



STAM01
VT 2010
Master thesis 15 ECTS

LUND UNIVERSITY
School of Economics and Management

Department of Statistics

Statistical analysis of experimental designs applied to biological assays

Student: Aldana Rosso

Supervisor: Peter Gustafsson

Abstract

Bioassays are methods employed to estimate the effect of a given substance in living matter, and therefore they are frequently used in the pharmaceutical industry.

The experimental design of bioassays has to take into account the intrinsic variability in the biological test units and other factors as operators, day variation, batch variation, etc. Thus, statistical models employed to analyse bioassays include both fixed and random effects. The sample size estimation in mixed models is a complicated issue and no general formulas can be applied. An alternative approach to estimate the sample size to obtain a certain estimate with a confidence interval of a specific width is to perform computer simulations.

In this thesis, simulations are performed to calculate the confidence interval of the logarithm of the effect of the test relative to the standard as a function of the number of replicates. The simulated data is compared with experimental data. The results obtained with the simulations agree well with the experimental ones. Furthermore, the method discussed here can also be used to analyse other experimental designs in which the size of the main sources of data variability are known.

Acknowledgments

I want to thank to my supervisor Peter Gustafsson from the Department of Statistics at the Lund University School of Economics and Management, for encourage and support.

I am also thankful to Novo Nordisk A/S for providing the data used in this work.

I am deeply grateful to Dr. Henrik Enqvist from MAX-lab, for his assistance with the MATLAB program.

1	Introduction.....	5
2	Principles of biological assays.....	7
2.1	Biological assays.....	7
2.2	Types of biological assays.....	7
2.3	The dose response relation.....	8
2.4	Parallel line assays.....	9
3	Statistical analysis of bioassays.....	11
3.1	Experimental design.....	11
4	Mixed models applied to bioassays.....	13
4.1	Data.....	13
4.2	Experimental design.....	13
4.3	Statistical models.....	15
4.3.1	Model 1: Full model.....	16
4.3.2	Model 2: Reduced model with two random factors.....	18
4.3.3	Model 3: Reduced model with one random factor.....	20
4.3.4	Investigation of non-constant variance.....	22
4.3.5	ANOVA F-test and the normality assumption.....	24
4.3.6	General considerations about the experimental design.....	24
5	Data simulations for bioassays.....	26
5.1	Relative log-potency calculation of the experimental data.....	26
5.1.1	Parallelism.....	27
5.1.2	Relative log-potency calculations.....	27
5.2	Data simulation.....	28
5.2.1	Simulated relative log-potency.....	28
5.2.2	Results.....	33
6	Conclusions.....	35
7	References.....	37
8	Appendix.....	38
8.1	MATLAB code to calculate log-relative potency.....	38
8.1.1	Function to used to simulate data.....	38
8.1.2	Program to calculate the logarithm of the relative potency between two substances.....	38

1 Introduction

The design of products and manufacturing processes are very important activities in most types of industries. The success of many companies relies on their capability to efficiently develop new products to meet the market demands and to deliver existing ones with the right level of quality. In this context quality means that the properties of the product are within a certain range. Some properties are easily measured, while the estimation of others requires sophisticated procedures. Experiments are used to confirm or reject a hypothesis, to characterize a process, or to evaluate the performance of a system. In general, the objective of an experiment is to determine how certain variables affect a system. The way to measure this effect is to study a response to some stimulus.

In industrial applications, experiments are often performed to determine the most influential variables in the manufacturing processes. If the most relevant variables are known, the variability of the process can be controlled. The response is affected not only by the input variables that the experimenter controls but also by other variables, for example, small changes in temperature, difference in operators, etc. The experimental design should take into account all the relevant variables that affect the response so they are represented in the data. The statistical analysis of the data takes these variables into account and therefore objective conclusions can be drawn from the data.

Statistical tools are widely used within the pharmaceutical industry. Statistical designed experiments are used to determine the properties of the drugs, like for example content and efficiency of the active ingredient. The concentration of a medicine can be determined by the effect of the active ingredient on the body. One of the experiments used to determine the effect of a drug relative to a known standard is the biological assay. In biological assays, the variable is the amount of drug, the system is a cell, and a response is a property of the cell that changes when the drug is given. The experimental design employed in biological assays is different from those typical used in chemical and physical experiments since the intrinsic variation in living matter increases the complexity of the experiment and the data analysis. The experimental design has not only to take into account the intrinsic variability in the biological test units but also other factors as operators, day variation, batch variation, etc.

In this thesis a discussion about how experimental designs are used in biological assays is presented. Furthermore, experimental data from a pharmaceutical company is presented and analysed. The data consist on a coagulation study performed using batches with different properties. The main objective of this study is to determine the biological activity of these batches. As mentioned, the experimental results are affected by external factors. Therefore, in order to assure that the correct value of response is obtained from the analysis it is of fundamental importance to characterize the variability of the experimental method.

One of the main objectives of the analysis presented here is to determine how the variability of the data affects the uncertainty of the reported results. By knowing the

uncertainty of the estimate, it is possible to decide in advance how many experiments are needed to estimate the response with certain precision. This is of fundamental practical importance since biological experiments are expensive and time consuming. The first part of the thesis presents an introduction to bioassays and the statistical methods used to analyse the experimental data. An estimation of the uncertainty of the estimate is obtained.

In the case of complicated experiments like biological assays with several factors, there is not a general formula to estimate how many experiments are needed in order to assure that the confidence interval of the estimate has a certain size. An alternative approach to estimate the how many experiments are needed is to perform computer simulations with data that has the same variability as the experimental data. The second part of this work shows the results from simulations based on the results of the analysis of the experimental data. By using this method, the number of experiments can be estimated. The procedures presented here are valid for other experiments in which the response is deeply affected by several variables that cannot be controlled.

The structure of the thesis is as follows: a brief description of biological assays is presented in Chapter 2 and a standard experimental design is described in Chapter 3. Data from a biological assay taken from the pharmaceutical industry is analysed in Chapter 4. Simulations based on the data analysis are discussed in Chapter 5. A summary of the work is presented in Chapter 6.

2 Principles of biological assays

2.1 Biological assays

A biological assay or bioassay is employed to determine the effect of a substance on a certain type of living matter [1, 2]. A biological test system, for example animals, tissue, etc, is exposed to a particular stimulus like a drug, whose concentration (dose) is usually varied. The magnitude of the response of the biological system depends on the dose. In contrast with physical or chemical methods, detailed information of the drug activity as a function of the dose is obtained.

A special characteristic of bioassays is that one of the largest sources of variation in the outcome is the difference between the test units, and since the response is dependent on living matter this introduces large variability between measurements obtained by identical operations [1-3].

The response can be a characteristic like body weight of the test units, or a change in a characteristic, change in body weight, or the occurrence of a certain phenomena, e.g. death. For example, in toxicological assays several doses of a substance are given to rats and the response variable is the survival of the rats [1].

Biological assays are usually comparative. The capacity of a substance to cause a specific effect is estimated relative to a standard. The standard and the test preparations are identical in their biological activity principle and differ only in extends to which they are diluted by solvents. This type of bioassay is used in the pharmaceutical industry to determine the potency of new batches relative to a standard sample [2, 4-5].

2.2 Types of biological assays

There are two types of quantitative biological assays: direct and indirect assays [1]. The principle of a direct assay is to measure the doses of the standard and the test preparations that produced a specified response. The potency of the test preparation relative to the standard is defined as the amount of standard equivalent in effect to one unit of the test [1]. In other words, the potency of the test is defined as how much standard it is needed to produce the same effect as with one unit of the test substance. A direct assay is of limited applicability since it requires measuring the exact dose needed, not merely one that is large enough to produce a certain effect. The response must be easily recognized and the dose must be administered in a manner that the exact amount needed to produce the response is recorded [1].

In an indirect assay, several doses are given to different biological units. The response is recorded for each dose (see Fig. 1). The relative potency of the test sample is determined relative to that of the standard by statistical analysis of the dose-response relation.

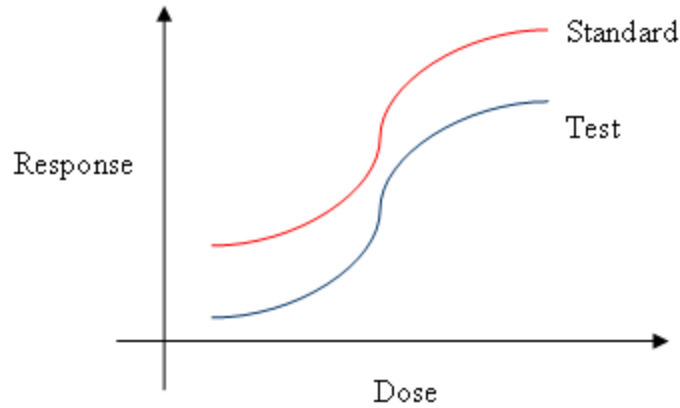


Figure 1 Typical bioassay data. A standard and a test substance are given to biological units. A response is recorded and the relative potency of the test substance can be determined relative to the potency of the standard.

2.3 The dose response relation

In order to compare the biological response caused by a test sample relative to a standard preparation, the test sample should contain the same active substances as the standard preparation. Furthermore, the test sample should not contain any other chemical, like for example impurities, which can affect the response variable. This implies that the test substance behaves as a dilution or concentration of the standard preparation.

A random sample of several identical biological units receives a dose z of the standard preparation, which produces a response of size U_s . The dose-response function is $F_s(z)$, which is a real single valued function in the dose range used in the assay. The same procedure is repeated for the test preparation. The response function for the test preparation is $F_t(z)$. For a certain response U , the doses are Z_s and Z_t (see Fig. 2).

Since the test preparation behaves as a dilution of the standard preparation, the mathematical form for the response is the same for both preparations. This condition is known as similarity, and it is a prerequisite of all dilution assays. For all doses Z is fulfilled that

$$Z_t * \rho = Z_s$$

This expression in log-scale is $\log(Z_t) + \log(\rho) = \log(Z_s)$. The logarithm of the relative potency is the horizontal distance between the two responses (see Fig. 2).

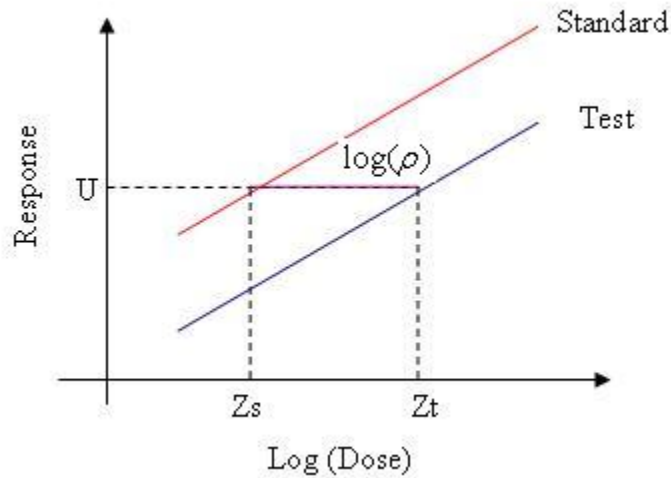


Figure 2 Dose-response curve for a bioassay. The relative potency is defined as the difference in dose between the standard and the test preparation that produces the same response.

2.4 Parallel line assays

In general, dose-response functions are not linear but they can be transformed to become linear. One of the most frequently used transformations is the log-transformation [2]. In the most widely used type of assay, the transformed response has a homoscedastic linear regression on log-dose, that is the variance of the log-response is independent of the log-dose. For such an assay the condition of similarity is equivalent to the condition that the test and the sample transformed response curves are parallel.

An example is presented in Fig. 2. The dose-response curves of the test and the standard preparation are

$$Y_t = a_t + b * \log(dose)$$

$$Y_s = a_s + b * \log(dose)$$

The relative potency is defined as the amount of standard sample needed to produce the same response as with the test preparation, therefore the relative log-potency is the horizontal distance between the two responses:

$$a_s + b * \log(dose_s) = a_t + b * \log(dose_t)$$

Therefore

$$\log(\rho) = \log(dose_t) - \log(dose_s) = \frac{a_s - a_t}{b}.$$

The value of the relative potency is obtained by taking the exponential of $\log(\rho)$.

3 Statistical analysis of bioassays

3.1 Experimental design

A typical experimental design used in bioassays is presented in Fig. 3. A sample of the test and standard drug are taken randomly from a production batch. The samples are diluted in different ways until reaching the final preparation. The final preparations are given to a certain biological system in several doses, and a numerical response is measured. Several response measurements are usually taken for each dose. The whole procedure is repeated several times.

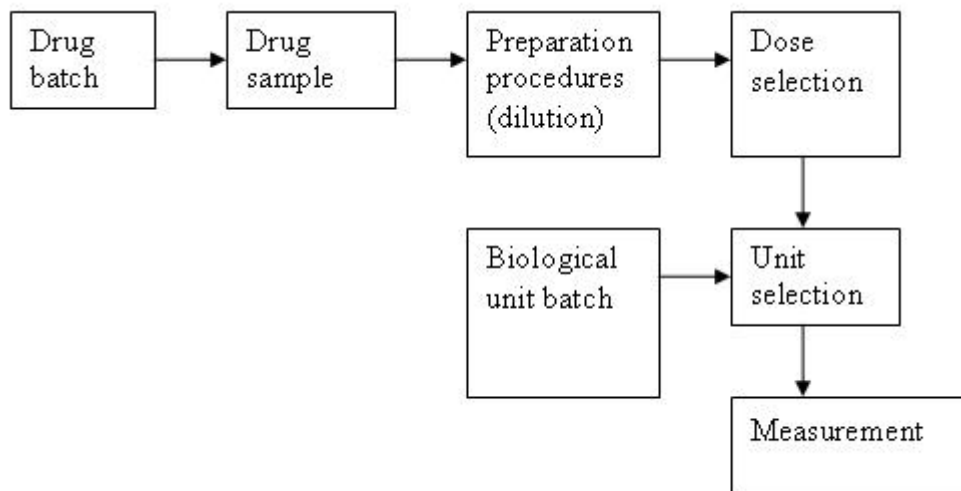


Figure 3 Schematic representation of the most common steps in a bioassay.

The relation between dose and response in biological systems is usually non-linear. However, if a small range of doses is used, it is usually possible to find a transformation that the response variable linear is this interval. The log-transformation is frequently applied [2], and a linear model is used to describe the log-response. Additionally, the dose axis is usually also log-transformed since it is easier to calculate the logarithm of the relative log potency.

In addition to the intrinsic variability due to differences between the biological test units, there is variability due to other experimental factors, due to variation in sample preparations, technicians, equipment, batches, etc. A model that describes the response variable as a function of the dose and sample type, must take into account all the factors that affect the response. The factors like sample type (test and standard) and dose levels are the specific levels of interest and they are denominated fixed factors. The conclusion of the experiments is only valid for the levels measured during the experiments and cannot be extended to other levels. It is desirable to extend the conclusions from the data analysis, not only to a certain machine but all the machines, and for all the batches, etc., therefore several factors included in the experimental design will be modelled as random

factors. This implies that it is assumed that the levels of these factors are a random sample of the entire population, and therefore the conclusions from the data analysis can be extended to all levels in this population. Statistical models that include fixed and random factors are denominated mixed models.

Since in the experimental set-up both random and fixed effects are included, the model used to analyse the data is a mixed model. The observed variance can be partitioned into components related to the experimental variables. This method is called analysis of variance (ANOVA) [7] and it allows attributing the variance to different experimental parameters. Therefore, it is used in this work to identify which parameters introduce the largest variation.

The observations must fulfil the following assumptions [2]:

- Independence
- Constant variance of the responses
- Normally distributed residuals with mean zero and constant variance.

During the sample preparation it is common to violate the independence assumption. For example, if a multichannel pipette is used, the observations are not really independent. Therefore it is important that the statistical model take into account blocks (e.g due to serial dilution) [2]. If these factors are not considered in the model, the estimates will not be as precise as they could be [2].

The ANOVA and linear regression methods are reasonably robust to mild departures from assumptions regarding constant variance or normality [7]. In many cases, the data can be transformed so the transformed response will be sufficiently close to constant variance and normality.

4 Mixed models applied to bioassays

4.1 Data

The data analysed in this work correspond to a coagulation assay and it is provided by Novo Nordisk A/S.

4.2 Experimental design

A bioassay is performed to compare the response of a biological system when exposed to four batches of various nature and quality.

The factors expected to affect the response of the biological system to the drug are:

- Biological unit
- Batch
- Operator
- Sample preparation process
- Dose
- Other factors (temperature, day of the week, etc.)

In order to assure that all the variation sources are included in the experiment, the set-up presented in Fig. 4 and Fig. 5 is implemented. A sample is selected randomly from each batch and treated in several steps to obtain five different concentrations. The final preparations are given to randomly selected biological units. Two measurements are taken for each unit. The experiment is replicated in 10 blocks during a period of one week. Therefore, the block effect and the date effect are highly confounded.

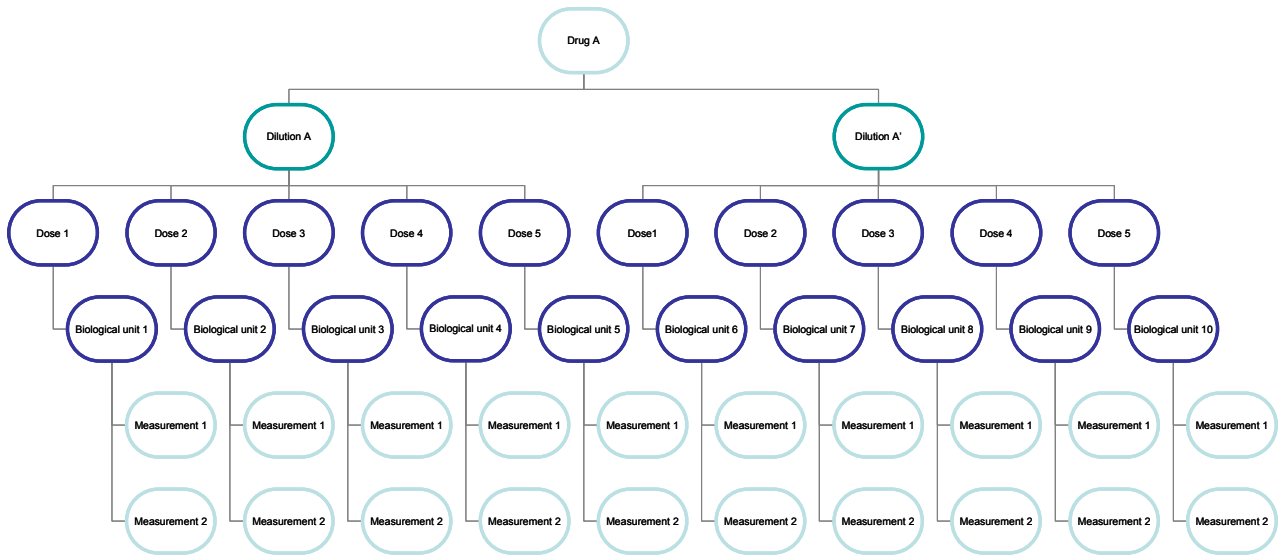


Figure 4: Experimental set-up used in a bioassay experiment. This procedure is performed for all samples taken from four batches. The experiment is replicated using new preparations and new biological units ten times during a week.

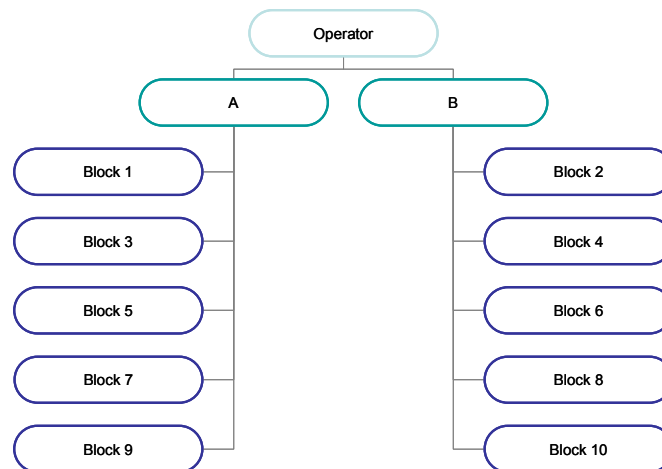


Figure 5 Experimental set-up used in the biological assay. Each operator was assigned to five replicates (blocks)

Typical data from the experiment is shown in Fig. 6. There it is observed that the dispersion in the data increase when increasing the dose¹. A log-transformation is applied to the data (see Fig. 7). The log-transformed data is linear in log-dose and an ANOVA model will be used to analyse the transformed data.

¹ Dose is defined as the inverse of the dilution.

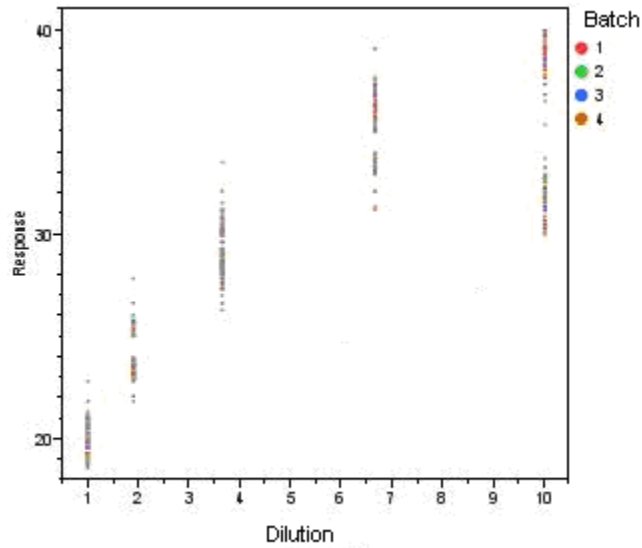


Figure 6: Typical results obtained in the bioassay.

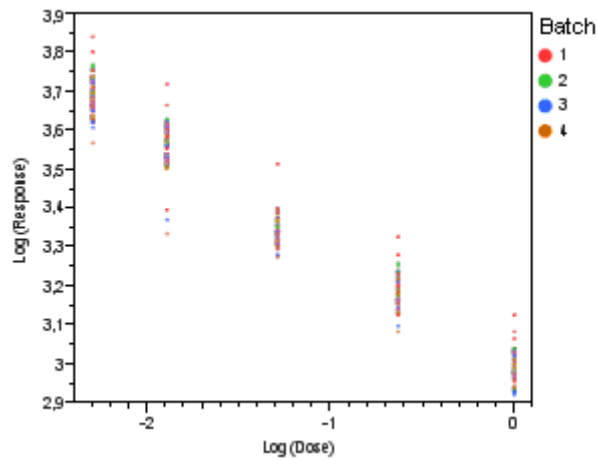


Figure 7 Log-transformed data from a bioassay.

4.3 Statistical models

In order to characterize the experimental set-up it is important to identify which are the largest variation sources.

A mixed model is applied to estimate the different variance component. The fixed effect of the model are: $\log(\text{dose})$ and batch. The following factors are modelled as random:

- Block
- Operator
- Sample preparation process

These factors are considered random since the conclusions of the assay are to be valid for all possible operators, sample processes and blocks. In other words, the levels of these factors are randomly selected from a larger population of possible levels.

The log-transformed data is assumed to follow a normal distribution. Furthermore, it is assumed that the variance is constant across the batches and log-dose's levels. Several models are fitted to the transformed data and these approximations and the results are discussed in the next section.

4.3.1 Model 1: Full model

The model is

$$\log(Y_{ijklms}) = \mu + \delta_i + \alpha_j + \gamma_k + \kappa_l + \psi_m + \varepsilon_{ijklms}$$

Where

μ is the overall mean

δ_i is the batch effect with i levels: batch 1,2, 3 and 4.

α_j is the block effect with j : 1,2,3...,10.

γ_k is the operator effect with k levels (T1 and T2).

κ_l is the log-dose effect with five l levels: 0, -0.6, -1.3, -1.9 and -2.3

ψ_m is the sample preparation effect with m levels: 1, 2.

ε_{ijklms} is the residual for each data point with s replicates (s :1,2)

The factors batch and dose are considered as fixed, therefore the inference from the experiment is only valid for the levels studied here.

The random effects are considered to be independently and normally distributed with mean zero and variance $\sigma_\psi^2, \sigma_\alpha^2$ and σ_γ^2 . The error term is assumed to be independently and normally distributed with mean zero and variance σ^2 . Furthermore, the random factors and the residuals are independent.

The model was fitted in SAS JMP using the “fit model” routine. The results are shown in Tabs. 1 and 2.

Table 1 ANOVA table for model 1. Both fixed effects are statistically significant at the 0.05 significance level. The degrees of freedom (DF) of the factors are indicated as well as the DF of freedom of the denominator of the test.

Source	DF	DF Denominator	F Ratio	P-value
log(dose)	1	717.40	57416.21	0.00
Batch	3	58.70	3.54	0.02

Table 2 Variance components for model 1. The random effects with larger contribution to the total variability are the residuals and the block effect.

Random Effect	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Block	0.0011451	0.0005909	-0.000013	0.0023033	52.97
Operator	0.0000979	0.0001534	-0.000203	0.0003986	4.53
Sample preparation process	0.0000656	2.679e-5	0.0000131	0.0001181	3.04
Residual	0.0008531	4.5053e-5	0.0007713	0.0009488	39.47
Total	0.0021617				100.00

As expected, both log-dose and batch are statistically significant effects. For the random effects, the block effect contributes most to the variability of the data after the residual contribution. The standard error of the estimate of the variability due to the factor operator is much larger than the estimate. This is due to the fact that only two levels of operator are tested in the experiment. In order to get a somewhat good estimator of the variability it is strongly recommended to include several levels of the random factor [7].

Different plots for the residual analysis of Model 1 are shown in Fig. 8. The normality assumption of the residuals can be checked by using the histogram and the normal plot (Fig. 8 (C)). Gross deviations from the normality assumptions are not observed. The plot of the residuals against the date in which the measurements are performed is useful to detect correlation between the residuals and violation of the constant variance assumption. From Fig. 8 (A) it seems that the residuals increase with time. This is a concern, because it indicates that some factor in the experiment changed overtime. Figure 8 (B) also provides information about the violation of the constant variance assumption. If the model is correct and the assumptions are satisfied, the residual should be structureless, and they should be uncorrelated to any other variable. It seems that the residuals for middle log-dose values are more negative than for the other log-dose values. Additionally several potential outliers are observed.

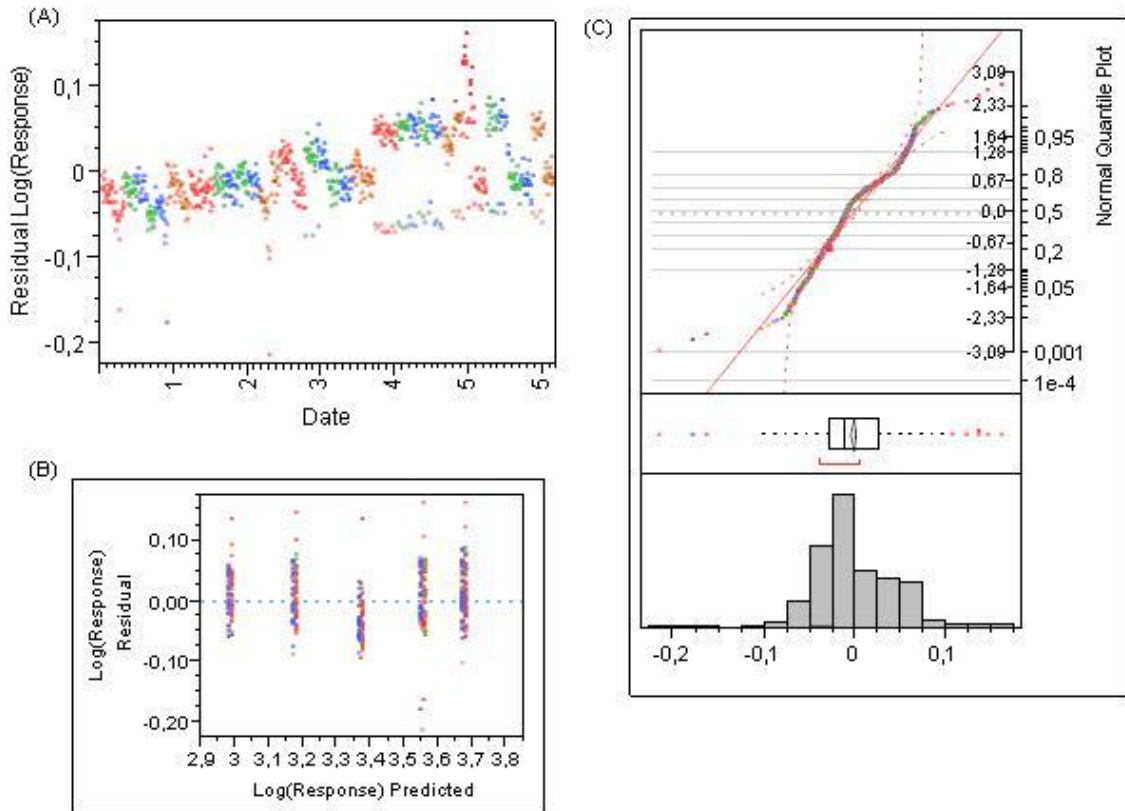


Figure 8 Residual of Model 1. (A) The residuals are plotted by day to identify potential correlation between the residuals. (B) The predicted values are plotted against the residuals to detect possible violation of the constant variance across group assumption. (C) Histogram and normal probability plot of the residuals are used to check the normality assumption.

4.3.2 Model 2: Reduced model with two random factors

It is seen in Tab. 2 that it is not possible to estimate the variation due to the factor operator. Therefore, a simpler model without this factor (reduced model) is fitted to the data. The model is

$$\log(Y_{ijlms}) = \mu + \delta_i + \alpha_j + \kappa_l + \psi_m + \varepsilon_{ijms}$$

Where

μ is the overall mean

δ_i is the batch effect with i levels: batch 1,2, 3 and 4.

α_j is the block effect with j : 1,2,3...,10.

κ_l is the log-dose effect with five l levels: 0, -0.6, -1.3, -1.9 and -2.3

ψ_m is the sample preparation effect with 2 m levels

ε_{ijms} is the residual for each data point with s replicates ($s:1,2$)

As in the full model, the only fixed factors are batch and log-dose.

The results are shown in Tabs. 3 and 4.

Table 3 ANOVA table for model 2. Both fixed effects are statistically significant at the 0.05 significance level. The degrees of freedom (DF) of the factors are indicated as well as the DF of freedom of the denominator of the test.

Source	DF	DF Denominator	F Ratio	P-value
log(dose)	1	719.10	56538.15	0.00
Batch	3	58.66	3.54	0.02

Table 4 Variance components for model 2. The random effects with larger contribution to the total variability are the residuals and the block effect.

Random Effect	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Block	0.0009054	0.0004622	-5.157e-7	0.0018113	49.31
Sample preparation process	6.4381e-5	2.6786e-5	1.188e-5	0.0001169	3.51
Residual	0.0008664	0.0000457	0.0007833	0.0009634	47.18
Total	0.0018362				100.00

Likewise in Model 1, both log-dose and batch are statistically significant effects. For the random effects, the block effect contributes most to the variability of the data after the residual contribution.

Different plots for the residual analysis of Model 2 are shown in Fig. 9. Also in this case, gross deviation from the normally assumptions are not observed. The assumption of non-constant variance is not well fulfilled in this model either.

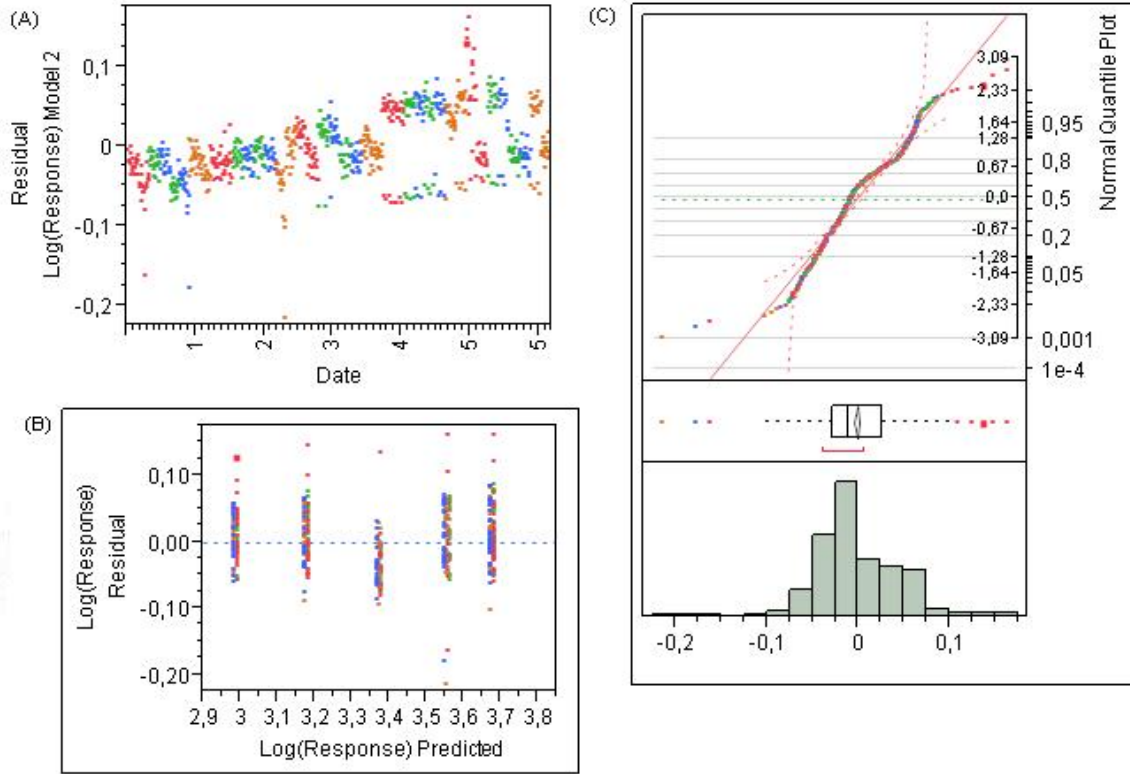


Figure 9 Residual of Model 2. (A) The residuals are plotted by day to identify potential correlation between the residuals. (B) The predicted values are plotted against the residuals to detect possible violation of the constant variance across group assumption. (C) Histogram and normal probability plot of the residuals are used to check the normality assumption.

4.3.3 Model 3: Reduced model with one random factor

Since approximately 50 % of the variation in the data is due to the blocks (see Tab. 2 and 4), thus a random model with only this factor is considered.

The model is

$$\log(Y_{iljs}) = \mu + \delta_i + \alpha_j + \kappa_l + \varepsilon_{iljs}$$

Where

μ is the overall mean

δ_i is the batch effect with i levels: batch 1,2, 3 and 4.

α_j is the block effect with j : 1,2,3...,10.

κ_l is the log-dose effect with five l levels: 0, -0.6, -1.3, -1.9 and -2.3

ε_{ijls} is the residual for each data point with s replicates ($s:1,2$)

Likewise in the full model, the only fixed factors are batch and log-dose.

The results are shown in Tabs. 5 and 6.

Table 5 ANOVA table for model 3. Both fixed effects are statistically significant at the 0.05 significance level. The degrees of freedom (DF) of the factors are indicated as well as the DF of freedom of the denominator of the test.

Source	DF	DF Denominator	F Ratio	P-value
log(dose)	1	778	53228.35	0.00
Batch	3	778	6.68	0.00

Table 6 Variance components for model 3. The random effects with larger contribution to the total variability are the residuals and the block effect.

Random Effect	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Block	0.0009108	0.0004612	6.8898e-6	0.0018146	49.735
Residual	0.0009205	4.667e-5	0.0008354	0.0010192	50.265
Total	0.0018312				100.000

As in the other cases, both log-dose and batch are statistically significant effect. The estimate for the variance component of block is similar to that of Model 2. In comparison with Models 1 and 2, Model 3 has larger degrees of freedom to estimate the fixed effects, since this model contains only one random effect.

Different plots for the residual analysis of Model 3 are shown in Fig. 10. Large deviations from the normality assumptions are not observed. The assumption of non-constant variance is not well fulfilled in this model either.

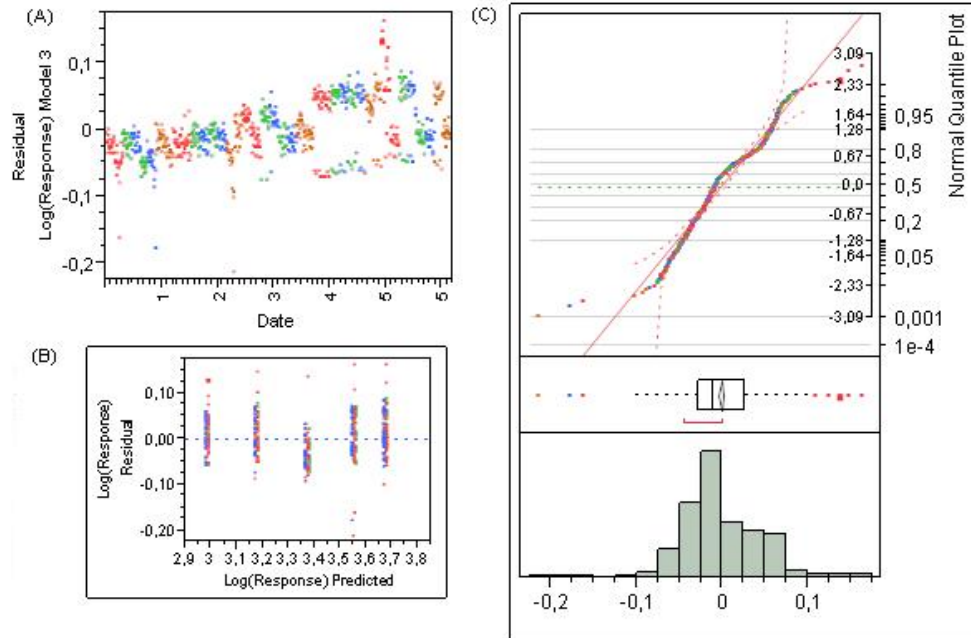


Figure 10: Residual of Model 3. (A) The residuals are plotted by day to identify potential correlation between the residuals. (B) The predicted values are plotted against the residuals to detect possible violation of the constant variance across group assumption. (C) Histogram and normal probability plot of the residuals are used to check the normality assumption.

4.3.4 Investigation of non-constant variance

As observed in Figs. 8-10, there are indications that the error variance is not constant. This is an important concern in models with random factors since the estimation of the variances of the random effects can be affected. Figure 11 show the data plotted by batch, block and by log-dose. The variability of the data is similar for all batch types, and for all block levels. The variability increases with log-dose, meaning that the detection and/or the sample preparation methods do not perform as well for large log-dose values.

A formal test of the equality of variance can also be performed. Bartlett's test is frequently used to test the equality of variance hypothesis [1, 7]. This test compares the weighted arithmetic average of the sample variances to the weighted geometric average of the sample variances [8]. Under the null hypothesis (the variances of all treatments are equal), the geometric mean is less than or equal to the arithmetic mean. A difference between the means indicates that the variances between treatments differ. The Bartlett's test is based on a function of these means that approximates a χ^2 -distribution. A drawback of this procedure is that it is very sensitive to the normality assumption.

The modified Levene test is robust to departure from normality [7]. This test calculates the absolute deviation of the observations in each treatment from the treatment median. Then it evaluates whether the means of these deviation are equal for all treatments. The results of the Bartlett's test and the modified Levene test are listed in Tabs. 7-9.

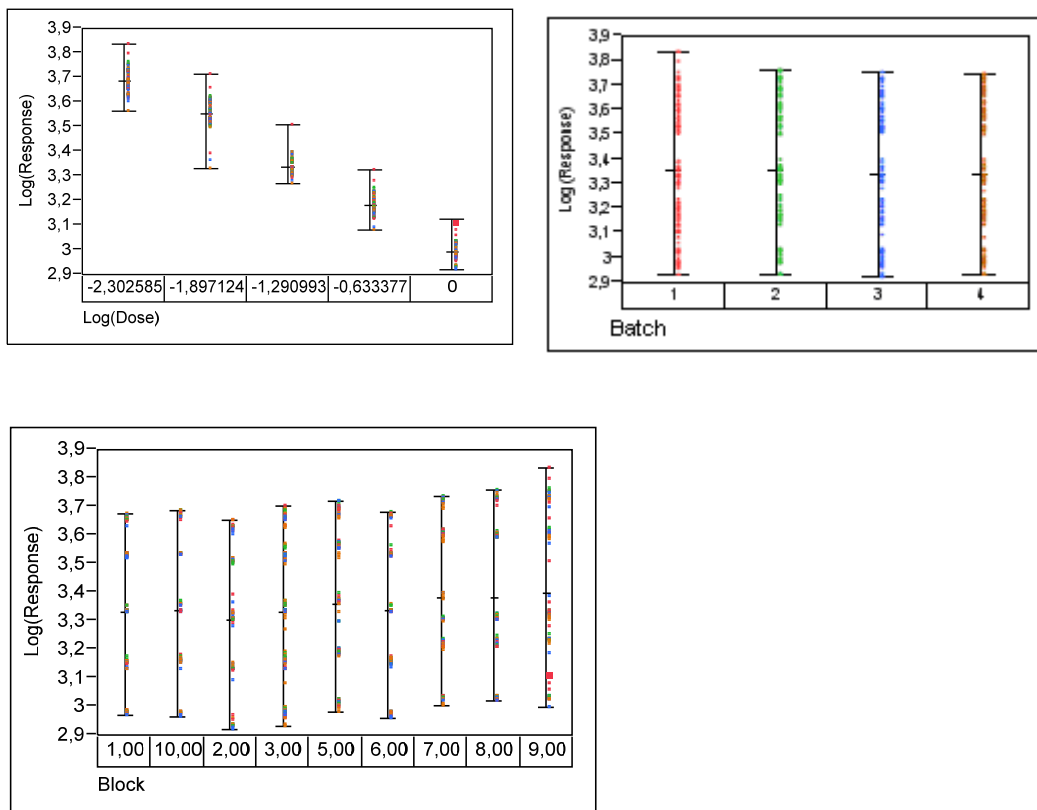


Figure 11 Log-response vs log-dose, batch and blocks. The variability seems to be constant across batch and block's levels. The variability of the response increases with log-dose indicating that the measuring and/or the sample preparation procedures are not as efficient for large dose levels as for lower levels.

Table 7 Results for the homogeneity of variance test for batch.

Test	Statistics	P-value
Levene	0.01	1.00
Bartlett	0.01	1.00

Table 8 Results for the homogeneity of variance test for block.

Test	Statistics	P-value
Levene	0.34	0.95
Bartlett	0.06	1.00

Table 9 Results for the homogeneity of variance test for log-dose.

Test	Statistics	P-value
Levene	11.65	0.00
Bartlett	11.93	0.00

The conclusions obtained from the plot analysis and the Levene and Bartlett's tests are similar, the variances for different blocks and batches can be assumed to be equal at the 0.05 significance level, but the variances for different doses are different. This

information is very valuable to evaluate the performance of the experimental method and plan possible improvements.

It is expected that the departure from the constant variance assumption should not affect the results considerably since ANOVA is a robust method. However in models with random effects, unequal error variances would disturb inference on the variance components [7].

4.3.5 ANOVA F-test and the normality assumption

The validity of the ANOVA F-test relies on the assumption that the random errors are normally and independently distributed random variables [7]. An alternative approach that does not rely on the normality assumption is to calculate the F randomization distribution [7]. Briefly, the data is randomly sorted between the different treatments. For example, if there is not difference between the blocks' means, all data allocation is equally likely. For each randomly selected data allocation, the F-values are calculated. These values are called the randomization distribution. Then, the actual F-value is compared to the randomization distribution.

The disadvantage of this approach is that for large amount of data is complicated to calculate the exact randomization distribution. However, this distribution is well approximated by the normal F distribution [7]. Therefore, even without the normality assumption, the ANOVA F-test can be seen as an approximation to the randomization test [7].

4.3.6 General considerations about the experimental design

There are several factors that must be considered when designing a bioassay. First of all the intrinsic variation between the biological units largely affect the results. Since the samples have to be manipulated before the measurements, factors like operator and sample preparation also increase the variability. Furthermore, additional “noise” is introduced by inhomogeneities in the drug's batches. Day-to-day variation also plays an important role.

In order to determine which factors affect the most this bioassay, several mixed models are fitted to the data. In all cases, it is observed that the larger contribution is the block-to-block variation, which reflects the day-to-day variation (the block and day effects are largely confounded in this design). It is also observed that the results obtained for larger log-dose show more variability, indicating that the experimental method does not perform as well for this log-dose range.

The objective of this study is to find an optimal set-up to obtain an estimate of the relative potency with a certain confidence interval. From the analysis presented in this chapter, it is concluded that the main variability factor are the blocks. In all cases the block-to-block variation accounted for almost 50 % of the total variability. Therefore, data simulations will be performed only considering a block-to-block variance of 0.001,

which is the rounded value corresponding to the block's variance component presented in Tab. 6. The results of the simulations are discussed in Chapter 5.

5 Data simulations for bioassays

5.1 Relative log-potency calculation of the experimental data

This chapter describes the calculations of the relative log-potency for four different batches. In the first case, we consider the drug of batch 1 as reference and the others as test. In the second case, we consider drug of batch 2 as reference and drug of batch 3 and 4 as test. Finally, drug of batch 3 is treated as a reference and drug of batch 4 as test. With this procedure six different log-potencies are calculated.

For each pair of samples, the following regression model is applied in each block [1]

$$\log(Y_i) = \delta_i * x_i + \kappa * \log(dose) \quad (5.1)$$

Where

x_i is the batch type ($i=1,2$).

δ_i is the intercept for each batch type.

κ is the slope of the line.

According to Ref. [1], the reportable value for the logarithm of relative potency ($\log(\rho)$) is the average of the $\log(\rho)$ values obtained for each block. A confidence interval is obtained by assuming that the distribution of the mean of $\log(\rho)$ is normal with mean and variance

$$\overline{\log(\rho)} = \frac{\sum_{b=1}^n \log(\rho)_i}{n}$$
$$\sigma_{\log(\rho)}^2 = \frac{\sum_{b=1, j=1}^{n, l} (\log(\rho)_{ij} - \overline{\log(\rho)_i})^2}{n}$$

where n is the total number of blocks and l is the total number of observations in each block.

As an example, the data corresponding to measurements in one block of batch 1 and 2 is shown in Fig. 12. This block contains 394 data points. Equation 5.1 is fitted to the data. The fit seems adequate and the residuals are approximately normally distributed.

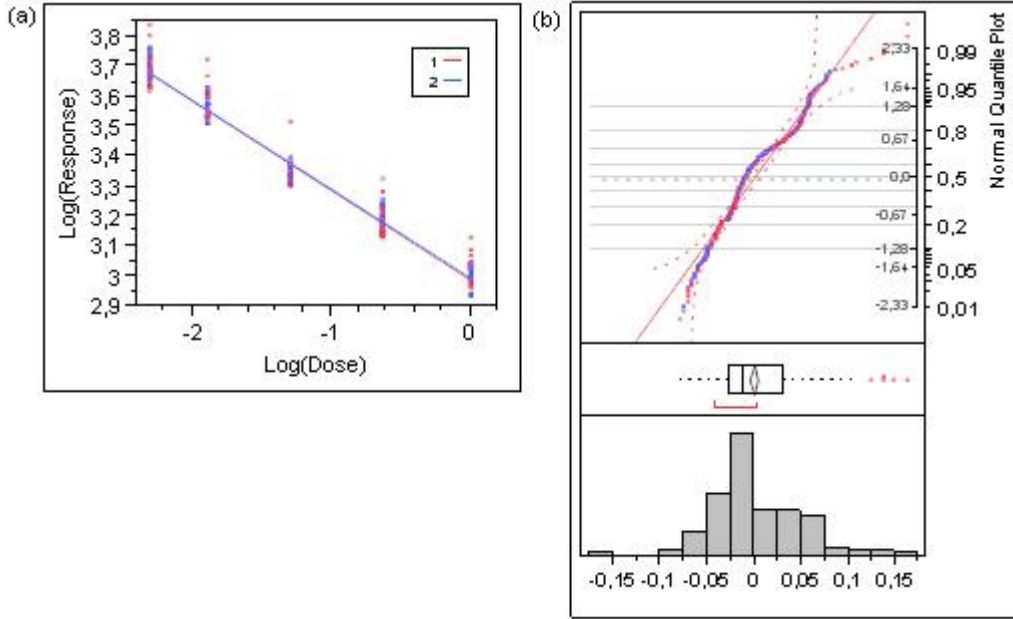


Figure 12 Linear model with residuals for batches 1 and 2 for block 1 (a) Data is plotted together with the line fit. (b) Histogram of the residuals.

5.1.1 Parallelism

For each pair of batches, a linear fit is applied to all blocks. The adequacy of the linear fit is used as a criterion to assess whether the parallelism approximation is fulfilled. No deviations from parallelism of practical importance are observed for the samples studied here.

5.1.2 Relative log-potency calculations

The relative potency is defined as the difference in which both drugs have the same response (see Fig. 2). Since the data is log-transformed the logarithm of the relative response is calculated.

The log-response curve is modelled using Eq. 5.1, thus the logarithm of the relative log-potency is calculated as

$$\log(\rho) = \log(\text{relative potency}) = \frac{(\delta_{TEST} - \delta_{REFERENCE})}{\kappa}$$

The distribution of $\log(\rho)$ corresponds to the distribution of the ratio of two correlated normally distributed variables [9]. This distribution has finite moments.

A value of $\log(\rho)$ is calculated for each block, and the final result is the average of $\log(\rho)$ values. The outcome of each experiment is an estimate of the mean value of $\log(\rho)$. The $\log(\rho)$ ratios are independent from each other, equally distributed and randomly selected. The central limit theorem [10] is applied to obtained ratios, thus the distribution of the mean of $\log(\rho)$ is normal with the same mean as the original distribution and a variance equal to the variance of the original distribution divided by the number of blocks.

5.2 Data simulation

In order to estimate how the number of blocks affects the obtained results, data was simulated in MATLAB. The simulated data reflects the block-to-block variation and the error made in the linear fit.

The value used to estimate the block-to-block variation is obtained from the mixed model analysis of the original data presented in Chapter 4 (see variance component corresponding to block listed in Tab 6). The uncertainty contribution of the linear fit in each block is estimated by fitting the original data and calculating the distribution of the error term for each fit.

The formula used to simulate data is

$$\text{Simulated } \log(\text{data}) = m + \kappa * \log(\text{dose}) + \text{Random Normal}_{\text{block}}(0, 0.001) + \text{Random Normal}_{\text{point}}(0, 0.009)$$

where *Random Normal* (a, b) is built-in function in MATLAB that randomly selects a normally distributed number with mean a and variance b . *Random Normal_{block}* takes into account the block-to-block variation, meaning that it is the same for all simulated points that belongs to the same block. *Random Normal_{point}* is the measurement error in each point and it is therefore different for all points. These values are taken from the fit of the original data. The constants k and m are the slope and the intercept.

5.2.1 Simulated relative log-potency

In order to calculate $\log(\rho)$, a linear fit for each pair of samples is performed for each block. In the original data, samples from four different batches are measured in ten blocks. The obtained $\log(\rho)$ are shown in Fig. 14.

The logarithm of the relative potency is calculated as the mean of the obtained $\log(\rho)$ divided by the number of blocks. Figure 15 shows how $\log(\rho)$ changes with the number of blocks. As expected, when several blocks are included a better estimate of $\log(\rho)$ is obtained.

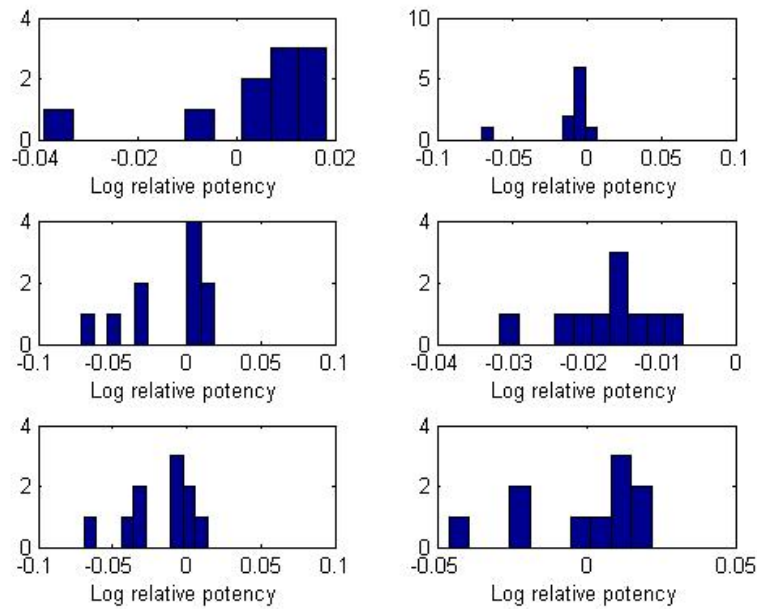


Figure 13 Histogram of the logarithm of the relative potency for the original data. Each subplot represents a different pair of batches.

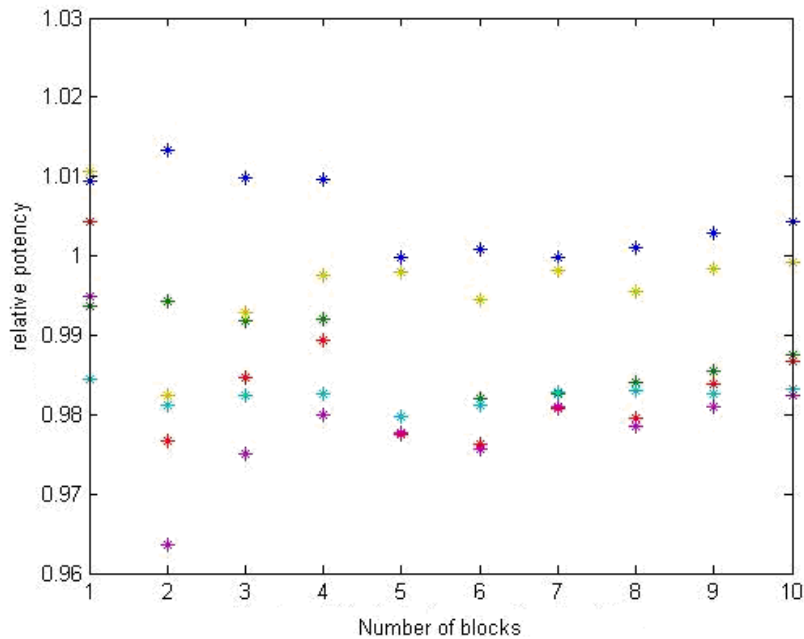


Figure 14 Relative potency (ρ) for the experimental data as a function of the number of blocks. The colors indicate ρ for different pairs of batches.

Simulations are performed to 10, 50, and 100 blocks. The results are shown in Figs. 16-19. The estimated value of ρ varies when a few blocks are included. However, for 50 blocks convergence is practically achieved.

The distribution of $\ln(\rho)$ becomes more normal as the number of blocks increases (see Figs 19-21).

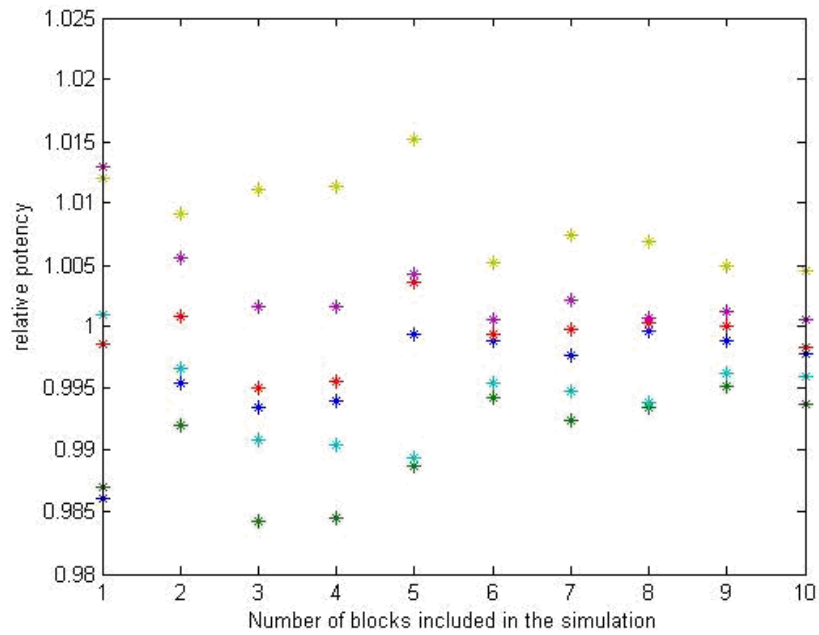


Figure 15 Relative potency for 10 simulated blocks as a function of the number of blocks. The colors indicate ρ for different pairs of batches.

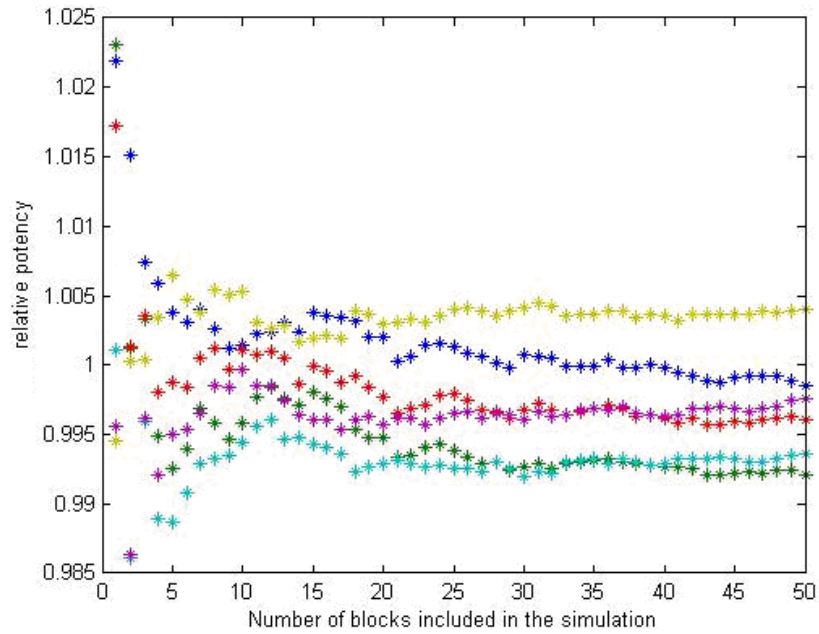


Figure 16: Relative potency for 50 simulated blocks as a function of the number of blocks. The colors indicate ρ for different pairs of batches.

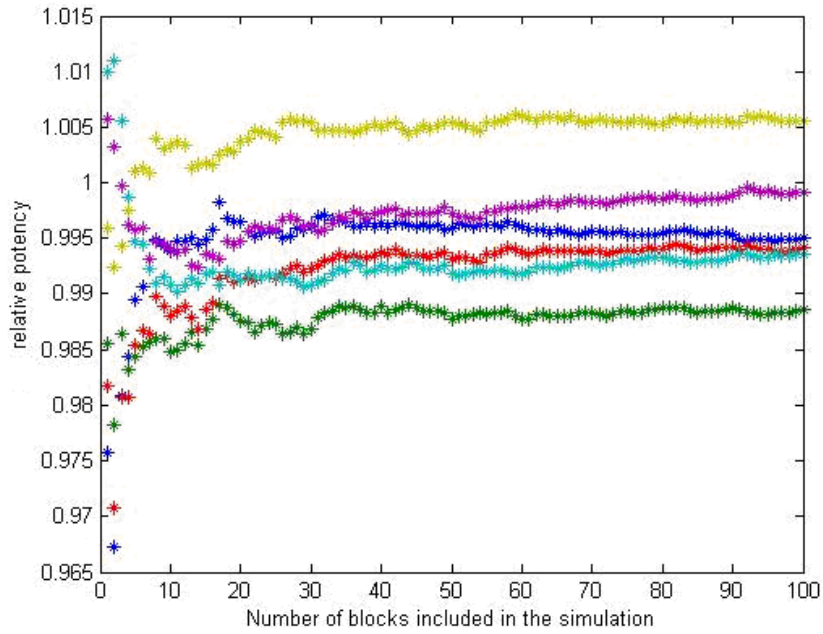


Figure 17 Relative potency (ρ) for 100 simulated blocks as a function of the number of blocks. The colors indicate ρ for different pairs of batches.

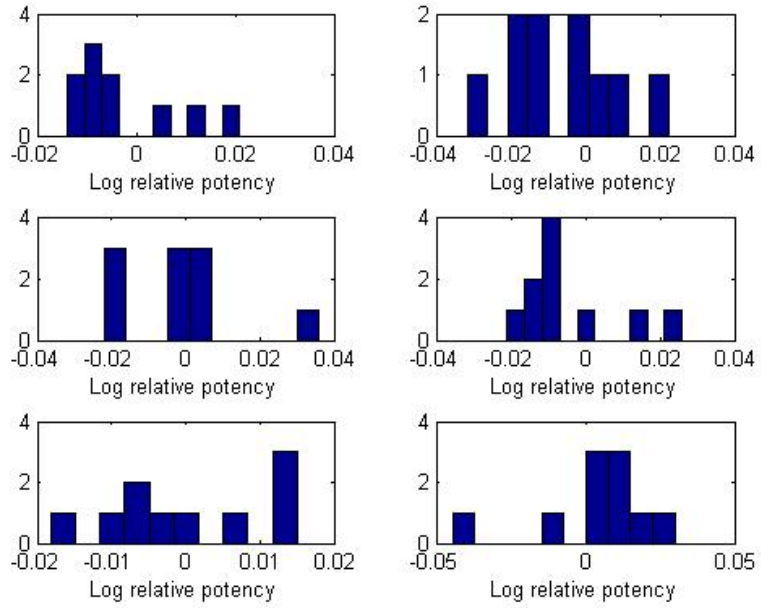


Figure 18 Histogram of the logarithm of the relative potency for the 10 blocks. Each subplot represents a different pair of batches.

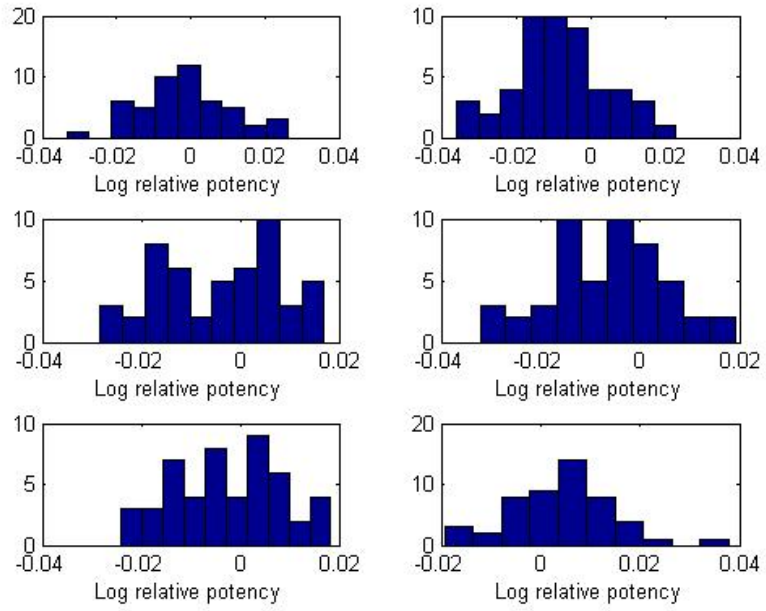


Figure 19 Histogram of the logarithm of the relative potency for the 50 blocks. Each subplot represents a different pair of batches.

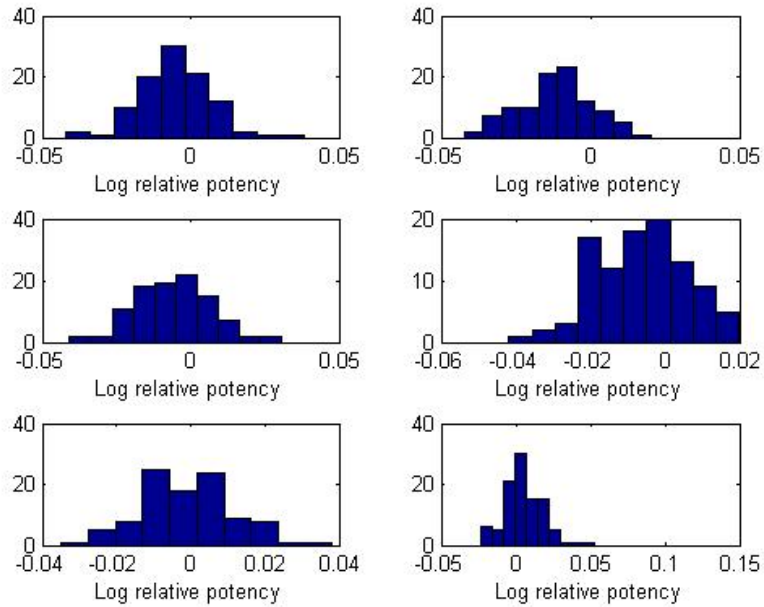


Figure 20 Histogram of the logarithm of the relative potency for the 100 blocks. Each subplot represents a different pair of batches.

5.2.2 Results

The results obtained for $\log(\rho)$ using the experimental and the simulated data are listed in Tab. 10. The relative 95% confidence intervals² in percent are also listed to facilitate the comparison between the different numbers of blocks. It is not possible to compare the results for $\log(\rho)$ between the simulations and the experimental data since the values of the slope and the intercept in the linear fit used in the simulations are arbitrary values. A good agreement for the variances is found between the results obtained from the experiment and from simulations.

The simulations can be used to analyse the applicability of the experimental set-up. For example, in order to compare batches 1 and 2, it is expected to measure 10 blocks to obtain a relative confidence interval of 99.24 - 100.76 %. However when comparing the batches 4 and 2, 50 blocks are needed to obtain a relative confidence interval of 99.74-100.26 %

² The relative confidence interval is calculated by dividing the confidence interval by the estimate.

Table 10 $\log(\rho)$ and ρ obtained from the simulations with their relative 95 % confidence intervals.

The results obtained with the experimental data are also listed. σ^2 is the variance of $\log(\rho)$.

Data type	Number of blocks	Batch	$\log(\rho)$	σ^2	ρ	95 % Relative lower CI for ρ	95 % Relative upper CI for ρ
Experimental	10	2 relative to 1		0.017		98.89	101.12
		3 relative to 1		0.021		98.45	101.57
		4 relative to 1		0.030		97.91	102.13
		3 relative to 2		0.007		99.51	100.49
		4 relative to 2		0.025		98.19	101.85
		4 relative to 3		0.023		98.02	102.02
Simulation	10	2 relative to 1	-0.002	0.012	0.99800	99.24	100.76
		3 relative to 1	-0.006	0.016	0.99402	98.90	101.12
		4 relative to 1	-0.002	0.016	0.99800	98.62	101.40
		3 relative to 2	-0.004	0.015	0.99601	98.97	101.04
		4 relative to 2	0.001	0.011	1.00100	98.64	101.38
		4 relative to 3	0.005	0.021	1.00501	98.40	101.63
	50	2 relative to 1	-0.001	0.012	0.99900	99.77	100.23
		3 relative to 1	-0.008	0.012	0.99203	99.64	100.36
		4 relative to 1	-0.004	0.012	0.99601	99.65	100.35
		3 relative to 2	-0.006	0.011	0.99402	99.70	100.30
		4 relative to 2	-0.002	0.011	0.99800	99.74	100.26
		4 relative to 3	0.004	0.010	1.00401	99.71	100.30
	100	2 relative to 1	-0.005	0.012	0.99501	99.76	100.24
		3 relative to 1	-0.011	0.012	0.98906	99.76	100.24
		4 relative to 1	-0.006	0.013	0.99402	99.73	100.27
		3 relative to 2	-0.006	0.012	0.99402	99.77	100.23
		4 relative to 2	-0.001	0.013	0.99900	99.72	100.28
		4 relative to 3	0.006	0.013	1.00602	99.72	100.28

6 Conclusions

In this thesis, a description of parallel bioassays is presented. Additionally, the main statistical tools used to analyse them, in particular mixed models are discussed. These concepts are applied to an example taken from a pharmaceutical company.

In the example presented here, the main sources of experimental variation are:

- Operator
- Block (day)
- Sample preparation process
- Batches

The data is analysed within the ANOVA framework. Since the conclusions of the experiment, meaning the effect of the batches, have to be valid for all blocks, operators and sample preparations, these effects are considered as random.

By reducing the model, it is concluded that approximately 50 % of the variation in the data is due to the day to day variation. The variance component associated with this is estimated to be 0.001.

In order to estimate how many experiments are needed to obtain a confidence interval for the batches' potency (biological effect), simulations are performed. The simulated data should have the same noise as the experimental data. Therefore, the simulated data has a normally distributed random component with a variance equal to the block variance. Additionally, the simulated data has also a normally distributed random component that takes into account the uncertainty in the measurement at each data point. Simulations are performed for several blocks. For each case, the potency and its relative confidence interval are calculated. The results from the simulations agree well with those obtained in the experiments.

The simulations show that different numbers of replicates are needed to obtain the same width in the confidence interval for the relative potency, depending on which type of batches are compared. For applications within the pharmaceutical industry, it is usually desired to report an estimate for potency with a relative confidence interval of 99 – 101 %. The simulation shows that when 10 blocks are used, only one comparison of batches has a confidence interval with this width. The simulations also indicate that when 50 measurements are performed, all batches are within this confidence interval.

The method presented in this work can be used to save time and experimental resources, since only the exact amount of needed data is collected. Furthermore, the analysis of variance provides valuable information about the performance of the experimental set-up. This information can be further used to improve the set-up by for example, response surface analysis.

The approach presented here can also be applied to other experiments. In industrial applications, historical data is usually available. This data can be used to estimate the size of the main sources of variation by an ANOVA analysis. The obtained variances are then used to simulate data that is analysed using the same procedures as the experimental data. The result of this analysis can be used to determine the sample size, and the accuracy of the simulation can be checked by comparison with the experimental data.

7 References

1. D. Finney, *Statistical method in biological assay*, Charles Griffin & Company LTD, 1978.
2. USP General Chapter, *Design and Analysis of Biological Assays <111>*, Pharm. Forum. 34 (3), 685, 2008.
3. D.M. Rocke, *Design and analysis of experiments with high throughput biological assay data*, Seminars in Cell & Developmental Biology 15, 703, 2004.
4. F. Tarding, P. Nielsen, B Keiser-Nielsen, and Aa. V. Nielsen, *Biological Assay of Glucagon in Rabbits*, Diabetologia 5, 146, 1969.
5. M. Pingel, Aa. Vølund, E. Sørensen, J.E. Collins and C.T. Dieter, *Biological potency of porcine, bovine and human insulins in the rabbit bioassay system*, Diabetologia 28, 862, 1985.
6. R. Buridik, C.M. Borrow, and D.C. Montgomery, *A Review of Methods for Measurement Systems Capability Analysis*, J. Quality Tech. 35, 342, 2003.
7. D.C. Montgomery, *Design and Analysis of Experiments*, John Willey & Sons INC., 2005.
8. NC Cary, *JMP® 8 Statistics and Graphics Guide*, SAS Institute Inc., 2009.
9. D.V. Hinckley, *Inference about the interception in two-phase regression*, Biometrika 56, 495, 1969.
10. R. Hogg and E.Tanis, *Probability and statistical inference*, Pearson Education International, 2006.

8 Appendix

8.1 *MATLAB code to calculate log-relative potency*

8.1.1 Function to used to simulate data

```
function [data] = simdata(nbrseries, repetition)

logdose=[0.00 -0.64 -1.29 -1.90 -2.30];

data=zeros(nbrseries*4*repetition*length(logdose), 9);

k=-0.34;
%m=linspace(1,1.03,4);
m=[3.318 3.316 3.310 3.314];

row=1;
for series=1:nbrseries
    randseries=0.03*randn(1);
    for sample=1:4
        for rep=1:repetition
            for dose=1:length(logdose)
                data(row,1)=series;
                data(row,2)=sample;
                data(row,8)=logdose(dose);

data(row,9)=k*logdose(dose)+m(sample)+0.03*randn(1)+randseries;
                row=row+1;
            end
        end
    end
end
end
```

8.1.2 Program to calculate the logarithm of the relative potency between two substances

```
%Program to calculate the relative potency between 2 drugs;
%
%Load the Excel file is data stored elsewhere
clear all
close all

f=0;%Index used to count combinations of A and B levels of X1;
s=0; %Index used to count number of series used in each calculations;

%Define total number of repetition (series);
total_number_series=100;
total_number_X1_levels=4;
```

```

total_number_repetitions= 4;

data=simdata(total_number_series,total_number_repetitions);

%Create vector for betas and ratio;
betas=zeros(3,total_number_series);
Ratio_B_A=zeros(1,total_number_series);
Final_ratio=zeros(1,total_number_series);
Total_Ratio= zeros(6,total_number_series); %6 is the number of
independent combinations of levels X1;

order=randperm(total_number_series);

for A=1:1:4; %For each level of X1 combination:
    for B=(A+1):1:4;

        f=f+1;
        index_A_B=((data(:,2)==A) | (data(:,2)==B));
        data_A_B=data(index_A_B,:);

        %Code X2 to -1 and 1;
        X2=data_A_B(:,8);
        X2_max=max(data_A_B(:,8));
        X2_min=min(data_A_B(:,8));
        b_X2=2/(X2_max-X2_min);
        a_X2=(-b_X2)*(X2_max+ X2_min)/2;
        coded_X2= a_X2+ (b_X2)*X2;

        %Code X1 to -1 and 1;
        X1 = data_A_B(:,2);
        coded_X1 = X1;
        coded_X1(coded_X1==A)=1;
        coded_X1(coded_X1==B)= -1;

        %New matrix with coded variables;
        coded_data= data_A_B;
        coded_data(:,2)=coded_X1;
        coded_data(:,8)=coded_X2;
        for number_series=1:1:total_number_series;
            %Do fitting for each series separately;

            for i=1:1:number_series

                index=(coded_data(:,1)==order(i));
                new_data=coded_data(index,:);
                X1=new_data(:,2);
                X2=new_data(:,8);
                Y=new_data(:,9);
                b=length(X2);
                X0=ones(b,1);
                X=[X0 X1 X2];
                betas=inv(X'*X)*X'*Y;
                y_approx=X*betas;
                Error=Y-y_approx;
                %hist(Error)
                %pause(0.2)
                %Sum of squares;
            end
        end
    end
end

```

```

        SSE= Error'*Error; %Sum squares residuals;
        SSR=betas'*X'*Y-(((sum(Y))^2)/length(Y));%Sum
of squares regression;
        MSR= SSR/3; %total number of regression
coefficients length Beta;
        MSE=SSE/(length(Y)-3-1);
        F_null=MSR/MSE;%To be used in test for beta;
        Ratio_B_A(i)= betas(2)/betas(3);%X2-X1;
    end;

    %Take mean value for ratio for all series included in
the calculations;

Final_ratio(1,number_series)=sum(Ratio_B_A)/number_series;
    SingleSerie_ratio(number_series)=Ratio_B_A(i);

    end;

    Total_Ratio(f,:)= Final_ratio;
    Total_SingleRatio(f,:) = SingleSerie_ratio;
    %Clear variables;
    Ratio_B_A=zeros(1,total_number_series);
    Final_ratio=zeros(1,total_number_series);

    end;
end;

%Calculate relative log relative potency
Rel_log_pot_1_2= Total_Ratio(1,:);
Rel_log_pot_1_3= Total_Ratio(2,:);
Rel_log_pot_1_4= Total_Ratio(3,:);
Rel_log_pot_2_3= Total_Ratio(4,:);
Rel_log_pot_2_4= Total_Ratio(5,:);
Rel_log_pot_3_4= Total_Ratio(6,:);

%Calculate relative log relative potency, singles series
RelSingle_log_pot_1_2= Total_SingleRatio(1,:);
RelSingle_log_pot_1_3= Total_SingleRatio(2,:);
RelSingle_log_pot_1_4= Total_SingleRatio(3,:);
RelSingle_log_pot_2_3= Total_SingleRatio(4,:);
RelSingle_log_pot_2_4= Total_SingleRatio(5,:);
RelSingle_log_pot_3_4= Total_SingleRatio(6,:);

%Plot histogram

figure
subplot(3,2,1)
hist(RelSingle_log_pot_1_2)
xlabel('Log relative potency');

subplot(3,2,2)
hist(RelSingle_log_pot_1_3)
xlabel('Log relative potency');

subplot(3,2,3)
hist(RelSingle_log_pot_1_4)
xlabel('Log relative potency');

```



```

subplot(3,2,4)
hist(RelSingle_log_pot_2_3)
xlabel('Log relative potency');

subplot(3,2,5)
hist(RelSingle_log_pot_2_4)
xlabel('Log relative potency');

subplot(3,2,6)
hist(RelSingle_log_pot_3_4)
xlabel('Log relative potency');

[rel_pot_1_2,var_1_2, CI_rel_pot_1_2,CI_var_1_2]=
normfit(RelSingle_log_pot_1_2);
[rel_pot_1_3,var_1_3, CI_rel_pot_1_3,CI_var_1_3]=
normfit(RelSingle_log_pot_1_3);
[rel_pot_1_4,var_1_4, CI_rel_pot_1_4,CI_var_1_4]=
normfit(RelSingle_log_pot_1_4);
[rel_pot_2_3,var_2_3, CI_rel_pot_2_3,CI_var_2_3]=
normfit(RelSingle_log_pot_2_3);
[rel_pot_2_4,var_2_4, CI_rel_pot_2_4,CI_var_2_4]=
normfit(RelSingle_log_pot_2_4);
[rel_pot_3_4,var_3_4, CI_rel_pot_3_4,CI_var_3_4]=
normfit(RelSingle_log_pot_3_4);

Mean_log_rel_pot_all_samples= [rel_pot_1_2 rel_pot_1_3 rel_pot_1_4
rel_pot_2_3 rel_pot_2_4 rel_pot_3_4 ]

Var_log_rel_pot_all_samples= [var_1_2 var_1_3 var_1_4 var_2_3 var_2_4
var_3_4 ]

Pct_CI_rel_log_pot_1_2= 100.*CI_rel_pot_1_2./rel_pot_1_2;
Pct_CI_rel_log_pot_1_3= 100.*CI_rel_pot_1_3./rel_pot_1_3;
Pct_CI_rel_log_pot_1_4= 100.*CI_rel_pot_1_4./rel_pot_1_4;
Pct_CI_rel_log_pot_2_3= 100.*CI_rel_pot_2_3./rel_pot_2_3;
Pct_CI_rel_log_pot_2_4= 100.*CI_rel_pot_2_4./rel_pot_2_4;
Pct_CI_rel_log_pot_3_4= 100.*CI_rel_pot_3_4./rel_pot_3_4;

Pct_CI_log_rel_pot_all_samples=[ Pct_CI_rel_log_pot_1_2
Pct_CI_rel_log_pot_1_3 Pct_CI_rel_log_pot_1_4 Pct_CI_rel_log_pot_2_3
Pct_CI_rel_log_pot_2_4 Pct_CI_rel_log_pot_3_4]

Result=exp(Total_Ratio);

figure
plot(Result','*')
xlabel('Number of blocks included in the simulation');
ylabel('Relative potency');

sprintf('%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n',Mean_log_rel_pot_all_samples)

sprintf('%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n%3.3f\n',Var_log_rel_pot_all_samples)

```

