

TIBCO Statistica®

Stability Analysis Formula Guide

Version 14.0

December 2020



Contents

- TIBCO Documentation and Support Services 3**
- Stability Analysis Overview5**
 - Model 1: Common Regression across All Batches 5
 - Model 2: Separate Intercepts and Common Slope 5
 - Model 3: Separate Intercepts and Separate Slopes 6
- Parameter Estimation 7**
 - t-Statistic7
 - Confidence Interval for a Single Parameter7
 - Predicted Values7
 - Confidence Interval 8
 - Prediction Interval8
- Model Selection 9**
- Shelf Life Estimation10**
- Residual Analysis 11**
- Legal and Third-Party Notices 12**

TIBCO Documentation and Support Services

How to Access TIBCO Documentation

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

The TIBCO Product Documentation website is updated frequently and is more current than any other documentation included with the product. To access the latest documentation, visit <https://docs.tibco.com>.

Product-Specific Documentation

Documentation for TIBCO Statistica® is available on the [TIBCO Statistica® Product Documentation](#) page.

The following documents for this product can be found on the TIBCO Documentation site:

- *TIBCO Statistica® Release Notes*
- *TIBCO Statistica® Installation*
- *TIBCO Statistica® Quick Reference*
- *TIBCO Statistica® Product Traceability*
- *TIBCO Statistica® Configuration for Windows Server 2019*
- *TIBCO Statistica® Data Entry Administration*
- *TIBCO Statistica® Server Administrator's Guide*
- *TIBCO Statistica® Options Configuration*
- *TIBCO Statistica® R Integration: Features and Options*
- *TIBCO Statistica® Security Guide*
- *TIBCO Statistica® User's Guide*
- *TIBCO Statistica® Enterprise Manager Guide*
- *TIBCO Statistica® Statistica Object Model Guide*
- *TIBCO Statistica® Logistic Regression Formula Guide*
- *TIBCO Statistica® Stability Analysis Formula Guide*
- *TIBCO Statistica® Stepwise Model Builder Formula Guide*
- *TIBCO Statistica® Weight of Evidence Formula Guide*

The following documents for TIBCO® Data Science for TIBCO Spotfire® Analyst can be found on the [TIBCO Documentation site](#):

- *TIBCO® Data Science for TIBCO Spotfire® Analyst Release Notes*
- *TIBCO® Data Science for TIBCO Spotfire® Analyst User Guide*

The following documents for TIBCO® Data Science Service for TIBCO Spotfire® can be found on the [TIBCO Documentation site](#):

- *TIBCO® Data Science Service for TIBCO Spotfire® Release Notes*
- *TIBCO® Data Science Service for TIBCO Spotfire® Installation and Configuration Guide*
- *TIBCO® Data Science Service for TIBCO Spotfire® User Guide*

How to Contact TIBCO Support

You can contact TIBCO Support in the following ways:

- For an overview of TIBCO Support, visit <http://www.tibco.com/services/support>.
- For accessing the Support Knowledge Base and getting personalized content about products you are interested in, visit the TIBCO Support portal at <https://support.tibco.com>.
- For creating a Support case, you must have a valid maintenance or support contract with TIBCO. You also need a user name and password to log in to <https://support.tibco.com>. If you do not have a user name, you can request one by clicking Register on the website.

How to Join TIBCO Community

TIBCO Community is the official channel for TIBCO customers, partners, and employee subject matter experts to share and access their collective experience. TIBCO 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 TIBCO products. In addition, users can submit and vote on feature requests from within the [TIBCO Ideas Portal](#). For a free registration, go to <https://community.tibco.com>.

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 can be 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

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

$i = 1, \dots, n$

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

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

$i = 1, \dots, n_j$

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

$j = 1, \dots, k$

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

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

$i = 1, \dots, n_j$

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

$j = 1, \dots, k$

k = number of batches

Parameter Estimation

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

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

where,

\mathbf{X} = N by p design matrix of coded predictor variables

\mathbf{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:

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

t-Statistic

The t-statistic associated with the null hypothesis that the i^{th} 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 i^{th} diagonal element of $\text{MSE}(\mathbf{X}^T \mathbf{X})^{-1}$

Confidence Interval for a Single Parameter

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

$$\hat{\beta}_i \pm t_{0.975, df_e} \hat{\sigma}_{\hat{\beta}_i}$$

where,

$t_{0.975, df_e}$ = 97.5th 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 \mathbf{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:

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

where,

df_e = error degrees of freedom

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

$t_{0.975, df_e}$ = 97.5th quantile of t distribution with $df = df_e$

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

Prediction Interval

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

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

where,

df_e = error degrees of freedom

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

$t_{0.975, df_e}$ = 97.5th quantile of t distribution with $df = df_e$

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

Model Selection

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.

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 Analysis

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 = $X(X^T X)^{-1} X^T$

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

$i = 1, \dots, N$

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.)

Legal and Third-Party Notices

SOME TIBCO SOFTWARE EMBEDS OR BUNDLES OTHER TIBCO SOFTWARE. USE OF SUCH EMBEDDED OR BUNDLED TIBCO SOFTWARE IS SOLELY TO ENABLE THE FUNCTIONALITY (OR PROVIDE LIMITED ADD-ON FUNCTIONALITY) OF THE LICENSED TIBCO SOFTWARE. THE EMBEDDED OR BUNDLED SOFTWARE IS NOT LICENSED TO BE USED OR ACCESSED BY ANY OTHER TIBCO SOFTWARE OR FOR ANY OTHER PURPOSE.

USE OF TIBCO SOFTWARE AND THIS DOCUMENT IS SUBJECT TO THE TERMS AND CONDITIONS OF A LICENSE AGREEMENT FOUND IN EITHER A SEPARATELY EXECUTED SOFTWARE LICENSE AGREEMENT, OR, IF THERE IS NO SUCH SEPARATE AGREEMENT, THE CLICKWRAP END USER LICENSE AGREEMENT WHICH IS DISPLAYED DURING DOWNLOAD OR INSTALLATION OF THE SOFTWARE (AND WHICH IS DUPLICATED IN THE LICENSE FILE) OR IF THERE IS NO SUCH SOFTWARE LICENSE AGREEMENT OR CLICKWRAP END USER LICENSE AGREEMENT, THE LICENSE(S) LOCATED IN THE "LICENSE" FILE(S) OF THE SOFTWARE. USE OF THIS DOCUMENT IS SUBJECT TO THOSE TERMS AND CONDITIONS, AND YOUR USE HEREOF SHALL CONSTITUTE ACCEPTANCE OF AND AN AGREEMENT TO BE BOUND BY THE SAME.

This document is subject to U.S. and international copyright laws and treaties. No part of this document may be reproduced in any form without the written authorization of TIBCO Software Inc.

TIBCO, the TIBCO logo, the TIBCO O logo, Statistica, Spotfire, Process Tree Viewer, Process Data Explorer, Predictive Claims Flow, Making the World More Productive, Live Score, Electronic Statistics Textbook, Decisioning Platform, Data Health Check, and Better Decisioning are either registered trademarks or trademarks of TIBCO Software Inc. and/or its subsidiaries in the United States and/or other countries.

Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

All other product and company names and marks mentioned in this document are the property of their respective owners and are mentioned for identification purposes only.

This 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. Please see the readme.txt file for the availability of this software version on a specific operating system platform.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS DOCUMENT COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THIS DOCUMENT. TIBCO SOFTWARE INC. MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED IN THIS DOCUMENT AT ANY TIME.

THE CONTENTS OF THIS DOCUMENT MAY BE MODIFIED AND/OR QUALIFIED, DIRECTLY OR INDIRECTLY, BY OTHER DOCUMENTATION WHICH ACCOMPANIES THIS SOFTWARE, INCLUDING BUT NOT LIMITED TO ANY RELEASE NOTES AND "READ ME" FILES.

This and other products of TIBCO Software Inc. may be covered by registered patents. Please refer to TIBCO's Virtual Patent Marking document (<https://www.tibco.com/patents>) for details.

Copyright © 1995-2020. TIBCO Software Inc. All Rights Reserved.