Weighting Data

Why all data are not equal

Weighting - Objectives

- To understand why all data are not equal
- To understand measures of data accuracy
- To understand how to use different weighting schemes
- To understand why different weighting schemes might be used

Weighting Schemes

- Equal Weight
- Reciprocal Variance
- Weighting Schemes
- Iteratively Reweighted Least Squares
- Extended Least Squares
Equal Weight

• Equal (same, similar) error in each data point
• Value of each data point similar
• Error in each data point small
Reciprocal Variance

- Large range of data values
- Error in data points variable
- Relatively larger error in data
- Magnitude of data in different sets are quite different

Reciprocal Variance

\[ \text{Weight} = \frac{1}{\text{Variance}} \]

- Estimate Variance or at least how Variance varies with Observed Value
Reciprocal Variance

Weighting Scheme

• Variance $\propto$ Observed Value

$$\text{Weight} = \frac{1}{\text{Observed Value}}$$

• Variance directly proportional to the measured value
  – For example with radioactive counting

Weighting Schemes

• Variance $\propto$ Observed Value$^2$

$$\text{Weight} = \frac{1}{\text{Observed Value}^2}$$

• Variance directly proportional to the square of the measured value
  – For example when the assay involves serial dilutions
Weighting Schemes

Weighting Schemes

Variance = a•Obs^b

- Plot Variance versus Observed Mean on log-log paper.
  - Determine a and b from intercept and slope, respectively


Estimate Variance from Data

Variance' = 0.0032•Cp^{1.96}
Weighting Schemes

\[ \text{Variance} = c^3 + a \cdot \text{Obs}^3 \]

- Assay sensitivity is \( c \) and \( a \) is a measure of the assay precision

Weighting Schemes

\[ \text{Variance} = a \cdot \text{Obs}^b \cdot e^{(c/t)} \]

- Older assay values, \( t \) smaller, have less weight
  - \( c = 1.05 \) (useful)
- Could be used for clinical samples collected over a number of days (weeks)

Iteratively Reweighted Least Squares

\[ \text{Variance} = f(\text{Calculated Value}) \]

- Very low observed value would be given very high weight with weight = \( 1/\text{Obs}^2 \), for example
Iteratively Reweighted Least Squares

• Variance = Calc^0
• Variance = Calc^1
• Variance = Calc^2

Extended Least Squares

• Weighting Scheme Parameters are obtained from the data DURING the fitting process
• Generally need more data since there are more parameters
• Different Fitting Algorithms needed and not universally available

Extended Least Squares

Objective function: \[ \sum \frac{(\text{Calc}_i - \text{Obs}_i)^2}{V} + \ln V \]

where \( V = f(\text{vp, Calc or Obs}) \)

E.g. \( V = a \cdot \text{Calc}^b \)

Or \( V = a + b \cdot \text{Calc} + c \cdot \text{Calc}^2 \)
Weighting Example

• Fitting Plasma and Urine Data together
• Magnitude of the Data quite different
• Error/Variance formula different
  – Plasma - try constant Coefficient of Variation
  – Urine - try constant Standard Deviation

Plasma Data

• Coefficient of Variation = 5%
  
  C.V. = 5% = \frac{\text{Std.Dev.}}{\text{Value}} \\
  \text{Std.Dev.} = CV \cdot \text{Obs} \\
  \text{Variance} = SD^2 = CV^2 \cdot \text{Obs}^2 \\
  \text{Weight} = \frac{1}{(0.05)^2 \cdot \text{Obs}^2} = \frac{1}{a \cdot \text{Obs}^2} \\
  a = 0.0025 \quad b = 2

Urine Data

• Standard Deviation = 5 mg
  
  \text{Std.Dev.} = 5 \cdot \text{Obs}^0 \\
  \text{Variance} = SD^2 = 5^2 \cdot \text{Obs}^0 \\
  \text{Weight} = \frac{1}{5^2 \cdot \text{Obs}^0} = \frac{1}{a \cdot \text{Obs}^0} \\
  a = 25 \quad b = 0
Boomer Output

Title:  Fit to two lines simultaneously
Input:  From Ch9905b.BAT
Output:  To Ch9905b.OUT
Data for [Drug] came from Ch9905bp.DAT
Data for Drug in Urine came from Ch9905bu.DAT
Fitting algorithm:  DAMPING-GAUSS/SIMPLEX
  Weighting for [Drug]  by 1/a*Cp(Obs)^b
  With a = 0.2500E-02 and b = 2.000
  Weighting for Drug in Urine by 1/a*Cp(Obs)^b
  With a = 25.00 and b = 0.0000
  Numerical integration method: 2) Fehlberg RKF45
  with 2 de(s)
  With relative error 0.1000E-03
  With absolute error 0.1000E-03
ST = 0.1000E-02  PC = 0.1000E-04  Loops = 1
Damping = 1