Spotfire Statistica®

Stability Analysis Formula Guide

Version 14.3.0 | February 2025

Contents

Stability Analysis Overview

Stability analysis is the study of how drug product potency degrades over time.

The primary statistical quantity of interest is the expiration date or shelf life. Typically, a drug product is manufactured in batches. When estimating the shelf life of a medication, it is necessary to evaluate how the batches differ with respect to the potency degradation of the drug product over time. Specifically, three different model types are considered.

Model 1: Common Regression across All Batches

This model is appropriate when the intercepts and slopes do not differ across batches and the data, therefore, are pooled in order to estimate a common intercept and common slope.

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where,

Y_i = response at i^{th} time point

β_0 = common intercept

β_1 = common slope

$x_i = i^{th}$ time point

$\epsilon_i \sim N(0, \sigma^2)$

$i = 1, \ldots, n$

n = total number of observations

Model 2: Separate Intercepts and Common Slope

This model is appropriate when the intercepts differ across batches, but the slopes are common across batches.

$$Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \epsilon_{ij}$$

where,

Y_{ij} = response at i^{th} time point and j^{th} batch

β_{0j} = intercept of j^{th} batch

β_1 = common slope across batches

x_{ij} = i^{th} time point for j^{th} batch

$\epsilon_{ij} \sim N(0, \sigma^2)$

$i = 1, \ldots, n_j$

n_j = number of time points for j^{th} batch

$j = 1, \ldots, k$

k = number of batches

Model 3: Separate Intercepts and Separate Slopes

This model is appropriate when the intercepts and slopes differ across batches. Additionally, the Mean Square Error (MSE) is not pooled across different batches, that is, homogeneity of variance across batches is not assumed.

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}$$

where,

Y_{ij} = response at i^{th} time point and j^{th} batch

β_{0j} = intercept of j^{th} batch

β_{1j} = slope of j^{th} batch

x_{ij} = i^{th} time point for j^{th} batch

$\epsilon_{ij} \sim N(0, \sigma_j^2)$

$i = 1, \ldots, n_j$

n_j = number of time points for j^{th} batch

$j = 1, \ldots, k$

k = number of batches

Parameter Estimation

Parameters are estimated by ordinary least squares. This can be accomplished by solving the normal equations:

$$\boldsymbol{X}^T \boldsymbol{X} \hat{\beta} = \boldsymbol{X}^T Y$$

where,

$\boldsymbol{X}$ = N by p design matrix of coded predictor variables
Y = N by 1 vector of observed values
N = Number of observations
p = Number of columns in design matrix

If the columns of the design matrix are linearly independent then the vector of parameter estimates is as follows:

$$\hat{\beta} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T Y$$

t-Statistic

The t-statistic associated with the null hypothesis that the ith parameter is equal to 0 is given by:

$$t = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}$$

where,

$\hat{\sigma}_{\hat{\beta}_i}$ = estimated standard error of $\hat{\beta}_i$
square root of ith diagonal element of MSE $(\boldsymbol{X}^T \boldsymbol{X})^{-1}$

Confidence Interval for a Single Parameter

A two-sided 95% confidence interval for the i^{th} parameter is given by:

$$\hat{\beta}_i \pm t_{0.975,\mathrm{df}_e} \hat{\sigma}_{\hat{\beta}_i}$$

where,

$t_{0.975,\mathrm{df}_e} = 97.5^{th}$ quantile of a t distribution with df_e degrees of freedom

Predicted Values

The predicted value or estimated mean response for a given covariate vector X_i is given by:

$$\hat{Y}_i = X_i^T \hat{\beta}$$

where,

$X_i = [1\ X_1\ \cdots\ X_{p-1}]^T$

Confidence Interval

The two-sided 95% confidence interval for the mean response for a given covariate vector X_i is given by:

$$\hat{Y}_i \pm t_{0.975,\mathrm{df}_e} \sqrt{MSE\left(X_i^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} X_i\right)}$$

where,

$\mathrm{df}_e =$ error degrees of freedom

$X_i = [1\ X_1\ \cdots\ X_{p-1}]^T$

$t_{0.975,\mathrm{df}_e} = 97.5^{th}$ quantile of t distribution with $\mathrm{df} = \mathrm{df}_e$

For a one-sided interval, then the 95^{th} quantile of *t* distribution is used.

Prediction Interval

The two-sided 95% prediction interval for a new observation X_i is given by:

$$\hat{Y}_i \pm t_{0.975,\mathrm{df}_e} \sqrt{1 + MSE\left(X_i^T \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} X_i\right)}$$

where,

$\mathrm{df}_e =$ error degrees of freedom

$X_i = \left[1 \; X_1 \; \cdots \; X_{p-1}\right]^T$

$t_{0.975,\mathrm{df}_e} = 97.5^{th}$ quantile of t distribution with $\mathrm{df} = \mathrm{df}_e$

For a one-sided interval, then the 95th quantile of *t* distribution is used.

Model Selection

Spotfire Statistica® selects the model by first evaluating if the slopes are equal across different batches.

If the slopes' p-value is significant (by default, less than 0.25), the separate intercepts, separate slopes model is used to estimate shelf life.

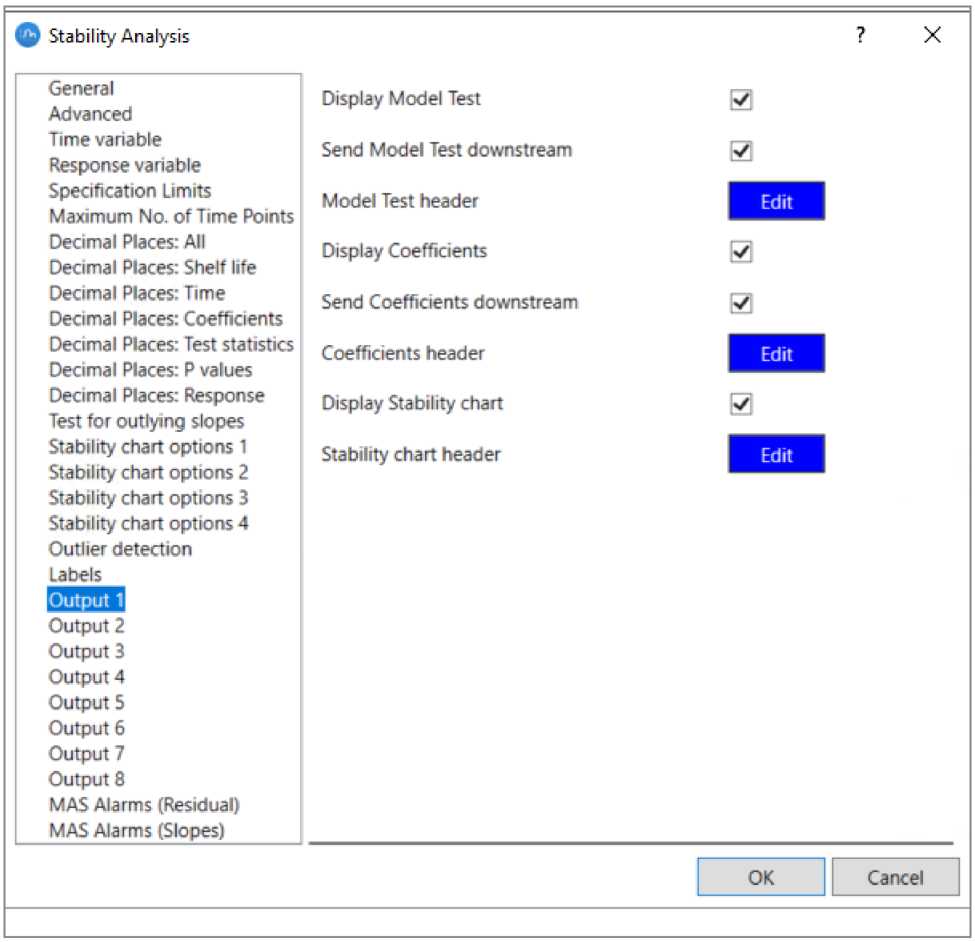
If the slopes' p-value is not significant, Statistica determines if the intercepts are equal across different batches. If the intercepts' p-value is significant (by default, less than 0.25), the separate intercepts, common slope model is used. If neither the slopes' p-value nor the intercepts' p-value are significant (by default, less than 0.25) the common intercept, common slope model is used.

The p-values are computed by using the type 1 sums of squares decomposition of the full linear model with effects placed into the model in the following order: time, batch, batch by time interaction. The slopes' p-value is associated with the batch by time interaction and the intercepts' p-value is associated with the batch effect.

Output Analysis

This example illustrates the steps followed in the Stability Analysis to select an appropriate model.

Display Model Test



The following image displays the output:

A: sep. intercep, sep slope vs. com intercep, com slope
B: sep. intercep, com slope vs. com intercep, com slope
C: sep. intercep, sep slope vs. sep intercep, com slope
D: Residual
E: Full Model

Poolability of intercepts alpha: 0.25
Poolability of slopes alpha: 0.25
Subset of ModelTest
Variables: *

	SS	DF	MS	F	P
A	2.57	12	0.21	2.74	0.01
B	2.12	6	0.35	4.52	0.00
C	0.45	6	0.07	0.95	0.47
D	2.58	33	0.08		
E	174.96	14	12.50		

The following table provides a description for each column.

Column	Full Form
SS	Sum of squares
MS	Mean of squares
DF	Degrees of Freedom
F	Variance Ratio
P	P value or significance ratio

The above results were created using representative data as input to the Stability Analysis node. The following sections analyze each line of the output and describe the process of selecting a model for producing the Stability graph and a shelf life estimate. The following sections describe the lines used by the Stability Analysis in its selection of the appropriate model.

Row C: Tests sep. intercep, sep slope vs. sep intercep, com slope

The row C test is to see if the user can assume that the batch does not have significant impact on the slope of the model. The null hypothesis is that these batches may be pooled together with a common slope unless evidence is present to reject this hypothesis.

Effect	Univariate Tests of Significance for Numeric Parameter Value (Subset) Over-parameterized model Type I decomposition; Std. Error of Estimate: 0.2797 SS	Degr. of Freedom	MS	F	p
Intercept	172.3404	1	172.3404	2202.963	0.000000
Batch Number	1.8089	6	0.3015	3.854	0.005058
Protocol Months On Stability	0.3630	1	0.3630	4.640	0.038641
Batch Number*Protocol Months On Stability	0.4461	6	0.0743	0.950	0.473259
Error	2.5816	33	0.0782		

If Batch and Months are included in the model as additive terms, then the intercept value differs depending on the batch. However, a common slope still exists.

The Batch * Time interaction term indicates the additional variance explained by having separate slopes instead of common slopes. The p-value of this term (.473259) is greater than the chosen value of .25. Hence, this hypothesis is not rejected.

In the event that the p-value is significant (below the chosen value of .25), you cannot pool the batches together as the slopes are different. In this case, select the Sep. Intercept, Sep Slope model.

The Error row also matches with Row D in the Model Test spreadsheet. Row D is not used in the selection of the model. It represents the unexplained variance that still exists when Time, Batch, and Batch*Time interaction are all in the model.

Row B: Tests sep. intercep, com slope vs. com intercep, com slope

The slopes can be considered equal for all of the batches. You must verify whether the intercepts can be considered equal as well. The null hypothesis is that the lines you draw for each of the batches cross the Y axis at a similar point that they are basically the same, unless you have evidence to reject this hypothesis.

Effect	Univariate Tests of Significance for Numeric Parameter Value (Subset) Over-parameterized model Type I decomposition; Std. Error of Estimate: 0.2786 SS	Degr. of Freedom	MS	F	p
Intercept	172.3404	1	172.3404	2219.924	0.000000
Protocol Months On Stability	0.0490	1	0.0490	0.631	0.431650
Batch Number	2.1229	6	0.3538	4.557	0.001361
Error	3.0277	39	0.0776		

You no longer need to include the interaction term in this model. Hence, **Time** and **Batch** are included.

If you only have **Time** in the model, you have a single regression line. All of the points contribute to the one line, regardless of the batch. There is one common slope and one common intercept.

Batch Term

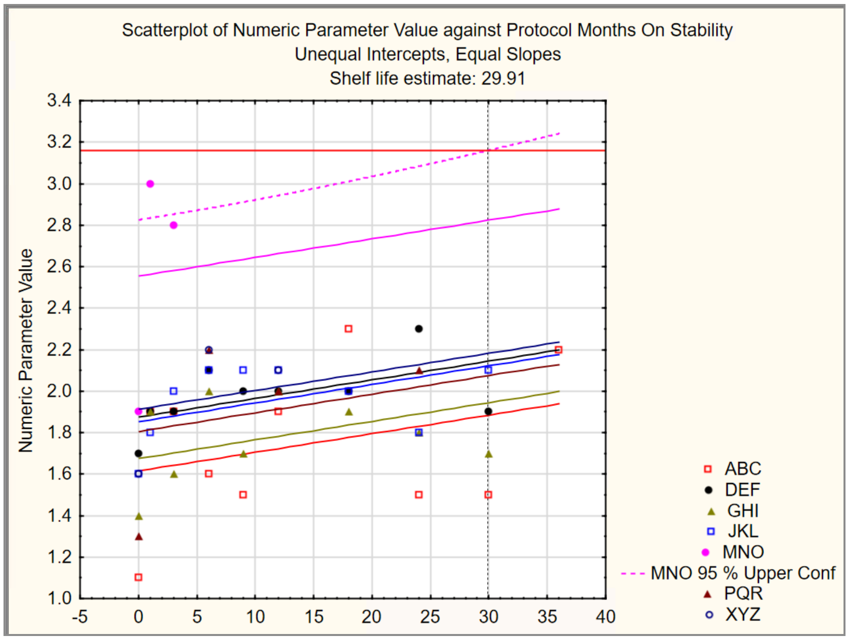
Batch term is the term that determines whether additional variance is explained by having intercepts rather than a common intercept.

The Batch term has a p-value of 0.001361 which is less than the chosen value of .25. Since the Batch term is significant, you must keep the lines separate. They are too different to combine. Select the Sep. Intercept, Common Slope model.

In a case where the p-value is not significant (below the chosen value of 0.25), you can pool all of the batches together into a single regression line. Select the Com. Intercept, Com. Slope model.

Stability Graph

A Stability graph with the chosen model is shown later in this section. The regression lines are drawn for each batch. The lines are parallel to one another (that is, have the same slope), but intersect the zero time line at different points. The red horizontal line represents the upper acceptable limit for this parameter value. The shelf life is determined by default by the time value where the first 95% confidence limit crosses the defined upper limit. In the following graph, the MNO batch line's confidence limit (pink dotted line) determines the shelf life of 29.91 months.



Row A: sep. intercep, sep slope vs. com intercep, com slope

Row A is not used in the selection of the Stability model.

It summarizes the variance indicated by both the batch and interaction term together.

Effect	Univariate Tests of Significance for Numeric Parameter Value (Subset) Over-parameterized model Type I decomposition; Std. Error of Estimate: 0.2797 SS	Degr. of Freedom	MS	F	p
Intercept	172.3404	1	172.3404	2202.963	0.000000
Protocol Months On Stability	0.0490	1	0.0490	0.627	0.434266
Batch Number	2.1229	6	0.3538	4.523	0.001896
Batch Number*Protocol Months On Stability	0.4461	6	0.0743	0.950	0.473259
Error	2.5816	33	0.0782		

Row A is a sum of these two effects with the MS, F, and P values based on these sums. These effects together represent the use of separate intercepts and separate slopes.

A: sep. intercep, sep slope vs. com intercep, com slope
B: sep. intercep, com slope vs. com intercep, com slope
C: sep. intercep, sep slope vs. sep intercep, com slope
D: Residual
E: Full Model

Poolability of intercepts alpha: 0.25
Poolability of slopes alpha: 0.25

	SS	DF	MS	F	P
A	2.57	12	0.21	2.74	0.01
B	2.12	6	0.35	4.52	0.00
C	0.45	6	0.07	0.95	0.47
D	2.58	33	0.08		
E	174.96	14	12.50		

If these two effects (assuming them together as a unit) are not in the model, only Time/Months is a predictor. Row A represents the amount of variance explained by including both Batch and Batch*Time effects in the model instead of having Time only as a predictor.

Row E: Full Model

Row E is not used to select a Stability model.

Univariate Tests of Significance for Numeric Parameter Value (Subset)
Over-parameterized model
Type I decomposition; Std. Error of Estimate: 0.2797

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	172.3404	1	172.3404	2202.963	0.000000
Protocol Months On Stability	0.0490	1	0.0490	0.627	0.434266
Batch Number	2.1229	6	0.3538	4.523	0.001896
Batch Number*Protocol Months On Stability	0.4461	6	0.0743	0.950	0.473259
Error	2.5816	33	0.0782		

Data: Spreadsheet211* (3v by 1c)

Univariate Tests of Significance for Numeric Parameter Value (Subset)
Over-parameterized model
Type I decomposition; Std. Error of Estimate: 0.2797

Effect	SS	Degr. of Freedom	**MS =SS/"Degr. of"**
SUM case 1-4	174.95837	14	12.4970265

This row contains all of the **SS Effect** values summed up, except **Error SS**. You must add the **Degrees of Freedom** for the same rows. MS for the full model is (SS Effect Sum)/(Sum Degrees Freedom).

Shelf Life Estimation

Shelf life estimation is performed by determining where the 95% lower and upper confidence or prediction interval intersects with the user-defined specification limits. If either Model 2 or Model 3 has been selected, shelf life is estimated separately for each batch. Shelf life is then defined as the shelf life associated with the worst-case batch, that is, the smallest shelf life estimate. By default, Statistica® bases shelf life computations on a two-sided 95% confidence interval.

Residual

A user can optionally perform a residual analysis to determine the accuracy of the selected model.

A residual is defined as the difference between the observed and predicted value, that is:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij}$$

where,

$Y_{ij} =$ observed value at i^{th} time point and j^{th} batch

$\hat{Y}_{ij} =$ predicted value at i^{th} time point and j^{th} batch

Additionally, studentized residuals are computed. The formula for the studentized residuals is given by:

$$s_{ij} = \frac{e_{ij}}{\sqrt{MSE(1 - h_{ii})}}$$

where,

$e_{ij} =$ residual for the i^{th} time point and j^{th} batch

$MSE =$ mean squared error associated with selected model

$h_{ii} = i^{th}$ diagonal element of hat matrix

hat matrix $= \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T$

$\boldsymbol{X} =$ design matrix of selected model

A normal probability plot of the (studentized) residuals is constructed as follows. First, the (studentized) residuals are rank ordered. From these ranks, Statistica® computes z values (that is, standardized values of the normal distribution) based on the assumption that the data come from a normal distribution.

$$z_i = \phi^{-1}\left[\frac{3i-1}{3N+1}\right]$$

where,

ϕ^{-1} is the inverse normal cumulative distribution function

N = Number of observations

$i = 1, \ldots, N$

These z values are plotted on the y-axis in the plot. If the observed (studentized) residuals (plotted on the x-axis) are normally distributed, all values should fall onto a straight line. If the residuals are not normally distributed, they deviate from the line. Outliers can also become evident in this plot. If there is a general lack of fit and the data seem to form a clear pattern (example, an S shape) around the line, the variable has to be transformed in some way (example, a log transformation to "pull-in" the tail of the distribution, etc.)

Spotfire Documentation and Support Services

For information about this product, you can read the documentation, contact Spotfire Support, and join Spotfire Community.

How to Access Spotfire Documentation

Documentation for Spotfire products is available on the Product Documentation website, mainly in HTML and PDF formats.

The Product Documentation website is updated frequently and is more current than any other documentation included with the product.

Product-Specific Documentation

The documentation for this product is available on Spotfire Statistica® Product Documentation page.

How to Contact Support for Spotfire Products

You can contact the Support team in the following ways:

- To access the Support Knowledge Base and getting personalized content about products you are interested in, visit our product Support website.
- To create a Support case, you must have a valid maintenance or support contract with a Cloud Software Group entity. You also need a username and password to log in to the product Support website. If you do not have a username, you can request one by clicking **Register** on the website.

How to Join Spotfire Community

Spotfire Community is the official channel for Spotfire customers, partners, and employee subject matter experts to share and access their collective experience. Spotfire Community offers access to Q&A forums, product wikis, and best practices. It also offers access to extensions, adapters, solution accelerators, and tools that extend and enable customers to gain full value from Spotfire products. In addition, users can submit and vote on feature requests from within the Spotfire Ideas Portal. For a free registration, go to Spotfire Community.

Legal and Third-Party Notices

Cloud SG software may be available on multiple operating systems. However, not all operating system platforms for a specific software version are released at the same time. See the “readme” file for the availability of a specific version of Cloud SG software on a specific operating system platform.