit linear effect of age in each age group
- Make the linear pieces 'meet' at the thresholds

The result is a **bent line:**

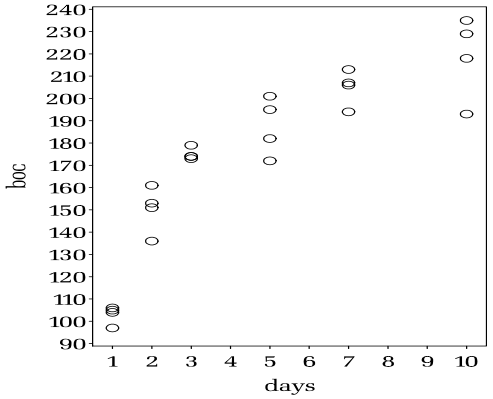


Example

Oxygen consumption (from earlier exercise)

days				
1	105	97	104	106
2	136	161	151	153
3	173	179	174	174
5	195	182	201	172
7	207	194	206	213
10	218	193	235	229

We want to give a description of the oxygen consumption (`boc`) over time (`days`)



The above plot shows that `boc` as a function of `days` is certainly **not linear**.

The biologists claim that the relation between $\texttt{boc}$ and $\texttt{days}$ can be described by a relation of the form

$$\texttt{boc} = \gamma \exp(-\beta/\texttt{days})$$

This relation is obviously nonlinear, but may be transformed to linearity using the (natural) logarithm:

$$\log(\texttt{boc}) = \log(\gamma) - \beta/\texttt{days}$$

With

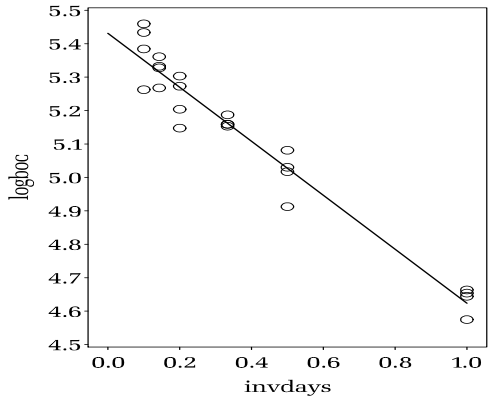
$$\begin{aligned} y &= \texttt{logboc} = \log(\texttt{boc}) \\ x &= \texttt{invdays} = 1/\texttt{days} \qquad \text{and} \\ \alpha &= \log(\gamma) \end{aligned}$$

we may write this equation as

$$y = \alpha - \beta x$$

i.e., a **linear relation**, just with a minus sign on the slope.

A scatter plot of these new variables



This plot seems reasonably linear. We get

$$\texttt{logboc} = 5.431 - 0.808 \times \texttt{invdays}$$

The linear regression model gives us the estimates:

$$\begin{aligned} \text{intercept: } \hat{\alpha} &= \log(\hat{\gamma}) = 5.431(0.019) \\ \text{slope: } \hat{\beta} &= -0.808(0.039) \end{aligned}$$

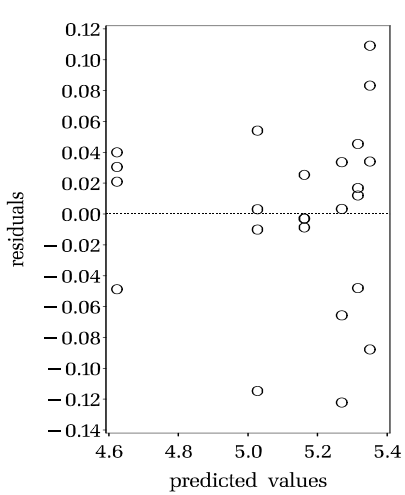
Noting that $\texttt{boc}(\infty) = \gamma = \exp(\alpha)$, we find the estimate of $\texttt{boc}(\infty)$ to be

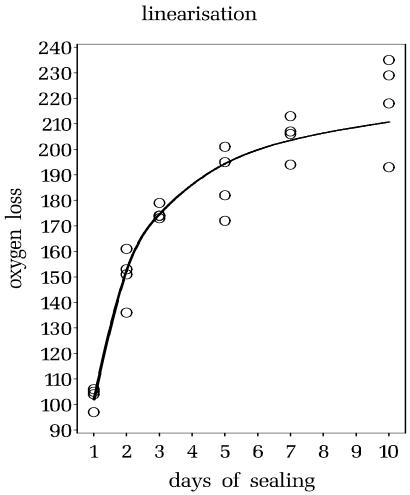
$$\exp(5.431) = 228.38$$

with the 95% confidence interval

$$(\exp(5.392), \exp(5.471)) = (219.6, 237.7)$$

Residual plot and fitted curve



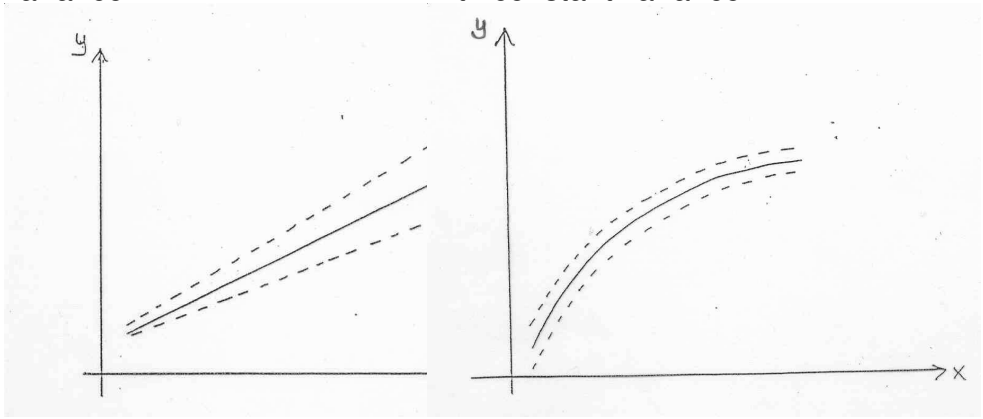


Why use non-linear regression?

- Transformation is necessary to obtain variance homogeneity, but transformation destroys linearity.
- Linearity does not fit, and the transformation seems to destroy other parts of the model assumptions, e.g. the assumption of variance homogeneity.
- Theoretical knowledge (e.g. from kinetics or physiology) indicates that the proper relation is intrinsically non-linear.
- Interest is in functions of the parameters that do not enter linearly in the model (e.g. kinetic rate constants or ED_{50} in dose-response studies)

Untransformed: Linearity, but with increasing variance

After a **logarithmic transformation:** Non-linearity, but with constant variance



Example

Quantification of the Reticuloendothelial cell system (**RES**) of the liver:

Concentration measurements y_i over the liver, following a bolus injection of radioactive tracer

First order kinetics implies

$$c(t) = \beta(1 - e^{-\gamma t})$$

No transformation to linearity possible!

$$y_i = \beta(1 - e^{-\gamma t_i}) + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

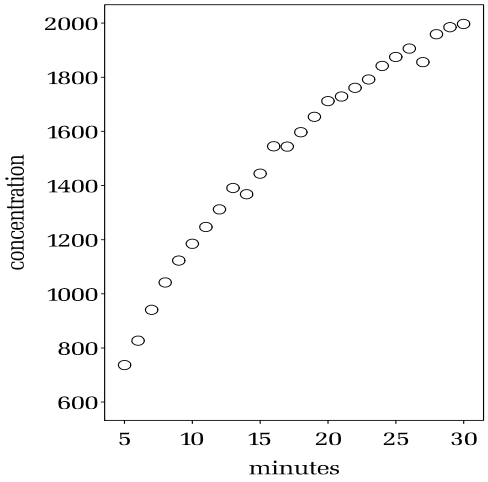
Least squares method

Minimize the sums of squares

$$SS(\beta, \gamma) = \sum (y_i - \beta(1 - e^{-\gamma t_i}))^2 = \sum \varepsilon_i^2$$

This requires an *iterative procedure*

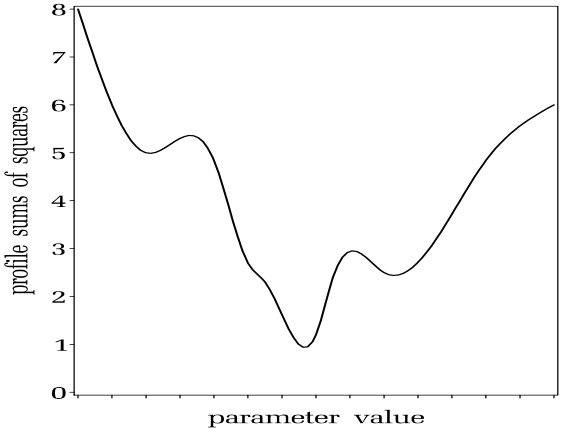
Example: RES in the liver



Starting values:

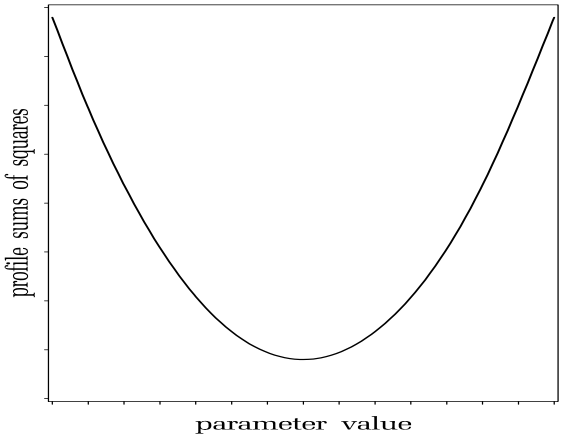
- $c(\infty) = \beta \approx 2000$
- $\frac{dc}{dt}(0) = \beta\gamma \approx 100$

Residual sum of squares, SS, as a function of only one parameter (hypothetical)

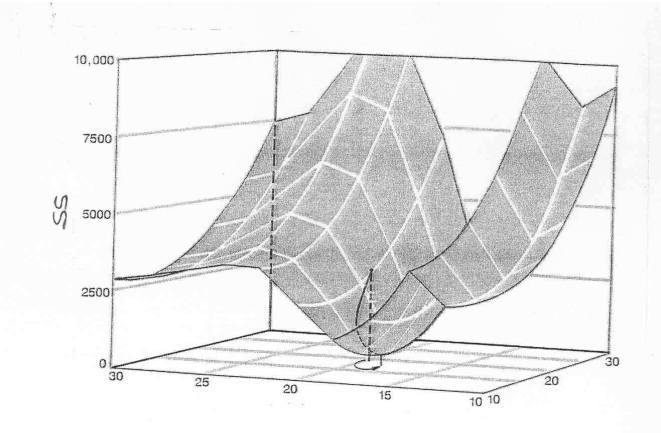


We have to determine the minimum

In case of **linear** regression:



Residual sum of squares, SS, as a function of two parameters (hypothetical)



There may be problems with **convergence** of the iteration procedure and the solution may be a **local minimum**

SAS program

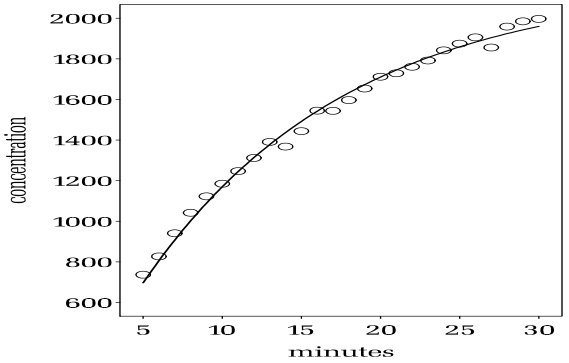
```
data reticulo;
infile 'kw_res.txt';
input tid conc;
run;

proc nlin data=reticulo;
  parms beta=2000
        gamma=0.05;
  model conc=beta*(1-exp(-gamma*tid));
run;
```

Output

```
Dependent Variable conc

              Iterative Phase
                                  Sum of
 Iter        beta       gamma     Squares

    0      2000.0      0.0500     4370990
    1      1958.6      0.0824      382995
    2      2169.2      0.0769     26355.0
    3      2174.2      0.0772     25286.7
    4      2174.0      0.0773     25286.7
    5      2174.0      0.0773     25286.7

NOTE: Convergence criterion met.

                                 Sum of        Mean                 Approx
Source                     DF    Squares      Square    F Value    Pr > F

Regression                  2   63048136    31524068    29920.0    <.0001
Residual                   24     25286.7      1053.6
Uncorrected Total          26   63073423

Corrected Total            25    3455129

                            Approx
Parameter      Estimate    Std Error    Approximate 95% Confidence Limits

beta             2174.0      28.3459      2115.5      2232.5
gamma            0.0773      0.00226      0.0726      0.0819
```

Example

The fit becomes:

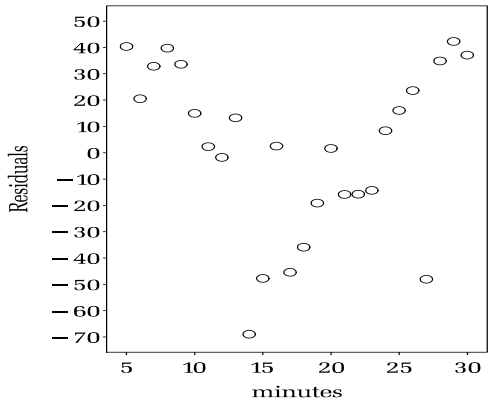


Estimates:

- $\hat{\beta} = 2174.0(28.3)$, CI=(2115.5, 2232.5)
- $\hat{\gamma} = 0.0773(0.0023)$, CI=(0.0726, 0.0819)

Residuals

Residual pattern shows **systematic behaviour:**



so the model does not fit particularly well

Asymptotic normality of parameter estimates

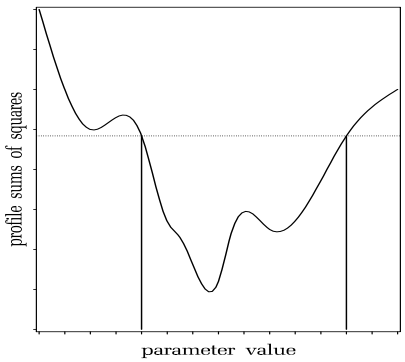
- The distribution of the estimates is in general unknown, we only have **approximate** normality
- The approximations may be *very poor* for small samples

Confidence regions:

- Asymptotically (i.e. for large sample sizes), the estimates are normally distributed
- The confidence areas will therefore be approximately elliptic
- For small sample sizes, this can become **extremely** misleading!

Alternative procedure

Determine confidence regions directly from the sum of squares *SS*, i.e. as those values of the parameter, which makes *SS* sufficiently small.



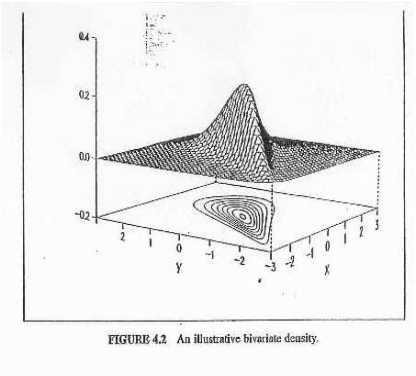


FIGURE 4.2 An illustrative bivariate density.

Determining the cutoff

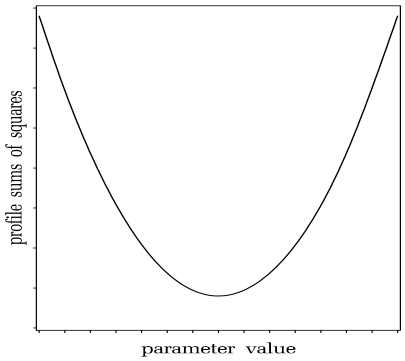
What is “sufficently small” SS to obtain a 95% confidence region?
Or: How do we determine the coverage probability for a given confidence region?

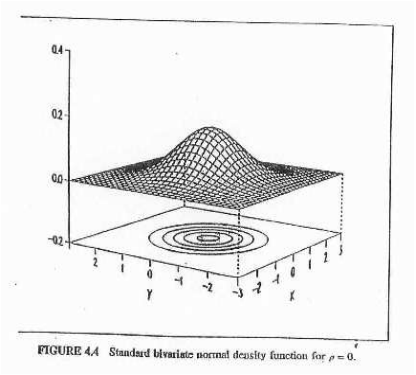
This is rather technical. . .

Good approximation to a (1-α)-sized region:

$$\{\theta | SS(\theta) \leq SS(\hat{\theta})(1 + \frac{p}{n-p} F_\alpha(p, n-p))\}$$

For **linear** regression, the situation is:

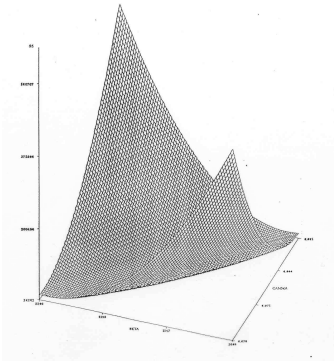




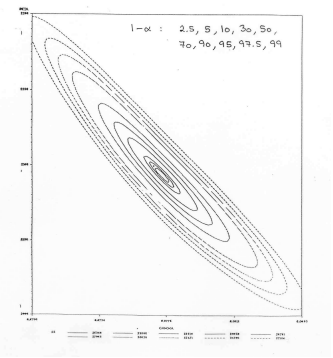
and confidence intervals become **symmetric/elliptic**

Example: RES in the liver

Residual sum of squares, *SS*:

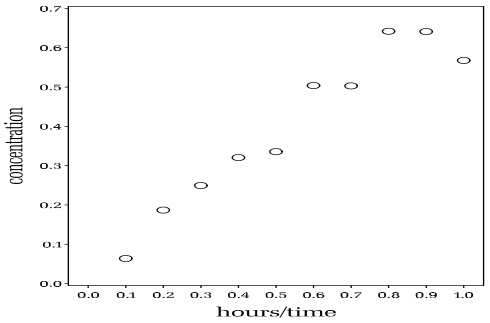


Confidence region based upon *SS*:



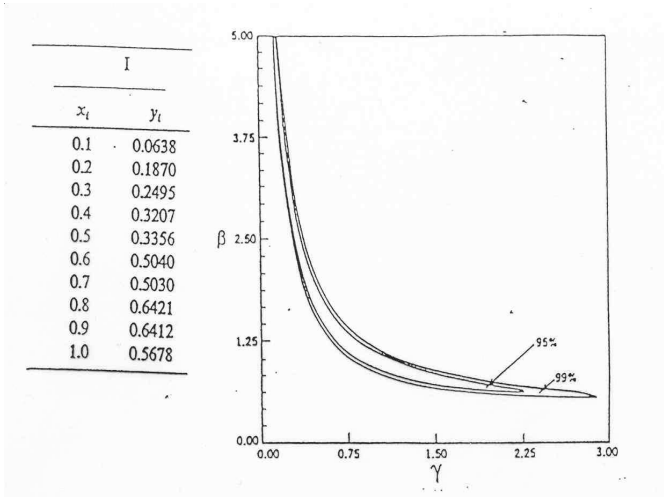
Same model as before: $c(t) = \beta(1 - e^{-\gamma t})$

Simulated data, ($\beta = 1, \gamma = 1, \sigma = 0.05$):



The design is not adequate, we only see the linear part of the concentration curve!

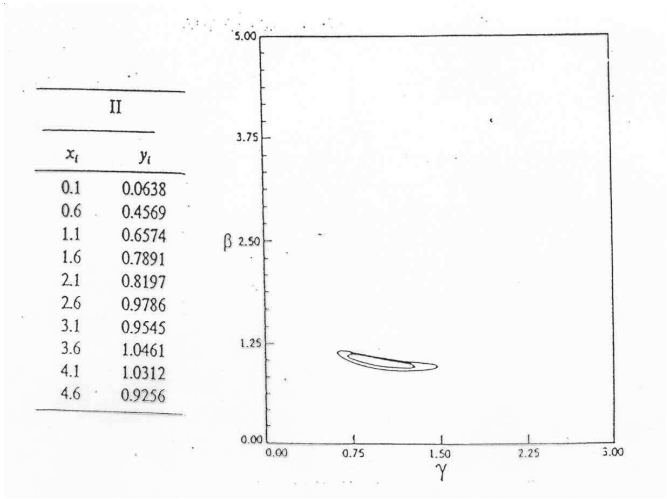
Confidence regions based on $SS(\beta, \gamma)$:



I

x_i	y_i
0.1	0.0638
0.2	0.1870
0.3	0.2495
0.4	0.3207
0.5	0.3356
0.6	0.5040
0.7	0.5030
0.8	0.6421
0.9	0.6412
1.0	0.5678

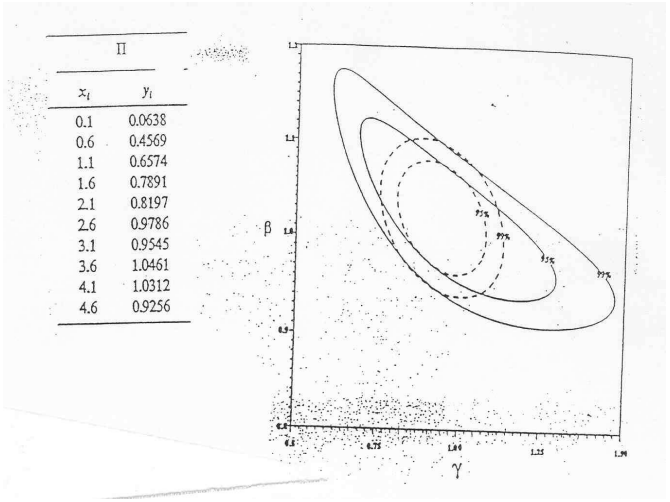
If we spread out the x's:



II

x_i	y_i
0.1	0.0638
0.6	0.4569
1.1	0.6574
1.6	0.7891
2.1	0.8197
2.6	0.9786
3.1	0.9545
3.6	1.0461
4.1	1.0312
4.6	0.9256

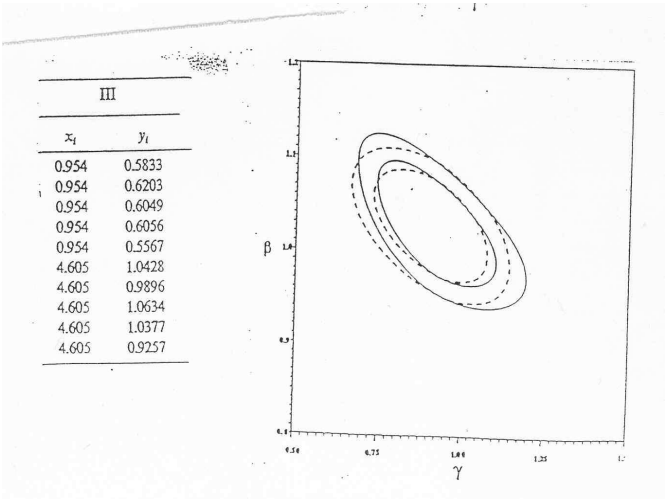
Enlarged picture, with superimposed normal approximation:



II

x_i	y_i
0.1	0.0638
0.6	0.4569
1.1	0.6574
1.6	0.7891
2.1	0.8197
2.6	0.9786
3.1	0.9545
3.6	1.0461
4.1	1.0312
4.6	0.9256

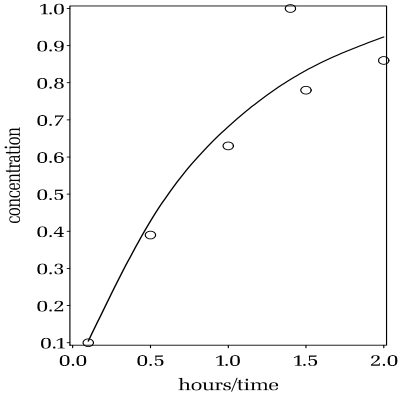
'*Optimal*' design (in terms of a normal confidence region):



III

x_i	y_i
0.954	0.5833
0.954	0.6203
0.954	0.6049
0.954	0.6056
0.954	0.5567
4.605	1.0428
4.605	0.9896
4.605	1.0634
4.605	1.0377
4.605	0.9257

Effect of **parametrization:**



$$y_i = \beta(1 - e^{-\gamma t_i}) + \varepsilon_i$$

Reparametrization:

$$\alpha = \beta\gamma$$

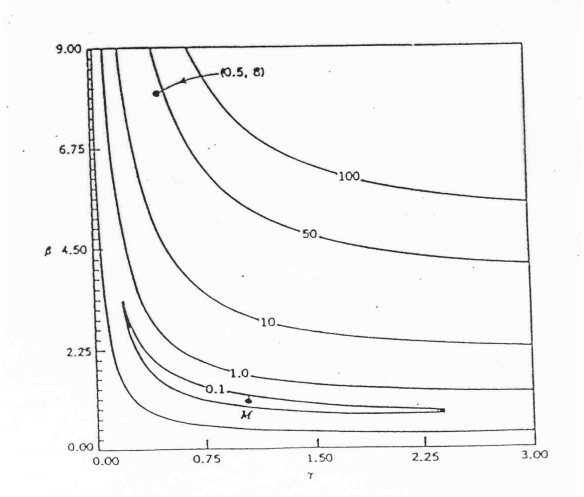
$$y_i = \frac{\alpha}{\gamma}(1 - e^{-\gamma t_i}) + \varepsilon_i$$

Parameters β and γ:

Parameter	Estimate	Approx Std Error	Approx. 95% Confidence Limits	
beta	1.0559	0.2407	0.3875	1.7244
gamma	1.0376	0.5202	-0.4067	2.4818

Approximate Correlation Matrix

	beta	gamma
beta	1.0000000	-0.9592479
gamma	-0.9592479	1.0000000

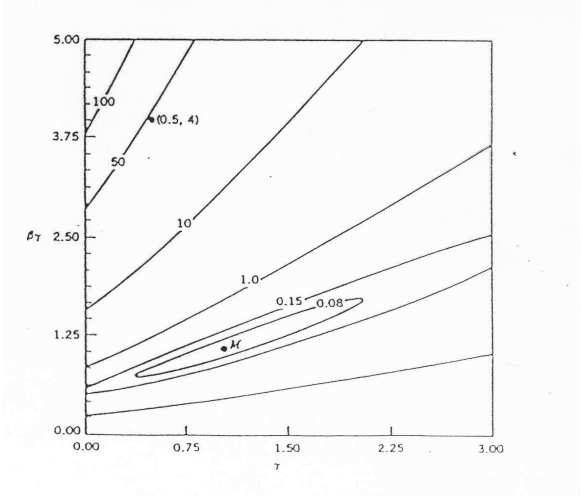


Parameters α and γ:

Parameter	Estimate	Approx Std Error	Approx. 95% Confidence Limits	
alfa	1.0956	0.3176	0.2138	1.9774
gamma	1.0376	0.5202	-0.4067	2.4818

Approximate Correlation Matrix

	alfa	gamma
alfa	1.0000000	0.9749950
gamma	0.9749950	1.0000000



Compartment models

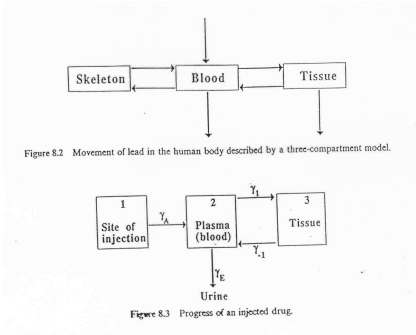


Figure 8.2 Movement of lead in the human body described by a three-compartment model.

Figure 8.3 Progress of an injected drug.

Differential equations $\Rightarrow$ multi-exponential curves

Often only measurements from a single compartment!

Identification problems

There may be problems with the **identification** of parameters, even with good quality data.

This will give rise to very unprecise and extremely correlated estimates.

Compartment models yield solutions as a sum of exponential curves:

$$f(t) = \alpha_1 \exp(-\lambda_1 t) + \alpha_2 \exp(-\lambda_2 t)$$

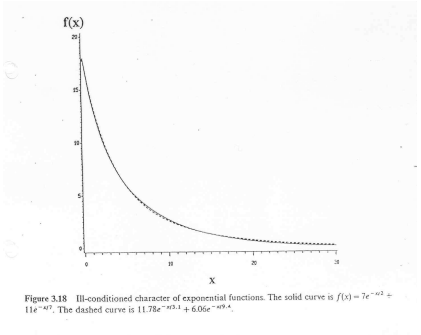


Figure 3.18 Ill-conditioned character of exponential functions. The solid curve is $f(x) = 7e^{-x/2} + 11e^{-x/7}$. The dashed curve is $11.78e^{-x/3.1} + 6.06e^{-x/9.4}$.

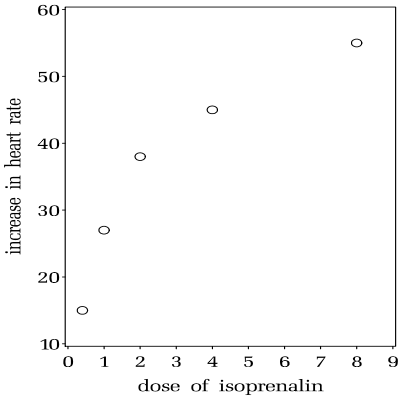
	α_1	λ_1	α_2	λ_2
1	7	$\frac{1}{2}$	11	$\frac{1}{7}$
2	11.78	$\frac{1}{3.1}$	6.06	$\frac{1}{9.4}$

Rule of thumb:
There must be a ratio of at least 5 in $\frac{\lambda_1}{\lambda_2}$

Example: Dose-effect

Effect of isoprenalin on heart rate

- E: Increase in heart rate
- D: Dose of isoprenalin



Michaelis-Menten relation: $E = \frac{E_{max} D}{k_d + D}$

Linearization (Lineweaver-Burk)

$$\frac{1}{E} = \frac{k_d + D}{E_{\max} D} = \frac{1}{E_{\max}} + \frac{k_d}{E_{\max}} \frac{1}{D}$$

Linear relation between the inverses:

$$\frac{1}{E} = \alpha + \beta \frac{1}{D}$$

with the reparametrisation:

$$\begin{array}{llll} \alpha & = 1/E_{\max} & \beta & = k_d/E_{\max} \\ E_{\max} & = 1/\alpha & k_d & = \beta/\alpha \end{array}$$

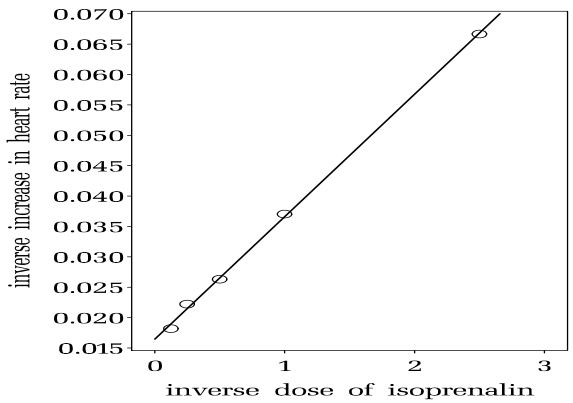
Ex. Isoprenalin:

α: 0.0165 (0.0004), β: 0.0202 (0.0004)

$E_{\max}$: $1/0.0165 = 60.6(57.8, 63.7)$

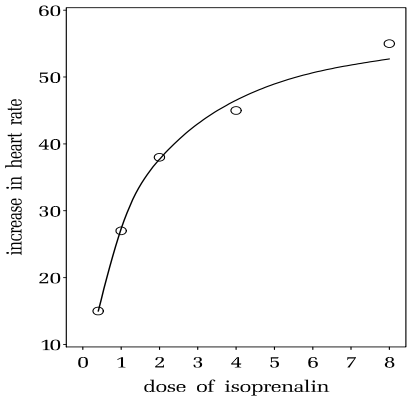
k_d : $0.0202/0.0165 = 1.22$

Lineweaver-Burk plot

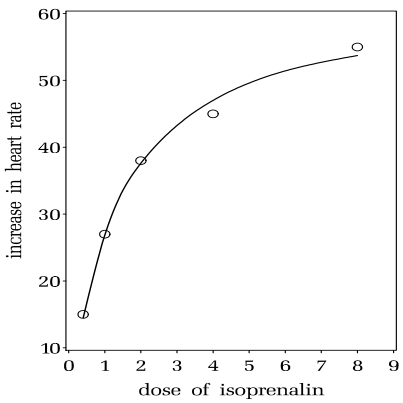


The model fits nicely **???**

Model fit using Lineweaver-Burk linearisation:



Model fit using direct non-linear regression:



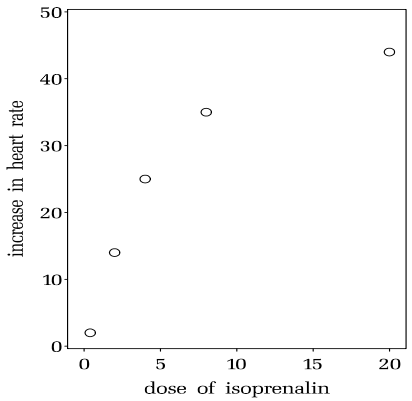
Estimates, isoprenalin

	$E_{\max}$	k_d
1/E linear	60.6	1.22
E, non-linear	62.67 (2.12)	1.33 (0.14)
log(E), non-linear	61.59 (1.82)	1.26 (0.08)

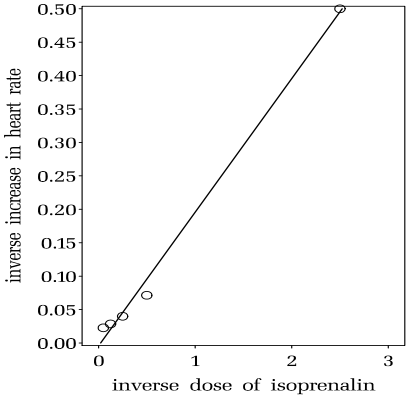
Isoprenalin following metropolol:

	E_{max}	k_d
1/E linear	-228.2	-45.62
E, non-linear	56.97 (3.47)	5.46 (0.85)
log(E), non-linear	80.73 (30.93)	11.71 (6.09)

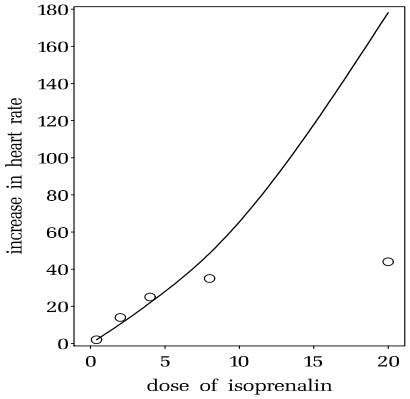
Isoprenalin following metropolol:



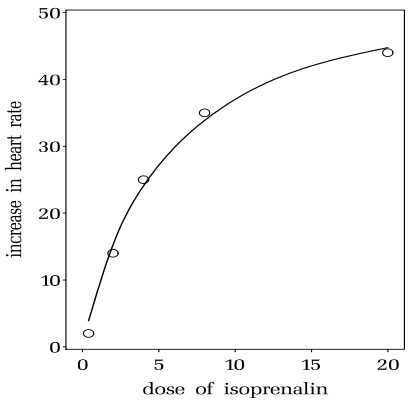
Lineweaver-Burk plot:



Model fit using Lineweaver-Burk linearisation:



Model fit using direct non-linear regression:



What is the difference between the two fits?

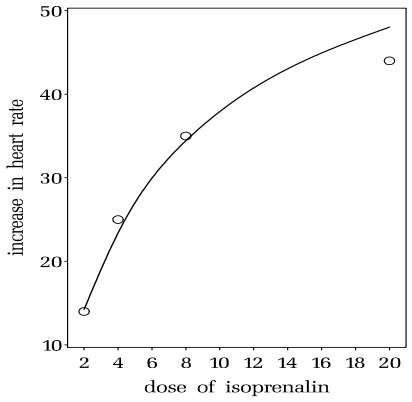
- It is **not** just a reparametrisation!
- We change the outcome from E to $\frac{1}{E}$
- If we have constant variance on the E scale, the variance on the $\frac{1}{E}$ scale will be proportional to $\frac{1}{E^4}$
- The assumption of constant variance on the $\frac{1}{E}$ scale corresponds to an assumption that the variance on the E scale is proportional to E^4, i.e. an SD proportional to E^2 – which more or less corresponds to *disregarding the observations with large outcomes!*

If the smallest concentration is omitted:

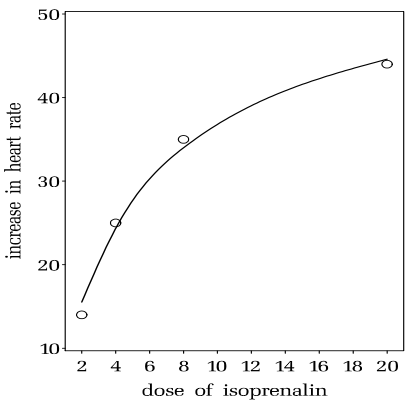
	E_{max}	k_d
1/E linear	65.19	7.14
E, non-linear	56.31 (3.04)	5.25 (0.74)
log(E), non-linear	59.49 (5.36)	6.06 (1.011)

If we omit the lowest concentration:

Model fit using Lineweaver-Burk linearisation:



Model fit using direct non-linear regression:

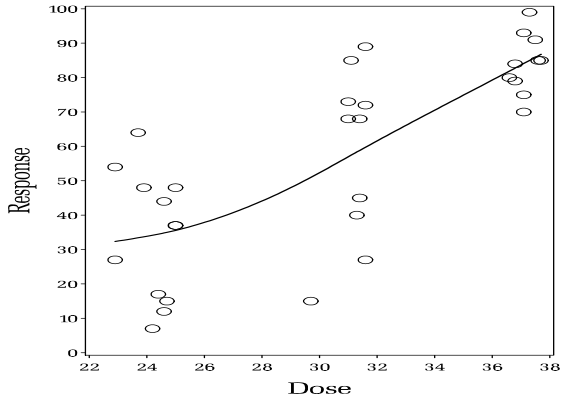


Why non-linear regression?

- Transformation is necessary to obtain variance homogeneity, but transformation destroys linearity.
- Linearity does not fit, and the transformation seems to destroy other parts of the model assumptions, e.g. the assumption of variance homogeneity.
- Theoretical knowledge (e.g. from kinetics or physiology) indicates that the proper relation is intrinsically non-linear.
- Interest is focused on functions of the parameters, that do not enter linearly in the model (e.g. kinetic rate constants or ED_{50} in dose-response studies)

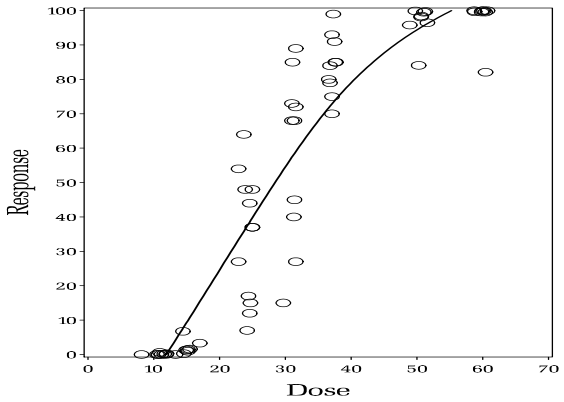
Example of a typical dose-response relation, for moderate doses

We *almost* have linearity in this dose range:

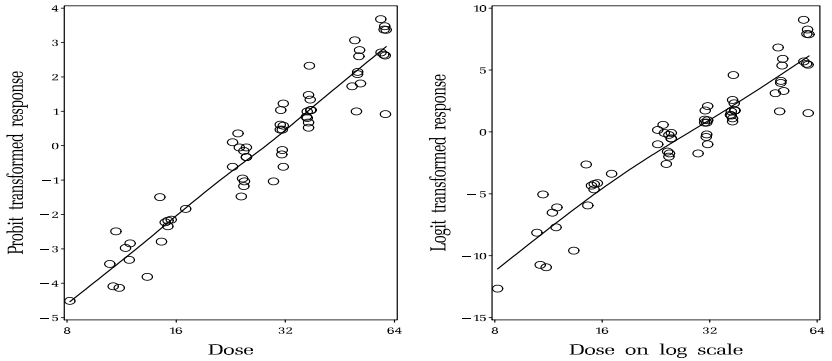


For *extreme* doses we see a *clear deviation* from linearity

and: smaller variation in the ends

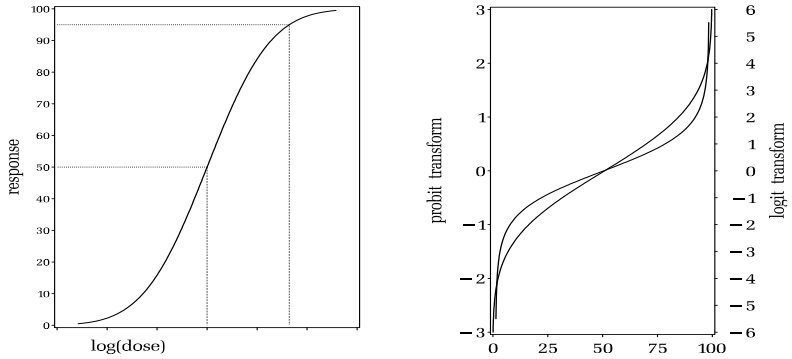


Y axis: Probit- or logit- transformed outcome
X axis: Logarithmic transformed dose



We get a reasonable linearity on these scales

Theoretical dose response relation:



Example from anaesthesia:

47 patients to be operated
with two different anesthetics

- Halothane
- Neurolept

Y: Twitch response at the ulnar nerve
(at the thumb), in %

X: Dose of muscle relaxantia

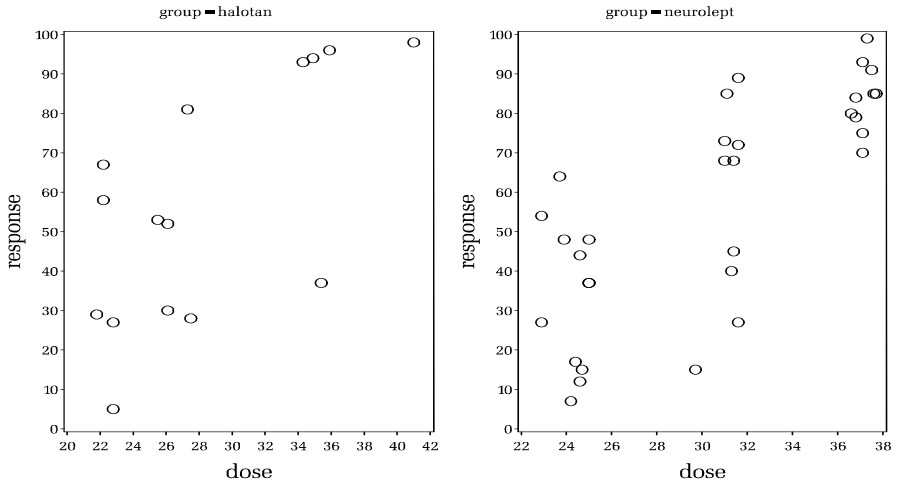
group=halothane

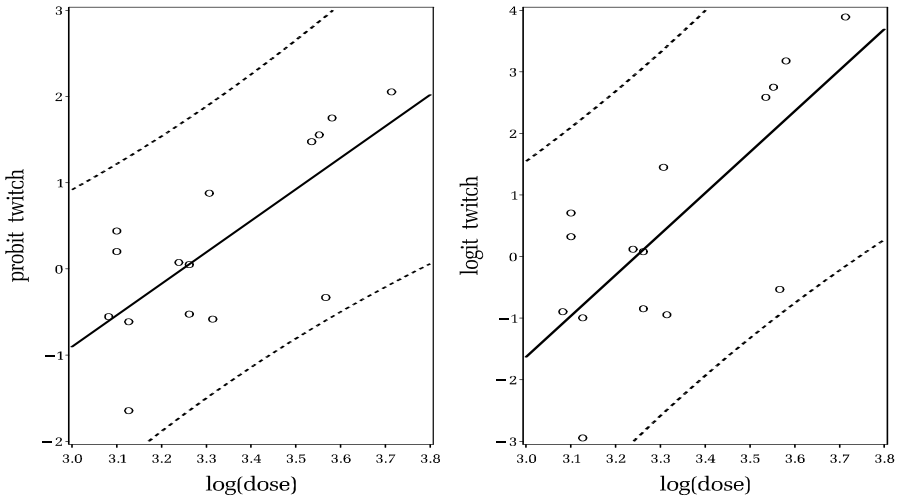
patient	dose	response
1	22.2	58
2	22.8	27
3	22.2	67
.	.	.
.	.	.
.	.	.
13	34.9	94
14	35.9	96
15	35.4	37

group=neurolept

patient	dose	response
1	22.9	27
2	25.0	48
3	24.6	12
.	.	.
.	.	.
.	.	.
30	37.1	93
31	37.7	85
32	37.1	70

Halothane, probit- and logit-transformed





Transformation to linearity

Using two different transformations:

$$y : \texttt{logit\_twitch} = \log\left(\frac{\text{twitch}}{100 - \text{twitch}}\right)$$

or

$$y : \texttt{probit\_twitch} = \text{probit}\left(\frac{\text{twitch}}{100}\right) = \Phi^{-1}\left(\frac{\text{twitch}}{100}\right)$$

$$x : \texttt{logdose} = \log(\texttt{dose})$$

Linear relation:

$\texttt{logit\_twitch} = \alpha + \beta\texttt{logdose}$

produces estimates of α and β, with corresponding standard errors (*s.e.*)

We get estimates of α and β from the equation

$\texttt{logit\_twitch} = \alpha + \beta\ \texttt{logdose}$

But: What about the parameters *of interest*, i.e. ED_{50} and ED_{90}?

$$\hat{ED}_{50} = \exp\left(-\frac{\hat{\alpha}}{\hat{\beta}}\right)$$

How do we calculate *s.e.*$(\hat{ED}_{50})$?

Reparameterization:

$\gamma_1 = \log(ED_{50})$

$\gamma_2 = \log(ED_{90})$

The model may then be written as:

y= `logit_twitch` = logit(0.9) $\times \frac{x-\gamma_1}{\gamma_2-\gamma_1} = 2.197 \times \frac{x-\gamma_1}{\gamma_2-\gamma_1}$

or, using the `probit`-transformation:

`probit_twitch` = Probit(0.9) $\times \frac{x-\gamma_1}{\gamma_2-\gamma_1} = 1.282 \times \frac{x-\gamma_1}{\gamma_2-\gamma_1}$

These functions are **nonlinear** in γ_1 and γ_2!

Direct estimation of γ_1 and γ_2 using
non-linear regression

Estimation

```
data twitch2;
set twitch;

logdose=log(dose);
logit_twitch=log(response/(100-response));
probit_twitch=probit(response/100);
run;

proc nlin data=twitch2;  by group;
  parms loged50=3.2
        loged90=3.6;
  model logit_twitch=probit(0.9)*(logdose-loged50)
                                  /(loged90-loged50);
run;
```

Halothane, output from logit analysis:

```
group=halotan

The NLIN Procedure
Dependent Variable logit_twitch

                                 Sum of       Mean                 Approx
Source                     DF   Squares     Square    F Value    Pr > F

Model                       1   27.6468    27.6468      15.80    0.0016
Error                      13   22.7511     1.7501
Corrected Total            14   50.3980

                                  Approx
Parameter       Estimate     Std Error     Approx. 95% Confidence Limits

loged50           3.2450        0.0551       3.1259       3.3641
loged90           3.5755        0.0814       3.3996       3.7513
```

Halothane, output from probit analysis:

```
group=halotan

The NLIN Procedure
Dependent Variable probit_twitch

                                 Sum of       Mean                 Approx
Source                     DF   Squares     Square    F Value    Pr > F

Model                       1    8.3329     8.3329      14.49    0.0022
Error                      13    7.4776     0.5752
Corrected Total            14   15.8105

                                  Approx
Parameter       Estimate     Std Error     Approx. 95% Confidence Limits

loged50           3.2473        0.0574       3.1234       3.3712
loged90           3.5984        0.0897       3.4045       3.7923
```

Halothane – results

from probit analysis:

Estimate of $\log(ED_{50})$: 3.247 (0.0574)

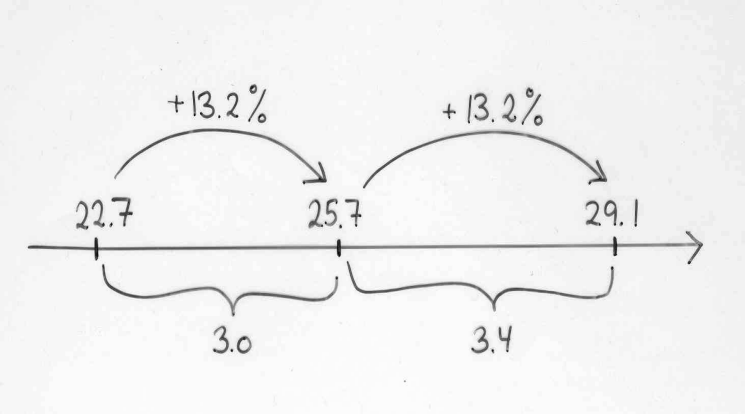
with **confidence interval:**

$$\begin{aligned} 3.247 &\pm 2.16 \times 0.0574 = \\ 3.247 &\pm 0.124 = (3.123, 3.371) \end{aligned}$$

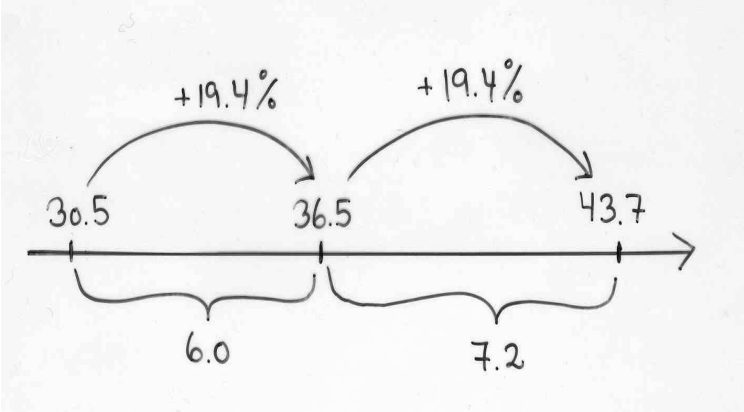
Transformed back to the original scale:
Estimate of ED_{50}: exp(3.247) = 25.7
with confidence interval:

$$(\exp(3.125), \exp(3.371)) = (22.7, 29.1)$$

The confidence interval for ED_{50} is **not symmetric** around 25.7!!

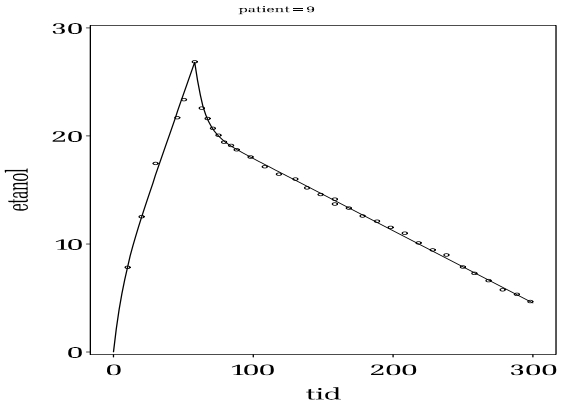


Similarly for ED_{90}:
$\log(ED_{90})$: 3.598 (0.090) ED_{90}: exp(3.598) = 36.5
with confidence interval (30.5,43.7)



A more complicated nonlinear model

Ethanol elimination: Infusion until time t_0:



Theoretical compartment model

- Blood compartment: V_B, c_B
- Peripheral compartment: V_E, c_E
- 1st order kinetics for interchange between compartments k
- 0th order elimination from blood: $v_{\max}$

Letting $T = \frac{V_E^2}{k(V_B+V_E)^2}$ and $\lambda = k(\frac{1}{V_B} + \frac{1}{V_E})$, we find that **until** t_0:

$$c_B = (I_0 - v_{\max})(T(1 - \exp(-\lambda t)) + \frac{t}{V_B + V_E})$$

After t_0:

$$\begin{aligned} c_B &= (I_0 - v_{\max})(T(1 - \exp(-\lambda t)) + \frac{t}{V_B + V_E}) \\ &\quad -I_0(T(1 - \exp(-\lambda(t - t_0))) + \frac{(t - t_0)}{V_B + V_E}) \end{aligned}$$

```
proc nlin;  by patient;
parms ve=18
      vb=6.2
      k=0.6
      vmax=2;
  v=ve+vb;
  t=ve**2/(k*v**2);
  lam=k*(1/vb+1/ve);
  b1=i0-vmax;
  if del=1 then do;
    model etanol=b1*(t*(1-exp(-lam*tid))+tid/v);
  end;
  if del=2 then do;
    model etanol=b1*(t*(1-exp(-lam*tid))+tid/v)
              -i0*(t*(1-exp(-lam*(tid-t0)))+(tid-t0)/v);
  end;
  output out=ny p=yhat r=resid;
run;
```

```
patient=9

The NLIN Procedure
Dependent Variable etanol
Method: Gauss-Newton

                      Iterative Phase
                                                        Sum of
 Iter          ve          vb           k        vmax     Squares

    0     18.0000      6.2000      0.6000      2.0000      1563.7
    1     21.6636      8.5737      0.8629      2.1047     86.1251
    2     22.4079      9.7563      1.0459      2.1391      3.8359
    3     22.4565      9.8477      1.1087      2.1416      2.6708
    4     22.4854      9.8174      1.1161      2.1417      2.6676
    5     22.4879      9.8142      1.1164      2.1418      2.6676
    6     22.4881      9.8140      1.1165      2.1418      2.6676

NOTE: Convergence criterion met.

                              Approx
Parameter       Estimate     Std Error    Approximate 95% Confidence Limits

ve               22.4881       0.5138     21.4402     23.5360
vb                9.8140       0.5867      8.6175     11.0106
k                 1.1165       0.0654      0.9830      1.2499
vmax              2.1418       0.0163      2.1085      2.1750
```