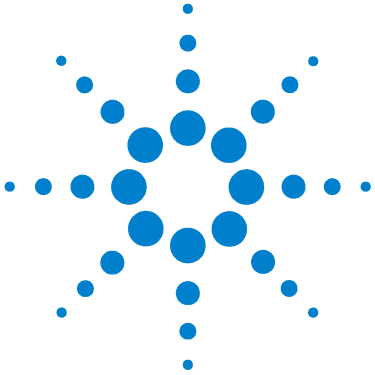


Agilent 85190A
IC-CAP 2008

Statistical Analysis

Agilent Technologies

Notices

Edition

March 2008

Printed in USA

Agilent Technologies, Inc.
5301 Stevens Creek Blvd.
Santa Clara, CA 95052 USA

Acknowledgments

Errata

The IC-CAP product may contain references to “HP” or “HPEESOF” such as in file names and directory names. The business entity formerly known as “HP EEsof” is now part of Agilent Technologies and is known as “Agilent EEsof.” To avoid broken functionality and to maintain backward compatibility for our customers, we did not change all the names and labels that contain “HP” or “HPEESOF” references.

Warranty

The material contained in this document is provided “as is,” and is subject to being changed, without notice, in future editions. Further, to the maximum extent permitted by applicable law, Agilent disclaims all warranties, either express or implied, with regard to this manual and any information contained herein, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Agilent shall not be liable for errors or for incidental or consequential damages in connection with the furnishing, use, or performance of this document or of any information contained herein. Should Agilent and the user have a separate written agreement with warranty terms covering the material in this document that conflict with these terms, the warranty terms in the separate agreement shall control.

Technology Licenses

The hardware and/or software described in this document are furnished under a license and may be used or copied only in accordance with the terms of such license.

Restricted Rights Legend

U.S. Government Restricted Rights. Software and technical data rights granted to the federal government include only those rights customarily provided to end user customers. Agilent provides this customary commercial license in Software and technical data pursuant to FAR 12.211 (Technical Data) and 12.212 (Computer Software) and, for the Department of Defense, DFARS 252.227-7015 (Technical Data - Commercial Items) and DFARS 227.7202-3 (Rights in Commercial Computer Software or Computer Software Documentation).

Safety Notices

CAUTION

A **CAUTION** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a **CAUTION** notice until the indicated conditions are fully understood and met.

WARNING

A WARNING notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a WARNING notice until the indicated conditions are fully understood and met.

Contents

1 Getting Started

The IC-CAP Statistics Package helps circuit designers, device engineers, and process engineers improve device and IC yields and design more robust products. IC-CAP Statistics provides you with the tools needed to identify and analyze the inter-relationships between device model parameters and electrical test data.

Both parametric and non-parametric analysis can be performed with IC-CAP Statistics. Parametric analysis features include principal component analysis, factor analysis and multiple linear regression. With these techniques, it is easy to select the best model parameters to track in electrical test or to build models that predict SPICE parameters from dominant parameters or independent factors.

The IC-CAP Statistics Package automatically generates model files from corner models or Monte Carlo analysis. In addition, an exclusive, new method for generating worst-case model candidates, called Non-Parametric Boundary Modeling is included.

Major Benefits

The IC-CAP Statistics Package:

- Minimizes the number of circuit simulations required to create robust designs
- Relates model parameters to manufacturing and process data
- Provides realistic worst-case models for arbitrary joint-probability densities

Major Features

The IC-CAP Statistics Package provides:

- Data management capability for handling large sets of extracted parameters
- Multivariate statistical analysis for data reduction and the generation of worst-case corner models, parametric boundary models, or Monte Carlo analysis
- Proprietary Agilent EEsof non-parametric analysis algorithms for identification of nominal models and worst-case-candidate models from arbitrary joint probability densities
- Statistical plotting for viewing data and determining relationships that exist between various parameters
- Statistical models that can be used within the IC-CAP environment for simulation and validation, or be imported directly into external SPICE or other simulators to perform worst-case analysis or design centering on large circuits incorporating the modeled device
- Flexibility to automate procedures using IC-CAP's PEL language.

IC-CAP Statistics can bridge the gap between manufacturing and design by assisting you in selecting an optimal set of parameters to be tracked from manufacturing through electrical test. Using factor analysis, a subset of model parameters for a particular model (e.g., BSIM3) can be identified as being responsible for the majority of the statistical variation in the device. Regression-based

equations can be generated by the program, which relate the dependent model parameters to this subset of “dominant” parameters.

Thus, a small set of parameters tracked in manufacturing can serve both as a means of process monitoring and a means of generating predictive models for use in circuit simulation.

User Interface

The IC-CAP Statistics user interface is similar to the main IC-CAP program. For an explanation of graphical user interface features such as toolbars and drop-down menus, refer to Chapter 2, “Program Basics,” in the *User’s Guide*.

Example—Building a Statistical Model

To introduce IC-CAP Statistics, let's go through the typical steps needed to build a parametric statistical model, using parameters for a common semiconductor device model. We will:

1. Measure and extract model parameters
2. Start IC-CAP Statistics and import data
3. Transform distributions to Gaussian
4. Eliminate outlier data
5. Perform correlation analysis
6. Perform factor or principal component analysis
7. Generate model equations
8. Generate models from parametric analysis
9. Test models

NOTE The example we use to introduce IC-CAP Statistics is based on a parametric analysis, which assumes a Gaussian distribution of the data. IC-CAP Statistics also contains non-parametric analysis, which can be used when the data is bimodal or otherwise non-Gaussian. This method is described briefly at the end of this chapter and in depth in the "Parametric Analysis Results Window" on page 66.

Measure and Extract Model Parameters

First you measure and extract the parameters needed for your device model using IC-CAP software or another parameter extraction program. This procedure is described in Chapter 5, "Making Measurements," in the *User's Guide*. The data is then imported into IC-CAP Statistics.

Start IC-CAP Statistics and Import Data

Start IC-CAP Statistics:

From the Main IC-CAP window choose the **Tools** drop-down menu and then choose **Statistics** (or click the Statistics icon). The Statistics package window is displayed.

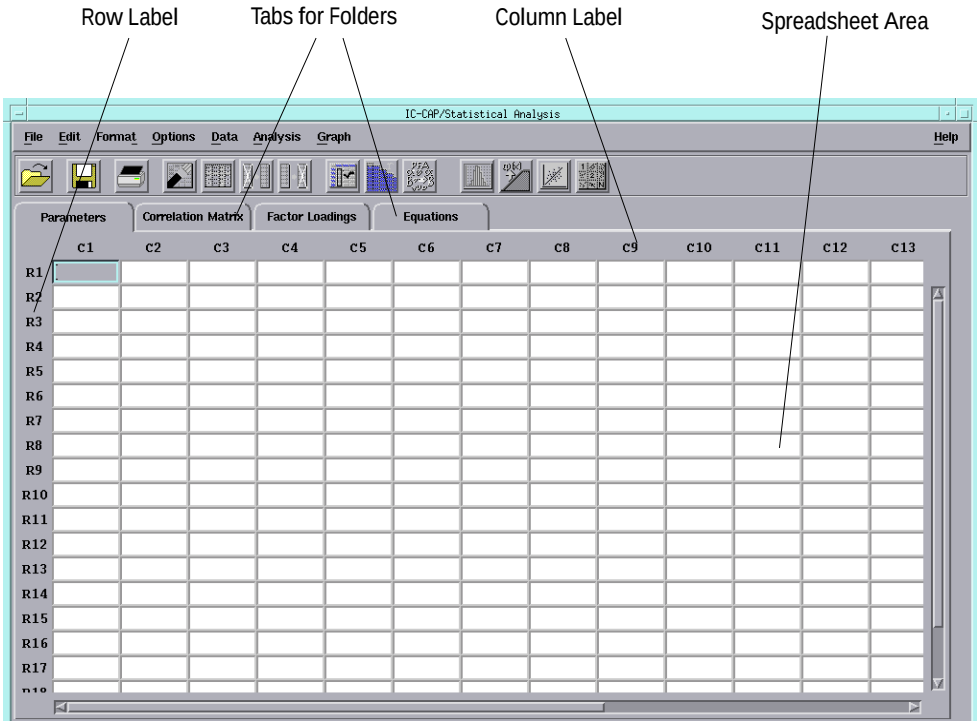


Figure 1 IC-CAP Statistics Window

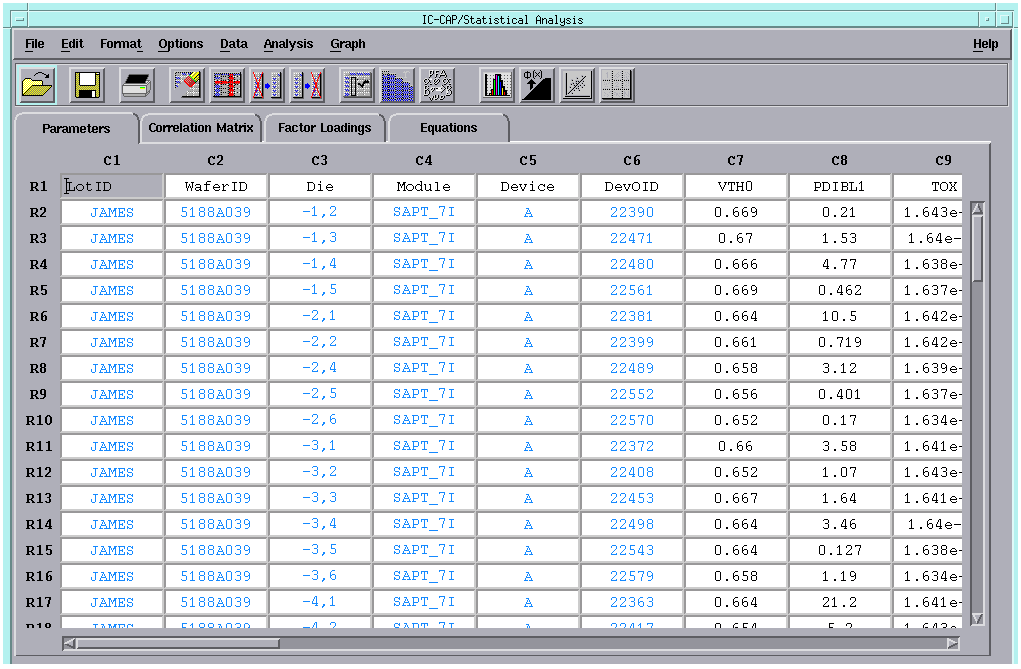
There are four ways to begin working with IC-CAP Statistics:

- Importing an ASCII text file containing your data (such as from Excel)
- Loading extraction data directly from IC-CAP
- Opening a file already in the IC-CAP Statistics data file (.sdf) format, which is based on the MDIF file format
- Manually typing the data in the Statistics spreadsheet

For this overview, we will use the third method, and open an example file called *bsim3.sdf*.

NOTE This BSIM3 data file is being used to teach you how to use the program *only*. It does not contain validated data. Do not be concerned if you primarily use other types of models.

1. From the File menu, choose **Examples**. The Examples Open dialog box appears.
2. Select **bsim3.sdf** from the list of files and choose **OK**.
3. The spreadsheet is loaded with data.



	C1	C2	C3	C4	C5	C6	C7	C8	C9
R1	LotID	WaferID	Die	Module	Device	DevOID	VTH0	PDIBL1	TOX
R2	JAMES	5188A039	-1,2	SAPT_7I	A	22390	0.669	0.21	1.643e-
R3	JAMES	5188A039	-1,3	SAPT_7I	A	22471	0.67	1.53	1.64e-
R4	JAMES	5188A039	-1,4	SAPT_7I	A	22480	0.666	4.77	1.638e-
R5	JAMES	5188A039	-1,5	SAPT_7I	A	22561	0.669	0.462	1.637e-
R6	JAMES	5188A039	-2,1	SAPT_7I	A	22381	0.664	10.5	1.642e-
R7	JAMES	5188A039	-2,2	SAPT_7I	A	22399	0.661	0.719	1.642e-
R8	JAMES	5188A039	-2,4	SAPT_7I	A	22489	0.658	3.12	1.639e-
R9	JAMES	5188A039	-2,5	SAPT_7I	A	22552	0.656	0.401	1.637e-
R10	JAMES	5188A039	-2,6	SAPT_7I	A	22570	0.652	0.17	1.634e-
R11	JAMES	5188A039	-3,1	SAPT_7I	A	22372	0.66	3.58	1.641e-
R12	JAMES	5188A039	-3,2	SAPT_7I	A	22408	0.652	1.07	1.643e-
R13	JAMES	5188A039	-3,3	SAPT_7I	A	22453	0.667	1.64	1.641e-
R14	JAMES	5188A039	-3,4	SAPT_7I	A	22498	0.664	3.46	1.64e-
R15	JAMES	5188A039	-3,5	SAPT_7I	A	22543	0.664	0.127	1.638e-
R16	JAMES	5188A039	-3,6	SAPT_7I	A	22579	0.658	1.19	1.634e-
R17	JAMES	5188A039	-4,1	SAPT_7I	A	22363	0.664	21.2	1.641e-

The spreadsheet displays the data in rows and columns. Each row contains one sample. Each column contains either a sample's attribute, such as the sample ID, lot number, date, or temperature; or is a sample's measured or extracted data, such as VSAT, VTH0, or TOX. *Attribute* information is displayed in blue, while *parameter* data is displayed in black.

Spreadsheet Format

The data may contain too many characters to fit in the cells. From the Format menu, choose **Column Width**, a dialog box appears. You enter a larger or smaller number in the field to fit your data. In this case, accept the default of 10 and choose **OK**.

Transform Data

One of the key assumptions made by multivariate techniques such as Factor Analysis is that the data set to be analyzed is a joint Gaussian distribution. If the data is not joint Gaussian, then the model generated from the analysis may not accurately reproduce the measured density.

One of the ways to help convert a data set to Gaussian is to perform a mathematical transformation.

You have to decide which data columns need to be transformed. Some columns may already be Gaussian. As described below, you can quickly plot the data to see if it is Gaussian. The next step, *Eliminate Outlier Data*, can be done before the data transformation step, depending on the look of the data.

Selecting Columns and Rows

The spreadsheet columns have the labels C1, C2, C3, etc., just above the columns. The rows have the labels R1, R2, R3, etc., just to the left of the rows. See Figure 1. To select an entire column or row, move the cursor to the column or row label you want and press the left mouse button.

Plot and Analyze the Data

To view the data before transforming it, we will plot the data for column C8 as a histogram.

1 Select column C8 (parameter PDIBL1) by clicking column label **C8**. The column is highlighted.

2 From the Graph menu choose **Histogram** (or click the Histogram icon from the toolbar). A plot window appears with the histogram for that column.

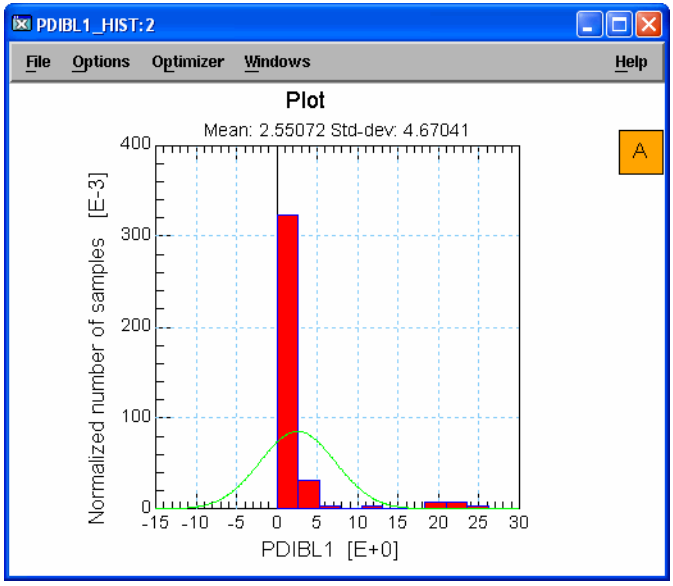


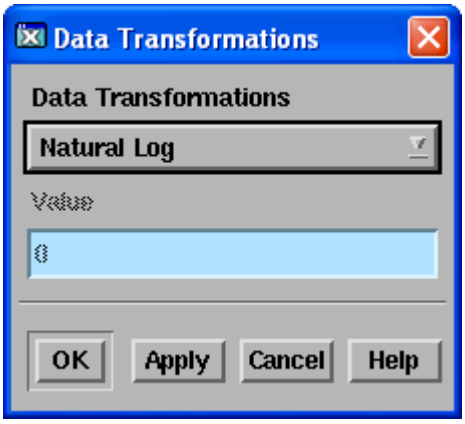
Figure 2 Histogram Before Data Transform

When you are done viewing plots, you can choose **File > Close** from the Plot window.

Transform the Data

1 Select column C8 again.

2 From the Data menu, choose **Data Transformations**. A dialog box appears.



3 To select a transformation type, click the drop-down list button and select the type you want.

4 For this example, choose **Natural Log** and choose **OK**. The data for column C8 is transformed.

The parameter name is appended with LN (for log natural) and becomes LN_PDIBL1.

Re-plot the histogram for column C8. Select column C8. From the Graph menu choose **Histogram**. Note that the data is now more Gaussian, but there is an outlier to the left.

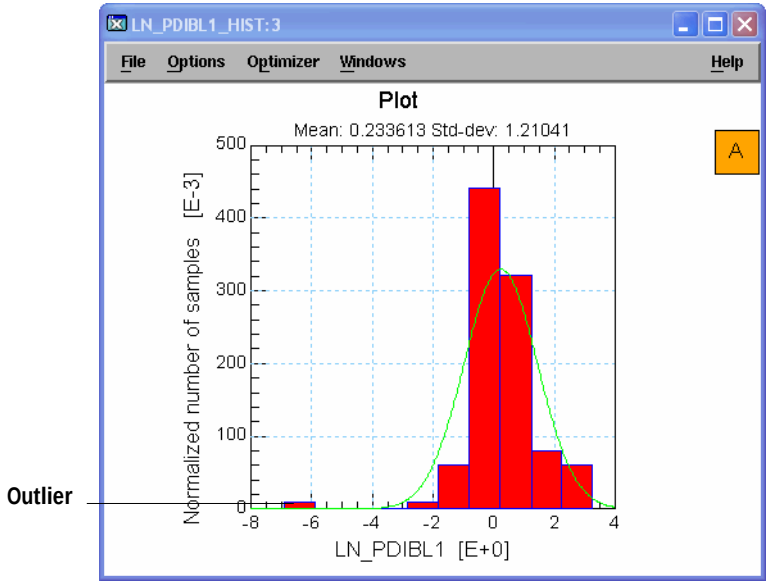


Figure 3 Histogram After Data Transform

Eliminate Outliers

There are several ways to eliminate outlier data or other invalid data. You can vary the order in which these methods are done. For example, you may immediately spot bad data and manually eliminate it, you can automatically filter the data to remove outliers, or you can plot the data in a histogram or scatter plot to help spot outliers. Often several iterations of these methods have to be performed until you're satisfied that the data is ready for correlation analysis.

Plot and Analyze the Data

To help spot outlier data, let's study the latest plot, above, for column C8. Note the that there appears to be an outlier at the far left of the plot, corresponding to a value of about -6.9. If you scan the data in the column, you will see that this value is in row R20.

Manually Eliminate Outliers

Let's assume that from a review of the data, you believe the sample in row R20 is a bad sample.

1 To select this row, click row label **R20**.

2 From the Edit menu choose **Deactivate**. The row's background color changes to gray indicating that this sample is deactivated (to re-activate it, choose **Edit > Activate**).

3 Select column C8 again. From the Graph menu choose **Histogram**. See that the plot is more Gaussian with the outlier eliminated.

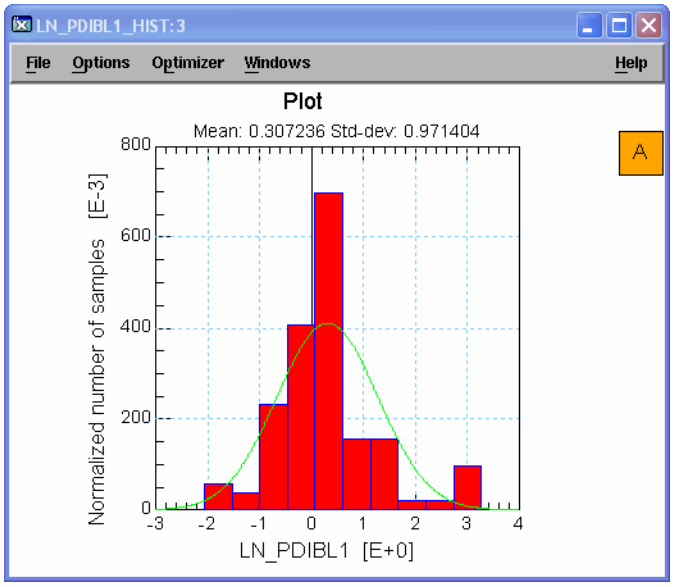


Figure 4 Histogram After Outlier Elimination

Automatic Data Filtering

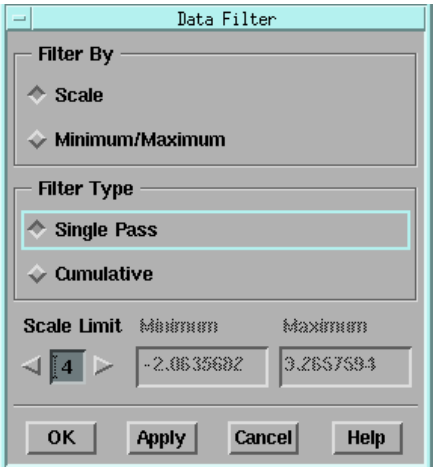
IC-CAP Statistics can automatically filter data based on minimum/maximum values or by a scale value. We will use a scale value. Scale is defined as the median absolute deviation (MAD) divided by a constant (approximately 0.6745). This standardizes MAD in order to make the scale

estimate consistent with the standard deviation of a normal distribution. The greater the scale value, the further from the median the filtering occurs.

1 Select column C8 again.

2 From the Data menu choose **Data Filter**. A dialog box is displayed.

3 Accept the default Scale option (near top) to filter by a scale value.



4 Change the Scale Limit (near bottom) to 4 by clicking the right arrow or typing in the field. Then choose **OK**.

The data is filtered based on this scale value. Note that eliminated rows are highlighted by a color change that indicates they have been filtered out. (The process can be undone by choosing Data > Undo Data Filtering.) Select column C8 once again. From the Graph menu choose **Histogram**. See that the plot is now more Gaussian with the data filtered.

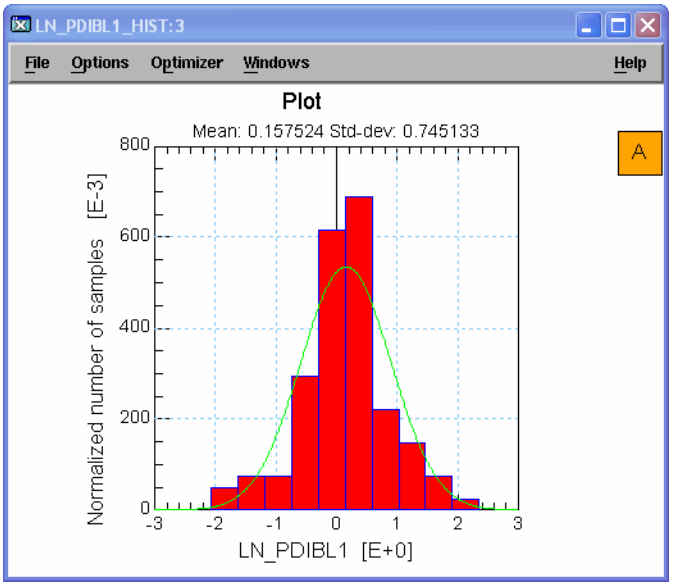


Figure 5 Histogram After Data Filtering

Choose Statistical Summary for a Numeric Display

Besides a variety of plots to help you analyze your data, IC-CAP Statistics also has a Statistical Summary window (Analysis > Statistical Summary), which shows you standard statistical data, such as mean, variance, standard deviation, skewness, kurtosis, etc.

Repeat Data Transformation and Outlier Elimination for Other Columns

Repeat the steps outlined in the last two sections for each column that is non-Gaussian. For this example, you can skip this step.

Perform Correlation Analysis

Correlation analysis provides a numerical measure of the amount of variation in one variable that is attributable to another variable. When an increase in the value of one variable

is associated with an increase in the value of the other variable, the correlation is positive. When the increase is associated with a decrease, the correlation is negative.

Correlation analysis is always performed before proceeding to factor analysis and the data used consists of *all* the rows in the spreadsheet that have not been *filtered*, *deactivated*, or *deleted*. To perform correlation analysis:

From the Analysis menu, choose **Correlation Analysis**. The Statistics window changes so that the Correlation Matrix folder is displayed. (If you want to go back to the parameter data before correlation analysis was performed, choose the folder tab labeled **Parameters**.)

Parameters | Correlation Matrix | Factor Loadings | Equations

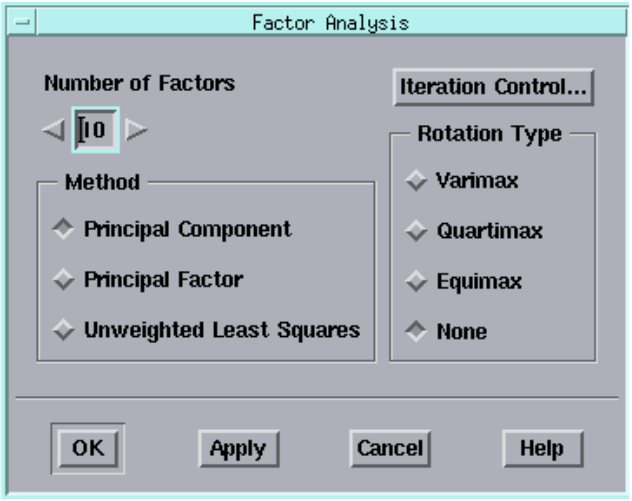
	C1	C2	C3	C4	C5	C6	C7	C8
R1	Parameters	VTH0	LN_PDIBL1	TOX	K1	K2	K3	UA
R2	VTH0	1	0.02301813	0.68508693	0.1079446	0.19678496	0.03376760	0.30100217
R3	LN_PDIBL1	0.02301813	1	.000422942	0.17145006	-0.1953400	-0.0586468	-0.2158907
R4	TOX	0.68508693	.000422942	1	0.23359816	0.10937899	-0.1279979	0.26975498
R5	K1	0.1079446	0.17145006	0.23359816	1	-0.8677680	-0.1814875	-0.6277974
R6	K2	0.19678496	-0.1953400	0.10937899	-0.8677680	1	0.1001625	0.66622811
R7	K3	0.03376760	-0.0586468	-0.1279979	-0.1814875	0.1001625	1	0.10394578
R8	UA	0.30100217	-0.2158907	0.26975498	-0.6277974	0.66622811	0.10394578	1
R9	UB	-0.2328701	0.16585016	-0.2162033	0.22932561	-0.1667067	0.01630693	-0.7420965
R10	UC	0.1901659	-0.1209297	0.21039859	-0.6108630	0.80073389	0.10678869	0.46041935
R11	VSAT	0.03458770	0.75046302	0.02422911	0.29249677	-0.3219798	0.04265632	-0.3064381
R12	VOFF	0.01443082	0.03026316	-0.0514615	-0.1132410	0.03670589	0.31890621	0.02077750
R13	DL	0.08582748	0.02851296	0.04693041	0.30660527	-0.3689783	0.04228276	-0.2096393
R14	DW	0.04627482	0.18672065	-0.1005797	0.63862369	-0.7076457	0.01348336	-0.5689850

The Correlation Matrix displays the same parameters down the rows and across the columns. The correlation coefficients for any two parameters are displayed where the rows and columns intersect. In the above example, the cell formed by R4 and C2 has a value of about 0.69, which shows moderate to strong correlation between parameters TOX and VTH0.

Perform Factor Analysis

Now that the correlation matrix is defined, the next step is to perform factor analysis.

1 From the Analysis menu, choose **Factor Analysis**, a dialog box is displayed.



You choose the method of factor analysis from three choices: Principal Component, Principal Factor, or Unweighted Least Squares.

The Principal Component and Principal Factor methods, while quite different in assumptions, are similar in the end effect; the main difference lies in their respective error terms. Unweighted Least Squares is a method of factor analysis using an iterative process. A detailed description of these methods can be found in Chapter 3, “Data Analysis.”

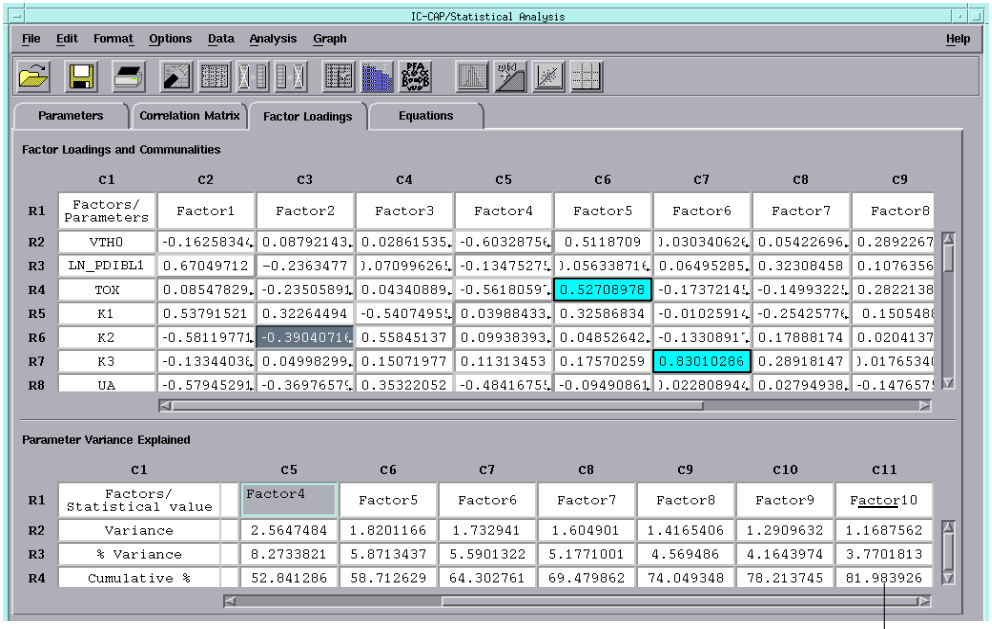
You choose a starting figure for the number of factors you want to be found in your analysis. After you see the results, which correspond to the percent variation that can be explained by this number of factors, you can increase or decrease the number and repeat the analysis.

2 For our example, enter `10` in the Number of Factors field.

3 In the Method field, choose the default **Principal Component** option button.

4 Accept the default Rotation Type of None.

5 Choose **OK** to perform the analysis. The screen changes to display the Factor Loadings folder.



Cumulative % with 10 factors

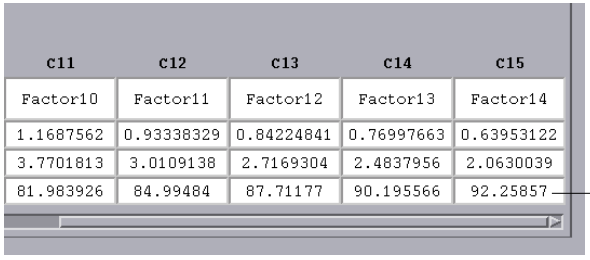
Two tables are generated in this window; the first contains the factor loadings. Factor loadings represent the correlations between each factor and the model parameters. The second table presents a summary of the variances associated with each factor as well as a report of the percentage error explained by each factor.

Note that the cumulative percent for 10 factors, shown in the lower right cell, is about 82% (you may have to use the scroll bar to see it). This means that if only 10 factors were used to

make a statistical model from this data, the model would explain 82% of the variance compared to using all of the parameters/factors.

6 Now we will re-analyze using 14 factors. Choose **Analysis > Factor Analysis** and enter 14 in the Number of Factors field.

With 14 factors, the cumulative percent is about 92%, as shown in the lower right cell below. You have to decide how high a figure is acceptable for your work.



C11	C12	C13	C14	C15
Factor10	Factor11	Factor12	Factor13	Factor14
1.1687562	0.93338329	0.84224841	0.76997663	0.63953122
3.7701813	3.0109138	2.7169304	2.4837956	2.0630039
81.983926	84.99484	87.71177	90.195566	92.25857

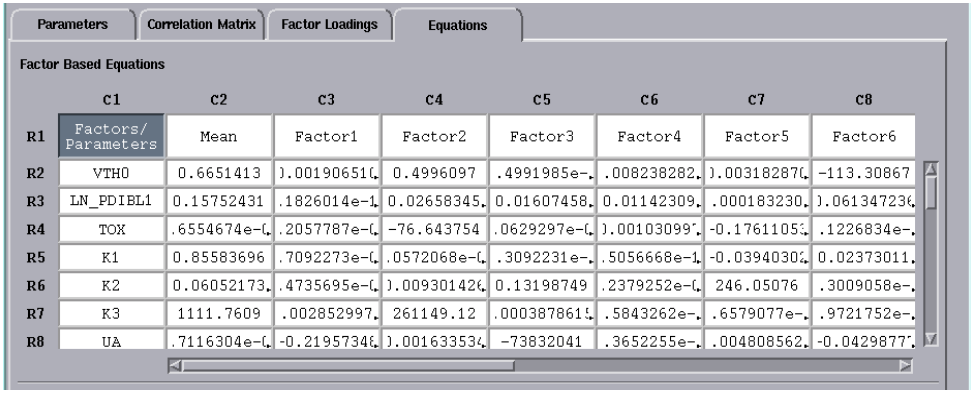
Cumulative % with 14 factors

The top portion of the Factor Loading folder displays the data in a color-coded format. Factor Group data, one group per row, is displayed in a red font. Dominant Parameter data, one dominant parameter per column, is displayed with a blue background. A detailed description of both can be found in "Perform Factor Analysis" on page 58.

Generate Equations

Next, we will generate equations from the factor analysis. You can generate equations from factors or dominant parameters. IC-CAP Statistics computes the equation coefficients that you use to build your SPICE model.

From the Analysis menu choose **Generate Equations**. A submenu with two choices appears to the right. Choose **Factors**. The screen changes to display the Equations folder



	C1	C2	C3	C4	C5	C6	C7	C8
R1	Factors/ Parameters	Mean	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
R2	VTH0	0.6651413	).001906510	0.4996097	.4991985e-.	.008238282.	).00318287(	-113.30867
R3	LN_PDIBL1	0.15752431	.1826014e-1	0.02658345.	0.01607458.	0.01142309.	.000183230.	).06134723(
R4	TOX	.6554674e-(	.2057787e-(	-76.643754	.0629297e-(	).00103099'	-0.17611054	.1226834e-.
R5	K1	0.85583696	.7092273e-(	.0572068e-(	.3092231e-.	.5056668e-1	-0.03940304	0.02373011.
R6	K2	0.06052173.	.4735695e-(	).00930142(	0.13198749	.2379252e-(	246.05076	.3009058e-.
R7	K3	1111.7609	.002852997.	261149.12	.000387861(	.5843262e-.	.6579077e-.	.9721752e-.
R8	UA	.7116304e-(	-0.21957348	).00163353(	-73832041	.3652255e-.	.004808562.	-0.04298777

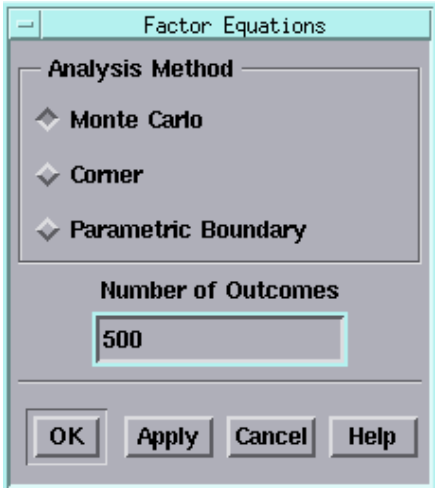
Generate a Parametric Model

Now that the equation coefficients are generated, you can build a variety of statistical models, or save the data in a SPICE equations format for use in circuit simulations. You can choose from Monte Carlo, Corner, or Parametric Boundary models. You can test your model, based on a reduced set of parameters, against the raw data to see how well it performs. At this point, IC-CAP Statistics has been designed for flexibility to work with your process.

For our example, we will perform Monte Carlo analysis and compare the results to the raw data.

Perform Monte Carlo Analysis, Plot Data, and Compare

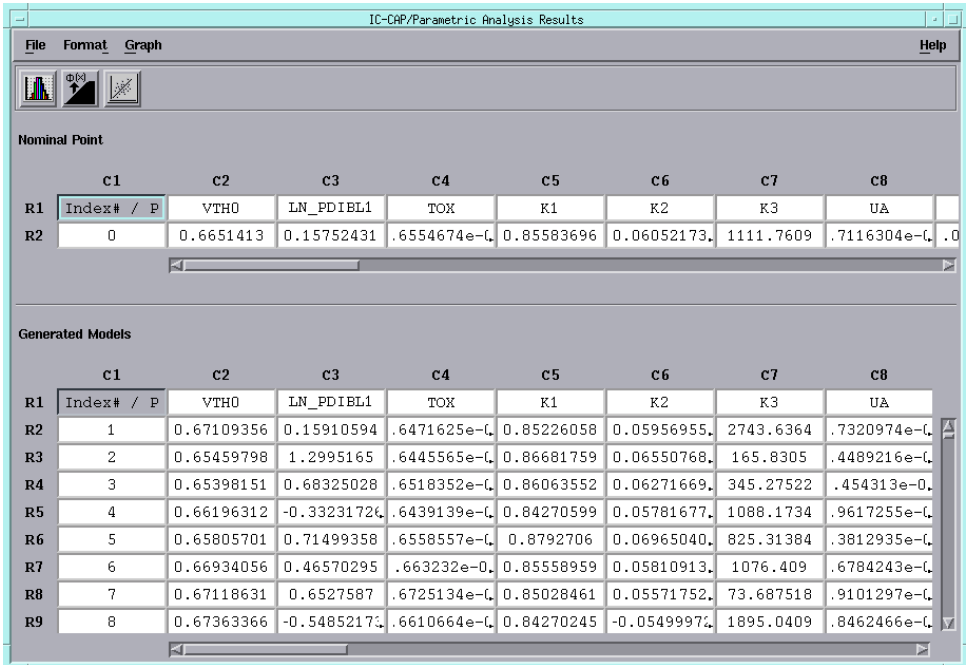
1 From the Analysis menu choose **Parametric Analysis**. A submenu is displayed to the right. Choose **Factor Equations**, a dialog box is displayed.



2 Choose the **Monte Carlo** method.

3 In the Number of Outcomes field, enter 500 and choose **OK**.

The results of the Monte Carlo analysis are displayed in the Parametric Analysis Results window. The number of rows is equal to the number of Monte Carlo outcomes, and the columns correspond to the parameters.



	C1	C2	C3	C4	C5	C6	C7	C8
R1	Index# / P	VTH0	LN_PDIBL1	TOX	K1	K2	K3	UA
R2	0	0.6651413	0.15752431	.6554674e-0.	0.85583696	0.06052173.	1111.7609	.7116304e-0.

	C1	C2	C3	C4	C5	C6	C7	C8
R1	Index# / P	VTH0	LN_PDIBL1	TOX	K1	K2	K3	UA
R2	1	0.67109356	0.15910594	.6471625e-0.	0.85226058	0.05956955.	2743.6364	.7320974e-0.
R3	2	0.65459798	1.2995165	.6445565e-0.	0.86681759	0.06550768.	165.8305	.4489216e-0.
R4	3	0.65398151	0.68325028	.6518352e-0.	0.86063552	0.06271669.	345.27522	.454313e-0.
R5	4	0.66196312	-0.33231726	.6439139e-0.	0.84270599	0.05781677.	1088.1734	.9617255e-0.
R6	5	0.65805701	0.71499358	.6558557e-0.	0.8792706	0.06965040.	825.31384	.3812935e-0.
R7	6	0.66934056	0.46570295	.663232e-0.	0.85558959	0.05810913.	1076.409	.6784243e-0.
R8	7	0.67118631	0.6527587	.6725134e-0.	0.85028461	0.05571752.	73.687518	.9101297e-0.
R9	8	0.67363366	-0.54852173	.6610664e-0.	0.84270245	-0.05499972	1895.0409	.8462466e-0.

Figure 6 Parametric Analysis Results Window

Earlier we used parameter PDIBL1 to plot as a histogram, and transformed the data to a more Gaussian fit, see Figure 5.

4 Now, select the column for parameter PDIBL1 (now labeled LN_PDIBL1 because we did a log transform of the data) in the Parametric Analysis Results window.

5 From the Graph menu choose **Histogram**.

6 Compare this plot (made from synthesized Monte Carlo data) with the earlier plot.

Details on Corner Models and Boundary Models are found in "Generate a Parametric Model" on page 64.

Non-Parametric Boundary Modeling

IC-CAP Statistics contains proprietary Agilent EEsof non-parametric analysis algorithms for identifying nominal models and worst-case-candidate models from arbitrary joint probability densities. This advanced feature, called Non-Parametric Boundary Modeling, differs from the parametric (joint Gaussian) methods described earlier, and can be used when the data is bimodal or otherwise non-Gaussian.

Details on Non-Parametric Boundary Modeling are found in "Parametric Analysis Results Window" on page 66.

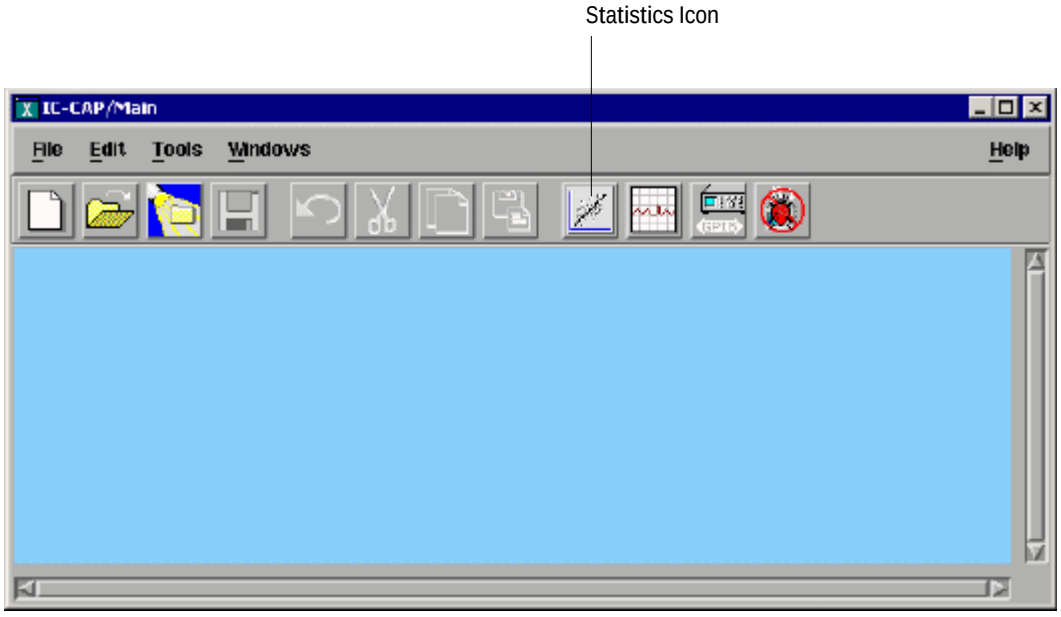
2
Program Basics

Chapter 1 introduced IC-CAP Statistics and used an example BSIM3 model file tutorial to show you how the program is typically used. If you haven't already read Chapter 1, you should do so before continuing. In this chapter we will show you how to use most of IC-CAP Statistics' basic functions.

Starting IC-CAP Statistics

To start the Statistics tool:

From the Main window, choose **Tools > Statistics** (or click the Statistics icon).



The Statistics window is displayed.

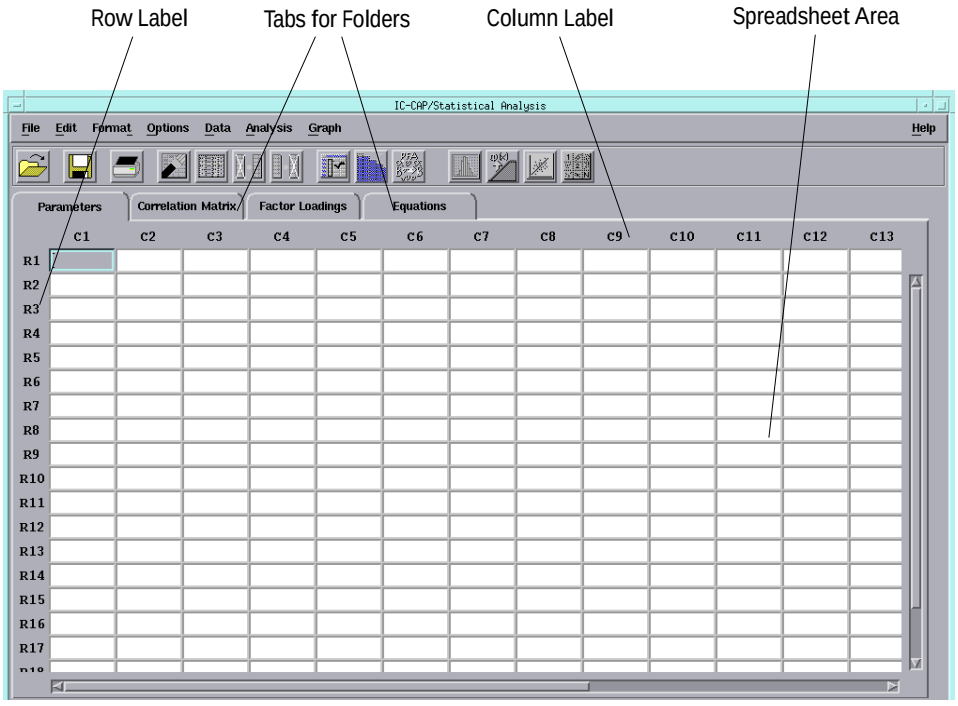


Figure 7 IC-CAP Statistics Window

Spreadsheet Format

The Main IC-CAP Statistics window appears in a spreadsheet format. The spreadsheet displays the data in rows and columns. Each row contains one sample. Each column contains either a sample's attribute, such as the sample ID, lot number, date, or temperature; or is a sample's measured or extracted data (parameters), such as Vsat, Voff, or DL. Attribute information is displayed in blue, while parameter data is displayed in black.

Changing Row Height or Column Width

Your data may contain too many characters to fit in the cells, or you may want narrower columns. The Format menu lets you change the number of characters for either rows or columns. Choose **Row Height** or **Column Width**, a dialog box appears. You enter a larger or smaller number in the field to fit your data. For columns, the number specifies the width in characters. This action changes the size of *all* the rows or columns.

If you want to change only *one* column or row, position the cursor along the cell edge line. The I-beam cursor turns into a pointer. Drag the pointer left or right to change columns, or up or down to change rows.

NOTE

If you manually change row height or column width by dragging the cell edge, those rows or columns cannot be altered with the Format menu command again during this Statistics session.

Selecting Rows and Columns

The spreadsheet columns have the labels C1, C2, C3, etc., just above the columns. The rows have the labels R1, R2, R3, etc., just to the left of the rows. To select an entire column or row, move the cursor to the column or row label you want and press the left mouse button.

Selecting Multiple Rows or Columns

You may want to delete, deactivate, or activate multiple rows (containing your samples) or columns (containing your parameters).

- To select *two or more* rows or columns, select the first row or column, then hold down the Control key before you select the next row(s) or column(s).
- To select *all* the rows or columns between two points, select the first row or column, then hold down the Shift key before you select the last row or column. Or you can select the first row or column and drag the mouse over the rest of the desired rows or columns. The two rows or columns you selected and all those in-between are selected.

Folders

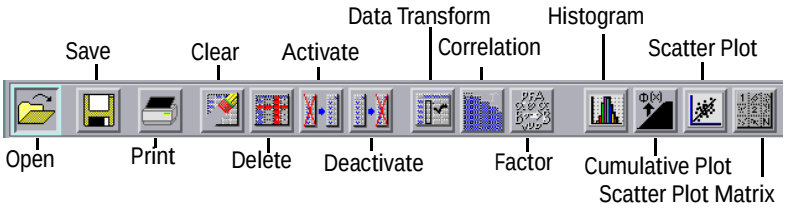
As shown in Figure 7, the Spreadsheet has tabs for four folders that contain data for:

- Parameters. This folder contains your raw data (measured or extracted), showing the parameters for each sample.
- Correlation Matrix. This folder shows how each parameter correlates with every other parameter.
- Factor Loading. This folder displays the loadings (correlations) that relate each model parameter to each of the derived factors as well as other data.
- Equations. This folder shows the equation coefficients derived from the factor analysis that you use to build your SPICE model.

Click the tab to display the data for each of these areas. A full description of each of these areas is found later in this chapter.

Icons

Near the top of the Statistics Window there is a toolbar with a group of icons. These icons provide one-click access to frequently used commands. Move the cursor just under each icon and leave it for a few seconds. A icon label appears that tells its function.



Opening or Creating a File

There are four ways to begin working with IC-CAP Statistics:

- Importing an ASCII text file containing your data
- Loading extraction data directly from IC-CAP
- Opening a file already in the IC-CAP Statistics data file (.sdf) format, which is based on the MDIF file format
- Manually typing the data in the Statistics spreadsheet

Importing ASCII Data

One way to begin work with IC-CAP Statistics is to import ASCII data saved in another program, such as Lotus 1-2-3 or Microsoft Excel. The data must be delimited with spaces or tabs (not commas). To import the data:

1. Choose **File > Import**, a dialog box is displayed. Select the path and name of the file you want to import and choose **OK**. The data is loaded into the spreadsheet. Complete details on using this type of dialog box are found in "Working with Dialog Boxes" in the *User's Guide*.
2. You have to specify which columns are attributes and which are parameters. The attribute columns have to be contiguous and on the left side of the spreadsheet. To specify that a column contains attributes, first select the columns. Then, choose **Edit > Parameter to Attribute**. If that column is not already at the left, it moves to the left and becomes an attribute column, denoted by a blue text font.
3. Repeat Step 2 for additional attribute columns. You can do this for multiple columns at one time by holding down the Shift or Ctrl key while selecting columns.
4. You can reverse this process if necessary by choosing Edit > Attribute to Parameter.

Loading Data from IC-CAP

There are two ways to load extracted model data from the Main IC-CAP program into Statistics:

- Using the supplied macro
- Manually executing PEL commands

The first method is preferable if you want automated file loading along with automated measurement and extraction. If you use the manual method, the IC-CAP Statistics window must be opened.

These procedures use IC-CAP's Parameter Extraction Language (PEL). For more information about PEL see Chapter 9, "Parameter Extraction Language," in the *Reference* manual.

Using the Supplied Macro

You will load the model file called *load_stat_data.mdl*, which contains macro files. This file is shipped with the software and found in the Examples area.

The macro called *Main* loads parameters from a collection of .mdl files into the Statistics Parameters spreadsheet. (The macro called *Load_Data* is invoked by *Main* and you should not use it.) Once the appropriate model variables have been set and the .mdl parameter files have been created, the *Main* macro is ready to be run. The macro is general, and can be run with any collection of .mdl files. To let you see how the macro works, a collection of example .mdl files are provided in the examples directory. The parameter values are automatically written to the Statistics Parameters spreadsheet through successive calls to the PEL function *icstat_from_partable*.

Before you use the macro, do the following:

1 In order to use the macro, your model files must be in the following format:

  <*base filename*>_<*n*>.mdl

  where *n* is an integer equal or greater than 1, and the first model file must start with 1. For example,

  bsim_1.mdl

  bsim_2.mdl

  bsim_3.mdl, or

eebjt_1.mdl

eebjt_2.mdl

eebjt_3.mdl, etc.

2 All the model files must reside in the same directory. You will indicate the pathname and the number of model files using model variables.

3 If you want attribute information to be automatically transferred from the .mdl files, then the attribute values and labels (attribute parameter names) will have to be stored in variable arrays in the .mdl files. The attribute labels should be stored in an ICCAP_ARRAY called ATTRIBUTE_LABELS. The attribute values should be stored in an ICCAP_ARRAY called ATTRIBUTE_VALS. The size of the arrays is arbitrary, but you must explicitly specify it in the variable NUM_ATTRIBUTES.

Now you are ready to use the macro.

1 From the Main IC-CAP program choose **File > Examples**. The Examples Open dialog box appears.

2 Select the **/ statistics / load_data** directory and then the **load_stat_data.mdl** file from the list of files and choose **OK**. The model file loads.

3 Choose the **Model Variables** tab. A table appears with four items. Ignore the first item (called Local_VAR). Enter data for the three remaining items, as follows:

Name	Value
numFiles	Enter the total number of model files you want to load into Statistics
filesDirPath	Enter the path to the directory that contains the model files
baseFileName	Enter the base filename for your model files, such as bsim or eebjt, as described above.

4 Choose the **Macros** tab and select **Main** in the Select Macro field.

5 Click the **Execute** button (left). The Statistics Parameters spreadsheet is filled with data.

NOTE Your data may contain parameter columns (such as nominal temperature) that contain constants. You should deactivate these columns before performing correlation analysis.

The following shows the macro.

```
! Open the statistics window
dummy = icstat_open()

! Prompt the user on whether he is using the example data or his own data

LINPUT “Use the example .mdl files (y/n)?”,”y”,selection

if selection == “y” then
  iccap_root=SYSTEM$(“echo $ICCAP_ROOT”)
  filesDirPath=VAL$(iccap_root) & “/examples/model_files/statistics/load_data/”
endif

! Prepend the path onto the file name and append _1.mdl to get the first
! file in the directory (all we care about is the parameter names, so any
! file in the directory would do).
filename = val$(filesDirPath) & val$(baseFileName) & “_1.mdl”

! Open the first file to get the parameter names
menu_func(“/”, “Add Model”, “temp”)
menu_func(“/temp”, “Open”, filename)

! Load the parameter names into the first row
dummy = icstat_set_param_column_labels(“/temp”,”MODEL PARAMETERS”)

! Compute the number of rows in the empty spreadsheet and then insert enough
! additional rows so that the total number of rows = numFiles+1.

numRows = icstat_num_rows(“PARAMETERS”)
if numRows < numFiles then
  numInsert = numFiles - numRows
  dummy = icstat_insert(numRows+1, numInsert, “ROW”)
endif

! Load the data into the statistics spreadsheet using the “Load_Data” macro

menu_func(“/load_stat_macro/Load_Data”, “Execute”)

! Use the value of the variable NUM_ATTRIBUTES to insert enough columns at
! the beginning of the parameters spreadsheet to store the attributes.
```

```
dummy = icstat_insert(1, /temp/NUM_ATTRIBUTES, “COLUMN”)

! Write the attribute names to the first row of the spreadsheet
! and convert these column to attribute columns.

index=1
while index <= /temp/NUM_ATTRIBUTES
   name = val$(/temp/ATTRIBUTE_LABELS[index-1])
   dummy = icstat_set_text_cell(1, index, name)
   dummy = icstat_parameter_2_attribute(index)
   index=index+1
end while

! Cycle through the .mdl files and read the attributes and parameters into
! the parameters spreadsheet

i=1
while i <= numFiles
   fileName = val$(filesDirPath) & val$(baseFileName) & “_” & val$(i) & “.mdl”
   print fileName
   menu_func(“/temp”, “Open”, fileName)
   k=i+1
   dummy = icstat_from_partable(k, “/temp”, “MODEL PARAMETERS”)
   index=1
   ! Here is the loop to set the attribute values
   while index <= /temp/NUM_ATTRIBUTES
      value = val$(/temp/ATTRIBUTE_VALS[index-1])
      dummy = icstat_set_text_cell(k, index, value)
      index=index+1
   end while
   i=i+1
end while
```

Manually Executing PEL Commands

The following manual procedure is recommended only for transferring one or two rows of data to Statistics.

1. Open the IC-CAP model file from which you want to load data into IC-CAP Statistics.
2. Open the Setup from the **DUT-Setups** folder.
3. Choose the **Extract / Optimize** folder from the setup.
4. Choose the **Browse** button (far right). The Function Browser dialog box appears.
5. From the Function Groups field (left side), scroll down and select **Statistical Analysis**.

NOTE

Before transferring any data, use the icstat_set_param_column_labels function to write the parameter name in the Parameters spreadsheet.

6 Then select **icstat_from_partable** from the functions listed on the right side. Documentation on how to use the function (such as the arguments) is listed in the middle part of the dialog box. Choose the **Select** button at bottom. The Transform window is displayed again.

7 Fill in the arguments needed to complete the function. Then choose **Execute**. The data from either the model or DUT parameter will appear in the Statistics Parameters spreadsheet.

8 Repeat this procedure for each additional row of data you want send.

Opening a File in IC-CAP Statistics Format

To open a file already saved in the IC-CAP Statistics format, choose **File > Open**. The File Open dialog box appears. Enter the path and filename you want or select the file. IC-CAP Statistics files use the *.sdf* extension.

IC-CAP Statistics is shipped with several example .sdf files. To open an example file, choose **File > Examples**. The Examples Open dialog box appears. Select the example file you want.

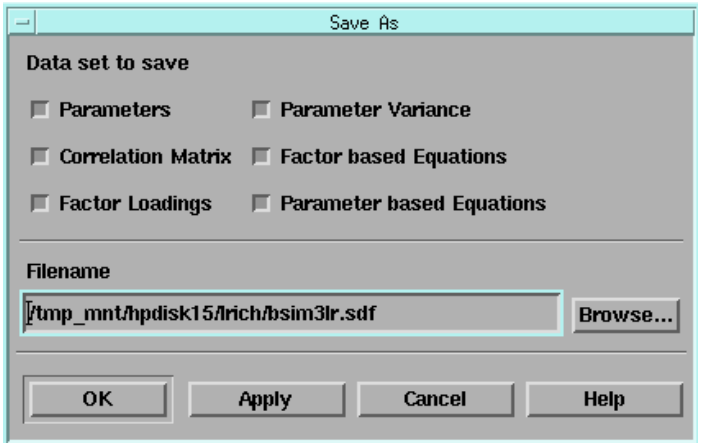
Manually Entering Data

Although this method would not typically be used, data can be entered (or modified) directly in the spreadsheet, attribute rows can be defined (see "Importing ASCII Data" on page 37, step 2) and the file saved. To manually enter data in a new file, choose **File > New**.

Saving Data

To save data:

Choose **File > Save As**, the Save As dialog box is displayed.



You can save portions of your file, such as only the parameters and correlation matrix, or all of the elements in your file. Select the check box next to the parts of your file you want to save. The default is set to save all portions.

If you want to keep the present file and save your data under a new filename, enter a new filename in the Filename field at bottom.

Exporting Data

To export a file, choose **File > Export**. A submenu is displayed. Choose the format you want. There are three file formats that you can export from IC-CAP Statistics:

- Text File. Exports the data currently contained in the Parameters folder (the raw data) in a row/column, space delimited (no commas or tabs) ASCII text format for use in other programs. If you choose this option, a dialog box is displayed. Enter a filename for the export file.
- SPICE Library File. Converts the data for desired samples into a form usable as a SPICE library. If you choose this option, a dialog box is displayed. Enter a filename for the export file.
- SPICE Equations. Converts the data contained in the two tables in the Equation folder into a form usable in a SPICE netlist. If you choose this option, a dialog box is displayed. You fill in the model file name, the sample row numbers you want to be saved, and a filename for the export file.

See Appendix A, “File Formats” for complete details on the structure of these file formats.

Printing

To print the contents of the current window or plot, choose **File > Print**, or click the Print icon. Choose **File > Print Set** to choose or alter printer setup parameters.

Closing and Exiting

Choose **File > Close** to close the file currently opened without leaving IC-CAP Statistics. Choose **File > Exit Statistics** to exit IC-CAP Statistics and return to the Main IC-CAP window.

Graphing Data

After you load the raw data, you may want to plot the data for a given parameter in one or more formats. Or, you may want to re-plot the data after performing a mathematical transformation or manipulation. To create a graph, choose **Graph** and then the type of plot you want (such as Histogram). IC-CAP Statistics contains the following type of plots:

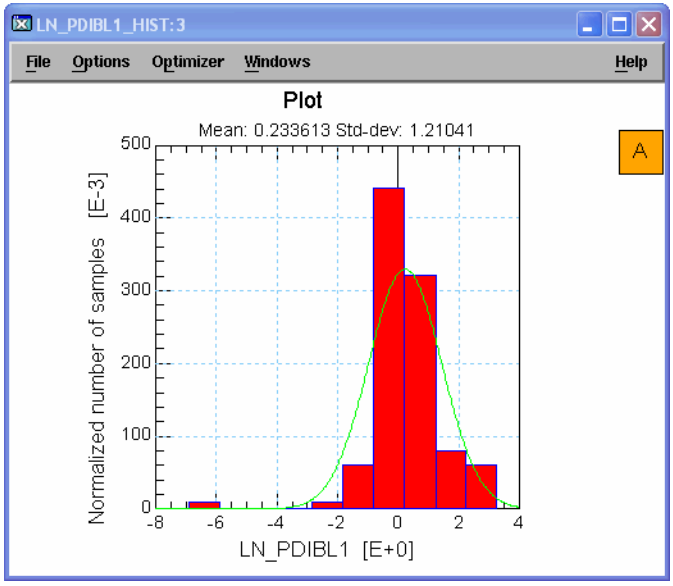


Figure 8 Histogram

IC-CAP's Statistic Analysis Histogram is normalized. The definition of the histogram normalization is as follows:

Assume:

Variable X indicated the width of the histogram bin,

Variable n indicated the number of the histogram bins.

Variable Y_i indicated the number of samples for the histogram bin with index of i

Variable $normalize\_Y_i$ indicated the normalized number of samples for the histogram bin with index of i

Then we have,

$$normalize\_Yi = (Yi) / \left(\sum_{i-1}^{n} (X \times Y_i) \right)$$

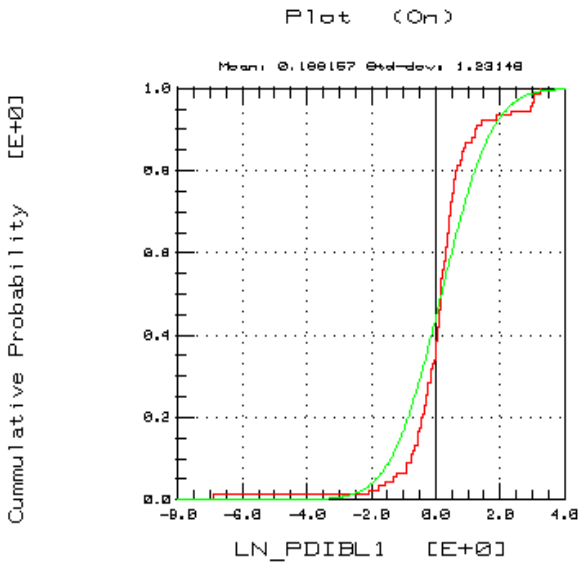


Figure 9 Cumulative Density Plot

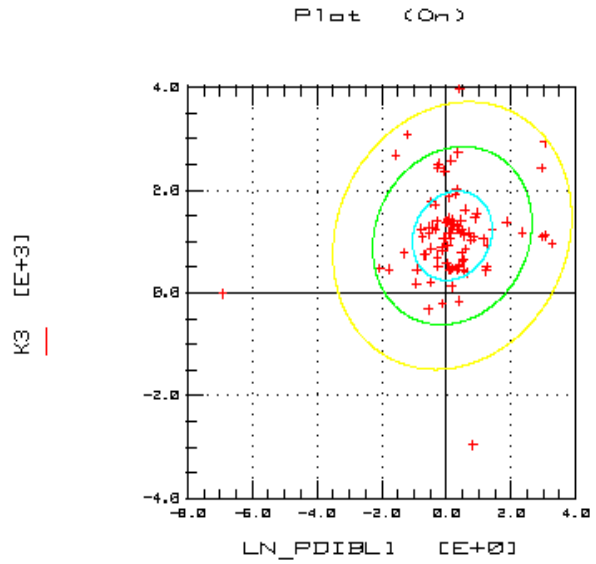


Figure 10 Scatter Plot

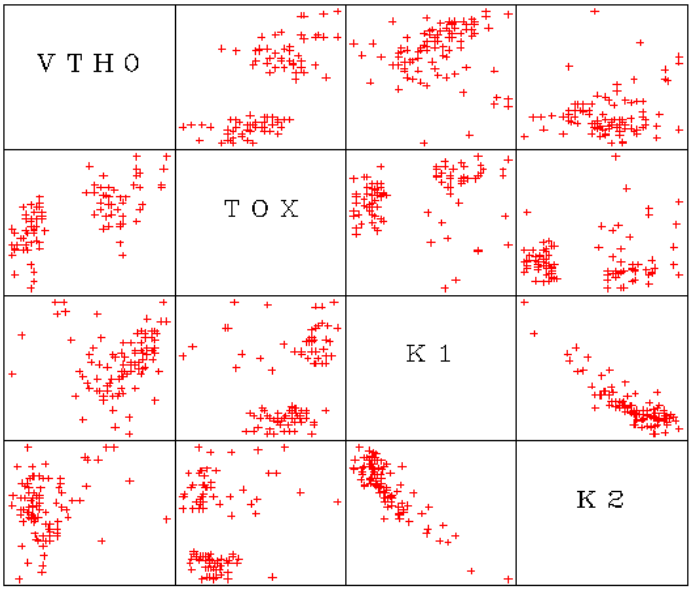


Figure 11 Scatter Plot Matrix

The scatter plot matrix contains multiple scatter plots, with each parameter plotted against every other parameter. For example, if your data had four parameters, P1, P2, P3, and P4, the matrix would contain 16 cells. First would be P1 against itself (only the label appears), then P1 vs. P2, P1 vs. P3, and then P1 vs. P4. The next row would contain P2 vs. P1, the P2 label, P2 vs. P3, and P2 vs. P4, and so on. If you have a large number of parameters, the plots will be quite small. Select the columns you want to plot and choose **Graph > Scatter Plot Matrix**.

More information on of each of these plots can be found in Chapter 4, “Data Visualization.”

Using the Edit Menu Commands

Refer to the table below for a list of features and functions found on the Edit menu.

Command	Function	Use
Swap	Reverses the order of any 2 rows or columns	Choose **Row** or **Column**. Enter the row or column number you want to swap from in the From field. Enter the row or column number where you want it to go in the To field. The two rows or columns will be exchanged.
Move	Moves one or more rows or columns to another place in the table	Choose **Row** or **Column**. Enter the row or column number you want to move from in the From field. Enter the row or column number where you want it to be moved to in the To field. The moved row(s) or column(s) will be inserted in the table.
Copy	Copies one or more rows or columns to another place in the table	Choose **Row** or **Column**. Enter the row or column number you want to copy from in the From field. Enter the number where you want it to go in the To field. The copied row(s) or column(s) will replace the data in one or more row(s) or column(s), as selected.
Insert	Inserts one or more rows or columns	Choose **Row** or **Column**. Select the row or column you want to insert more rows or columns before. Enter the number of rows or columns you want to insert.
Sort	Sorts the data in a column	Select the column in which you want to sort data. Choose **Ascending** or **Descending**. You can't sort row data.
Clear	Deletes the data but leaves the row or column	Select a row or column. Choose **Clear** and the data is erased.
Delete	Deletes the data along with the row or column	Select a row or column. Choose **Delete** and the row or column is removed from the table.

Command	Function	Use
Activate	Activates a row or column that has been deactivated	Select a row or column. Choose **Activate** to reactivate a row or column that has been deactivated. A deactivated row's or column's background color is gray.
Deactivate	Deactivates a row or column	Select a row or column. Choose **Deactivate** to deactivate a row or column. The row's or column's background color changes to gray. You can deactivate a row to eliminate an outlier sample or for other reasons.
Flip Active to Deactive	Activates rows (or columns) which are not active; deactivates rows (or columns) which are active	The Flip feature is provided to make it easier to reach the desired goal when more than half of the rows or columns need to have their active status reversed. In this case, just change the ones that are correct to the other status, and then Flip all the rows (or columns) at once.
Hide Attributes	Controls whether or not the Attribute columns are visible in the table	This functions as a toggle. When the Attribute columns are hidden, an option marker appears next to this menu item. Use this feature to view only the parameters.
Attribute to Parameter	Tells the program to change a column from Attribute to Parameter	Select a row or column. Choose **Attribute** to Parameter to change a column from an Attribute column to a Parameter column.
Parameter to Attribute	Tells the program to change a column from Parameter to Attribute	Select a row or column. Choose **Parameter** to Attribute to change a column from a Parameter column to an Attribute column.

NOTE

Several of the Edit menu functions have toolbar icons. See "Icons" on page 36.

Using the Format Menu Commands

Refer to the table below for a list of features and functions found on the Format menu.

Command	Function	Use
Row Height	Allows you to change the row height	Enter a number to increase or decrease the height of all rows. The number corresponds to the height of a row.
Column Width	Allows you to change the column width	Enter a number to increase or decrease the width of all columns. The number corresponds to the number of characters in a column.
Color	Allows you to change the color settings for the spreadsheet	You can select colors for seven areas: Table Background, Table Foreground, Selection Background, Deactivate, Parameter Attribute, Data Filter, and Attribute Filter. To change colors, click on the down arrow.

Building a Parametric Model

In Chapter 1 we went through the typical steps needed to take raw extraction data and develop a parametric statistical model. This exercise was done as a tutorial on IC-CAP Statistics. If you haven't read this section, please refer to the procedure, which begins with "Transform Data" on page 15. In this chapter, we will provide expanded information on some of the steps related to building a statistical model.

The general steps needed to build a parametric statistical model are:

1. Measure and extract model parameters
2. Start IC-CAP Statistics and import data
3. Transform distributions to Gaussian
4. Eliminate outlier data
5. Perform correlation analysis
6. Perform factor or principal component analysis
7. Generate model equations
8. Generate models
9. Test models

Transforming Distributions / Eliminating Outliers

As stated in Chapter 1, one of the key assumptions made by multivariate techniques such as Factor Analysis is that the data set to be analyzed is a joint Gaussian distribution. If the data is not joint Gaussian, the model generated from the analysis may not accurately reproduce the measured density.

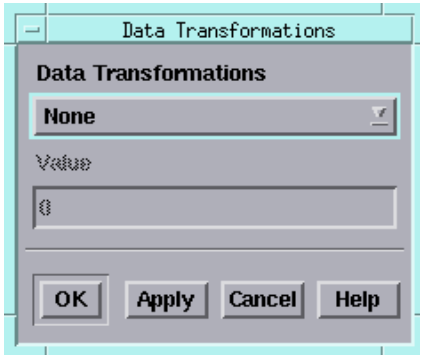
Therefore steps 3 and 4 in our procedure to build a statistical model can be done in any order. The elimination of outlier data can be done before or after data transformation. Outliers can be eliminated manually or through automatic data filtering.

Data Transformation

You have to decide which data columns need to be transformed. Generally, only some of the columns contain key parameters. Some columns may already be Gaussian. You can quickly plot the data to see if it is Gaussian.

To transform data, select a column and then choose **Data > Data Transformations**, a dialog box appears with the mathematical transforms available, as shown in the table below. To show the transformation, in some cases the parameter name is appended, also as shown in the table. For example, parameter TOX would become LN_TOX in a natural log transform.

Transformation	Label Appended to Name
None	
Exponential	EXP
Natural Log	LN
Square Root	SQRT
Square	SQR
Constant Value	
Mean	



To select a transformation type, click the drop-down list button and select the type you want. If you select Constant Value, the Value field (center) becomes active, and you can enter the constant you want to substitute for your data.

Pairs of the transforms work as opposites. If you transform your data using Natural Log and want to undo the change, choose **Exponential**. Similarly, Square and Square Root work the same way. To undo either Constant Value or Mean, however, choose **Undo Constant / Mean**. More complex transformation types are available in the Main IC-CAP program. You can export data to be transformed and then import it back into Statistics.

Manually Eliminating Outliers

You can manually eliminate outliers by deactivating any row you believe contains outlier data. Select the row that contains the sample you want to deactivate and choose **Edit > Deactivate** or click the icon on the toolbar.

Automatic Data Filtering

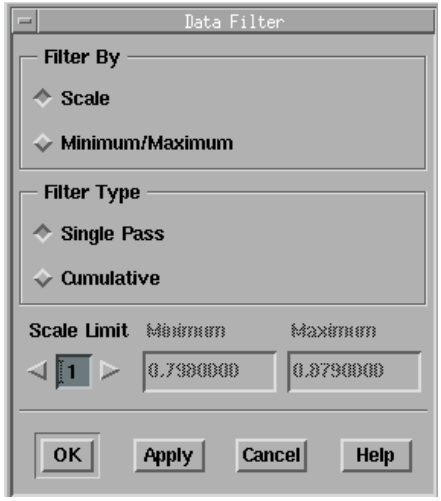
You can automatically filter your data to eliminate outliers.

IC-CAP Statistics can filter data based on minimum/maximum values or by a scale value. Scale is defined as the median absolute deviation (MAD) divided by a constant (approximately 0.6745). This standardizes MAD in order to make the scale estimate consistent with the standard deviation of a normal distribution. The greater the scale value, the further from the median the filtering occurs. If too much data is eliminated using, for example, a scale value of 4, you can retrieve some of the data by repeating a single-pass filtering operation with a scale value of 3.

To use the Data Filter feature:

1 Select a column you want to filter.

2 Choose **Data > Data Filter**. A dialog box is displayed.



3 To filter by a scale value, choose the **Scale** option button. Change the Scale Limit (near bottom) to the number you want by clicking the right arrow or typing in the field.

4 To filter by minimum/maximum values, choose the **Minimum / Maximum** option button and then enter your minimum and maximum values in the lower portion of the dialog box.

5 The Filter Type field (center) allows you to filter samples by a single pass or cumulatively. Choosing Cumulative means that rows eliminated during one filter pass remain deactivated during the subsequent filtering passes, regardless of the filtering criteria of the subsequent passes. Single Pass filtering deactivates rows based on only one filtering operation.

6 When done, choose **OK**.

The data is filtered based on the parameters you chose. Note that eliminated rows are highlighted by a color change that indicates they have been deactivated.

To undo Data Filtering, choose **Data > Undo Data Filtering**.

Repeat Data Transformation and Outlier Elimination for Other Columns

Repeat the above steps for each column of data you want to make Gaussian through data transformation, manual outlier elimination, or data filtering.

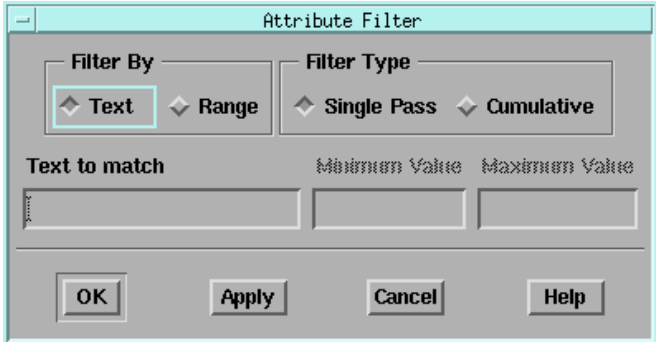
Attribute Filter Options

Besides Data Filter and Undo Data Filtering, there are the companion choices of Attribute Filter and Undo Attribute Filtering. These choices work in the same manner as parameter data filtering but apply only to attribute columns.

To use the Attribute Filter feature:

1 Select an attribute column you want to filter.

2 Choose **Data > Attribute Filter**. A dialog box is displayed.



3 If you want to filter by a character string that matches an attribute label (such as “sample1”), choose **Text** in the Filter By field, and enter the text in the Text to match field.

4 If you want to filter by a range of values (such as sample23 to sample29), choose **Range** in the Filter By field and enter the minimum and maximum values in the right-center part of the box.

5 The Filter Type field allows you to filter samples by a single pass or cumulatively. Choosing Cumulative means that rows eliminated during one filter pass remain deactivated during the subsequent filtering passes, regardless of the filtering

criteria of the subsequent passes. Single Pass filtering deactivates rows based on only one filtering operation.

6 When done, choose **OK**.

Performing Correlation Analysis

As stated in Chapter 1, correlation analysis provides a numerical measure of the amount of variation in one variable that is attributable to another variable. When an increase in the value of a variable is associated with an increase in the value of the other variable, the correlation is positive. When the increase is associated with a decrease the correlation is negative.

Correlation analysis is always performed before proceeding to factor analysis and the data used consists of all the rows in the spreadsheet that have not been filtered, deactivated, or deleted.

To perform correlation analysis, choose **Analysis > Correlation Analysis**. The Statistics window changes so that the Correlation Matrix folder is displayed. (If you want to go back to the parameter data before correlation analysis was performed, choose the folder tab labeled Parameters.)

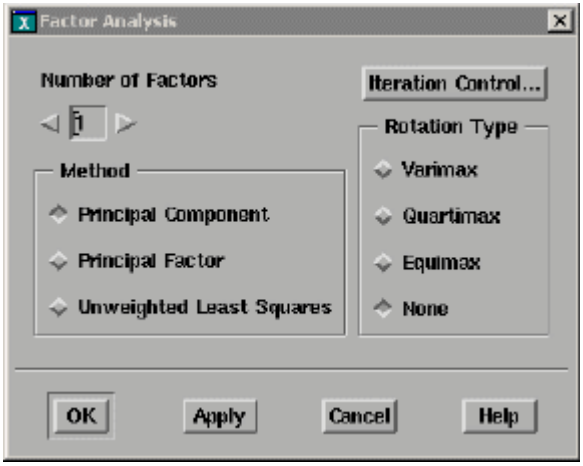
Parameters | Correlation Matrix | Factor Loadings | Equations

	C1	C2	C3	C4	C5	C6	C7	C8
R1	Parameters	VTH0	LN_PDIBL1	TOX	K1	K2	K3	UA
R2	VTH0	1	0.02301813	0.68508693	0.1079446	0.19678496	-0.03376760	0.30100217
R3	LN_PDIBL1	0.02301813	1	.000422942	0.17145006	-0.1953400	-0.0586468	-0.2158907
R4	TOX	0.68508693	.000422942	1	0.23359816	0.10937899	-0.1279979	0.26975498
R5	K1	0.1079446	0.17145006	0.23359816	1	-0.8677680	-0.1814875	-0.6277974
R6	K2	0.19678496	-0.1953400	0.10937899	-0.8677680	1	0.1001625	0.66622811
R7	K3	-0.03376760	-0.0586468	-0.1279979	-0.1814875	0.1001625	1	0.10394578
R8	UA	0.30100217	-0.2158907	0.26975498	-0.6277974	0.66622811	0.10394578	1
R9	UB	-0.2328701	0.16585016	-0.2162033	0.22932561	-0.1667067	0.01630693	-0.7420965
R10	UC	0.1901659	-0.1209297	0.21039859	-0.6108630	0.80073389	0.10678869	0.46041935
R11	VSAT	0.03458770	0.75046302	0.02422911	0.29249677	-0.3219798	-0.04265632	-0.3064381
R12	VOFF	0.01443082	0.03026316	-0.0514615	-0.1132410	0.03670589	0.31890621	0.02077750
R13	DL	-0.08582748	-0.02851296	0.04693041	0.30660527	-0.3689783	0.04228276	-0.2096393
R14	DW	-0.04627482	0.18672065	-0.1005797	0.63862369	-0.7076457	-0.01348336	-0.5689850

The Correlation Matrix displays the same parameters down the rows and across the columns. The correlation coefficients for any two parameters are displayed where the rows and columns intersect. In the preceding example, the cell R4 C2 has a value of about 0.69, which shows moderate to strong correlation between parameters TOX and VTH0.

Perform Factor Analysis

Now that the correlation matrix is defined, the next step is to perform factor analysis. To perform factor analysis, choose **Analysis > Factor Analysis**, a dialog box is displayed.



You choose the method of factor analysis from three choices:

- Principal Component –The principal component model of factor analysis. Direct calculation; no iterations.
- Principal Factor–Principal factor analysis. Direct calculation; no iterations.
- Unweighted Least Squares–A method of factor analysis using an iterative process. This methods is also knows as minres or the minimum residual method.

Which of these three methods you should use depends on the model you are building. If your goal is to generate a set of equations with factors as the independent variables, then the

principal component method, without factor rotation, is preferred. If factor analysis is being used as an interim step in building a regression model, then principal factor analysis or unweighted least squares should be used with one of the three rotation types.

A detailed description of each of these three methods can be found in Chapter 3, “Data Analysis.”

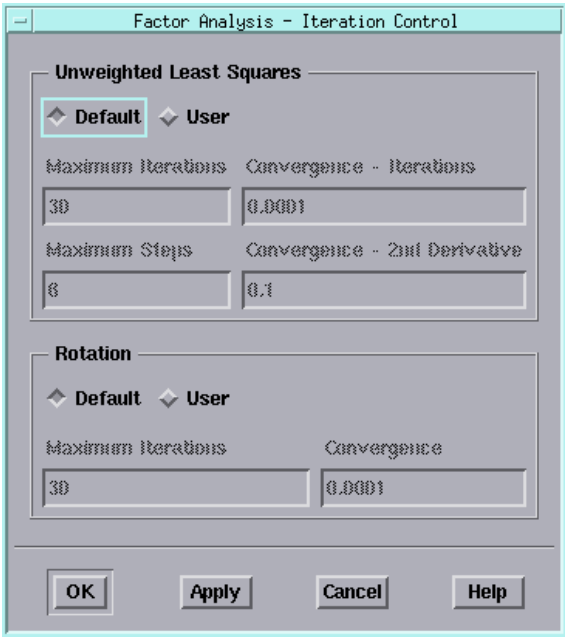
You choose a starting figure for the Number of Factors you want to be found in your analysis. After you see the results, which correspond to the percent variation that can be explained by this number of factors, you can increase or decrease the number and repeat the analysis.

The Rotation Type field has four choices.

- Varimax
- Quartimax
- Equimax
- None (default)

For more information on factor rotation, refer to “Factor Rotation” on page 98.

The Iteration Control button (upper right) controls fine tuning for the Unweighted Least Squares and Rotation functions.



Therefore, if you select the Unweighted Least Squares method and click **Iteration Control**, a dialog box is displayed. You accept the default values (as shown) or choose **User** and type in the values for the following four parameters, as desired:

- Maximum Iteration. The maximum number of iterations that will be attempted before arriving at a solution.
- Convergence - Iterations. When the relative change in the criterion function is less than this number from one iteration to the next, convergence is assumed.
- Maximum Steps. The maximum number of step halvings allowed during any one iteration.
- Convergence - 2nd Derivative. When the largest relative change in the unique standard deviation vector is less than this number, exact 2nd derivatives are used.

Likewise, if you select a rotation type other than None in the Factor Analysis dialog box, and click **Iteration Control**, the same dialog box is displayed. For Rotation, you can accept the default values (as shown) or choose **User** and type in the values for the following two parameters, as desired:

- Maximum Iterations. The maximum number of iterations that will be attempted before arriving at a solution.
- Convergence. When the relative change in the criterion function is less than this number from one iteration to the next, convergence is assumed.

When done, choose **OK** to return to the Factor Analysis dialog box and OK again perform the factor analysis. The screen changes to display or update the Factor Loadings folder.

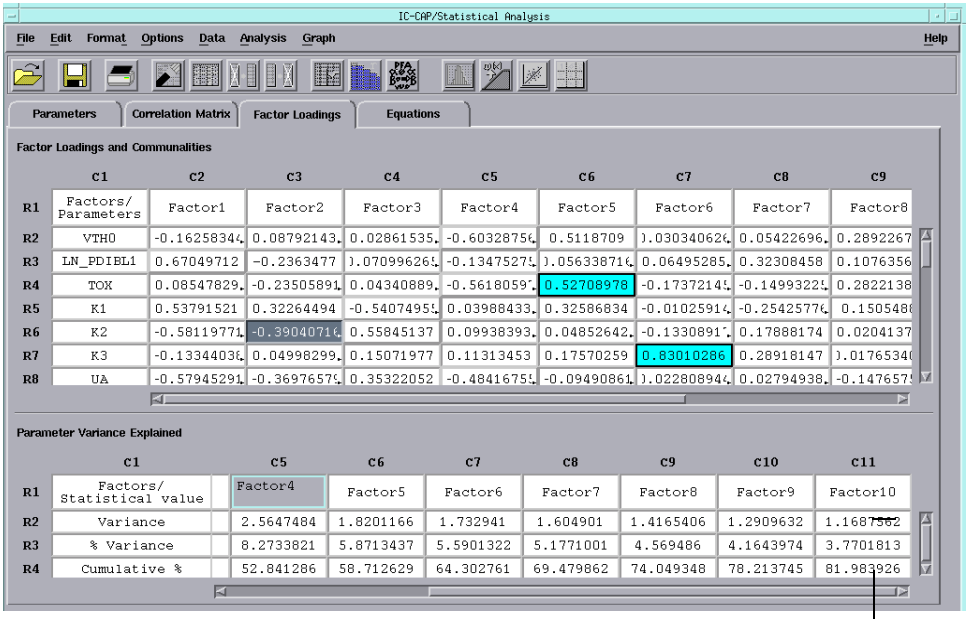


Figure 12 Factor Analysis Folder

Factor Loadings Folder

Two tables are generated in the Factor Loadings folder. The first table contains the factor loadings, which represent the loadings (correlations) that relate each model parameter to each of the derived factors. This table also has a column called

Communality, always displayed at the far right part of the table. This field shows the variance explained by all of the factors for a single parameter.

The top portion of the Factor Loading folder displays the data in a color-coded format. Factor Group data, one group per row, is displayed in a red font. Dominant Parameter data, one dominant parameter per column, is displayed with a blue background. These terms are defined as follows:

Factor Group. The factor group shows, for a given parameter, which factor the parameter is most highly correlated with.

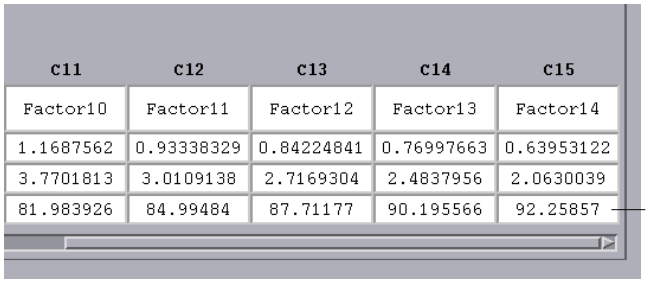
Dominant Parameter. The dominant parameter is the one parameter in the column that has the highest loading (correlation).

The second table has three fields:

- Variance. Presents a summary of the variances associated with each factor. For example, a variance of 3.45 indicates that the factor accounts for as much variance in the data collection as would 3.45 variables, on average.
- % Variance. Shows how much of the variance of all the parameters is explained by a single factor.
- Cumulative %. Shows how much of the variance of all the parameters is explained cumulatively by from one to all of the factors. That is, as you move left to right in the table, the percentage increases as more and more factors are included.

Note that the cumulative percent for the example shown in Figure 12, which was analyzed for 10 factors, is about 82%. See the lower right cell in the figure. This means that if only 10 factors were used to make a statistical model from this data, the model would explain 82% of the variance compared to using all of the parameters/factors.

With 14 factors, the cumulative percent is about 92%, as shown in the lower right cell below. You have to decide how high a figure is acceptable for your work.



Factor Loadings Options on the Analysis Menu

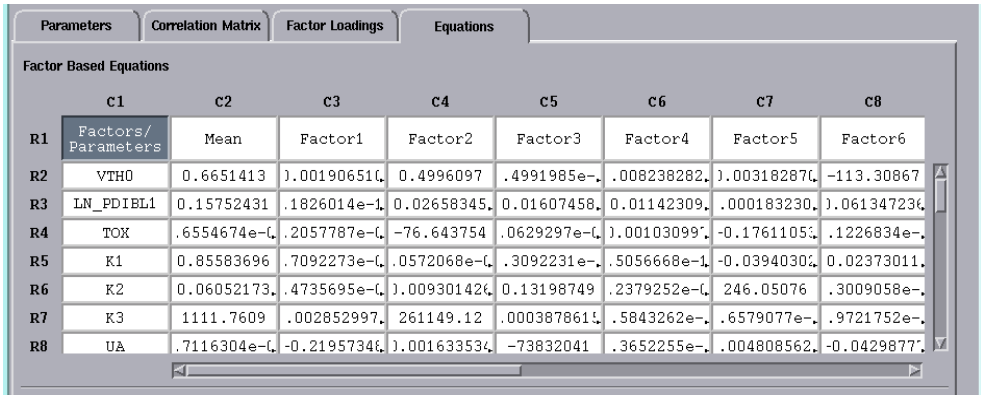
The last four options on the Analysis menu are specifically for the Factor Loadings folder, and are only active when there is factor analysis data present. The options are as follows:

- Change Dominant Parameter. Select a cell that you want to become the dominant parameter for that column and then choose **Analysis > Change Dominant Parameter**. The program replaces the calculated dominant parameter with the one you chose. To revert back to the calculated dominant parameter, choose **Analysis > Default Dominant / Grouping**.
- Change Factor Group. Select a cell that you want to become the factor group for that parameter and then choose **Analysis > Change Factor Group**. The program replaces the calculated factor group with the one you chose. To revert back to the calculated factor group, choose **Analysis > Default Dominant / Grouping**.
- Default Dominant/Grouping. If you changed either the dominant parameter or the factor group, and you want to revert back to the one calculated by the program, choose **Analysis > Default Dominant / Grouping**.
- Factor/Parameter Groups. Choose **Analysis > Factor / Parameter Groups** to display a summary table, which shows the dominant parameter for each factor and its value, as well as the factor group data. The dominant parameter data is shown first. For a given factor, all parameters that belong to that group are shown.

Generate Equations

The next step is to generate equations from the factor analysis. IC-CAP Statistics supplies the equation coefficients that you use to build your SPICE model.

To generate equations, choose **Analysis > Generate Equations**. A submenu with the two choices by which the equations will be generated appears to the right. Choose either **Factors** or **Dominant Parameters**. The screen changes to display the Equations folder, which, has two tables. The upper table displays factor based equations, the lower table displays dominant parameter based equations.



	C1	C2	C3	C4	C5	C6	C7	C8
R1	Factors/ Parameters	Mean	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
R2	VTH0	0.6651413	).00190651(	0.4996097	.4991985e-.	.008238282.	).00318287(	-113.30867
R3	LN_PDIBL1	0.15752431	.1826014e-1	0.02658345.	0.01607458.	0.01142309.	.000183230.	).061347236
R4	TOX	.6554674e-(	.2057787e-(	-76.643754	.0629297e-(	).00103099'	-0.17611053	.1226834e-.
R5	K1	0.85583696	.7092273e-(	.0572068e-(	.3092231e-.	.5056668e-1	-0.03940304	0.02373011.
R6	K2	0.06052173.	.4735695e-(	).00930142(	0.13198749	.2379252e-(	246.05076	.3009058e-.
R7	K3	1111.7609	.002852997.	261149.12	.000387861!	.5843262e-.	.6579077e-.	.9721752e-.
R8	UA	.7116304e-(	-0.21957348	).00163353(	-73832041	.3652255e-.	.004808562.	-0.0429877'

Figure 13 Equations Folder Showing Factor Based Equations

Generate a Parametric Model

Now that the equation coefficients are generated, you can build a variety of statistical models, or save the data in a SPICE equations format for use in circuit simulations. You can test your model, based on a reduced set of parameters, against the raw data to see how well it performs. At this point, IC-CAP Statistics has been designed for flexibility to work with your process.

Parametric analysis can be performed from either:

- Factor Equations

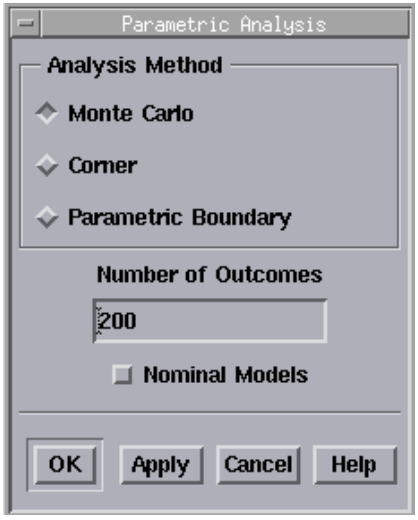
- Regression Equations (see "Parametric Analysis Results Window" on page 66)

Factor Equations

To build the parametric model from factor equations, choose **Analysis > Parametric Analysis > Factor Equations**, a dialog box appears with three choices:

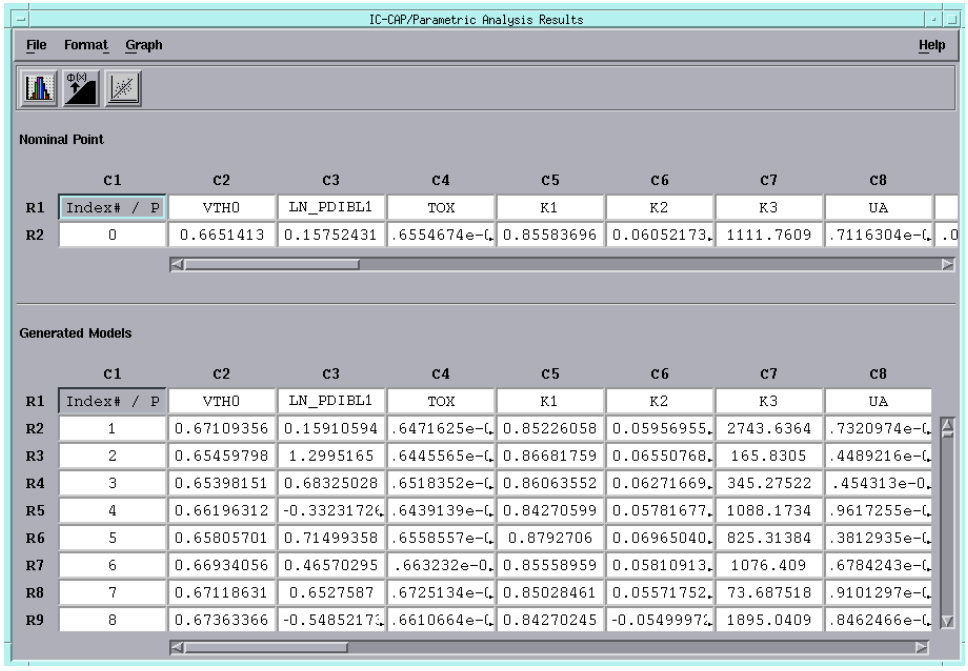
- Monte Carlo
- Corner
- Parametric Boundary

The advantages and disadvantages of these methods is described beginning with "Parametric Analysis Results Window" on page 66.



If you select Monte Carlo, you fill in the number of outcomes you want in the field below. If you select Corner or Parametric Boundary, you fill in the number of +/- sigmas you want. When done, choose **OK**. The Parametric Analysis Results window is displayed.

Parametric Analysis Results Window



	C1	C2	C3	C4	C5	C6	C7	C8
R1	Index# / P	VTH0	LN_PDIBL1	TOX	K1	K2	K3	UA
R2	0	0.6651413	0.15752431	.6554674e-0.	0.85583696	0.06052173.	1111.7609	.7116304e-0.

	C1	C2	C3	C4	C5	C6	C7	C8
R1	Index# / P	VTH0	LN_PDIBL1	TOX	K1	K2	K3	UA
R2	1	0.67109356	0.15910594	.6471625e-0.	0.85226058	0.05956955.	2743.6364	.7320974e-0.
R3	2	0.65459798	1.2995165	.6445565e-0.	0.86681759	0.06550768.	165.8305	.4489216e-0.
R4	3	0.65398151	0.68325028	.6518352e-0.	0.86063552	0.06271669.	345.27522	.454313e-0.
R5	4	0.66196312	-0.3323172(	.6439139e-0.	0.84270599	0.05781677.	1088.1734	.9617255e-0.
R6	5	0.65805701	0.71499358	.6558557e-0.	0.8792706	0.06965040.	825.31384	.3812935e-0.
R7	6	0.66934056	0.46570295	.663232e-0.	0.85558959	0.05810913.	1076.409	.6784243e-0.
R8	7	0.67118631	0.6527587	.6725134e-0.	0.85028461	0.05571752.	73.687518	.9101297e-0.
R9	8	0.67363366	-0.5485217(	.6610664e-0.	0.84270245	-0.0549997(	1895.0409	.8462466e-0.

Figure 14 Parametric Analysis Results Window

When you perform a parametric analysis, the results are displayed in a window as shown in the previous figure. The upper spreadsheet displays the nominal point of a corner or parametric boundary analysis. The lower spreadsheet contains the rows for the samples (or Monte Carlo outcomes) and columns for the parameters. This window also has menu functions similar to those available in the Parameters folder, and include:

- File menu. Save As, Export, Print, and Close. For more information, refer to “Saving Data” on page 43, “Exporting Data” on page 44, “Printing” on page 44, and “Closing and Exiting” on page 44.

- Format menu. Row Height and Column Width. For more information, refer to "Changing Row Height or Column Width" on page 34.
- Graph menu. Histogram, Cumulative Plot, Scatter Plot, and Close All. For more information, refer to "Graphing Data" on page 45.
- Toolbar Icons. Histogram, Cumulative Plot, and Scatter Plot. For more information, refer to "Icons" on page 36.

Monte Carlo Analysis

Monte Carlo analysis can be used to randomly generate outcomes for the independent parameters in a statistical model (factors or dominant parameters). These random outcomes are then substituted in the model equations to generate a set of correlated models whose statistics (such as means, variances, and correlations) resemble those of the original extracted parameter set. Monte Carlo analysis is very useful for validating a statistical model against the original extracted parameters.

To build Monte Carlo models, choose **Analysis > Parametric Analysis**. A dialog box appears. Select **Monte Carlo** and fill in the number of outcomes you want and choose **OK**. The results of the Monte Carlo analysis are displayed in a spreadsheet window similar to the Parameters folder. The number of rows are equal to the number of Monte Carlo outcomes, and the columns correspond to the model parameters.

You can select a column of data in the Monte Carlo spreadsheet and plot a histogram. Compare this plot (made from synthesized Monte Carlo data) with a plot made from the raw data. Do they agree? (To go back to the raw data, click the **Parameters** folder tab. Then select a column and choose **Graph > Histogram**.)

To test the Monte Carlo models, you have to simulate each model and compare the results.

Corner Model Analysis

Corner models are used to synthesize worst-case candidate models for a given model parameter set. This method computes the dependent parameters of a data set to arrive at a set of correlated parameters. Corner models are generated by taking all the +/- n-sigma limits of the independent factors (or parameters). The dependent parameters are then computed by substituting these values into the derived equations. This yields a set of correlated parameters.

To build corner models, choose **Analysis > Parametric Analysis**. A dialog box appears. Select **Corner** and enter the number of +/- sigmas (standard deviations) you want. The results are displayed in a spreadsheet window similar to the Parameters folder. The number of rows are equal to 2^n, where n is equal to the number of factors you chose during Factor Analysis. For example, if you chose 10 factors, the table will contain 1024 (2^{10}) rows. The columns correspond to the parameters. Each row contains one corner model.

NOTE To keep the data set manageable, it is recommended that the number of factors you use to generate corner models be limited to 10.

To test the corner models, you have to simulate each model and compare the results.

Parametric Boundary Analysis

NOTE Do not confuse the parametric boundary models, described in this section, with the non-parametric boundary models discussed in the next section.

Traditional worst-case modeling uses corner models. Corner modeling chooses a set of extreme values at the outside of the real multi-dimensional probability density function (PDF) and requires 2^n simulations for an n-dimensional problem.

In contrast, parametric boundary modeling chooses those extreme values at the boundary of the real multi-dimensional PDF, and only needs 2n simulations for an n-dimensional problem. For example, if you chose 10 factors, the number of simulations would be 20 compared to 1024 using corner models.

Boundary modeling circumvents the typically large simulation times required for a Monte Carlo analysis. However, its results are only useful in assessing worst-case behavior, and it does not provide a yield estimate.

To build parametric boundary models, choose **Analysis > Parametric Analysis**. A dialog box appears. Select **Parametric Boundary** and enter the number of +/- sigmas (standard deviations) you want. The results are displayed in a spreadsheet window similar to the Parameters folder. The number of rows are equal to 2n, where n is equal to the number of factors you chose during Factor Analysis. Each row contains one boundary model. The columns correspond to the parameters.

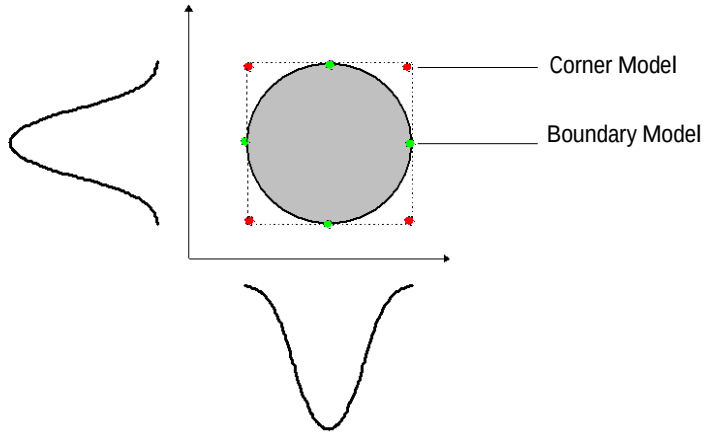


Figure 15 Parametric Boundary Models vs. Corner Models In Two Dimensions

The previous figure compares corner models to boundary models in two dimensions. Using boundary models, you can select four evenly spaced models that lie on an n-sigma boundary of the joint probability density function.

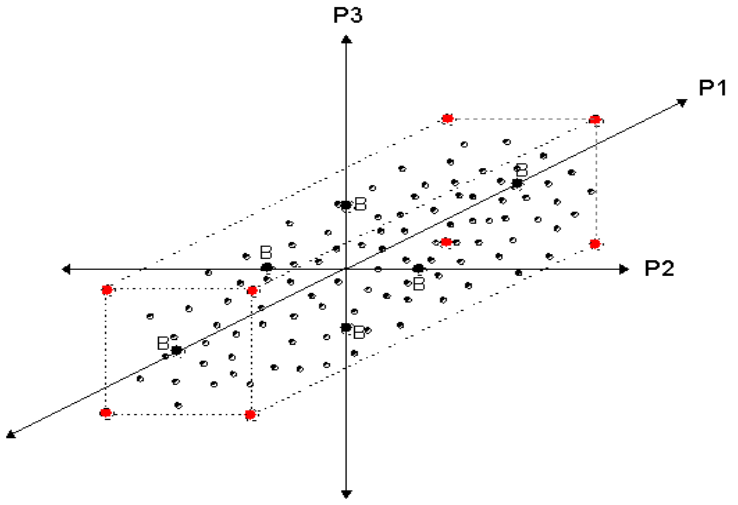


Figure 16 Parametric Boundary Models (Labeled “B”) in Higher Dimensions

The previous figure shows a graphical representation of parametric boundary models in three dimensions, for parameters P1, P2, and P3. Note that each pair of boundary models lies on the parameter axis (orthogonal space). The values of the parameters on the other axes are at their nominal values. For a three-dimensional probability density function (PDF), you need six boundary models.

Regression Analysis

Unlike correlation analysis, which provides a summary coefficient of the extent of relationship between two variables, regression analysis provides an equation describing the nature of the relationship between two variables.

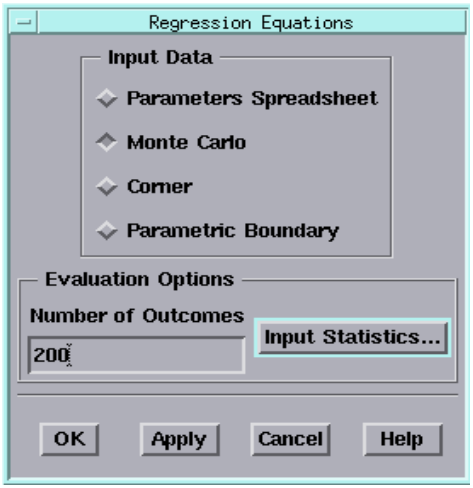
Regression Equations is one of two methods of performing a parametric analysis; the other method is Factor Equations, refer to "Perform Factor Analysis" on page 58. For more information on Regression Analysis, see "Regression Analysis" on page 100.

NOTE

The regression equations method can *only* be done using Dominant Parameters. If you generated your equations using factors, choose **Analysis > Generate Equations > Dominant Parameters** to generate equations using dominant parameters.

To evaluate the regression equations, choose **Analysis > Parametric Analysis > Regression Equations**, a dialog box appears with four choices:

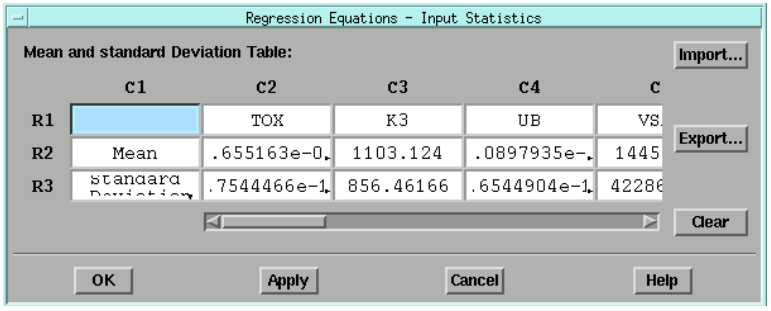
- Parameters Spreadsheet
- Monte Carlo
- Corner
- Parametric Boundary



These four choices all evaluate the same regression equations. However , each option uses different input values for the dominant parameters (independent variables) in the regression equations.

To use the values of the dominant parameters contained in the Parameters spreadsheet as the input, select **Parameters Spreadsheet**. These values will be substituted into the regression equations and values of the dependent variables will be computed. Upon completion of analysis, the Parametric Analysis Results window will be displayed with three columns of data for each parameter: the raw or measured data (this is an exact copy of the data in the Parameters spreadsheet), the simulated data (computed from the regression equations), and the residual (difference) between the two. The presence of these three data types (measured, modeled and residual) in a single spreadsheet window makes model validation very convenient. For example, measured versus modeled results can be compared with one another or with the residual error, either graphically or directly from the tabular listing.

The remaining three choices, Monte Carlo, Corner and Parametric Boundary, all require you to provide information on the statistics of the dominant parameters in the regression equations. This information consists of means and standard deviations for every dominant parameter in the regression equations. To enter this information, choose the **Input Statistics** button to display a dialog box, as shown below.



This dialog box displays the dominant parameters as well as default values for the mean and standard deviation of those parameters. The default values are computed from the data in the Parameters folder. From this dialog box you can change this data by typing in the cells, or choose the **Import** button to import this data from a text file.

When you are ready to build your model, and you select Monte Carlo, fill in the number of outcomes you want in the field below. If you select Corner or Parametric Boundary, you fill in the number of +/- sigmas you want.

When done, choose **OK**. The Parametric Analysis Results window appears, which shows the results based on the regression equations method using the data you chose. See "Parametric Analysis Results Window" on page 66 for details about the menu and toolbar options in this window.

Building a Non-Parametric Model

IC-CAP Statistics contains an exclusive feature called Non-Parametric Boundary Analysis. Unlike other statistical analysis tools, which only handle Gaussian distributions, non-parametric boundary analysis uses a new technique to handle arbitrary data distributions, Gaussian or non-Gaussian, and selects nominal and boundary models. Non-parametric analysis works effectively on data from any stochastic process. The data can be unimodal or multimodal, residing in a single cluster or multiple clusters, with no dimensional limitations.

NOTE Do not confuse the non-parametric boundary models, described in this section, with the parametric boundary models discussed earlier.

The non-parametric analysis starts by selecting a nominal point and choosing boundary points from an arbitrary data collection. The nominal point is the point that has the highest estimated local density and the boundary points are those that have an estimated local density greater than some threshold value. The threshold value is determined by specifying the enclosure percentage that is, under certain circumstances, related to the yield.

To use Non-Parametric Boundary Modeling, choose **Analysis > Non-Parametric Analysis**. A dialog box is displayed; the next section explains its use.

Non-Parametric Boundary Analysis Example

In Chapter 1, we performed a step-by-step tutorial using IC-CAP Statistics for parametric analysis. Now we will use the same example file to learn how to use Non-Parametric Boundary Analysis.

1 If it's not already loaded, load the example file bsim3.sdf: Choose **File > Examples**. In the dialog box that appears, select

bsim3.sdf from the list of files that appears and choose **OK**. The data is loaded into the spreadsheet.

2 Optional step (not needed for this example). If your data contains outliers, you should manually deactivate those samples or use automatic data filtering to do the same. Refer to "Manually Eliminating Outliers" on page 54.

3 Optional step (not needed for this example). If a given parameter naturally ranges over more than one decade of values (e.g., 10-100), you should choose **Data > Data Transformations**, and select **Natural Log** from the dialog box that appears, to transform your data before proceeding.

4 Choose **Analysis > Non-Parametric Analysis**. The Non-Parametric Analysis dialog box is displayed.

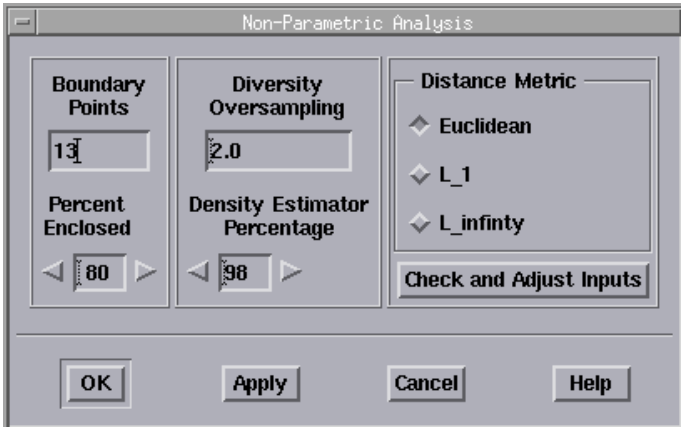
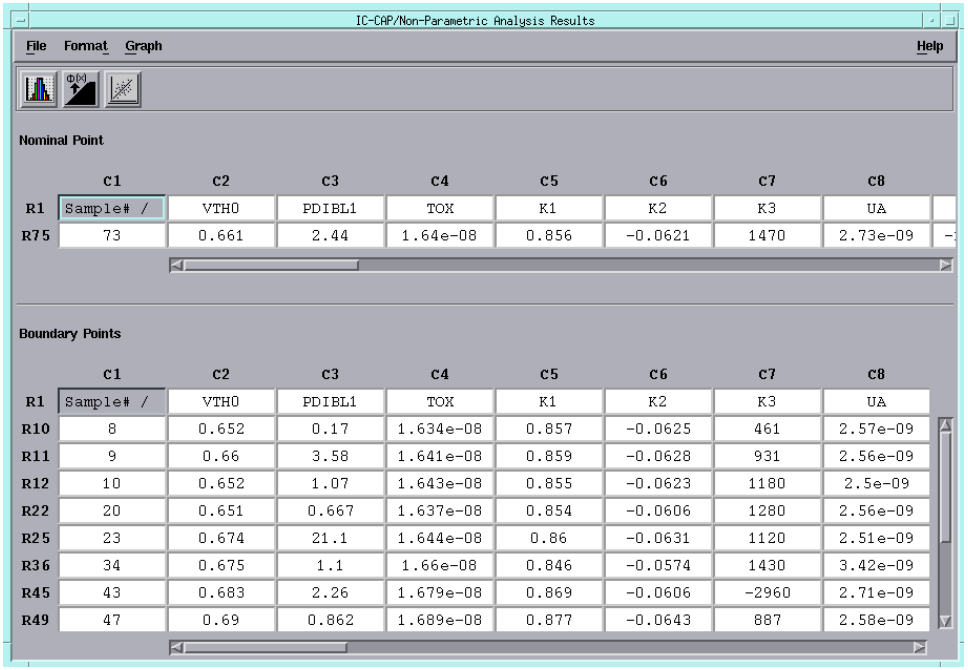


Figure 17 Non-Parametric Analysis Dialog Box

The available controls are Boundary Points, Percent Enclosed, Diversity Oversampling, Density Estimator Percentage, and Distance Metric. These controls will be described fully after the example.

5 For this example, accept the default values. Choose **OK**. An analysis on the full 31 dimension data set is performed, as shown in the following figure.



Nominal Point

	C1	C2	C3	C4	C5	C6	C7	C8
R1	Sample# /	VTH0	PDIBL1	TOX	K1	K2	K3	UA
R75	73	0.661	2.44	1.64e-08	0.856	-0.0621	1470	2.73e-09

Boundary Points

	C1	C2	C3	C4	C5	C6	C7	C8
R1	Sample# /	VTH0	PDIBL1	TOX	K1	K2	K3	UA
R10	8	0.652	0.17	1.634e-08	0.857	-0.0625	461	2.57e-09
R11	9	0.66	3.58	1.641e-08	0.859	-0.0628	931	2.56e-09
R12	10	0.652	1.07	1.643e-08	0.855	-0.0623	1180	2.5e-09
R22	20	0.651	0.667	1.637e-08	0.854	-0.0606	1280	2.56e-09
R25	23	0.674	21.1	1.644e-08	0.86	-0.0631	1120	2.51e-09
R36	34	0.675	1.1	1.66e-08	0.846	-0.0574	1430	3.42e-09
R45	43	0.683	2.26	1.679e-08	0.869	-0.0606	-2960	2.71e-09
R49	47	0.69	0.862	1.689e-08	0.877	-0.0643	887	2.58e-09

Figure 18 The Non-Parametric Analysis Results Window

6 The Non-Parametric Analysis Results window appears, which has two spreadsheets, one for the nominal point and the other for boundary points. The menu and toolbar options in this window are the same those described in "Parametric Analysis Results Window" on page 66.

Since we chose the default of 13 Boundary Models for this data set, 13 worst-case models are displayed in the rows of the lower spreadsheet.

The first row of each spreadsheet contains parameter labels. The row labels are numbered to correspond to the raw data's row number. The first column contains indexes used in plotting. The remainder of each of the spreadsheets contains data specifying each sample's location.

7 Next, we will perform a new non-parametric boundary analysis, but for only two parameters, VTH0 and TOX. To do

this we will deactivate all parameter columns except these two. Select the **IC-CAP / Statistical Analysis** window, then the **Parameters** folder.

8 Use the horizontal scroll bar to go to the far right column, and select column **C37**. Scroll back to column C10 and while holding down the Shift key, select column **C10**.

9 Now hold down the Ctrl key and select column **C8**. This selects columns C8 and C10 through C37.

10 Select **Edit > Deactivate**.

We are deactivating most of the parameters for this example so we can plot a two-parameter analysis and look at a boundary. With higher dimensional analyses, where all or most of your model parameters are used, there is no way to visually examine the boundary. However, the program calculates the boundaries and provides you with a selection of worst-case models to simulate.

11 Select **Analysis > Non-Parametric Analysis** again and enter the following data in the Non-Parametric Analysis dialog box:

Boundary Points: 9

Percent Enclosed: 80%

Diversity Oversampling: 3.0

Density Estimator Percentage: 45

We changed the Density Estimator Percentage because the distribution is bimodal, with approximately half of the distribution being in each mode. The previous value of 98 is valid for unimodal, Gaussian distributions only.

12 Choose **OK**. The Non-Parametric Analysis Results window updates based on the new parameters specified.

13 Select the **IC-CAP / Statistical Analysis** window, and then the Parameters folder. Use the horizontal scroll bar to bring columns C7 and C8 back into view. Select column **C7** and continue pressing the mouse button while dragging to column **C8**.

14 Choose **Graph > Scatter Plot**. A scatter plot of data from the two highlighted columns appears. Maximize the plot.

The scatter plot, shown in the following figure, directly illustrates the analysis results. The raw data is marked with crosses, the nominal point with a diamond, and the boundary points with squares. The indirect parametric results can be seen from the ellipses that mark the one-, two-, and three-sigma parametric boundaries.

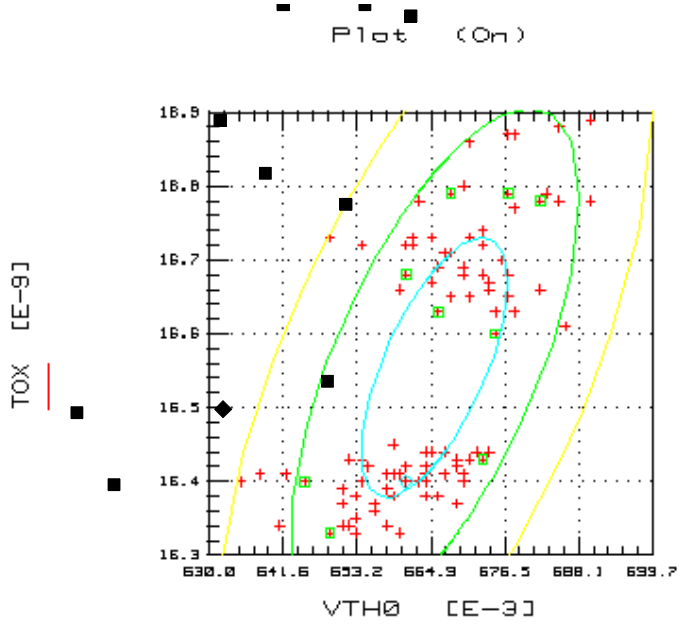


Figure 19 Scatter Plot Distribution for Values VTH0 and TOX

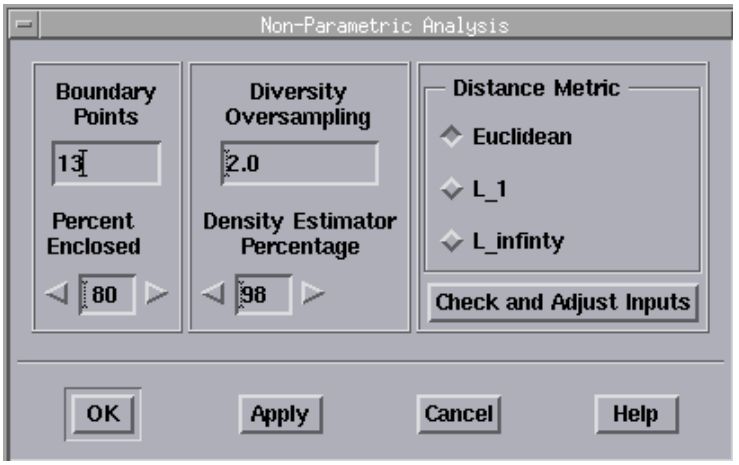
The non-parametric nominal point is appropriate because it's in the center of a region that is densely populated, and the non-parametric boundary points form a boundary around the center of both modes or clusters.

The ellipses illustrate the difficulties that parametric boundary analysis has with multimodal, non-Gaussian distributions. Depending of the sigma limit specified by the user, the parameter boundary modeling returns the points where the corresponding ellipse crosses its major and minor axes. In this case, for any of the ellipses drawn, two of the parametric

boundary models would be in regions where no data exists. Also the parametric nominal model is at the intersection of the minor and major axes of the ellipses. This point is also in region of no data.

Using the Non-Parametric Analysis Dialog Box

The preceding example was performed to help learn how the use the Non-Parametric Boundary Modeling feature. In this section we will describe the controls you have in Non-Parametric Boundary Modeling and their use. The Non-Parametric Analysis dialog box has the following fields and usage:



Boundary Points

Choose the number of boundary points you want. The default is a calculated maximum number based on your data set. The number corresponds to the number of worst-case models that will be generated. The minimum is 1. As a rule of thumb, choose a value up to twice the number of parameters you have. Obviously, too high a value will require an excessive number of simulations when you utilize your worst-case models.

Diversity Oversampling

This feature is used to make sure you get an even distribution of points along the boundary. The oversampling value (limit 1.2 to 5.0) multiplied by the number of boundary points equals the number of worst-case candidate models generated for subsequent selection. From these candidates, the program picks a representative set of boundary models.

Percent Enclosed

Enter a number corresponding to the percentage of your distribution you want enclosed by the boundary. Limit 10 to 100. If most of the data points are clustered near the center, with a few outliers near the edges, you might want the boundary to enclose only 50%, for example.

Density Estimator Percentage

The density estimator is a percentage of sample points that are to be used as the nearest neighbors for computing density. The program dynamically sets the limits that can be entered in this field. If you use the left and right arrow keys to enter a number, the value will wrap around the acceptable limits.

Distance Metric

This field has three choices:

- Euclidean
- L_1
- L_infinity

These refer to the formula used to calculate the distance between data points. The default is Euclidean. Choosing either of the other options generally will result in slightly different results.

Check and Adjust Inputs

Click this button to check and adjust the inputs of three of the fields in this dialog box, which are constrained together:

- Boundary Points
- Percent Enclosed
- Diversity Oversampling

If you select this button and the values are acceptable to the program, there is no change. If one or more of the values are out of range with respect to each other, the values in the Boundary Points and Diversity Oversampling fields will be adjusted. The action taken depends on the last field in focus. The Percent Enclosed is never altered.

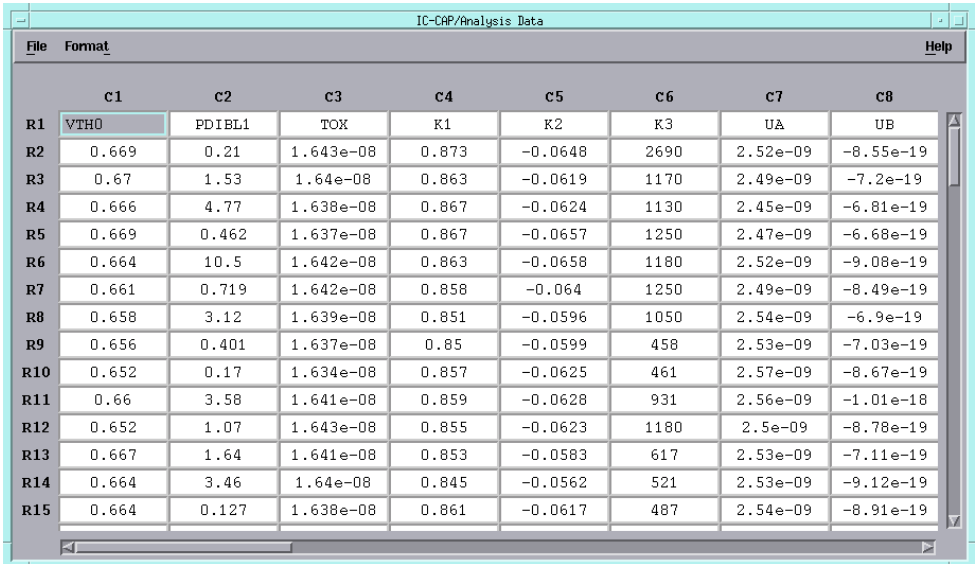
Other Analysis Menu Options

There are several other useful options on the Analysis Menu, including the following:

Analysis Data

This feature shows you only the active parameters and samples that are being used for your analyses. For example, if you have deactivated several parameter columns (because they are not of interest at this time) and filtered out a portion of the rows (because they represent bad samples), the Analysis Data window will display only the data that will be used for the current analysis.

To use Analysis Data, choose **Analysis > Analysis Data**. The Analysis Data window is displayed.

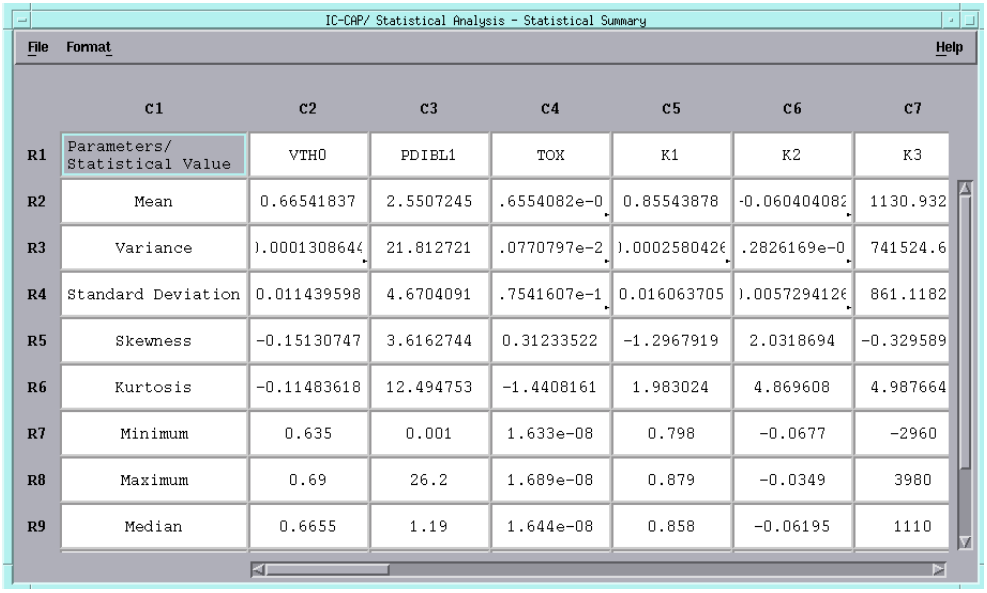


	C1	C2	C3	C4	C5	C6	C7	C8
R1	VTHO	PDIBL1	TOX	K1	K2	K3	UA	UB
R2	0.669	0.21	1.643e-08	0.873	-0.0648	2690	2.52e-09	-8.55e-19
R3	0.67	1.53	1.64e-08	0.863	-0.0619	1170	2.49e-09	-7.2e-19
R4	0.666	4.77	1.638e-08	0.867	-0.0624	1130	2.45e-09	-6.81e-19
R5	0.669	0.462	1.637e-08	0.867	-0.0657	1250	2.47e-09	-6.68e-19
R6	0.664	10.5	1.642e-08	0.863	-0.0658	1180	2.52e-09	-9.08e-19
R7	0.661	0.719	1.642e-08	0.858	-0.064	1250	2.49e-09	-8.49e-19
R8	0.658	3.12	1.639e-08	0.851	-0.0596	1050	2.54e-09	-6.9e-19
R9	0.656	0.401	1.637e-08	0.85	-0.0599	458	2.53e-09	-7.03e-19
R10	0.652	0.17	1.634e-08	0.857	-0.0625	461	2.57e-09	-8.67e-19
R11	0.66	3.58	1.641e-08	0.859	-0.0628	931	2.56e-09	-1.01e-18
R12	0.652	1.07	1.643e-08	0.855	-0.0623	1180	2.5e-09	-8.78e-19
R13	0.667	1.64	1.641e-08	0.853	-0.0583	617	2.53e-09	-7.11e-19
R14	0.664	3.46	1.64e-08	0.845	-0.0562	521	2.53e-09	-9.12e-19
R15	0.664	0.127	1.638e-08	0.861	-0.0617	487	2.54e-09	-8.91e-19

Statistical Summary

The Statistical Summary window shows you standard statistical data, such as mean, variance, standard deviation, skewness, kurtosis, etc. For each parameter, you get a statistical summary that includes all samples that have not been deactivated, filtered out, or deleted.

This feature can be selected from any place in the Statistics program. To select Statistical Summary, choose **Analysis > Statistical Summary**. The Statistical Summary window is displayed.



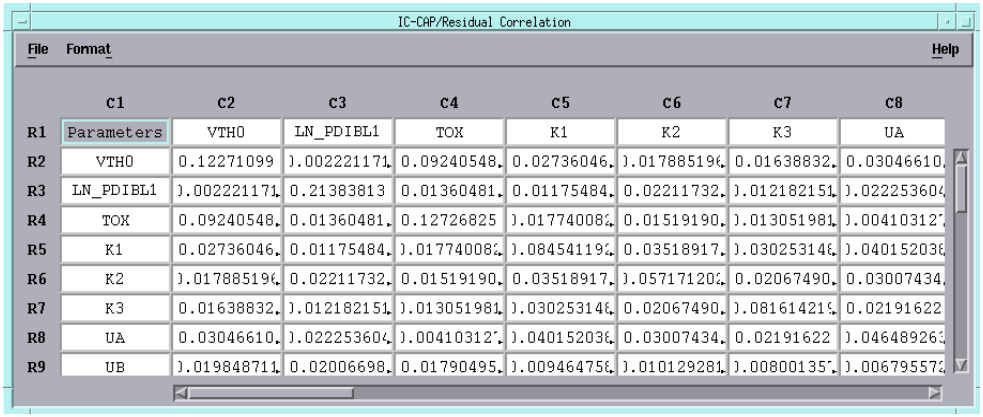
	C1	C2	C3	C4	C5	C6	C7
R1	Parameters/ Statistical Value	VTH0	PDIBL1	TOX	K1	K2	K3
R2	Mean	0.66541837	2.5507245	.6554082e-0	0.85543878	-0.060404082	1130.932
R3	Variance	).000130864	21.812721	.0770797e-2	).000258042	.2826169e-0	741524.6
R4	Standard Deviation	0.011439598	4.6704091	.7541607e-1	0.016063705	).005729412	861.1182
R5	Skewness	-0.15130747	3.6162744	0.31233522	-1.2967919	2.0318694	-0.329589
R6	Kurtosis	-0.11483618	12.494753	-1.4408161	1.983024	4.869608	4.987664
R7	Minimum	0.635	0.001	1.633e-08	0.798	-0.0677	-2960
R8	Maximum	0.69	26.2	1.689e-08	0.879	-0.0349	3980
R9	Median	0.6655	1.19	1.644e-08	0.858	-0.06195	1110

Residual Correlation

For a principal component analysis or factor analysis of less than full rank, the residual correlation shows where the unexplained variance is located.

Residual correlation can be used to see if you have chosen enough factors for your factor analysis, and if not, it can help you spot specific problem areas. You can also use residual correlation to compare principal component analysis to factor analysis by examining their respective errors.

To use Residual Correlation, choose **Analysis > Residual Correlation**. The Residual Correlation window is displayed.



	C1	C2	C3	C4	C5	C6	C7	C8
R1	Parameters	VTH0	LN_PDIBL1	TOX	K1	K2	K3	UA
R2	VTH0	0.12271099	).002221171	0.09240548	0.02736046	).01788519	0.01638832	0.03046610
R3	LN_PDIBL1	).002221171	0.21383813	0.01360481	0.01175484	0.02211732	).012182151	).022253604
R4	TOX	0.09240548	0.01360481	0.12726825	).01774008	0.01519190	).013051981	).00410312
R5	K1	0.02736046	0.01175484	).01774008	).08454119	0.03518917	).030253146	).040152038
R6	K2	).01788519	0.02211732	0.01519190	0.03518917	).05717120	0.02067490	0.03007434
R7	K3	0.01638832	).012182151	).013051981	).030253146	0.02067490	).081614219	0.02191622
R8	UA	0.03046610	).022253604	).00410312	).040152036	0.03007434	0.02191622	).046489263
R9	UB	).019848711	0.02006698	0.01790495	).00946475	).01012928	).008001357	).006795573

3
Data Analysis

In Chapters 1 and 2, we learned how to use IC-CAP Statistics and most of the programs's basic functions and features. This chapter presents general statistical information (such as definitions and advantages/limitations) on the program's features as well as some advanced functions.

General Statistics

The IC-CAP Statistics statistical summary window provides one of the key sets of information that assist in arriving at informed statistical inferences—numerical measures that describe the two aspects of descriptive statistics, central tendency and variability or dispersion. Measures of central tendency describe the center of the distribution of measurements and measures of variability describe how the measurements vary about the center.

Mean

A mean is a measure of central tendency. It is the sum of all the values divided by the total number of values. For example, 3 is the mean of 1, 2, 3, 4, and 5. A mean is the true arithmetic "average" of a set of values, and it is computed as follows:

$$\bar{X} = \frac{1}{n}\sum_{i}^{n} X_i$$

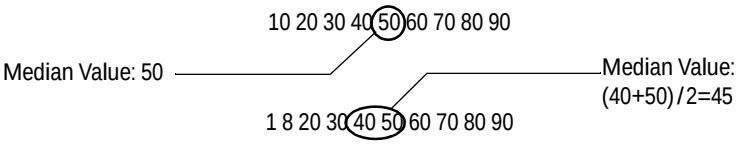
Only one mean value exists for any given set of values, and is applicable only to quantitative data. The mean is subject to distortion when extreme values or outliers are present in the set. In such cases, the mean ceases to be an accurate measure of the central value.

Median

The median is another measure of central tendency. A median is the middle value when all the values in the data set are arranged from lowest to highest. For example, 3 is the median of 1, 2, 3, 4, and 5. It is the true midpoint of a set of values.

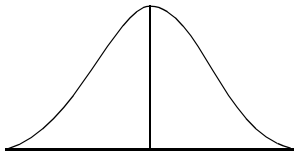
Only one median value exists for any given set of values, it is not influenced by extreme values within the set, and is applicable only to quantitative data.

In a data set with an odd number of values, the median is the middle value; in a data set with an even number of values, the median is an average of the two middle values. In either case, an equal number of values may be found above and below the median

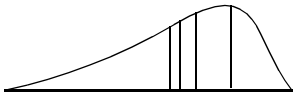


Skewness

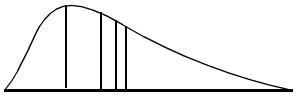
Skewness indicates the degree of asymmetry of a distribution. It is a measure of how much the sample varies from a normal distribution where the mean, median, mode, and trimmed mean have the same value



Normal Distribution: Identical Mean, Median, Mode, and Trimmed Mean



Left-Skewed Distribution: Varying Mean, Median, Mode, and Trimmed Mean



Right-Skewed Distribution: Varying Mean, Median, Mode, and Trimmed Mean

In a symmetric distribution, the value of skewness is 0. A positive skewness value indicates that relatively more values are above the mean than below it; a negative skewness value indicates that relatively more values are below the mean than above it. However skewness is highly variable and difficult to interpret for small samples.

Skewness is the average cubed deviation about the mean, and it is computed as follows:

$$sk(X) = \left(\frac{n}{n-2}\right)\left(\frac{1}{n-1}\right)\sum_{i}^{n}\left(\frac{X_i - \bar{X}}{S_x}\right)^3$$

Variance

Variance indicates the dispersion or variability of the data. It is a measure of the spread of values about the mean. When the values are clustered around the mean, the variance is small. When the values are widely scattered, the variance is large.

The variance of a set of n measurements can be expressed simply as the sum of the squared deviations of the measurements from their mean divided by $n-1$.

$$S_x^{\,2} = \frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2$$

Standard Deviation

Standard deviation is another measure of the dispersion or variability of the data. It is a measure of the “average” deviation about the mean. When the values are clustered around the mean, the standard deviation is small. When the values are widely scattered, the standard deviation is large.

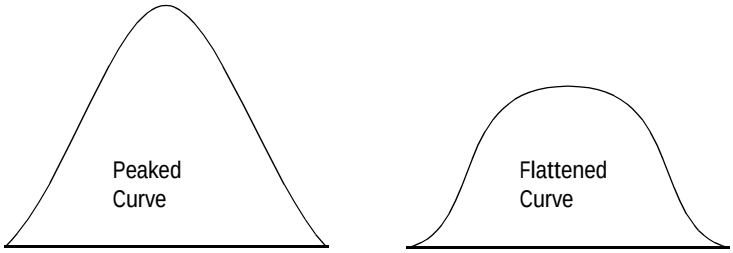
The standard deviation of a set of measurements can be expressed as the positive square root of the variance.

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2}$$

Kurtosis

Kurtosis indicates the heaviness of the tails in relation to the middle of the distribution. It is a measure of the peakedness of the distribution.

For example, when a large number of the values are clustered around the middle, the curve is peaked, resulting in a greater kurtosis value. When the values are spread around more evenly the curve is flattened, resulting in smaller kurtosis values.



In a normal distribution, the value of kurtosis is 3. A greater kurtosis value indicates that relatively less values are distributed around the tails; a smaller kurtosis value indicates that relatively more values are distributed around the tails. However kurtosis is highly variable and difficult to interpret for small samples.

Kurtosis is the average of the fourth power of the deviations about the mean, and is always a positive value that can be expressed as:

$$Kur(X) = \left(\frac{n(n+1)}{(n-2)(n-3)}\right)\left(\frac{1}{n-1}\right)\sum_{i}^{n}\left(\frac{X_i - \bar{X}}{S_x}\right)^4$$

Minimum

A minimum is the lowest value in the data set. The difference between the minimum and the maximum value provides the simplest measure of data variation, the range.

Maximum

A maximum is the greatest value in the data set. The difference between the maximum and the minimum value provides the simplest measure of data variation, the range.

Median Absolute Deviation

The median absolute deviation (MAD) provides a scale estimate. The scale value is used in the Data > Data Filter operation. Scale is defined as the median absolute deviation (MAD) divided by a constant (approximately 0.6745). This standardizes MAD in order to make the scale estimate consistent with the standard deviation of a normal distribution.

MAD = median $\{|x_1 - \text{median}\{x_j\}|\}$

The scale constant (approximately 0.6745) is the inverse of the standard normal distribution function evaluated at 3/4.

Covariance

Covariance is the expected value of the product of the deviations of two random variable from their respective means. The covariance of two variables, x and y, can be expressed as:

$$S_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n(n-1)}$$

Probability Density Function (PDF)

PDF is function of a continuous random variable whose integral over an interval gives the probability that its value will fall within the interval.

The significance of the PDF $f(x)$ is that $f(x)\,dx$ is the probability that the random variable is in the interval $(x,x+dx)$, written as:

$$P(x \le x' \le x + dx) = f(x)dx$$

Specifically, the probability that the random variable is in the interval (a, b) is:

$$P(a \le x \le b) = \int_a^b f(x)dx$$

where the PDF has the following properties:

$$f(x) \ge 0$$

for all x in the open interval (-∞, ∞), and

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Correlation Analysis

The correlation coefficient provides a numerical measure of the amount of variation in a variable that is attributable to another variable. The correlation analysis function in IC-CAP Statistics generates a complete matrix of correlation coefficients (one for each pair-wise combination of parameters). The correlation values can vary between 1 (perfect positive correlation) and -1 (perfect negative correlation). Uncorrelated variable have a correlation coefficient of 0. The correlation between two variables x and y can be expressed as:

$$r = \frac{S_{xy}}{S_x S_y}$$

where S_{xy} is the covariance between the variables and S_x and S_y are the standard deviations of variables x and y respectively.

The correlation matrix is closely related to the covariance matrix of a data set. In fact, the correlation matrix is really nothing more than a covariance matrix with the variances (matrix diagonals) normalized to unity. Under this condition, the denominator in the above equation reduces to one, and the relationship

$$r = S_{xy}$$

holds true.

Factor analysis can be applied to either the covariance matrix (unnormalized) or the correlation matrix (normalized covariance matrix). For applications involving semiconductor device model parameters where parameter values may vary over many orders of magnitude, the correlation matrix is preferable, since it weights each parameter variance equally. For this reason, IC-CAP Statistics uses the correlation matrix exclusively as input to factor analysis.

Correlation analysis is always performed before proceeding to factor analysis.

Factor Analysis

Factor analysis is used to explain relationships among several, difficult to interpret, and correlated variables using a few, conceptually meaningful, and relatively independent factors.

Factor analysis focuses on one or both of the following tasks—an assessment of underlying relationships or dimensions in the data (or what is being measured), and the replacement of original variables with fewer, new variables.

Qualitative Description

Factor analysis is used in IC-CAP Statistics to reduce a set of model parameters to a smaller set of parameters that are statistically independent. This reduced parameter set is called the factors, or in the case of the principal component model, principal components. Henceforth, we will refer to the reduced parameter set as *factors*, regardless of the particular method used to compute them. These factors can be used to build a mathematical model that predicts parameter values from the independent factors. Alternatively, the independent factors can be used to assist in the identification of a set of dominant parameters. These dominant parameters can then be used as the regressors in a multiple regression model that predicts the remaining model parameters from the dominant (subset) parameters.

Depending on the number of factors that are chosen for a particular model, some percentage of the overall variance in the original parameter set will be accounted for in this factor-space model. For example, given a set of parameter values for an n-parameter model, an n-factor model can be built from the data that will explain all (100%) of the variance in the data. Another way of looking at this is that the correlation matrix of the original parameter set is reproduced exactly by the factor-space model. For an n-parameter problem, a factor model that retains fewer than n factors will explain < 100% of the variance in the original parameter set. In terms of the correlation matrix, this means that the correlations in the original correlation matrix of parameters will not be

reproduced exactly by the factor-space model. The distribution of this error within the correlation matrix depends on the particular method used to compute the factors. The three methods available in IC-CAP statistics are: Principal Factor Analysis (PFA), Principal Component Analysis (PCA) and Unweighted Least Squares (ULS). Specifics about each of these methods are discussed later.

Mathematical Description

The primary quantity computed in a factor analysis is the factor loadings matrix, V. The factor loadings matrix is an n by m matrix, where n = number of parameters and m = number of factors, whose individual coefficients represent the correlations between the factors and the parameters. The factor loadings matrix V is related to the predicted correlation matrix R_{mod} through the following relation:

(1) $R_{mod} = VV^T$

where V^T is the transpose of matrix V. As explained in the previous section, the matrix R_{mod}, is *not* equal to the correlation matrix R (computed from the extracted model parameter values), when the number of factors is less than the number if model parameters. Under these conditions, the error in the factor model or the residual correlation is given by

(2) $\Delta R = R - R_{mod}$

Substituting (1) into (2) and solving for R, we can also write

(3) $R = VV^T + DR$.

The key underlying base to common factor analysis is that the chosen model parameters can be transformed into linear combinations of an underlying set of hypothesized or unobserved factors f. The original model parameters are related to the factors through the following relation:

(4) $z = Vf + e$,

where

z is the vector of standardized parameters of length n,

f is the factor vector of length m,

e is the error in the model.

Finally, the standardized parameters are related to the actual parameters x, through the relation

(5) $z = (x - \mu)/\sigma$

where

μ is the vector of means of the parameters

σ is the vector of standard deviations of the parameters

Principal Component Method

Principal component analysis (PCA) is the optimum method for explaining maximum variability in a parameter data set with m factors (principal components). In other words, with m factors, a factor analysis performed with PCA will always return a larger value of total variance explained than either PFA of ULS. From this standpoint, it is the preferred method when building a model to predict parameter values from factors. This is especially true for the case of corner or boundary modeling, where we are attempting to accurately predict the extreme values of a set of correlated parameters.

Recall, that in the previous two sections, the subject of error in the correlation matrix was discussed. Also, recall from the section on "Correlation Analysis" on page 92 that the diagonal of the correlation matrix represents the normalized variances of the model parameter data. Since PCA is the optimal method for explaining total variability, it should come as no surprise that the residual correlation matrix, ΔR in PCA, will have its smallest entries along the diagonal. In other words, the PCA model is better at reproducing the variances in the correlation matrix than it is at reproducing interrelationships or the off-diagonal elements in the correlation matrix. You are encouraged to validate this on a set of measured data, by viewing the residual correlation matrix after a PCA is performed and comparing it with the results obtained from either a PFA or ULS analysis.

Principal Factor Analysis and Unweighted Least Squares

In IC-CAP Statistics, the principal component analysis (PCA), principal factor analysis (PFA), and unweighted least squares (ULS) methods have been lumped under the heading of Factor Analysis. This has been done to avoid confusion with disparate terminology, such as factors and components. In addition, from a functional standpoint, the methods accomplish much the same task: They reduce the data to a smaller set of uncorrelated parameters. However, the literature often treats PCA as a separate technique from factor analysis. Indeed, there are significant differences in the techniques used to compute the factor loadings for the principal component and factor models. For applications related to the analysis of device model parameters, however, the only relevant difference between PCA and the “pure” factor methods, PFA and ULS, lies in the way the respective techniques model the correlation matrix.

Recall that the PCA model is optimal for reproducing the diagonals of the correlation matrix. The “pure” factor methods differ from PCA in that they are better at reproducing the off-diagonal elements of the correlation matrix. In other words, the “pure” factor methods explain correlations between parameters or interrelationships better than they explain variability.

In selecting dominant parameters, the desire is to locate parameters that have a high loading on a particular factor. Since PFA and ULS are better at reproducing correlations between factors and parameters, these methods are preferred when your ultimate goal is to generate a set of dominant parameter equations. In this case, the factor analysis is not being used for model construction, but rather, as a means of identifying actual parameters that can be used in place of the factors for model construction (see “Regression Analysis” on page 100). The IC-CAP Statistics module is very flexible and unrestrictive on this point, however. You can construct a model that predicts parameters from factors for any of the three methods: PCA, PFA or ULS. However, as stated previously, the PCA method is preferred for model construction when the factors are the independent variables.

Until now, no distinction has been made between PFA and ULS. All we have said is that they differ from PCA in the way they account for residual errors in the correlation matrix. The ULS (also known as minres or minimum residuals) method is a more recent algorithm than the PFA method. The ULS method operates by minimizing a quantity that is the sum of squared discrepancies between the observed values and the modeled values. Generally speaking, it is thought to be superior to the PFA method for minimizing the off-diagonal elements in the residual correlation matrix. In practice, however, the two methods give very similar results. Computationally, either method is fast enough on a modern computer to allow virtually any device modeling problem to be evaluated with both methods in span of a few seconds. The residual correlation matrices for the two cases can then be compared, and the better performing method can be selected for a given application.

Decisions on Running a Principal Component or Factor Analysis

To analyze data with either principal component analysis or principal factor analysis, three key decisions must be made:

- The factor extraction method
- The number of factors to extract
- The transformation method to be used

Interpretation of a Factor Analysis

There are several areas to be aware of, including:

- Magnitude of eigenvalues. Assess the amount of original variance accounted for. Retain factors whose eigenvalues are greater than 1. (Ignore those with eigenvalues less than one as the factor is accounting for less variance than an original variable.)
- Substantive importance. This is an absolute test of eigenvalues in a proportional sense. Retain any eigenvalue that accounts for at least 5% of the variance.
- Interpretability. This is a battery of tests where the above heuristics may all be applied.

Factor Rotation

The distribution of variance among factors in an unrotated factor analysis follows a definite pattern. The first factor always accounts for the most variance, the second factor accounts for the second largest portion of the variance, and so on. In this model, the factors are independent and orthogonal. Unfortunately, it is difficult to gain an understanding of the underlying meaning of the factors when most of the model parameters load heavily on the first few factors and load very little on the remaining factors.

Factor rotation attempts to redefine the position of the factors such that the loadings tend to be either very high (near 1 or -1) or very low (near 0). This eliminates as many of the medium-sized loadings as possible and aids in the interpretation of the factors. The rotation also results in a more even distribution of variance explained between the factors.

Some examples of applications where the rotation of factors would be helpful include: 1. trying to identify dominant parameters for use in a regression model; 2. attempting to associate a particular processing step with a given factor; and 3. selecting electrical test parameters that could be used as predictors in a regression model. Unfortunately, one of the side effects of rotating the factors is that they are no longer uncorrelated. Consequently, for applications that require building a model using factors as the independent variable, rotation offers little benefit and should be avoided.

Rotation Type

The Rotation Type field in the Factor Analysis dialog box has four choices.

- Varimax
- Quartimax
- Equimax
- None (default)

All three rotation types are orthogonal. Varimax (originated in 1958) is the most popular of these rotation types, with the constant, c, in the orthogonal expression, equal to 1. Quartimax is a slightly older variation, with c = 0. The Equimax variation (1961) treats c = k/2. You should experiment with these rotation types to see which one works best for your work, or choose None.

Regression Analysis

The intent of regression analysis is to provide a quantitative description of the relationship between two or more variables. In the case of simple regression, the objective is to predict the value of a response variable from a single predictor, or regressor. In multiple regression, the value of the response variable is dependent on two or more predictors. For problems involving more than one response variable that is dependent on the same set of predictors, we have what is called multivariate linear regression. Essentially, a multivariate linear regression amounts to nothing more than a collection of multiple regression problems. However, when properly solved, the multivariate problem requires only one matrix inversion operation (as does the single multiple regression operation). Hence, the computational burden of the multivariate problem does not scale in proportion to the number of responses.

For compact semiconductor device modeling problems, the ability to predict model parameters from a smaller set of predictor parameters has some compelling advantages. The smaller set of parameters can be electrical test parameters that are tracked in manufacturing. These predictors can then serve the dual role of providing predictive SPICE models for simulation as well as being used for process tracking. The predictors can also be a subset of the compact model parameters. A regression model can be built using any appropriate predictor(s).

Choosing Predictors

Using factor analysis, it is a fairly straightforward procedure to identify a set of dominant (predictor) device model parameters. These are selected automatically as the parameters that have the highest loading on a given factor. The multivariate regression feature in IC-CAP Statistics takes the dominant parameters from the previously run factor analysis and builds a regression model using these dominant parameters as the regressors. The remaining parameters are taken as the response variables. If for physical reasons, you want to change the dominant parameter for a given factor, choose **Analysis > Change**

Dominant Parameter. By controlling the number of factors in the factor analysis and by changing the default dominant parameters, you have complete control over the predictor (regressor) variables used in the regression.

Solving for the Regression Coefficients

Mathematically, the regression problem requires the solution of the p (number of predictors) by q (number of responses) matrix B of regression coefficients.

$Y = XB + \varepsilon$

where

Y is the n by q matrix of response parameters values

X is the n by p matrix of predictor parameter values

ε is the n by q matrix of errors

n is the number of data samples

the solution B of this equation can be expressed as:

$B = [X^T X]^{-1} X^T Y$

IC-CAP Statistics normalizes the matrix of inputs X prior to evaluating the matrix B, by subtracting the mean and dividing by the standard deviation of the predictor parameters. This results in a better conditioned matrix, since once normalized, entries in the matrix will no longer vary over the many orders of magnitude typically encountered with device model parameters. Upon completion of the solution, the B matrix is un-normalized and is output to the spreadsheet entitled Dominant Parameter Based Equations.

Evaluating the Regression Model

Once the regression coefficients have been computed, the regression model can be used to predict parameter values based on measured or generated values of the predictor variables. The various options for evaluating the regression model can be

found by choosing Analysis > Parametric Analysis > Regression Equations. Refer to "Parametric Analysis Results Window" on page 66 for more details.

Eigenvalues

An eigenvalue is a scalar associated with a given linear transformation of a vector space and having the property that there is some nonzero vector which when multiplied by the scalar is equal to vector obtained by letting the transformation operate on the vector.

Eigenvalues present a summary of the variances associated with each factor. For example, an eigenvalue of 3.45 indicates that the factor accounts for as much variance in the data collection as would 3.45 variables, on average.

Residual Correlation

From the factor loadings matrix V, a modeled correlation matrix can be computed from the matrix product VV^T, where V^T is the transpose of V. The difference between the actual correlations (those in the Correlation Matrix folder) and the modeled correlation VV^t is the residual correlation. For a full-rank principal component analysis or factor analysis, the residual correlation will be 0 for all entries (that is, 100% explained variance). For a principal component analysis or factor analysis of less than full rank, the residual correlation shows where the unexplained variance is located.

R_{mod} = modeled correlation matrix = VV^t

R = correlations computed from extracted parameters

$\Delta R = R - VV^t$

Data Transformations

Data transformation is a process in which the measurements on the original scale are converted to a new scale. A transformation is the re-expression of a variable, typically the dependent variable. It is used to stabilize the variance of the variable, to normalize the variable, and to linearize the regression model. For example, if a quantity y grows at a rate r per unit time then:

$$y_t = y_0 e^{rt}$$

Data transformation presents one way to simplify this nonlinear model into a linear model. In this case taking the natural logarithms of both sides leads to a linear model:

$$\log y_t = \log y_0 + rt$$

IC-CAP Statistics provides a number of ways to transform data where each method is better suited to a particular data set. As a whole the available methods provide a comprehensive range of data transformation options.

Exponential

The exponential transformation is used to stabilize random variance, and to approximate solutions to difficult distributions. One of the situations in which this works is when time has no effect on the future outcomes. That is, the future distribution is not affected by the past.

Natural Log

The log transformation is used to stabilize marked increases in variance, to normalize the dependent variable in a positively skewed distribution of residuals, and to linearize the regression model in the case of models with a consistently increasing slope.

Square Root

The square root transformation is used to stabilize a variance that is proportional to the mean. It is a particularly appropriate transformation if the dependent variable has the Poisson distribution, which describes the number of occurrences for a given interval.

Square

The square transformation is used to stabilize a variance that decreases with the mean, to normalize the dependent variable in a negatively skewed distribution of residuals, and to linearize the model if the dependent variable consistently decreases with any increase in the independent variable.

Constant Value

The constant value transformation simply replaces your data with the specified constant.

Mean

The mean transformation replaces your data with the calculated mean value for this column of data.

Outlier Screening

Outliers are points that represent unusual, implausible, or impossible values. Typically an outlier is as many as three or more standard deviations from the mean. The presence of an outlier can adversely affect the estimates for means and correlations.

Multivariate techniques such as Factor Analysis assume that the data set to be analyzed is a joint Gaussian distribution. If the data is not joint Gaussian, then the model generated from the analysis may not accurately reproduce the measured density.

The elimination of outlier data can be done before or after data transformation. Outliers can be eliminated manually or through automatic data filtering. For more information refer to "Transforming Distributions/Eliminating Outliers" on page 52.

Dominant Parameter

The dominant parameter is the one parameter in the column that has the highest loading (correlation).

Factor Group

The factor group shows, for a given parameter, which factor the parameter is most highly correlated with.

Generating Equations

From the factor loading matrix and the parameter means and standard deviations, equations can be constructed that will predict parameter values from independent factors. In matrix notation, these equations can be written

X = M + DVf

where

V = factor loadings matrix (n by m)

M = mean vector (n values)

D = diagonal matrix of standard deviations (n values), and

f = vector of independent factors (m values) with a mean of 0 and a variance of 1.

Comparison of IC-CAP's Analysis Methods

IC-CAP has several statistical analysis options for different situations: direct Monte Carlo, generated Monte Carlo, corner modeling, parametric boundary modeling, and Non-Parametric Boundary modeling.

The difference between direct and generated Monte Carlo is that direct uses the original extracted data and skips fitting a multidimensional Gaussian PDF.

The strength of direct Monte Carlo is that it can be used with the smallest amounts of data, it provides a yield estimate, and it is non-parametric. While a Monte Carlo yield estimate is noisy, the level of noise goes essentially as $S^{-0.5}$, where S is the number of samples. However, one gets some feel for robustness even if the estimate is corrupted. The disadvantage, which it shares with Non-Parametric Boundary analysis, is that it does not provide factor (independent variable) reduction or allow for discarding the raw data set. For large enough data sets and given sufficient time and computing resources, direct Monte Carlo gives increasingly accurate yield estimates.

Generated Monte Carlo shares the primary strength of direct Monte Carlo —a yield estimate. Additionally, generated Monte Carlo allows for factor reduction via factor analysis or principal component analysis. Additionally, the method requires the calculation of a Gaussian PDF covariance matrix and mean vector from which the Monte Carlo samples are generated. Generated Monte Carlo requires more raw data than direct Monte Carlo because a sufficiently accurate covariance matrix and mean vector are required. Given that the data really is Gaussian and given enough data to calculate the covariance matrix, then the raw data can be discarded. As with direct Monte Carlo the compute time and resources can be prohibitive.

Corner and parametric boundary analyses both allow for factor reduction and discarding the raw data. The previous caveats regarding data volume/covariance accuracy and appropriateness of the Gaussian PDF also apply. Given the exponential increase of corner models as the number of factors (independent variables) increases, parametric boundary

analysis is more desirable. In fact, corner modeling can lead to such high numbers of generated model points that parametric modeling is almost always more desirable. Additionally, parametric boundary modeling has the potential for providing a lower bound on the circuit yield. Corner modeling could be made to do the same thing, but the relationship between placement of generated points and the enclosure fraction is not direct. Corner analysis is included as an option for users of traditional analysis methods. We strongly recommend parametric or Non-Parametric Boundary analysis instead.

Parametric boundary analysis is a shortcut compared to a full-blown Monte Carlo analysis. Computing resources are saved by judiciously picking a much smaller set of component parameter points to simulate. These points are likely to be the extremes that cause the overall design not to meet its performance specifications. As mentioned earlier the enclosure percentage corresponding to a particular set of boundary points can, under the proper circumstances, place a lower bound on the yield of the design.

Non-parametric boundary analysis is also a shortcut to a full-blown Monte Carlo analysis. Non-parametric shares all of the strengths of parametric analysis discussed except factor reduction is lost and the raw data must be retained. Further advantages of Non-Parametric Boundary analysis are that it works on any distribution of data since the method does not require the fitting of a PDF. This quality alleviates the effort involved and the inherent loss of accuracy arising from fitting any function to real data. Non-parametric analysis also avoids the problem of returning component parameter values that cannot physically exist; the method picks actual instances (or samples) from the raw data set provided.

All of the parametric methods have the potential for returning non-physical component values. If, for various reasons, the fitted PDF is not appropriate for the actual data, then this problem of unrealistic returned values becomes quite serious. A simple example of this is shown at the end of this section. The non-parametric method even gives better results when the amount of raw data is small and the dimension of the data is moderate to large. This circumstance arises from the limited

accuracy of the covariance matrix in high dimensional problems with limited numbers of samples. Non-parametric boundary analysis has no dimensional limit. Furthermore, the computational effort of Non-Parametric Boundary analysis does not suffer from the curse of dimensionality. (The curse of dimensionality is the phenomena that an algorithm, which operates on a problem of varying dimension, exhibits an exponential increase in computational effort as a function of problem dimension.)

Parametric Analysis

Once equation coefficients are generated in IC-CAP Statistics, you can build a variety of statistical models or load the equation data directly into IC-CAP and perform a simulation. You can test your model, based on a reduced set of parameters, against the raw data to see how well it performs.

There are three choices for parametric analysis:

- Monte Carlo
- Corner
- Parametric Boundary

Monte Carlo Analysis

Monte Carlo analysis provides an efficient solution to problems involving elements of uncertainty that are too complex to be solved by strict analytic methods. Instead of calculating all possible combinations, this method uses a small set of randomly generated values to approximate a solution.

Corner Modeling

Corner modeling is used to select worst-case models from a given data set. This method is computes the dependent parameters of a data set to arrive at a set of correlated parameters. Traditional worst-case modeling uses corner models. Corner modeling chooses a set of extreme values at the outside of the real multi-dimensional probability density function (PDF) and requires 2^n simulations for an n-dimensional problem.

Parametric Boundary Modeling

Parametric boundary modeling chooses a set of extreme values at the boundary of the real multi-dimensional PDF, and only needs 2n simulations for an n-dimensional problem.

For example, if you chose 10 factors, the number of simulations with parametric boundary models would be 20 compared to 1024 using corner models.

Parametric boundary modeling is used to extract a nominal model from a given data set. Although most effective with a large number of data points, boundary modeling can be used for data sets as small as 20 points.

Boundary modeling is used to generate an estimate for the density of data about every point in a data set. These density estimates are then sorted to determine the nominal points and the boundary points for the data set.

Boundary modeling circumvents the typically large simulation times required for a Monte Carlo analysis. However, its results are qualitative in comparison to a full Monte Carlo analysis, and it does not provide a yield estimate.

Boundary modeling provides a superior alternative to traditional worst-case modeling. It does this by using a statistically robust method to isolate existing deviant models that stress the design rather than constructing models that may not occur in practice.

Non-Parametric Analysis

IC-CAP Statistics contains an exclusive feature called Non-Parametric Boundary Analysis. Unlike other statistical analysis tools, which only handle Gaussian distributions, Non-Parametric Boundary analysis uses a new technique to handle arbitrary data distributions, Gaussian or non-Gaussian, and selects nominal and boundary models. Non-parametric analysis works for any data from any arbitrary stochastic process.

Because Non-Parametric Boundary Analysis is a new technique and not described in statistical textbooks, we will describe this method here in some detail. For a basic description of this method and step-by-step instructions to use it, refer to "Parametric Analysis Results Window" on page 66.

Applying Non-Parametric Boundary Analysis

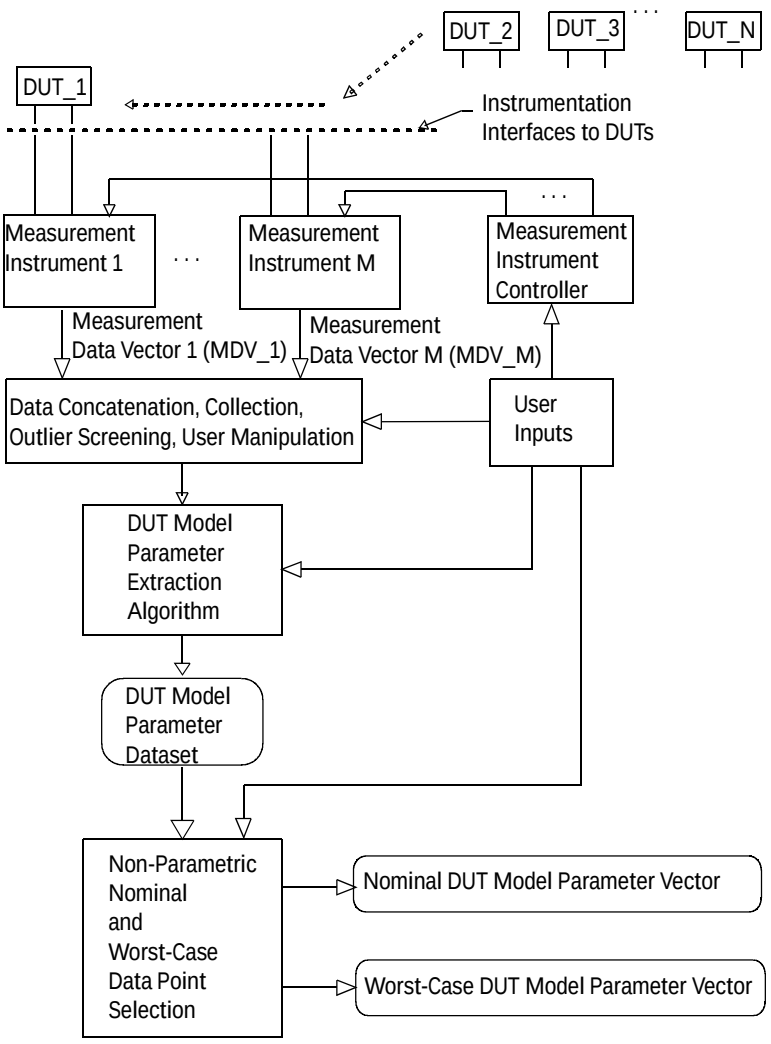


Figure 20 Using Non-Parametric Boundary Analysis—One View

The previous figure illustrates one mode of applying Non-Parametric Boundary analysis that depicts a flow of DUTs across various instrumentation interfaces. Each instrument gathers its data as dictated by the controller, which takes its inputs from the user. The gathered data is passed, in the form of a measurement data vector, into the data collection function. In the data collection block the various measurement data vectors are concatenated and manipulated. One common manipulation of the data is to discard outliers (those measurements taken from obviously defective devices, or measurements that are far removed from the prevailing data distributions). Each concatenated measurement vector is then passed to the DUT model parameter extraction algorithm. After extracting every concatenated measurement vector, a DUT model parameter data set is formed. The Non-Parametric Boundary analysis then processes the DUT model parameter data set. The result of this processing is a nominal DUT model parameter vector and a set of worst-case vectors.

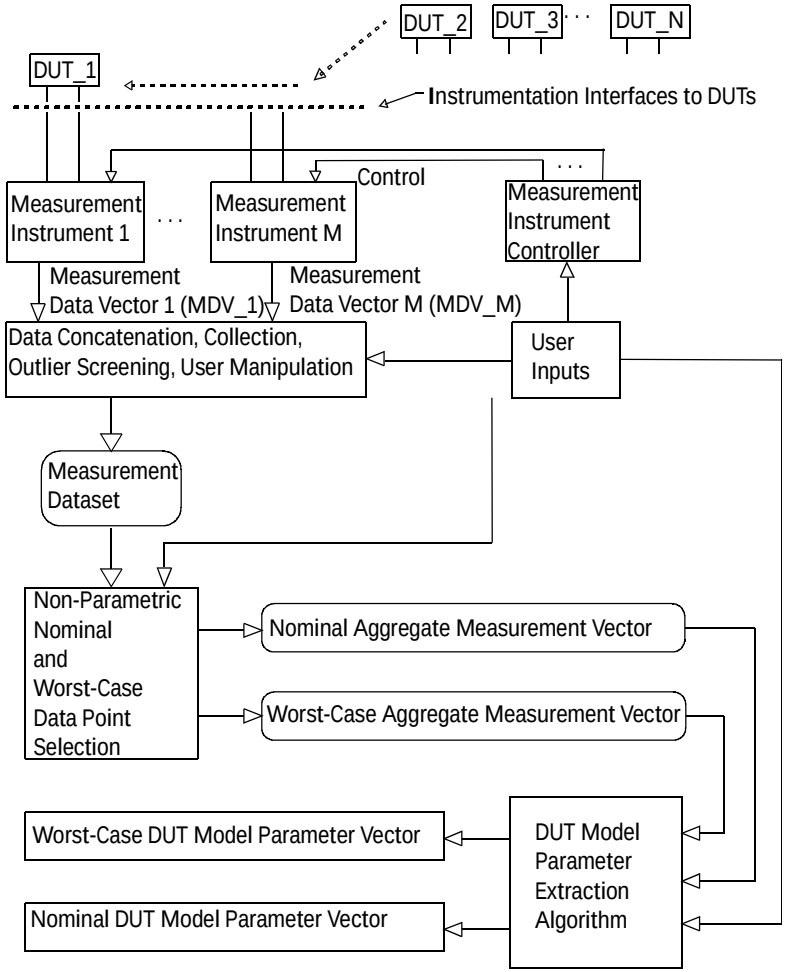


Figure 21 Using Non-Parametric Boundary Analysis—Alternate View

The previous figure shows an alternate and important variation on the data flow presented in Figure 20—the difference is the sequencing of DUT model parameter extraction. In Figure 20, the DUT model parameter extraction procedure is performed on all valid measurement data vectors drawn from the instruments, and boundary analysis is then done on the model parameter data set. Since extracting the DUT model can be

computationally intense, performing the boundary analysis directly on the measurements first can lead to significantly less computing time. The flow illustrated in Figure 21 has the added benefit that the analysis is done on the electrical behavior of the DUTs. The measured electrical behavior will potentially correlate more closely to the DUT's worst-case performance in its final circuit environment. The method shown in Figure 21 is then the preferred course.

Algorithm Description

Like all algorithms, Non-Parametric Boundary analysis has inputs, outputs, and internal processes. Inputs to the Non-Parametric Boundary algorithm are:

- A real valued data set of *S* sample vectors with each vector having
  P parameters
- *B*, number of boundary points
- *E*, percent enclosed
- *O*, diversity oversampling
- *D*, density estimator percentage
- *M*, distance metric

The data set is a 2-dimensional collection of real data. Each row is a single sample with the columns containing the parameter values for each sample. The desired number of boundary points is self explanatory, as is the percent enclosed. These inputs are primary to the algorithm.

The remainder of the inputs merely influence the algorithm's behavior. Diversity oversampling allows the user some control over a process called spatial diversification. This process was developed to prevent spatial clustering of the returned boundary models. This spatial clustering manifests itself as the algorithm picking two or more boundary points that are very close to one another. The problem arises because the data set has a finite number of samples. The finite number of samples means that the raw data set exhibits clustering on some regions of the boundary. This clustering of the raw data leads to clustering of the returned boundary points.

The density estimator percentage is the percentage of sample points that are to be used as the nearest neighbors for computing density. A percentage is used (instead of an absolute number) because it is more appropriate and consequently more convenient for the user.

The outputs from the algorithm are the nominal data vector and the set of boundary vectors. It is important to note that the vectors returned are selected from the data set. This circumvents any problems with the algorithm provided data that is unrealistic.

The internal processes of the algorithm are: normalize the data set, calculate density estimates, select the nominal vector, and select the boundary vectors.

The data set must be normalized because the parameters can have vastly different magnitude ranges. Since the algorithm makes a distance calculation to estimate density, normalizing the data ensures that each parameter has equal weight in the density estimate. The data is normalized such that it is bounded by a unit hypercube. The user must perform outlier screening on the data before it is fed to the algorithm so as to ensure equal weighting of each parameter. If a given parameter naturally ranges over more than one decade of values, the user must logarithmically transform that parameter before applying Non-Parametric Boundary analysis.

The density estimate for every point in the data set is performed by finding the specified number of nearest neighbors (density estimator percentage of the number of samples) and calculating the average distance to those points. A suitable density estimate is then one over the average distance. Calculating a dimensionally correct density estimate is problematic because the Monte Carlo noise in the average distance calculation is amplified to the point that the estimate has insufficient signal to noise ratio.

Selecting the nominal point is easy—it is the point with the highest density estimate. (The algorithm returns the mode of the data.) If the data is multimodal, then the algorithm picks the mode with the highest estimated density. If there is a tie, then the returned point is arbitrarily selected between the ties.

The boundary points are selected by sorting the data points per their density estimate. The sorted data points are then partitioned into two groups. The first group includes the points with the higher density estimates. The second group includes the points with the lower density estimates. The cut between the two groups is specified such that the first group holds the enclosure percentage of the entire data set.

The boundary models are now ideally those datapoints in group one with the lowest density estimates (those points just above the cut line in the sorted list). This procedure locates a set of points for which the density estimate is approximately constant.

The final step of the algorithm is to enforce spatial diversity on the points returned. During the development of the algorithm, it became clear that the points returned were generally not evenly distributed over the boundary. The phenomena occurred for all trial validation data sets, even synthesized multidimensional Gaussian sets with a very large number of samples (>10,000). As mentioned previously, the main reason for this non-uniform spatial distribution over the boundary is that the starting data points exhibit local clustering. This situation is just a fact of life for a finite number of samples drawn from any stochastic process. The lack of spatial diversity in the boundary points arises because two or more points in a local cluster near the boundary will have nearly identical density estimates, resulting in their being adjacent in the sorted listed associated with picking the boundary points.

The spatial diversification process uses oversampling above the cut line followed by selective discarding to determine the final set of boundary points. The oversampling is done by taking some multiple of points more than the desired number of points just above the cut line. All possible pairs are formed while searching for the pair closest together. The member of the pair that is closest together and that has the higher density estimate is discarded. This process of pairing and discarding is continued until only the desired number of boundary points remain. The net result is that the points are approximately evenly distributed over the boundary.

Users will discover that three of the inputs are constrained by a formula. This formula accounts for the fact that only a finite number of points are available from which to pick the final boundary points. The formula is

$$int\langle BO\rangle \le int\langle SR\frac{E}{100}\rangle$$

where

int is the truncating integer operator
B is the number of requested boundary models
O is the diversity oversampling factor ($1.2 \le 0 \le 5$)
S is the number of samples
R is a heuristic constraint on the range of density values over which boundary
models can be picked ($R = 0.35$)
E is the percent enclosed

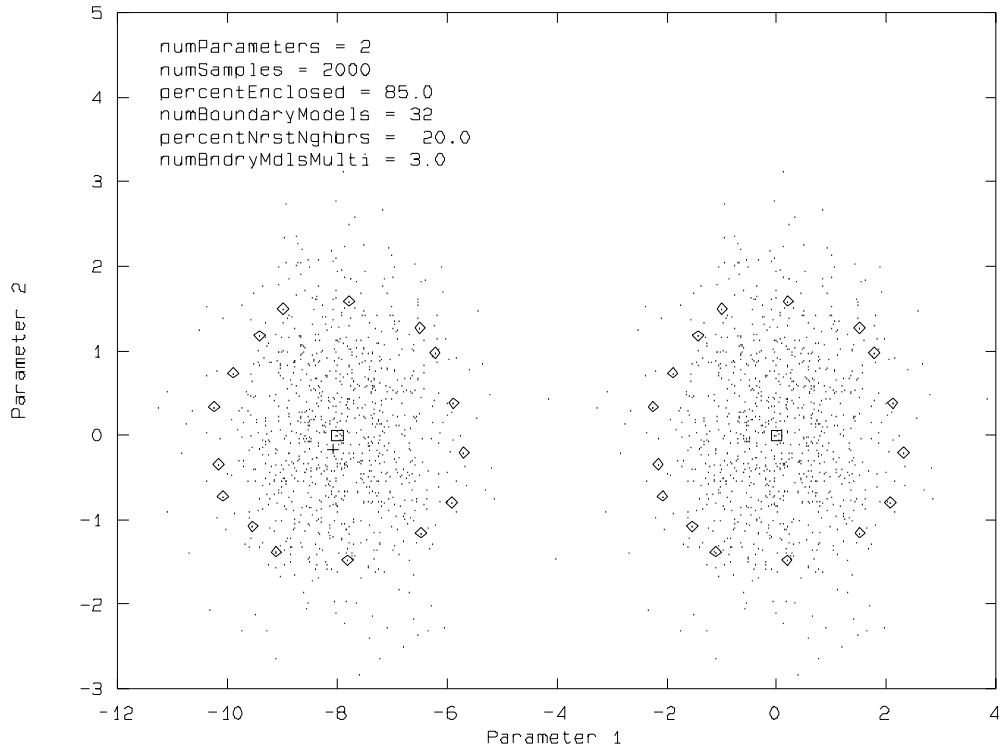
The user has direct control (via the input dialog box) over *B*, *O*, and *E*. *S* is set by the provided data set. *R* is hardwired from the standpoint of the graphical user interface, but can be changed by modifying via a configuration file.

Algorithm Validation

A variety of methods were used to validate the functionality of the Non-Parametric Boundary analysis algorithm. Only one of these methods, 2-D scatter plots, is available in IC-CAP. The other methods were implemented in stand-alone software for the sole purpose of validation. Because there are no visualization tools for general data sets with high dimensionality, the other validation methods presented in this section use simplified data sets with specific properties.

The most obvious validation method is to take some samples from a known 2-D stochastic process and look for the expected boundary by viewing scatter plots of the samples and the selected boundary points. A good governing PDF for this stochastic process is a 2-D Gaussian. The expected boundary is

then a circle or an ellipse. It is important to note that this graphical confirmation works only for a 2-D distribution. Higher numbers of dimensions require other, less direct, methods.



Notes:
Dots are dataset points; diamonds are selected boundary points; the cross is the nominal point; squares are the known means.

Figure 22 Scatter Plot of Bimodal Gaussian Distribution with Selected Boundary Points Overlaid

The previous figure gives an example of a 2-D scatter plot for a bimodal Gaussian distribution with the boundary points overlaid. As anticipated, the boundaries for each of the modes is approximately elliptical. Furthermore, the algorithm has no

difficulty with this multimodal data. At this point, we could view a 2-D scatter plot for data distributions that are non-Gaussian. Multimodal, non-analytic distributions are the forte of this algorithm.

In higher dimensions, we employ two other validation aids. The first aid relies on making high dimensional Gaussian distributions spherical. (Set the covariance matrix for the PDF function equal to a constant times the identity matrix.) For this special type of distribution, the boundaries of constant density are shells or hypershells.

Validation proceeds by overlaying two histograms: one of the radial distance from the mean for all of the sample points and one of the radial distance to the boundary points. The expected result is for the boundary points to fall in a shell. That is, the boundary points will fall into some narrow band of histogram bins. The limitation of this tool is that the distribution must be spherical, and it must be unimodal. However, the tool works for low to modest numbers of points with any dimensionality.

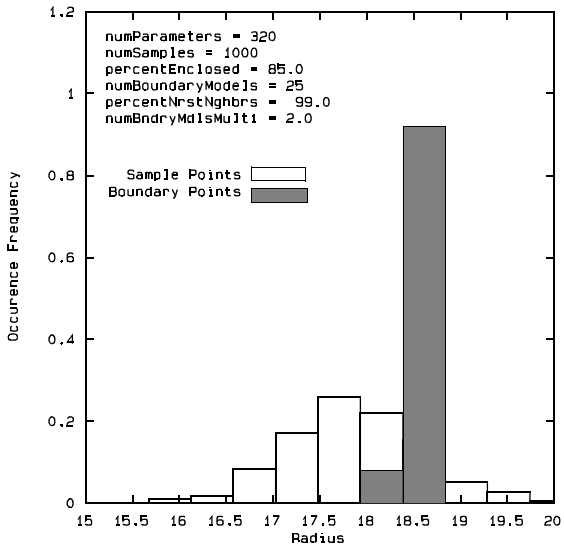


Figure 23 Overlay of Histograms of Radial Distance to Sample Points and Radial Distance to Boundary Points for a 320 Dimensional Data Set

The previous figure is an example of a radial distance histogram for a 320 dimensional data set (320 is arbitrary). Note the histogram for the entire data set is given by the unfilled bins, and the filled bin histogram is for the boundary points. The histograms indicate that boundary points do, in fact, reside in hypershell.

The second tool for validation in higher dimensions does not require the unimodal constraint. However, as the dimension of the space goes up, the number of sample points needed to see a result grows, probably at an exponential rate. The second tool is conditional density histograms. This tool involves taking repeated slices of the space that are parallel with each of the coordinate axes until only one dimension (independent variable) remains. A histogram is formed over some range of values for the remaining variable.

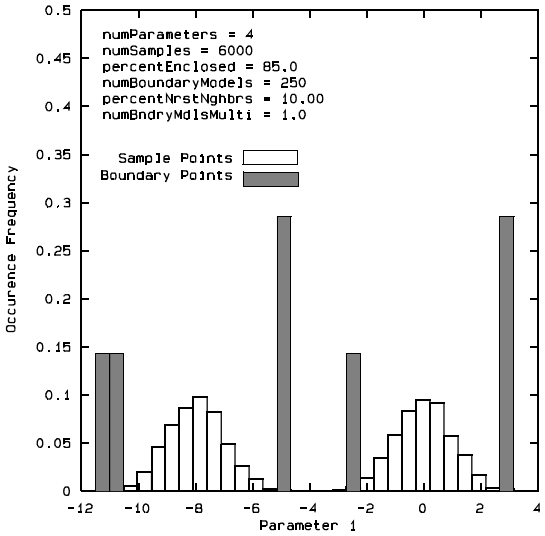


Figure 24 Conditional density histogram of a 4-D multimodal Gaussian data set

The previous figure provides a graphical affirmation of the algorithm's correct performance on a 4-D bimodal Gaussian distribution. We expect to find that the boundary points are in the tails of the two distributions—this is, in fact, what we observe.

Unfilled bins are histogram of points from entire data set that reside in the slice. Filled bins are the histogram of boundary points found in the slice. The means of the two Gaussian distributions are (-8,0,0,0) and (0,0,0,0). The covariance matrix for both distributions is the 4 x 4 identity matrix. The slice is the region of the space (-1.2 < P2 < 1.2, -1.2 < P3 < 1.2, -1.2 < P3 < 1.2), where P2, P3 and P4 are coordinate axes labels.

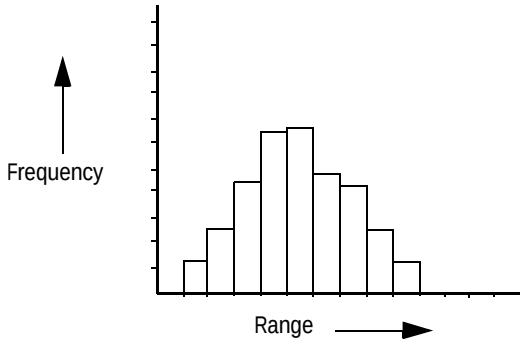
4
Data Visualization

Data visualization uses abstract, non-representational pictures to depict numerical data sets. At the very least, it is meant to be more than a mere substitute for statistical tables; at best, it can be a medium for exploring, analyzing, and displaying quantitative information.

IC-CAP Statistics offers four standard methods for data visualization, each of which is best suited for analyzing a particular aspect of the data—convergence, variance, correlation, and outlier identification. As a group, the four methods provide a solution for a broad range of data visualization needs.

Histogram

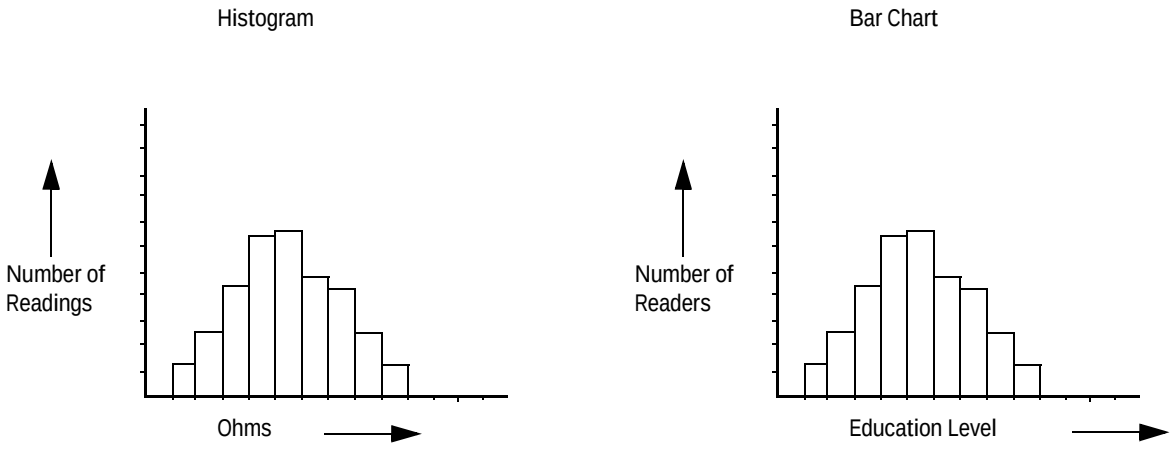
A histogram depicts either the frequency or the relative frequency of data distribution. It is typically used to identify both the ranges around which most of the data converges and the outliers of the data set. In a histogram, the x-axis displays the value ranges while the y-axis plots the frequency (or relative frequency) for each range.



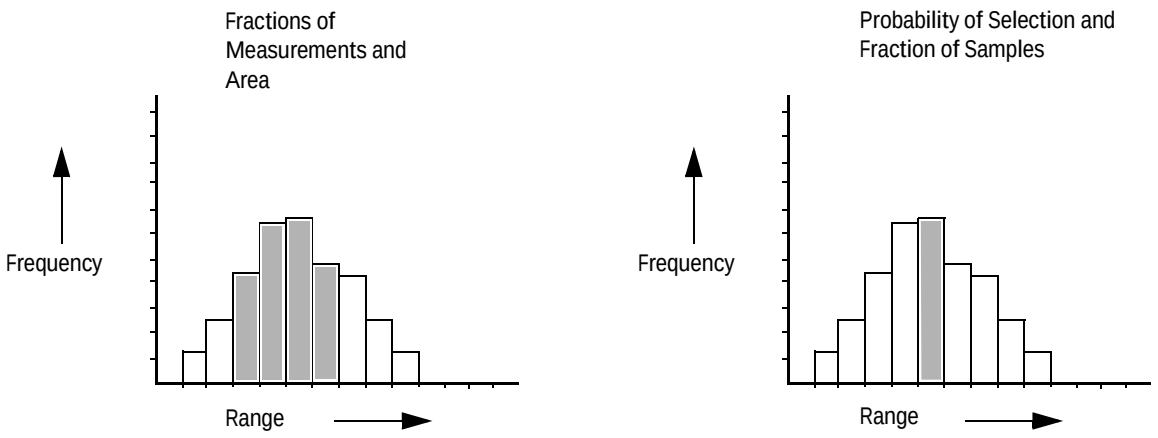
A histogram is mainly used for the visualization of data for a single variable. It is different from a bar chart primarily because of the kind of data it displays, quantitative as opposed to qualitative. While a bar chart is used to display frequency data for qualitative variables, a histogram is used to display frequency data for quantitative variables.

Values of quantitative variables are based upon actual units of measure, and hence they are measurements. Values of qualitative variables, on the other hand, vary in kind but not in degree, and hence they are not measurements.

For example, a histogram best depicts the frequency distribution of recorded parasitic gate resistance for a given setup because each data variable has a quantitative value, based upon a unit of measure. On the other hand, a bar chart is better suited to depict the distribution of readership for a magazine when each range represents a qualitative distinction, the education level of each respondent.

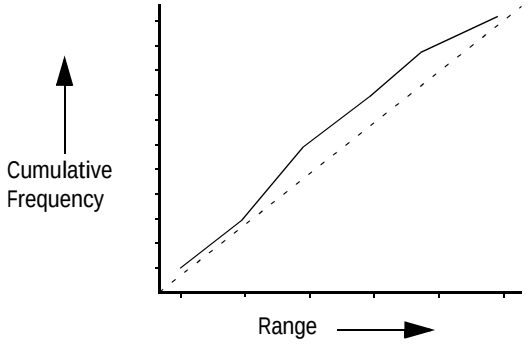


The number of measurements in an interval as a fraction of all the measurements is equal to the area of the interval as a fraction of the total area of the histogram. In addition, the probability of a random selection lying within an interval is equal to the fraction of the total number of samples within that interval.



Cumulative Density Plot

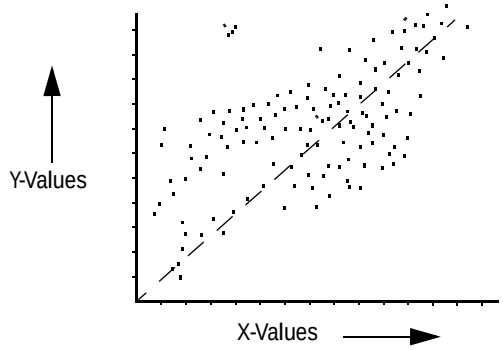
A cumulative density plot depicts either the cumulative frequency or the cumulative relative frequency of data distribution. It is typically used to track variations between contemporaneous observations by highlighting changes in mean levels. In a cumulative density plot the x-axis displays the value ranges while the y-axis plots the cumulative frequency or cumulative relative frequency for each range.



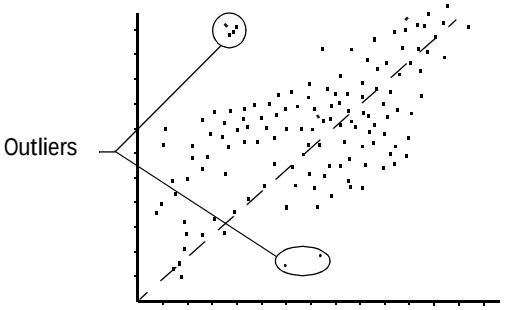
A cumulative density plot offers a comparative visualization of current results within the context of a reference distribution. It is ideal for situations where the rate of specific changes needs to be tracked.

Scatter Plot

A scatter plot depicts the distribution of pairs of values on a rectangular, two-dimensional plane. It is the most effective method of analyzing the correlation between two variables. Values for one variable are tracked on the x-axis while values for the other variable are tracked on the y-axis.

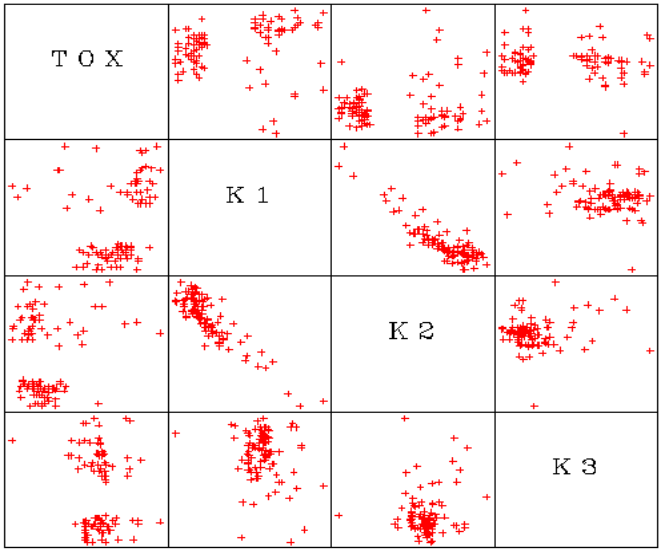


In addition to its ability to depict information about the relationships between variables, a scatter plot is an effective method of discovering outliers, those points that represent unusual, implausible, or impossible values.



Scatter Plot Matrix

The scatter plot matrix contains multiple scatter plots, with each parameter plotted against every other parameter. For example, if your data had four parameters, P1, P2, P3, and P4, the matrix would contain 16 cells. First would be P1 against itself (only the label appears), then P1 vs. P2, P1 vs. P3, and then P1 vs. P4. The next row would contain P2 vs. P1, the P2 label, P2 vs. P3, and P2 vs. P4, and so on. If you have a large number of parameters, the plots will be quite small. You may want to deactivate parameter columns to reduce the number of plots of interest, or only select the ones of interest.



A
File Formats

This Appendix describes technical details on three file formats used in IC-CAP Statistics:

- Statistical Data Format File (.sdf)
- SPICE Equation File
- SPICE Library File

Statistical Data Format File (.sdf)

SDF file is an ASCII format file used save the data in the different spreadsheets used in the Statistical Analysis window, including:

- Parameters
- Correlation
- Factor analysis
- Parameter variance explained
- Factor-based equations
- Dominant parameter-based equations

so that they can be used in a different sessions of the Statistical analysis.

A typical SDF file has different data blocks corresponding to each of the spreadsheet mentioned above. Each data block has an identifier to mark the beginning and the end. The beginning is marked by BEGIN_<data-type> and the end by END. The begin identifiers are BEGIN_PARMDATA, BEGIN_CORRELATION, BEGIN_FACTOR_LOADINGS, BEGIN_PARAMETER_VARIANCE, BEGIN_EQN_FACTOR, BEGIN_EQN_DOMPARM.

Comments begin with the character “!”.

Parameters Data Block

- Each line or row in the data block represent a single sample of a particular device.
- Each column specifies either an Attribute or a Model Parameter.
- The first line in the block represents the header and contains information regarding the data block.
- Each attribute or parameter value can be either space or tab separated.
- "~#A" appended to an attribute name, specifies that the data in the column rep resent and attribute as opposed to a data.

- Deactivated columns are identified by "~#C" appended to the parameter name.
- An attribute column cannot be a deactivated column.
- Data filtered rows are identified by "~#F" in the first column of the row.
- Attribute filtered rows by "~#AF".
- Deactivated rows by "~#R".

Correlation Data Block

- The first line in the block has the correlated parameter names.
- The first cell or word in a row also has the correlated parameter name.
- All other values represent a correlation value.

Factor Loading Data Block

This data block is divided into two sections:

- Dominant Parameter/Factor Group details
- Values data

The first section has 6 lines, identifying:

- Default Dominant Parameter row number
- Current Dominant Parameter row number
- Dominant parameter type: 0 - Default, 1 - User defined
- Default Factor Group column number
- Current Factor Group column number
- Factor Group type: 0 - Default, 1 - User defined

The second section contains the factor values and the parameter names.

Parameter Variance Explained Data Block

This data block always has only 4 lines excluding the markers.

Factor-based Equations and Parameter-based Equations

These two data blocks are similar and contain the equation coefficient values. The first line is the Parameter line.

Example

Portion of the bsim3.sdf file in `${ICCAP_ROOT}/examples/icstat` is reproduced below:

```
! @(#) $Source: /wlv/src/iccap500/mdl/icstat_examples/bsim3.sdf,v $ $Revision: 1 .4 $
! $Date: 97/06/09 16:24:24 $
! Date/Time: Wed May 21 11:17:46 1997
! Filename: /tmp_mnt/hpdisk15/lrich/bsim3new.sdf
!*****************************************************************************
! IC-CAP Statistical Analysis Data Set File
! File Format:
! * Comments begin with !
! * A begin and end of a member-set data block is marked by BEGIN_<data-set Type >
!   and END.  Eg. BEGIN_PARMDATA, END.
! * Parameter rows are identified by the keyword PARAMETER
! * Identifiers for the different data-sets
! * Raw data data-set: BEGIN_PARMDATA
! * Correlation data-set: BEGIN_CORRELATION
! * Factor Loadings data-set: BEGIN_FACTOR_LOADINGS
! * Parameter Variance Explained data-set: BEGIN_PARAMETER_VARIANCE
! * Equation data-set: EQUATIONS
! Parameters Data Set:
! ---------------------
! * Each Row in the raw data-set matrix represent a single sample
! * Each field in a single sample data-set are separated by one or more space(s) /tab
! * deactivated rows are identified by the Delimiter: ~#R in the first column
! * deactivated columns are identified by the Delimiter: ~#C appended to the <pa ram-name>
! * Attribute Columns are identified by the Delimiter: ~#A appended to the <para m-name>
! * Data filtered rows are identified by the Delimiter: ~#F in the first column
! * Attribute filtered rows are identified by the Delimiter: ~#AF in the first column
!*****************************************************************************
BEGIN_PARMDATA
PARAMETER LotID~#A WaferID~#A Module~#A VTH0~#C TOX K1 K2 PDIBL1 ~#R JAMES 5188A039
SAPT_7I 0.669 1.643e-08 0.873 -0.0648 0.21 ~#F JAMES 5188A039 SAPT_7I 0.67 1.64e-08 0.863
-0.0619 1.53
~#AF JAMES 5188A039 SAPT_7I 0.666 1.638e-08 0.867 -0.0624 4.77 JAMES 5188A039 SAPT_7I 0.664
1.642e-08 0.863 -0.0658 10.5 JAMES 5188A039 SAPT_7I 0.661 1.642e-08 0.858 -0.064 0.719 JAMES
5188A039 SAPT_7I 0.658 1.639e-08 0.851 -0.0596 3.12 JAMES 5188A039 SAPT_7I 0.656 1.637e-08
0.85 -0.0599 0.401 JAMES 5188A039 SAPT_7I 0.652 1.634e-08 0.857 -0.0625 0.17 JAMES 5188A039
SAPT_7I 0.66 1.641e-08 0.859 -0.0628 3.58
...
END
```

SPICE Equations File (.eqn)

This file contains equations in the SPICE Equation format. Each file has some comments at the top of the file.

A sample of a file created from the Statistical Analysis is give below:

```
*  IC-CAP  Statistical Analysis -  SPICE Equations File:  bsim3.eqn
* Generated From SDF Data File:  bsim3.sdf
* Generated Date:  Mon Jun 09 18:46:59 1997
* ------------------------------------------------------------------------
VTH0 = 0.66541837 + F1 * -0.0016484802 + F2 * -1.4804296e-11 + F3 * 0.0071357339
+ F4 * -0.0027095029 + F5 * 2.095063
TOX = 1.6554082e-08 + F1 * -41.243159 + F2 * -1.4446406e-10 + F3 * 1.1719242e-19
+ F4 * -0.0021408843 + F5 * 3754992.9
K1 = 0.85543878 + F1 * 0.00032398078 + F2 * -1.1656058e-09 + F3 * 1.8302239e-08 + F4 *
-2.1728014e-05 + F5 * 5.0762303e-11
K2 = -0.060404082 + F1 * -0.025936659 + F2 * -0.015696001 + F3 * -0.010483001 + F4 *
-0.00014561763 + F5 * 0.063589613
PDIBL1 = 2.5507245 + F1 * -0.049678299 + F2 * 0.1045004 + F3 * 0.46793438 + F4 *
0.018056057 + F5 * -2.8387574e+08
K3 = 1130.9328 + F1 * -7.4849785e-05 + F2 * -0.011869139 + F3 * -0.20928796 + F4
* 1.4278207e-05 + F5 * -119.79896
UA = 2.7067347e-09 + F1 * 7.5152648e-09 + F2 * -0.0013190211 + F3 * -4.0657294e- 11 + F4 *
0.0089069807 + F5 * -0.0036095769
UB = -1.0752857e-18 + F1 * -1.1436865 + F2 * -165.23484 + F3 * -1.3560184e-10 + F4 *
4.2622828e-20 + F5 * -0.0032652145
UC = 0.025313265 + F1 * -1060845.9 + F2 * -0.00051317033 + F3 * 3.0105432e-08 + F4 *
1.94928e-08 + F5 * -2.0728674e-05
VSAT = 15066837 + F1 * 4.563842e-11 + F2 * -0.024208771 + F3 * -0.014704545 + F4
* -0.0099990684 + F5 * -7.5277773e-05
VOFF = -0.063031633 + F1 * -0.055642135 + F2 * 0.028192398 + F3 * -0.49288422 + F4 *
-0.34715499 + F5 * -0.0074837098
DL = 1.8572857e-07 + F1 * 2.2536854e+08 + F2 * 5.3864784e-05 + F3 * -0.009459965
+ F4 * 0.11948863 + F5 * 2.9853918e-05
DW = 1.5183571e-07 + F1 * 235.34875 + F2 * 6.2129472e-09 + F3 * 0.00040430508 + F4 *
1.2869758e-11 + F5 * 0.0074758302
CJ = 0.00062490612 + F1 * -0.0026761127 + F2 * -0.01479755 + F3 * -132.97546 + F 4 *
-7.4866106e-11 + F5 * -3.1814582e-20
CJSW = 1.5317449e-10 + F1 * -0.0027828084 + F2 * -191523.01 + F3 * -0.0003807477 4 + F4 *
4.2958317e-09 + F5 * 1.6064242e-08
PB = 0.75327551 + F1 * 5.3624279e-05 + F2 * -1.3008101e-10 + F3 * 0.066217104 + F4 *
0.03925663 + F5 * 0.011609906
MJ = 0.40983878 + F1 * -8.0454919e-05 + F2 * 0.03346715 + F3 * -0.004038763 + F4
* 0.15106436 + F5 * 0.034714099
MJSW = 0.15991735 + F1 * -0.0022275619 + F2 * -20726473 + F3 * -9.5245228e-06 + F4 *
-0.0018714502 + F5 * -0.0061978962
W0 = 0.0012067959 + F1 * 1.786684e-05 + F2 * 6.05232 + F3 * 1.0382995e-08 + F4 *
-0.0064290892 + F5 * -9.7930725e-11
DVT2 = -0.22997551 + F1 * -0.0011420485 + F2 * -0.00025342926 + F3 * 0.19827535 + F4 *
97.000921 + F5 * -1.1566666e-10
```

```
DVT1 = 0.56420408 + F1 * 2.213371e-19 + F2 * 0.00061279851 + F3 * -430444.26 + F 4 *
0.00060900233 + F5 * -2.2689368e-08
DVT0 = 3.2722449 + F1 * -1.8741288e-09 + F2 * 5.4030315e-06 + F3 * -1.7628352e-1 1 + F4 *
0.0085192897 + F5 * 0.0046996512
PCLM = 1.8097959 + F1 * -0.0028554436 + F2 * -1.8542061e-05 + F3 * -0.016602481 + F4 *
-0.050208481 + F5 * -0.0939772
PDIBL2 = 0.11476122 + F1 * -0.0012308386 + F2 * -0.0061667199 + F3 * 87283799 + F4 *
1.7832732e-05 + F5 * -0.013191715
PSCBE1 = 4.6253265e+08 + F1 * 0.080299784 + F2 * -1.5420951e-05 + F3 * -98.50338 3 + F4 *
-1.9360527e-08 + F5 * 0.0029385782
PSCBE2 = 8.0151684e-05 + F1 * 3.1128467e-11 + F2 * -0.00057252505 + F3 * 0.00053 950547 + F4
* 2.492631 + F5 * 222.55912
NFACTOR = 0.87427551 + F1 * -1.3964716e-11 + F2 * 1.1790193e-19 + F3 * 0.0017370 088 + F4 *
2173860 + F5 * 0.00080899241
DROUT = 0.32879796 + F1 * 1.5053356e-08 + F2 * -4.4544711e-09 + F3 * 8.8803028e- 06 + F4 *
-2.4705077e-11 + F5 * 0.011843057
CDSC = 4.9278061e-05 + F1 * 0.0068528745 + F2 * 0.00024778847 + F3 * -5.1706045e -05 + F4 *
-0.063039521 + F5 * 0.0029217923
RDSW = 835.93886 + F1 * -0.54772615 + F2 * 1.1144781 + F3 * 0.015988799 + F4 * 8 2175519 + F5
* 2.2747554e-05
NLX = 6.664051e-08 + F1 * -0.003186587 + F2 * -0.0013447802 + F3 * 6.7265878e-05
+ F4 * 134.85442 + F5 * -5.8133275e-09

* ----------------------------------------------------------------
VTH0 = 0.64625363 + UB * -7.2696318e+15 + VSAT * 1.7258753e-10 + PB * 0.00674746 67 + PCLM *
0.00033850608 + RDSW * 3.6511423e-06
TOX = 1.6333945e-08 + UB * -1.0078683e+08 + VSAT * -7.5323763e-21 + PB * 1.38744 32e-10 +
PCLM * 1.3133432e-11 + RDSW * -1.9625579e-14
K1 = 0.83280799 + UB * 5.7576654e+15 + VSAT * 6.6506778e-10 + PB * 0.017791722 +
PCLM * -0.00067931295 + RDSW * 7.9297691e-06
K2 = -0.050069315 + UB * -1.0491259e+15 + VSAT * -3.1249522e-10 + PB * -0.006064 8888 + PCLM
* 0.00038695215 + RDSW * -3.4528059e-06
PDIBL1 = -9.1485286 + UB * 4.3806074e+17 + VSAT * 5.3638758e-07 + PB * 4.1940073
+ PCLM * 0.63953244 + RDSW * -0.00027279117
K3 = 512.50115 + UB * 2.0728e+19 + VSAT * 2.7145265e-05 + PB * 408.67876 + PCLM * -13.006188
+ RDSW * -0.062902083
UA = 2.4595269e-09 + UB * -4.9335968e+08 + VSAT * -7.488236e-18 + PB * -1.646977 7e-10 + PCLM
* 1.2916168e-11 + RDSW * -8.3479318e-14
UC = 0.039975633 + UB * 4.7854919e+15 + VSAT * -2.8218705e-10 + PB * -0.00504656 34 + PCLM *
0.00043872542 + RDSW * -2.7005336e-06
VOFF = -0.05993201 + UB * 3.6253177e+15 + VSAT * 6.4933915e-11 + PB * 0.00138654 61 + PCLM *
-0.00031937972 + RDSW * -7.7297252e-07
DL = 6.0788068e-08 + UB * -6.7396362e+09 + VSAT * 2.4764115e-15 + PB * -1.074261 4e-08 + PCLM
* -2.0167433e-09 + RDSW * 1.1020402e-10
DW = 8.6283044e-08 + UB * -7.0107411e+08 + VSAT * 2.374413e-15 + PB * 2.9674645e -08 + PCLM
* -3.0907009e-09 + RDSW * 1.467126e-11
CJ = 9.3472742e-06 + UB * -4.4411136e+12 + VSAT * 3.0503087e-14 + PB * 0.0008120 2965 + PCLM
* -1.8719197e-07 + RDSW * -1.2195564e-09
CJSW = 1.6488351e-09 + UB * -3851835.2 + VSAT * -4.0304614e-19 + PB * -1.9741282 e-09 + PCLM
* 6.166705e-13 + RDSW * -9.3109972e-15
MJ = -0.038611163 + UB * -1.0604103e+15 + VSAT * 5.1238188e-12 + PB * 0.59451407
+ PCLM * -0.00015151467 + RDSW * -3.9019696e-07
MJSW = 0.049913741 + UB * -1.6308312e+16 + VSAT * -1.178243e-09 + PB * 0.1624270 1 + PCLM *
0.00088518397 + RDSW * -1.6429894e-05
```

```
W0 = 0.0019480706 + UB * -1.855009e+14 + VSAT * -1.9577002e-11 + PB * -0.0004185 4263 + PCLM
* -2.9798242e-05 + RDSW * -3.3085073e-07
DVT2 = -0.31705233 + UB * -3.0316066e+16 + VSAT * 5.6578983e-09 + PB * 0.1824430 1 + PCLM *
-0.0047479316 + RDSW * -0.00019092941
DVT1 = 0.37375386 + UB * -8.8031637e+16 + VSAT * -1.2884102e-09 + PB * -0.008618 1762 + PCLM
* -0.0041475023 + RDSW * 0.00015455833
DVT0 = 2.9075918 + UB * -3.5757306e+17 + VSAT * 2.6189947e-09 + PB * 1.4513995 +
PCLM * -0.0004892043 + RDSW * -0.0013777543
PDIBL2 = 0.030562809 + UB * -6.3022416e+15 + VSAT * 4.8224389e-09 + PB * 0.00010 75759 + PCLM
* 0.0011282501 + RDSW * 3.1579693e-06
PSCBE1 = 1.0123365e+09 + UB * 1.4967827e+26 + VSAT * -44.624434 + PB * -1.411856 7e+08 + PCLM
* 32913618 + RDSW * 395097.01
PSCBE2 = 0.00026089839 + UB * 2.7018006e+13 + VSAT * -1.1921656e-11 + PB * -8.75 78136e-05 +
PCLM * 1.029354e-05 + RDSW * 9.004041e-08
NFACTOR = 0.85553793 + UB * -3.6167858e+16 + VSAT * -1.2501827e-09 + PB * 0.0053 127194 +
PCLM * 0.0016425063 + RDSW * -9.9187245e-06
DROUT = 0.93229074 + UB * 1.4507241e+17 + VSAT * -4.7587999e-08 + PB * 0.0335755 13 + PCLM *
0.024926013 + RDSW * 0.00023817474
CDSC = -0.00029792003 + UB * 3.4062683e+13 + VSAT * 9.2197205e-12 + PB * 0.00021 223444 +
PCLM * -4.0384623e-06 + RDSW * 1.1047575e-07
NLX = -7.6614936e-09 + UB * -1.8957838e+10 + VSAT * 1.1915555e-15 + PB * 2.96887 16e-08 +
PCLM * -8.9066509e-10 + RDSW * 1.8197598e-11

*
```

SPICE Library File (.lib)

The SPICE Library file can be created either from the Parameters spreadsheet data or the generated models spreadsheet data. An example of a file generated using the IC-CAP's Statistical analysis is shown below:

```
* Date/Time Created: Mon Jun 02 13:00:44 1997
* Statistical File: /tmp_mnt/home17/mperoomal/ic50/data/bsim3.sdf
* ------------------------------------------------------
.SUBCKT CGaas1_7 1 2 3
JCGAAS 11 22 33
+ NMF1
* Series Inductors & Gate Resistance LD 1 11 3.5E-12
LG 2 20 3.5E-12
RG 20 22 1
LS 3 33 3.5E-12
* N-Channel MESFET - Curtice Model
* for  HPSPICE
.MODEL NMF1 RCAY NJF
+ MODEL = 1
+ BETA = 0.0001
+ VTH0 = 0.8
+ TOX = 1.642E-08
+ K1 = 0.858
+ CGSO = 0
+ CGDO = 0
+ FC = 0.5
+ VBR = 100
+ IS = 1E-14
+ N = 1
+ TAU = 0
+ A0 = 0.01
+ A1 = 0.001
+ A2 = -0.001
+ A3 = -0.0001
+ GAMMA = 0.5
+ VDSO = 5
+ RDSO = 1E+12
+ VTO = -2
+ ALPHA = 2
+ LAMBDA = 0
+ R1 = 0
+ R2 = 0
+ RF = 0
+ CGS = 0
+ CGD = 0
+ CDS = 0
+ RIN = 0
+ A5 = 0
```

```
+ VDSDC = 0
.ENDS
* ------------------------------------------------- .SUBCKT CGaas1_8 1 2 3
JCGAAS 11 22 33
+ NMF1
* Series Inductors & Gate Resistance LD 1 11 3.5E-12
LG 2 20 3.5E-12
RG 20 22 1
LS 3 33 3.5E-12
* N-Channel MESFET - Curtice Model
* for  HPSPICE
.MODEL NMF1 RCAY NJF
+ MODEL = 1
+ BETA = 0.0001
+ VTH0 = 0.8
+ TOX = 1.639E-08
+ K1 = 0.851
+ CGSO = 0
+ CGDO = 0
+ FC = 0.5
+ VBR = 100
+ IS = 1E-14
+ N = 1
+ TAU = 0
+ A0 = 0.01
+ A1 = 0.001
+ A2 = -0.001
+ A3 = -0.0001
+ GAMMA = 0.5
+ VDSO = 5
+ RDSO = 1E+12
+ VTO = -2
+ ALPHA = 2
+ LAMBDA = 0
+ R1 = 0
+ R2 = 0
+ RF = 0
+ CGS = 0
+ CGD = 0
+ CDS = 0
+ RIN = 0
+ A5 = 0
+ VDSDC = 0
.ENDS
```

A File Formats

B
Creating and Accessing Data

Creating a New File

To create a new file for data:

1. Choose **File > New** to display the New File dialog box.
2. Enter a name and path for the new data file.
3. Click **OK** to create and display a new data file in the specified location.

Only one data file can be open at a time. Before you create and display the new data file, you will be presented with an option to save and close the currently open data file.

Opening an Existing File

To open an existing data file:

1. Choose **File > Open** to display the File Open dialog box.
2. Enter a path and name for the file to be opened, or select the file from the list.
3. Click **OK** to open the specified data file.

Only one data file can be open at a time. If a file is currently opened, you will be presented with an option to save and close the currently open data file.

Opening an Example File

To open an example data file:

1. Choose **File > Examples** to display the Examples Open dialog box.
2. Select the example file you want from the list.
3. Click **OK** to open the specified data file.

Only one data file can be open at a time. If a file is currently opened, you will be presented with an option to save and close the currently open data file.

Loading Data

To load data directly from Main IC-CAP into Statistics you should load the model file called load_stat_data.mdl, which contains macro files. This file is shipped with the software and found in the Examples area.

Before you use the macro, do the following:

1 In order to use the macro, your model files must be in the following format:
`<base filename>_<n>.mdl`

where n is an integer equal or greater than 1, and the first model file must start with 1. For example, bsim_1.mdl, bsim_2.mdl, bsim_3.mdl, or, eebjt_1.mdl, eebjt_2.mdl, eebjt_3.mdl, etc.

2 All the model files must reside in the same directory. You will indicate the pathname and the number of model files using model variables.

3 If you want attribute information to be automatically transferred from the .mdl files, then the attribute values and labels (attribute parameter names) will have to be stored in variable arrays in the .mdl files. The attribute labels should be stored in an ICCAP_ARRAY called ATTRIBUTE_LABELS. The attribute values should be stored in an ICCAP_ARRAY called ATTRIBUTE_VALS. The size of the arrays is arbitrary, but you must explicitly specify it in the variable NUM_ATTRIBUTES.

Now you are ready to use the macro.

1 From the Main IC-CAP program choose **File > Examples**. The Examples Open dialog box appears.

2 Select the **/ statistics / load_data** directory and then the load_stat_data.mdl file from the list of files and choose **OK**. The model file loads.

3 Choose the **Model Variables** tab. A table appears with four items. Ignore the first item (called Local_VAR). Enter data for the three remaining items, as follows:

Name	Value
numFiles	Enter the total number of model files you want to load into Statistics
filesDirPath	Enter the path to the directory that contains the model files
baseFileName	Enter the base filename for your model files, such as bsim or eebjt, as described above.

4 Choose the **Macros** tab and select **Main** in the Select Macro field.

5 Choose the **Execute** button (left). The Statistics Parameters spreadsheet is filled with data.

Saving Data

To save data in a new or existing file:

1 Choose **File > Save As** to display the Save As dialog box.

2 Select the data set(s) you wish to save, such as Parameters, Correlation Matrix, etc. The default is to save all of the data sets.

3 Enter the filename and path for the data to be saved.

4 Click **OK** to save the data in the specified file.

Click **Apply** instead of OK to repeat the process and save one or more data sets in separate files.

When you choose **File > Close** or **File > Exit Statistics**, a dialog box appears with three choices:

- Choose the **Yes** button to save your changes, as described starting in step 2, above.
- Choose the **No** button to discard your changes.
- Choose the **Cancel** button to continue without closing your file or exiting Statistics.

Importing Data

To import data into IC-CAP Statistics:

1 Choose **File > Import** to display the Import File dialog box.

2 Use the dialog box to locate the file you wish to import.

3 Click **OK** to import the data.

Exporting Data

To export data from IC-CAP Statistics:

1 Choose **File > Export**. A submenu is displayed. Choose the format you want. from three choices:

Text File. Data from the Parameters folder is exported in a standard row/column, space delimited ASCII text format that may be used in other programs. Use the dialog box to specify a filename and location for the data you wish to export.

SPICE Library File. Converts the data for desired samples into a form usable as a SPICE library. Use the dialog box to specify a filename and location for the data you wish to export.

SPICE Equations. Converts the data in the Equations folder into a form usable in a SPICE netlist. Use the dialog box to fill in the model file name, the sample row numbers you want to be saved, and a filename for the export file.

2 Click **OK** to export the data.

B Creating and Accessing Data

C
Analyzing Data

Correlation Analysis

Correlation analysis is used to predict the value of a variable based upon information on an independent variable.

Correlation analysis is the first step in creating a parametric model. Parametric analysis assumes that your data is Gaussian. You can help make your data Gaussian by eliminating outliers using data filtering (see "Filtering Data" on page 175) and/or performing data transformation (see "Transforming Data" on page 176). After correlation analysis, you perform factor analysis and generate equations. Then you choose Parametric Analysis to build your model.

To perform a correlation analysis of the data:

1 Choose **Analysis > Correlation Analysis**.

2 The data is analyzed, reorganized, and presented in the Correlation Matrix folder.

Click the **Parameters** tab if you wish to view the data as it was presented before the correlation analysis.

Factor Analysis

Factor analysis is used to explain relationships among several, difficult to interpret, and correlated variables using a few, conceptually meaningful, and relatively independent factors. It does so by assessing the underlying relationships or dimensions in the data and replacing them with fewer variables.

To perform a factor analysis of the data:

1 Choose **Analysis > Factor Analysis** to display the Factor Analysis dialog box.

2 Enter an initial figure for the number of factors you want to use to explain the relationships. This number can be changed to repeat the analysis, as desired.

3 Select the factor analysis method from one of three choices:

4 Principal Component –The principal component model of factor analysis. Direct calculation; no iterations.

Principal Factor–Principal factor analysis. Direct calculation; no iterations.

Unweighted Least Squares–A method of factor analysis using an iterative process. This methods is also knows as *minres* or the minimum residual method.

5 Optionally, select a rotation type or accept the default of None.

6 Optionally, choose the **Iteration Control** button (upper right) to fine tune the Unweighted Least Squares and Rotation functions.

7 Click **OK** to perform the factor analysis.

The results of the analysis are automatically displayed in the Factor Loadings folder.

Two tables are generated in the Factor Loadings folder. The first table contains the factor loadings, which represent the loadings (correlations) that relate each model parameter to each of the derived factors. This table also has a column

called Communality, always displayed at the far right part of the table. This field shows the variance explained by all of the factors for a single parameter.

The top portion of the Factor Loading folder displays the data in a color-coded format. Factor Group data, one group per row, is displayed in a red font. Dominant Parameter data, one dominant parameter per column, is displayed with a blue background.

The second table has three fields:

Variance. Presents a summary of the variances associated with each factor. For example, a variance of 3.45 indicates that the factor accounts for as much variance in the data collection as would 3.45 variables, on average.

% Variance. Shows how much of the variance of all the parameters is explained by a single factor.

Cumulative %. Shows how much of the variance of all the parameters is explained cumulatively by from one to all of the factors. That is, as you move left to right in the table, the percentage increases as more and more factors are included.

Generating Equations

To generate equations from the data after factor analysis:

1 Choose **Analysis > Generate Equations > Factors** (or Dominant Parameters).

2 The equations are generated and presented in the Equations folder.

The equations are displayed in two tables. The upper table displays factor based equations and the lower table displays dominant parameter based equations.

Parametric Analysis

Parametric analysis assumes that your data is Gaussian and requires three prerequisites: Performing correlation analysis, factor analysis, and generating equations. See "Correlation Analysis" on page 152 for starting information.

Parametric analysis can be performed from either:

- Factor Equations
- Regression Equations

To build a parametric model based on Factor Equations:

1. Choose **Analysis > Parametric Analysis > Factor Equations**.
2. A dialog box is displayed with three choices: Monte Carlo, Corner, and Parametric Boundary. Choose one.
3. If you select Monte Carlo, you fill in the number of outcomes you want in the field below. If you select Corner or Parametric Boundary, you fill in the number of +/- sigmas you want. When done, choose **OK**.
4. The Parametric Analysis Results window is displayed. The upper spreadsheet displays the nominal point of a corner or parametric boundary analysis. The lower spreadsheet contains the rows for the samples (or Monte Carlo outcomes) and columns for the parameters.

To build a parametric model based on Regression Equations:

1. Choose **Analysis > Parametric Analysis > Regression Equations**.
2. A dialog box is displayed with four choices: Parameters Spreadsheet, Monte Carlo, Corner, and Parametric Boundary. Choose one.
3. Choose **Parametric Spreadsheet** to display the Parametric Analysis Results window with three columns for each parameter: the raw or measured data, the simulated data and the residual (difference) between the two.
4. Before you build a Monte Carlo, Corner, or Parametric Boundary model, you can choose the **Input Statistics** button to display a dialog box.

5 The Input Statistics dialog box displays the dominant parameters as well as the calculated mean and standard deviation for those parameters. From this dialog box you can change this data by typing in the cells, or choose the **Import** button to import this data from a text file.

6 When you are ready to build your model, and you select Monte Carlo, fill in the number of outcomes you want in the field below. If you select Corner or Parametric Boundary, you fill in the number of +/- sigmas you want. When done, choose **OK**.

7 The Parametric Analysis Results window is displayed. The upper spreadsheet displays the nominal point of a corner or parametric boundary analysis. The lower spreadsheet contains the rows for the samples (or Monte Carlo outcomes) and columns for the parameters.

Non-Parametric Analysis

To build a non-parametric model using a given data set:

1 Load the data and filter out the outliers either manually or automatically.

For details on eliminating outliers through data filtering, refer to "Filtering Data" on page 175.

2 Transform the data if a given parameter naturally ranges over more than one decade of values (e.g., 10-100).

For details on transforming data, refer to "Transforming Data" on page 176.

3 Choose **Analysis > Non-Parametric Analysis** to display the Non-Parametric Analysis dialog box.

4 Enter the desired number of boundary points and the percentage of data to be enclosed.

For background details on modifying these options, refer to "Using the Non-Parametric Analysis Dialog Box" on page 79.

5 Enter the desired diversity oversampling and the density estimator percentage. For background details on modifying these options, refer to "Using the Non-Parametric Analysis Dialog Box" on page 79.

6 Click **Check and Adjust Inputs** to check and adjust the inputs for boundary points, percent enclosed, and diversity oversampling.

7 Click **OK** to automatically analyze and display the results in a separate window.

Click **Apply** instead of OK to repeat the process and perform an analysis using different settings.

Analysis Data

The Analysis data feature displays the data, active parameters and samples, currently being used for analysis.

To display the data currently being used for analysis:

- Choose **Analysis > Row Analysis Data** to display the data in a separate window.

Statistical Summary

The Statistical summary feature displays the standard statistical measures of the data currently being used for analysis.

To display the statistical summary for the data:

- Choose **Analysis > Statistical Summary** to display the statistical summary in a separate window.

  The measures calculated and displayed in this window include both measures of central tendency and of dispersion. For each parameter, the measures are calculated using samples that have not been deactivated, filtered, or deleted.

  Mean, Variance, Standard Deviation, Skewness, Kurtosis, Minimum, Maximum, Median, Median Absolute Deviation, and Robust Estimate of Scale are the measures calculated.

Residual Correlation

Residual correlation is used to examine nonconstant variance between residual and predicted values during regression analysis.

To display the residual correlation for the data:

- Choose **Analysis > Residual Correlation** to display the residual correlation in a separate window.

  When the time sequence of residual values is known, similar values that are adjacent in time constitute a positive serial correlation; dissimilar values that are adjacent in time constitute a negative serial correlation.

D
Editing and Formatting

Swapping Data

To swap data from one row or column to another:

1. Choose **Edit > Swap** to display the Swap dialog box.
2. Select **Row** or **Column** to identify the data.
3. Enter the source (From) location of the data in terms of the row or column number.
4. Enter the target (To) location of the data in terms of the row or column number.
5. Click **OK** to move data from the source to the target location. Any data that was previously displayed in the target location is moved to the source location.

   Click **Apply** instead of OK to repeat the process and swap more than one row or column of data.

Moving Data

To move data from one row or column to another:

1. Choose **Edit > Move** to display the Move dialog box.
2. Select **Row** or **Column** to identify the data.
3. Enter the source (From) location of the data in terms of the row or column number.
4. Enter the target (To) location of the data in terms of the row or column number.
5. Click **OK** to move data from the source to the target location.

   Click **Apply** instead of OK to repeat the process and move more than one row or column of data.

Copying Data

To copy data from one row or column to another:

1. Choose **Edit > Copy** to display the Copy dialog box.
2. Select **Row** or **Column** to identify the data.
3. Enter the source (From) location of the data in terms of the row or column number.
4. Enter the target (To) location of the data in terms of the row or column number.
5. Click **OK** to copy data from the source to the target location. Any data that was previously displayed in the target location is replaced by the copied data.

   Click **Apply** instead of OK to repeat the process and copy more than one row or column of data.

Inserting Rows or Columns

To insert one or more rows or columns:

1. Select the row or column before which you wish to insert one or more rows or columns.
2. Choose **Edit > Insert** to display the Insert dialog box.
3. Select **Row** or **Column** to define what is to be inserted.
4. Enter the number of rows or columns to be inserted.
5. Click **OK** to insert the specified number of rows or columns.

   Click **Apply** instead of OK to repeat the process and insert rows or columns at more than one location.

Sorting Data

To sort data within a row or column:

1. Select the row or column containing the data to be sorted.
2. Choose **Edit > Sort** to display the Sort dialog box.
3. Select **Ascending** or **Descending** to identify the sort order.
4. Click **OK** to sort the data.

   Click **Apply** instead of OK to repeat the process and sort more than one row or column of data.

Changing Row Height

To change the height of a single row:

Position the pointer on the row divider and drag in the desired direction to change (increase or decrease) the height of an individual row.

To change the height of all rows:

1 Choose **Format > Row Height** to display the Row Height dialog box.

2 Enter a number to specify the desired row height.

3 Click **OK** to change the row height.

Click **Apply** instead of OK to repeat the process and change the row height more than once.

Changing Column Width

To change the width of a single column:

Position the pointer on the column divider and drag in the desired direction to change (increase or decrease) the width of an individual column.

To change the width of all columns:

1. Choose **Format > Column Width** to display the Column Width dialog box.
2. Enter a number to specify the desired column width.
3. Click **OK** to change the column width and close the dialog box.

   Click **Apply** instead of OK to repeat the process and change the column width more than once.

Changing Colors

To change the colors used in a spreadsheet:

1. Choose **Format > Color** to display the Change Color dialog box.
2. Choose the desired color from the drop-down list associated with each of the six areas: Table background color, Table foreground color, Selection background color, Deactivate color, Attribute color, and Data Filter color.
3. Click **OK** to change the color and close the dialog box.

   Click **Apply** instead of OK to repeat the process and change the color more than once using one or more different colors.

Configuring Plots

To configure the options for a plot:

1 Choose **Options > Plot Configurations** to display the Plot Configuration dialog box.

2 Select the available options for the desired plot type.

Histogram: Enter the desired number of bins for organizing the data, and select one or both, normalized and Gaussian fit.

Cumulative Density Plot: Select to display using a Gaussian fit.

Scatter Plot: Select two or more columns of data to display a scatter plot.

3 Click **OK** to configure the plot and close the dialog box.

Click **Apply** instead of OK to repeat the process and configure the same or a different plot.

E
Filtering and Transforming Data

Filtering an Attribute

To filter data for an attribute:

1. Select one or more columns of attribute data (such as ID, lot number, date, etc.).
2. Choose **Data > Attribute Filter** to display the Attribute Filter dialog box.
3. Select the filtering criteria.

   If you select to filter by Text, enter the text that should be matched. If you select to filter by range, enter the minimum and maximum values of range to be matched.
4. Select the type of filtering.

   Use single pass filtering to deactivate data for one filtering criterion. With cumulative filtering, data deactivated for one operation remains deactivated during subsequent filtering passes, regardless of the criterion.
5. Click **OK** to filter the data.

   Click **Apply** instead of OK to repeat the process.

   To undo the last or all attribute data filtering, choose **Data > Undo Attribute Filtering**.

Filtering Data

To filter data:

1 Select one or more columns of data (by clicking on the column label).

2 Choose **Data > Data Filter** to display the Data Filter dialog box.

3 Select the filtering criteria.

If you select to filter by Scale, specify the scale that should be matched. If you select to filter by minimum and maximum, enter the minimum and maximum values of range to be matched.

4 Select the type of filtering.

Use single pass filtering to deactivate data for one filtering criterion. With cumulative filtering, data deactivated for one operation remains deactivated during subsequent filtering passes, regardless of the criterion.

5 Click **OK** to filter the data.

Click **Apply** instead of OK to repeat the process and filter more data or filter the data some more.

To undo the last or all data filtering, choose **Data > Undo Data Filtering**.

Transforming Data

To transform data:

1. Select one or more columns of data (by clicking on the column label).
2. Choose **Data > Data Transformations** to display the Data Transformations dialog box.
3. Choose the desired transform from the Data Transformations drop-down list.

   If you choose Constant Value as the transform, enter a value.
4. Click **OK** to transform the data.

   Click **Apply** instead of OK to repeat the process and transform more data or further transform the data.

Index