OPTIMIZATION AND CV ESTIMATION OF A PLATE COUNT ASSAY USING JMP®

Author: Marianne Toft, Statistician, Novozymes A/S, Denmark

ABSTRACT

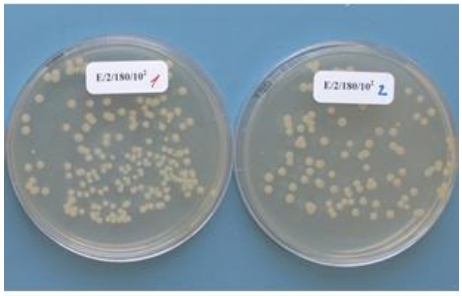
Some of our products are bacterial spores, for which the assay used from early development to product release involves counting bacterial colonies on agar plates – the response being the number of colony forming units (CFU) per gram of product. This presentation will show some studies made during optimization of such an assay. The statistical distribution of CFU data is discussed, and CV estimation involving several sources of variation is presented. This involves mixed linear models with a hierarchical structure of nested random factors.

HOW TO ASSAY A MICROBIAL PRODUCT

The product consists of microbial cells or spores mixed with one or more excipients and the primary assay is enumeration of viable cells or spores, also known as "Colony Forming Units". Secondary assays could be detection of unwanted bacterial strains like Salmonella and E. coli.

Principle of analysis:

- Weigh out a certain amount of product
- Dilute with buffer to a certain dilution factor
- Spread on an agar plate
- Incubate for e.g. 24 hours at 35°C
- Count number of colonies on plate
- Calculate CFU/g as
    - CFU/g = Count*Total dilution



CFU: Colony Forming Units

Calculation example: Count is 178; total dilution is $1 \cdot 10^7$; product has $178 * 1 \cdot 10^7 = 1.78 \cdot 10^9$ CFU/g.

DEVELOPMENT OF A MICROBIAL ASSAY

Bacterial strains are different so it should be tested what the optimal conditions are for the CFU assay. We have made many studies to investigate the design factors and decided on specific conditions.

During development, we also estimate noise factors, because if these are not understood correctly, conclusions about the design factors may be wrong due to a design not taking the noise factors into account.

Examples of Design factors:

- Amount of product weighed out
- How to do sampling
- Buffer to use for dilutions
- Mixing/blending – type and time
- Heat shock temperature
- Agar type
- Incubation temperature
- Method for counting plates

Examples of Noise factors:

- Days
- Weighings
- Plates
- Counting time
- Technicians
- Laboratories

In this paper two studies are used to illustrate how we handle and investigate assay variation during development of a new assay.

ANALYSIS OF STUDY 1

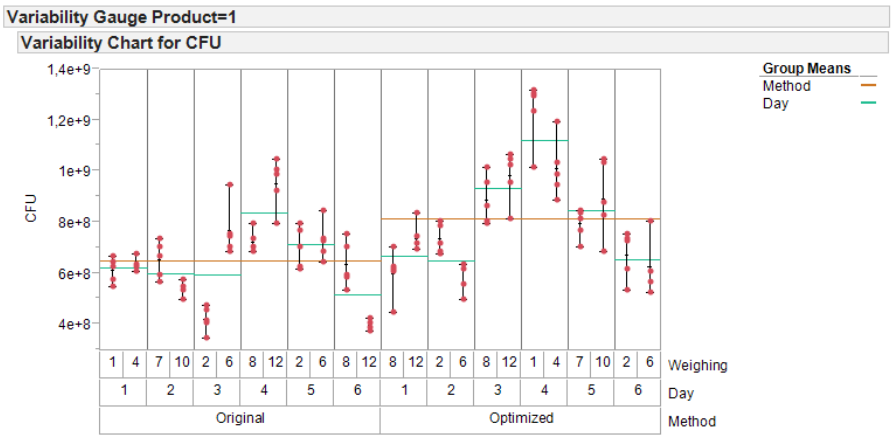
Setup*:

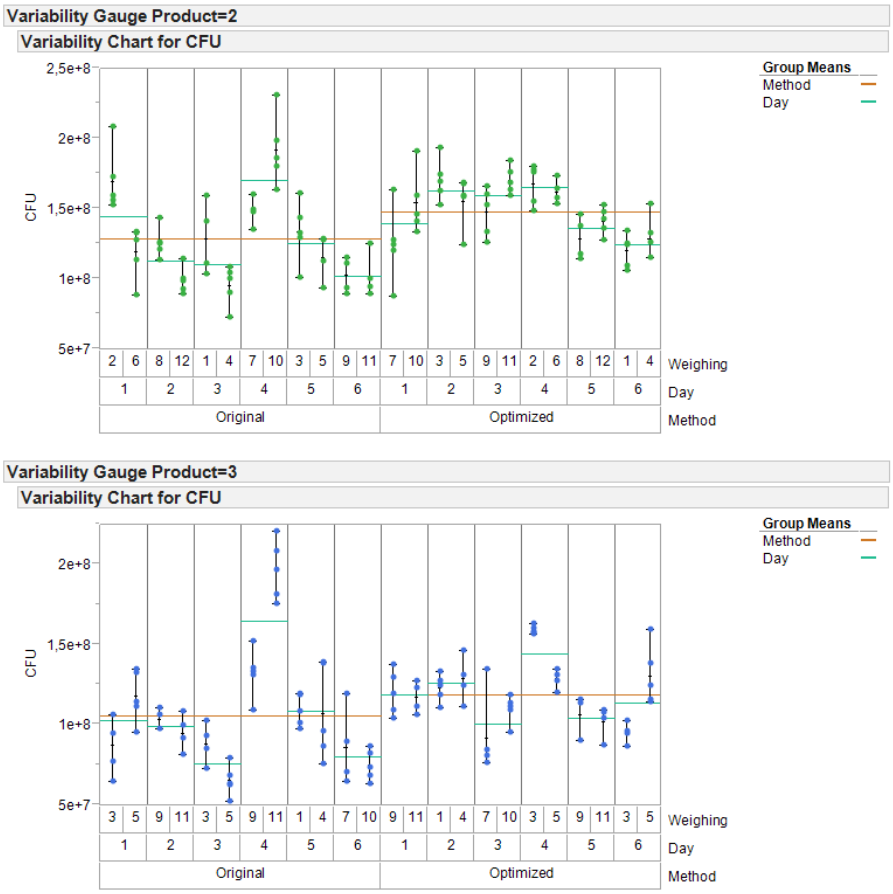
- Three products
- Two methods
- Six days for each method
- Two weighings each day
- Five plates from each weighing
- Response is CFU/g

* The data shown and analyzed here are based on a true study but levels and factor names have been changed for proprietary reasons.

Raw data plots:

The structure of the design is seen in the raw data plots. No obvious outliers are seen.





Modeling:

Product and Method and their interaction are modeled as fixed factors. Day, Weighing and Plate number as modeled as random effects.

How the factor Day is modeled has large influence on the tests for the fixed effects as Day is the upper factor in the hierarchy of random terms. Below are shown three different versions – which of them is most correct depends on a deeper understanding of the assay and what the "day-to-day effect" really represents. Here the different models are shown to emphasize the importance of understanding noise factors of an assay.

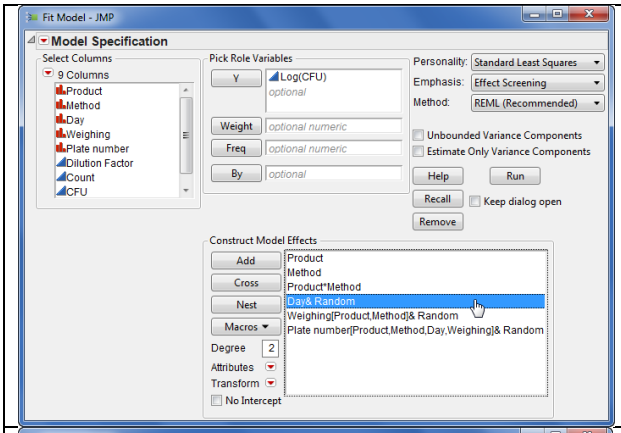
Note that Plate is the lowest level of the random effects hierarchy, so it is confounded with the residual term, which is therefore removed by JMP. Thus removing the term Plate[…] would give the exact same model as below, only the bottom row in the REML table would be named Residual instead of Plate[…]. I like to add it anyway as a check of whether I got the hierarchy right – if there is both a Plate[…] and a Residual term in the REML table, I know I got some nesting wrong (assuming Plate is the lowest level, which it usually is in our studies in this project).

Also note that the response is modeled as Log(CFU). This is not due to the data not being normally distributed (see next section), but due to the fact that the three products have very different levels and since the variation of this assay is relative, the three products do not have the same variance and thus the assumption of variance

homogeneity is violated if modeled without transformation. The Log (natural logarithm in JMP) transformation is thus used to ensure variance homogeneity in the model.

As can be seen from the Fixed Effect Tests output, the term Method has very different p-values in the three different models. This is because the DFDen changes from model to model and the error estimate used in the test differs for the different models.

From the JMP Help: DFDen gives the degrees of freedom for the synthesized denominator. These are constructed using Satterthwaite's method (Satterthwaite, 1946).



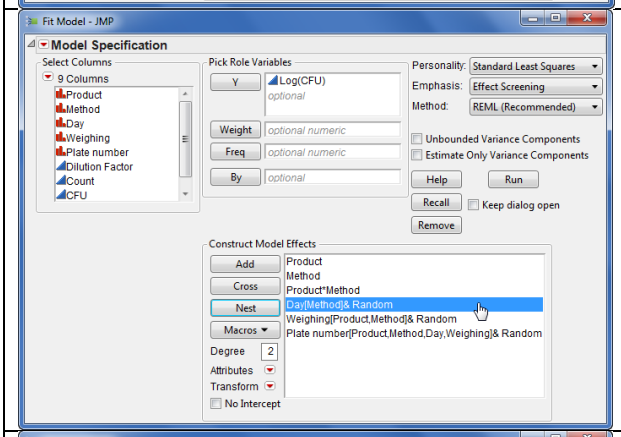
Fixed Effect Tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Product	2	2	43,36	1065,527	<,0001*
Method	1	1	43,14	27,7091	<,0001*
Product*Metho	2	2	43,33	0,3996	0,6730

REML Variance Component Estimates

Random Effect	Var Component	95% Lower	95% Upper	Pct of Total
Day	0,0217021	0,0082198	0,1448572	35,022
Weighing[Product,Method]	0,0121899	0,0073309	0,0241797	19,671
Plate number[Product,Method,Day,Weighing]	0,0280758	0,0239935	0,0333025	45,307
Total	0,0619678	0,0407446	0,1055299	100,000

-2 LogLikelihood= -139,5758731
Note: Total is the sum of the positive variance components.
Total including negative estimates = 0,0619678
Residual is confounded with Plate number
[Product,Method,Day,Weighing] and has been removed.



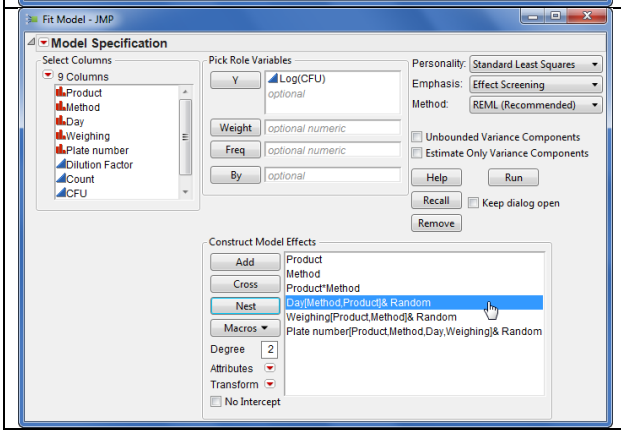
Fixed Effect Tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Product	2	2	39,7	966,9762	<,0001*
Method	1	1	11,66	3,3000	0,0951
Product*Metho	2	2	39,7	0,8913	0,4182

REML Variance Component Estimates

Random Effect	Var Component	95% Lower	95% Upper	Pct of Total
Day[Method]	0,0279039	0,013179	0,0934151	42,205
Weighing[Product,Method]	0,0145632	0,0089037	0,0280462	22,027
Plate number[Product,Method,Day,Weighing]	0,0236478	0,0201717	0,028112	35,768
Total	0,0661149	0,045392	0,1052117	100,000

-2 LogLikelihood= -173,7925691
Note: Total is the sum of the positive variance components.
Total including negative estimates = 0,0661149
Residual is confounded with Plate number
[Product,Method,Day,Weighing] and has been removed.



Fixed Effect Tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Product	2	2	34,9	300,7411	<,0001*
Method	1	1	34,96	7,2534	0,0108*
Product*Metho	2	2	34,9	0,1326	0,8762

REML Variance Component Estimates

Random Effect	Var Component	95% Lower	95% Upper	Pct of Total
Day[Product,Method]	0,0321298	0,0191037	0,0650713	49,440
Weighing[Product,Method]	0,0097101	0,0053031	0,0232184	14,942
Plate number[Product,Method,Day,Weighing]	0,0231469	0,0197668	0,0274797	35,618
Total	0,0649868	0,0488211	0,0908028	100,000

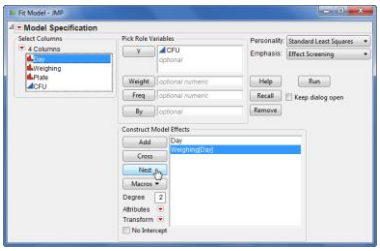
-2 LogLikelihood= -165,9171153
Note: Total is the sum of the positive variance components.
Total including negative estimates = 0,0649868
Residual is confounded with Plate number
[Product,Method,Day,Weighing] and has been removed.

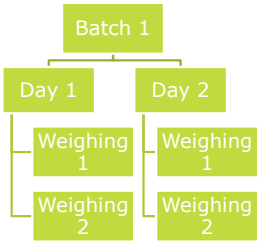
Conclusion:

Modeling and understanding the assay variation is not only important when validating an assay but also during development as the tests for the design factors may be wrong if the wrong model is used.

Fact box: Hierarchy of model effects

In some cases we have to specify factors to JMP as nested under other factors. This happens e.g. when weighings are numbered 1&2 on day 1 and also 1&2 on day 2 instead of in a continuous manner, i.e. 1&2 on day 1 and 3&4 on day 2. Weighing 1 on day 1 has nothing to do with weighing 1 on day 2, so in JMP we nest Weighing under Day as shown below.





Fact box: Linear Mixed Models

A Linear Mixed Model is a model that contains both fixed and random effects.

Example:

$$\text{Response} = \mu + \beta_1(\text{Treatment}) + \beta_2(\text{Temp}) + \beta_{12}(\text{Treatment} * \text{Temp}) + b_3(\text{Day}) + \varepsilon$$

Treatment and Temperature are fixed factors while Day is a random factor.
We are interested in the specific levels of Treatment and Temperature, while we are only interested in the variance of Day.

So this model contains two variance components; Day and residual

$b_3 \sim N(0, s^2_{day})$ and $\varepsilon \sim N(0, s^2)$

INTERMEZZO

What is actually the distribution of CFU results?

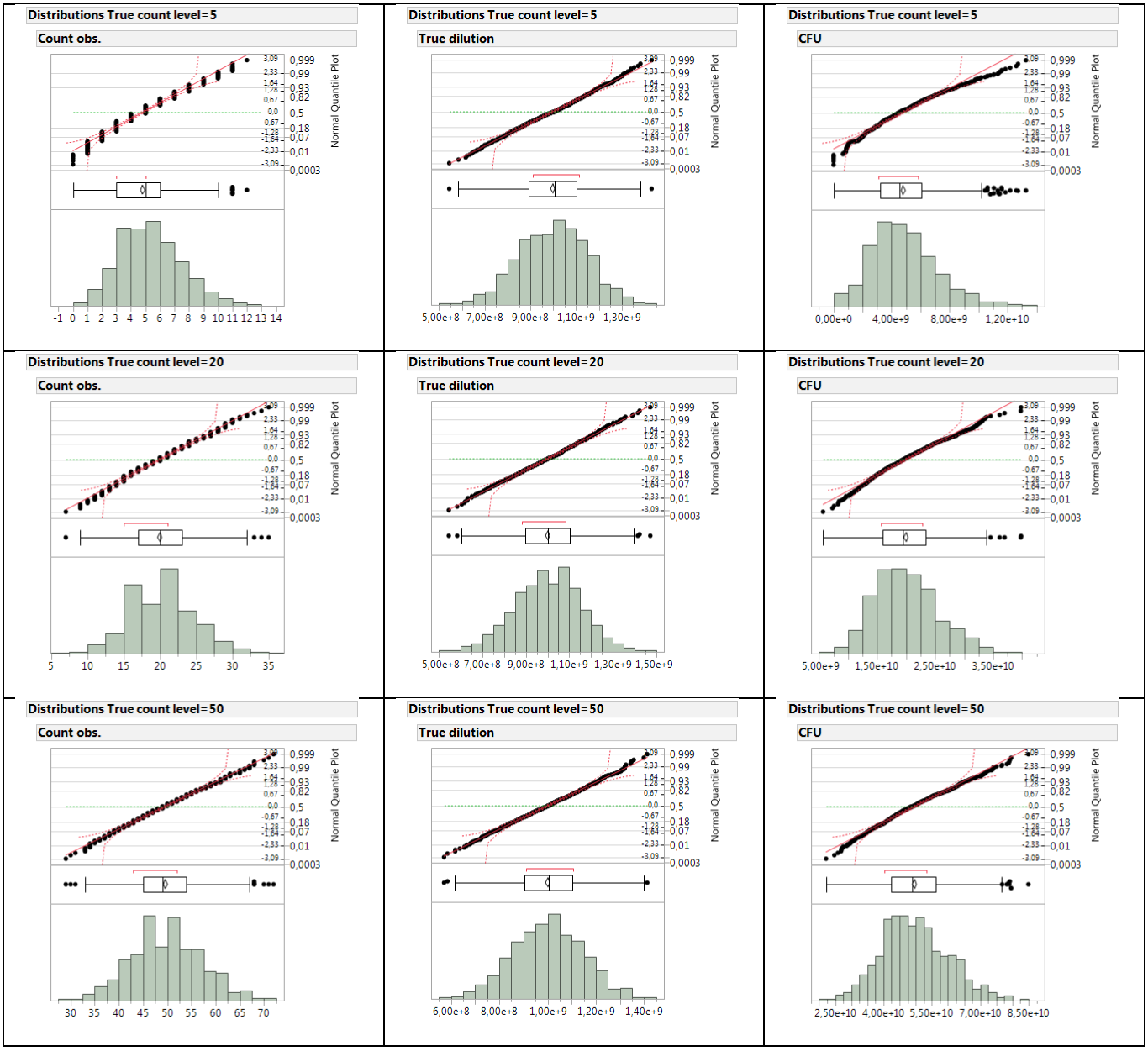
The result CFU/g is calculated from a plate count and a dilution factor; CFU = Count*Dilution. Counts follow a Poisson distribution, while Dilution most probably follows a Normal distribution.

Our experience is that CFU results most often can be considered to be approximately normally distributed.

Simulation shows that if Count is very low (i.e. below 20) the contribution from Count will be larger than the contribution from Dilution and the CFU results will not be normally distributed.

Below are shown some examples from the simulation – top row represents Average Count = 5, middle row Average Count = 20, and bottom row Average Count = 50. Left column shows the distribution of Observed Count, Middle column shows the distribution of Dilution, and right column shows the distribution of CFU. For each setting of Average Count, 1000 rows were simulated.

When Count is very low, the CFU is clearly non-normally distributed, while CFU can be considered approximately normally distributed at higher Counts.



ANALYSIS OF STUDY 2

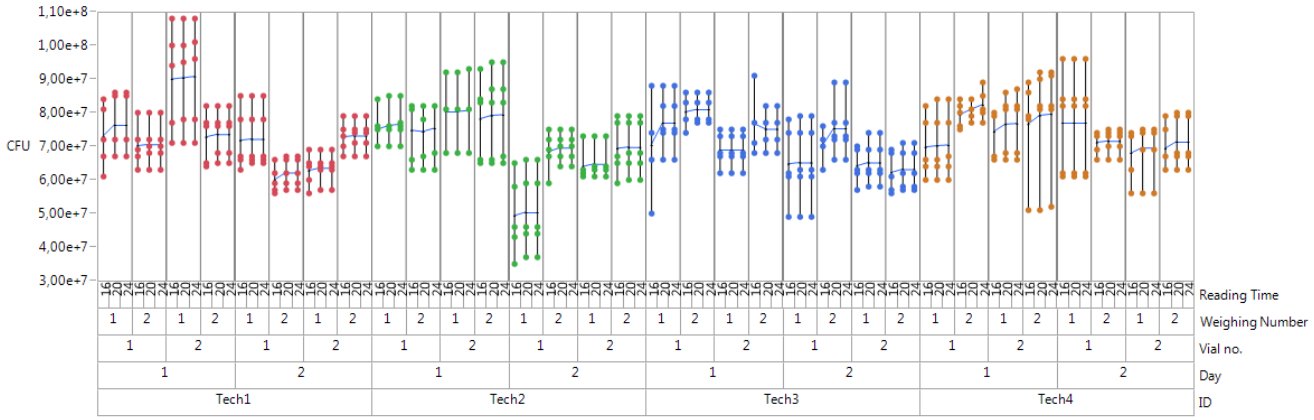
Setup*:

- Four technicians
- Two days each
- Two vials per day
- Two weighings per vial
- Five plates per weighing
- Three reading times per plate
- Response is CFU/g

* The data shown and analyzed here are based on a true study but levels and factor names have been changed for proprietary reasons.

Raw data plot:

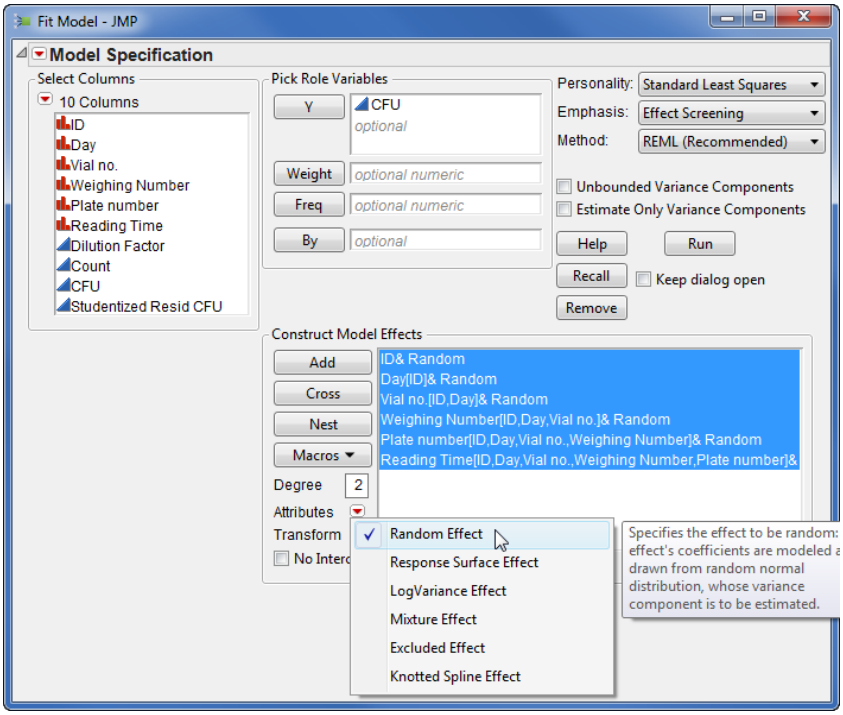
The structure of the design is seen in the raw data plot. No obvious outliers are seen.



Modeling:

All factors are modeled as random effects.

Note that Log is not necessary as transformation of CFU as this study contains only one sample.

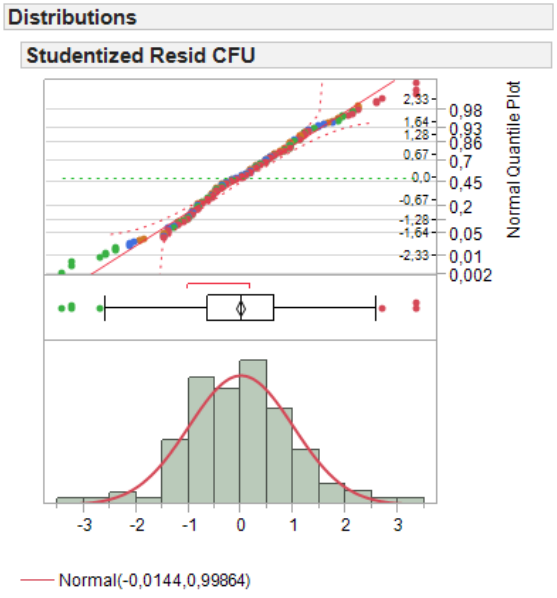


REML Variance Component Estimates

Random Effect	Var Component	95% Lower	95% Upper	Pct of Total
ID	0	0	0	0,000
Day	3,87e+13	7,327e+12	8,579e+16	28,292
Vial no.[ID,Day]	0	0	0	0,000
Weighing Number[ID,Day,Vial no.]	1,952e+13	9,236e+12	6,508e+13	14,271
Plate number[ID,Day,Vial no.,Weighing Number]	7,476e+13	5,9e+13	9,783e+13	54,647
Reading Time[ID,Day,Vial no.,Weighing Number,Plate number]	3,818e+12	3,284e+12	4,494e+12	2,791
Total	1,368e+14	6,824e+13	4,001e+14	100,000

The variance from technicians and vials are estimated to zero.

The studentized residuals from this model can be investigated to check for outliers and to check the assumption of normally distribution of residuals. The residuals look fairly normally distributed, although the tails are a bit heavy. The most extreme residual could be investigated in more detail.



Using the 'Mean Of Response' from the Summary of Fit output, the CV's can be calculated including 95% confidence intervals based on Restricted Maximum Likelihood estimation.

The Total CV is estimated to 15%, with Plate as the largest contributor.

	Random Effect	Var Component	95% Lower	95% Upper	Lower 95%	CV estimate in %	Upper 95%
1	ID	0,00e+0	0,00e+0	0,00e+0	0%	0%	0%
2	Day[ID]	2,02e+13	6,70e+12	2,39e+14	4%	6%	22%
3	Vial no.[ID,Day]	0,00e+0	0,00e+0	0,00e+0	0%	0%	0%
4	Weighing Number[ID,Day,Vial no.]	2,10e+13	9,48e+12	7,95e+13	4%	6%	12%
5	Plate number[ID,Day,Vial no.,Weighing Number]	7,49e+13	5,91e+13	9,80e+13	11%	12%	14%
6	Reading Time[ID,Day,Vial no.,Weighing Number,Plate number]	3,82e+12	3,28e+12	4,49e+12	3%	3%	3%
7	Total	1,20e+14	8,97e+13	1,68e+14	13%	15%	18%

Conclusion:

Estimation of assay variability and variance breakdown is important to be able to understand the strengths and weaknesses of an assay. In JMP Linear Mixed Models are easy to run and interpret.

Fact box: Two approaches to CV estimation

"Top-down" or Variance breakdown approach

- One big study
- Use DOE to be able to estimate main noise factors and let the other noise factors vary
- Estimate total variation from this study
  - Including estimates of contributions from main noise factors

"Bottom-up" or measurement uncertainty approach

- Many small studies
- Estimate variance contribution from each noise factor separately
- Add them all together
- Example Density
  - $\rho = \frac{m}{V}$
  - $\frac{u(\rho)}{\rho} = \sqrt{\left(\frac{u(m)}{m}\right)^2 + \left(\frac{u(V)}{V}\right)^2}$