SIGMAPLOT®
Exact Graphs and Data Analysis

Statistical Analysis using SigmaPlot

Statistical Analysis using SigmaPlot

- Statistics are not difficult at all. Statistics are mathematical methods of interrogating data.
- You will be surprised to learn how familiar statistics are to us.
- Without mental statistics we would not know when to leave our house in the morning and how much money we need for shopping or a night out.
- Statistics is a mathematical science pertaining to the collection, analysis, interpretation or explanation, and presentation of data. It is applicable to a wide variety of academic disciplines, from the natural and social sciences to the humanities, government and business.

Statistics in a few simple steps

1. Make an initial appraisal of your data.
2. Select the type of test you require based on the question you are asking.

3. Select the actual test you need to use from the appropriate key.

4. Determine any preliminary tests you need to carry out prior to performing the statistical test

5. If your data are suitable for the test chosen based on the results from 4 proceed to the test

6. If your data do not meet the demands of the chosen test go back to 3 and choose the non-parametric equivalent.

Data Types

What type of data do you have?

Some statistical tests are sensitive to the type of data you have and it is necessary to identify which data you have before you choose a test.

Firstly all measurements fall into one of two overriding categories continuous or discontinuous.

- Continuous - Measurements of length are continuous whereby fractions can be included. If you have rounded up to whole figures you can consider your data to be continuous.

- Discontinuous - Generally counts of things (frequencies) whereby a fraction of an individual is impossible (although you may only have part of that individual).

Data Types

- Data then falls into several other categories in order of complexity starting with the simplest.

- Nominal - according to categories i.e. species of plant or colour of eyes. These data are most often associated with the Chi2 and contingency table. But they are not excluded from other tests.

- Ordinal - categories but in ascending or descending ranks.

- Interval - describes a nominal data set in which the units are the same size throughout the scale i.e. the difference between 21 and 27 is the same as between 1 and 7. Temperature is a good example.

- Ratio - This is the highest level of data complexity and can include all the previous categories. Such scales have zeros and are continuous. Linear dimensions are a good example.

Data Types

In addition to the previous slides there are other categories that the data may also fall into. It is not necessary to apply these to statistical tests unless stated.

- Qualitative - do your data describe a quality rather than a measurement? i.e. species, male/female or colour describe qualities and normally the data that follow are counts for these categories.

- Quantitative - measurements of a variable normally indicate a quantitative variable. Length of femur or weight of gonads are quantitative values.

- Derived variables - these are data that have not been measured directly but have been calculated from measurements. The most common derived variables are proportions (ratios and percentages).

Descriptive Statistics

Descriptive statistics are used simply to describe the sample you are concerned with.

They are used in the first instance to get a feel for the data, in the second for use in the statistical tests themselves, and in the third to indicate the error associated with results and graphical output.

For instance, when have you taken a trip to see a friend without a quick estimate of the time it will take you to get there (= mean)? Very often you will give your friend a time period within which you expect to arrive "say between 7.30 and 8.00 traffic depending". This is an estimate of the standard deviation or perhaps standard error of the times taken in previous trips. The more often you have taken the same journey the better the estimate will be.

Comparisons

1. Have you got more than two samples?

No : go to 2

Yes : go to 8

2. Have you got one or two samples?

One : Single sample t-test

Two : go to 3

3. Are your data sets normally distributed (K-S test or Shapiro-Wilke)

No : go to 4

Yes : go to 5

4. Do your data sets have any factor in common (dependence), i.e. location or individuals?

No : Mann Whitney U Test

Yes : Wilcoxon Matched Pairs

5. Do your data sets have any factor in common (dependence), i.e. location or individuals?

No : go to 6

Yes : paired sample t-test

Comparisons

6. Do your data sets have equal variances (f-test)?

    No : unequal variance t-test

    Yes : go to 7

7. Is n greater or less than 30?

    <30 : equal variance t-test or ANOVA

    >30 : z-test or ANOVA

8. Are your samples normally distributed and with equal variances?

    No : Kruskal-Wallis non-parametric ANOVA

    Yes : go to 9

9. Does your data involve one factor or two factors?

    One : One-way ANOVA

    Two : Two-way ANOVA

Testing of Hypothesis

Hypothesis testing is one of the most important tools of application of Statistics to real life problems.

There are two types of Hypothesis:

Null Hypothesis (Ho)
Alternative Hypothesis (H1)

Example:

Ho : There are no differences between the means of two samples.

H1 : There is a difference between the means of two samples

The Null Hypothesis, H0

Typically represents the hypothesis that there is “no association”or “no difference”

It represents current “state of knowledge”(i.e., no conclusive research exists)

For example, there is no association between alcohol intake and blood pressure$H0$: μ= 0

Normality Tests

Kolmogorov-Smirnov Test

Shapiro-Wilke Test

The simplest method of assessing normality is to look at the frequency distribution histogram. The most important things to look at are the symmetry and peakiness of the curve.

ANOVA

ANOVA stands for ANalysis Of VAriance. In general, this statistical procedure compares the variance of scores in each of the groups of an experiment. With the two sample t-tests, we can have experiments with only two groups. With ANOVA, we can have as many groups as we would like to have.

A one-way ANOVA has one independent variable with a number of groups. For example, the IV might be Prozac and there might be three groups - one group is given no pill each day (the control group), one group is given one pill a day and a third group is given two pills a day.

Limitations of the ANOVA test

if you do not find a significant difference in your data, you cannot say that the samples are the same

ANOVA will only indicate a difference between groups, not which group(s) are different. For the latter you will need to use a multiple comparison test.

A two-way ANOVA has two independent variables. This would be the case if a researcher looked at the effect of caffeine and sugar on mood. There are two independent variables - caffeine and sugar. This experiment might have four groups:

Group 1: low level of caffeine & low level of sugar
Group 2: low level of caffeine & high level of sugar
Group 3: high level of caffeine & low level of sugar
Group 4: high level of caffeine & high level of sugar

A three-way ANOVA would have three independent variables.

In general, ANOVA compares the variance of scores within a group (people that got the same level of the IV) to the variance between the groups (people that got different levels of the IV).

Paired t-test

Test: The paired t-test is actually a test that the differences between the two observations is 0. So, if D represents the difference between observations, the hypotheses are:

Ho: D = 0 (the difference between the two observations is 0)

H1: D 0 (the difference is not 0)

The test statistic is t with n-1 degrees of freedom. If the p-value associated with t is low (< 0.05), there is evidence to reject the null hypothesis. Thus, you would have evidence that there is a difference in means across the paired observations.

Repeated Measures ANOVA

As with any ANOVA, repeated measures ANOVA tests the equality of means. However, repeated measures ANOVA is used when all members of a random sample are measured under a number of different conditions.

As the sample is exposed to each condition in turn, the measurement of the dependent variable is repeated. Using a standard ANOVA in this case is not appropriate because it fails to model the correlation between the repeated measures: the data violate the ANOVA assumption of independence.

Keep in mind that some ANOVA designs combine repeated measures factors and non-repeated factors. If any repeated factor is present, then repeated measures ANOVA should be used.

Correlation

Correlation (often measured as a correlation co-efficient), indicates the strength and direction of a relationship between two random variables.

Correlation has direction and can be either positive or negative. With a positive correlation, individuals who score high (or low) on one measure tend to score similarly on the other measure. The scatterplot of a positive correlation rises.

With negative relationships, an individual who scores high on one measure tends to score low on the other (or vise verse). The scatterplot of a negative correlation falls.

Correlation

A correlation can differ in the degree or strength of the relationship (with the Pearson product-moment correlation coefficient that relationship is linear).

Zero indicates no relationship between the two measures and r = 1.00 or r = -1.00 indicates a perfect relationship. The strength can be anywhere between 0 and + 1.00. Note:

The symbol r is used to represent the Pearson product-moment correlation coefficient for a sample.

The stronger the correlation--the closer the value of r (correlation coefficient) comes to + 1.00--the more the scatterplot will fall along a line.

Regression Analysis

Regression analysis is a technique used for the modeling and analysis of numerical data consisting of values of a dependent variable (response variable) and of one or more independent variables (explanatory variables).

The dependent variable in the regression equation is modeled as a function of the independent variables, corresponding parameters ("constants"), and an error term. The error term is treated as a random variable. It represents unexplained variation in the dependent variable.

The parameters are estimated so as to give a "best fit" of the data. Most commonly the best fit is evaluated by using the least squares method, but other criteria have also been used.

Regression Analysis

Regression can be used for prediction (including forecasting of time-series data), inference, hypothesis testing, and modelling of causal relationships.

These uses of regression rely heavily on the underlying assumptions being satisfied.

Regression Analysis

Multiple Logistic Regression

Use a Multiple Logistic Regression when you want to predict a qualitative dependent variable, such as the presence or absence of a disease, from observations of one or more independent variables, by fitting a logistic function to the data.

The independent variables are the known, or predictor, variables. When the independent variables are varied, they produce a corresponding value for the dependent, or response, variable. SigmaStat?'s Logistic Regression requires that the dependent variable be dichotomous or take two possible responses (dead or alive, black or white) represented by values of 0 and 1.

Survival Analysis

Use survival analysis to generate the probability of the time to an event. For example, a survival curve shows as a function of time the probability of surviving lung cancer.

Survival analysis studies the variable that is the time to some event. The term survival originates from the event death. But the event need not be death; it can be the time to any event. This could be the time to closure of a vascular graft or the time when a mouse footpad swells from infection.

Of course it need not be medical or biological. It could be the time a motor runs until it fails. For consistency we will use survival and death (or failure) here.

Survival Analysis

Types of Survival Analysis

SigmaPlot provides three types of Kaplan-Meier survival analysis and uses two data formats. The types are the analysis of a single curve, the comparison of multiple curves using the LogRank test and the comparison of multiple curves using the Gehan-Breslow test.

. Single Group: Use this to analyze and graph one survival curve.

. LogRank: Use this to compare two or more survival curves. The LogRank test assumes that all survival time data is equally accurate and all data will be equally weighted in the analysis.

Survival Analysis

. Gehan-Breslow Use this to compare two or more survival curves when you expect the early data to be more accurate than later. Use this, for example, if there are many more censored values at the end of the study than at the beginning.

If the LogRank or Gehan-Breslow statistic yields a significant difference in survival curves then you have the option to use one of two multiple comparison procedures to determine exactly which pairs of curves are different. These are the Bonferroni and Holm-Sidak tests and are described for each test.

Survival Analysis

Data Format for Survival Analysis

Your survival data will consist of three variables:
. Survival time
. Status
. Group

The survival times are the times when the event occurred. They must be positive and all non-positive values will be considered missing values. The data need not be sorted by survival time or group.

The status variable defines whether the data is a failure or censored value. You are allowed to use multiple names for both failure and censored. These can be text or numeric.

The group variable defines each individual survival data set (and curve).

Transforms

Transforms are used to modify the existing data or calculating new data from already existing data.

There are various transformations available in SigmaPlot:

You can add, subtract, multiply, divide, etc...

Also you can standardize, center, generate random numbers, filter, rank data and so on using the transform feature.

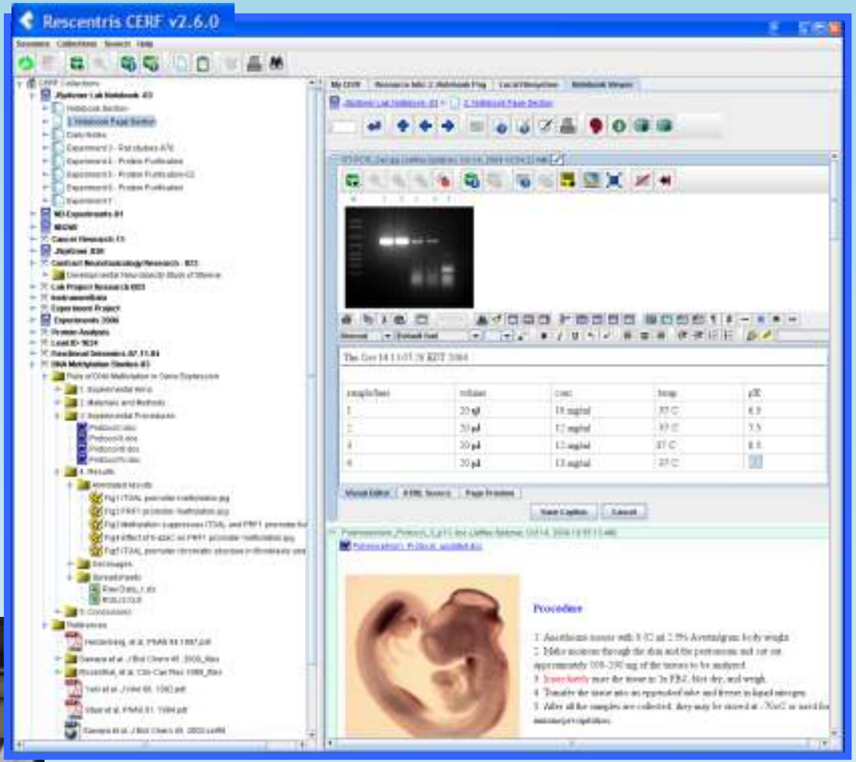
Apart from this you can also use user defined transform where you specify formulas and if conditions.

More on this feature will be demonstrated.

Special Offer: SigmaCERF boosts R&D performance –increases research efficiency, quality, collaboration

Observations & Discovery

SigmaCERF unifies the scientist's information world

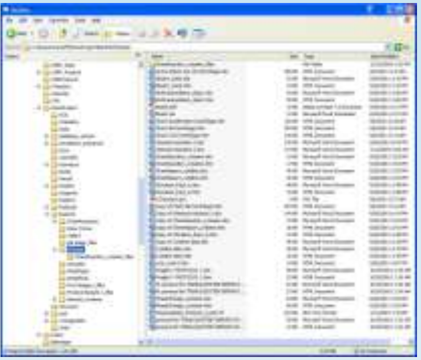


Paper & digital archives

research content in compliance with regulatory requirements.

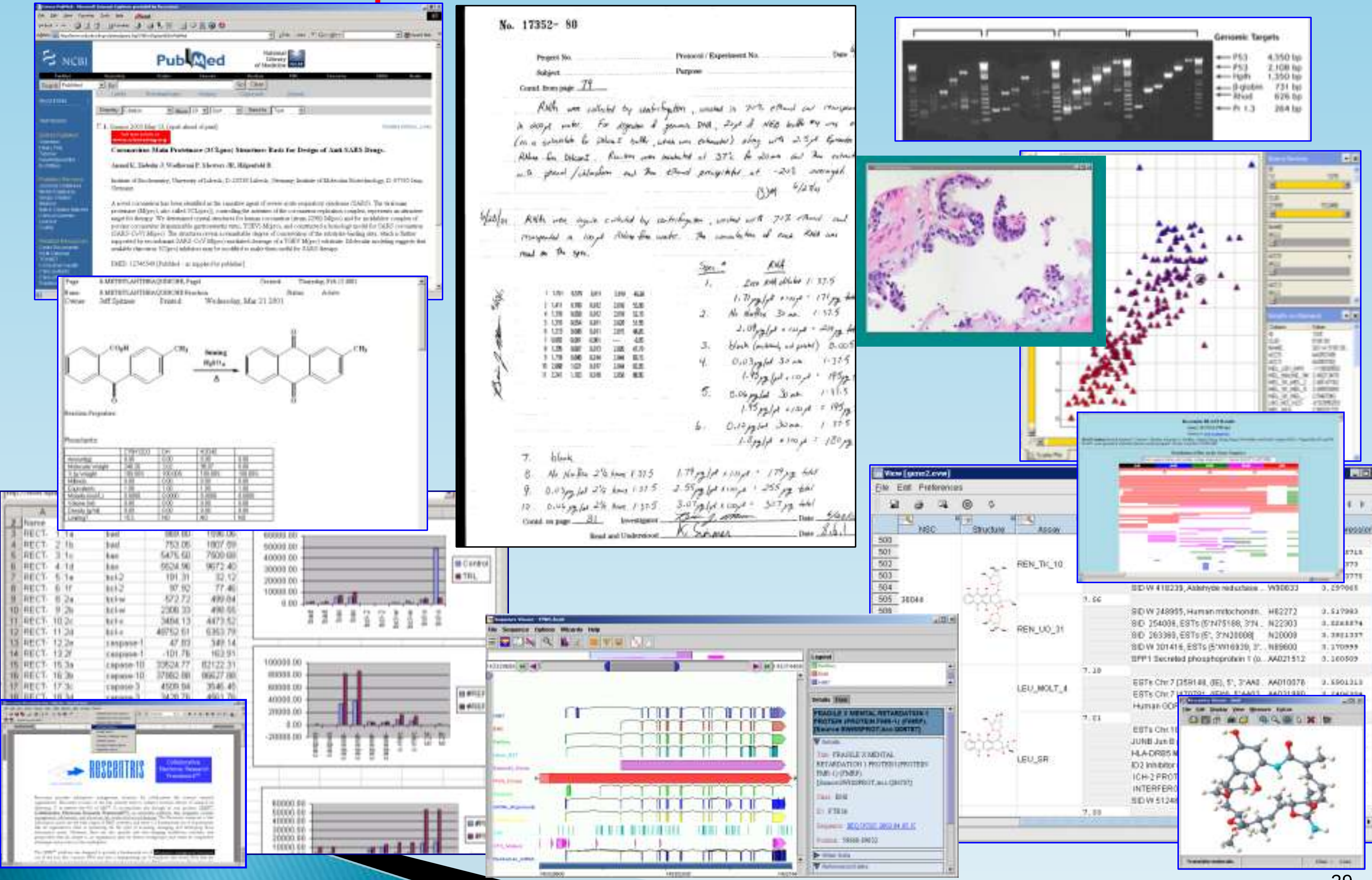
Printing electronic data - Excel, etc.

Currently: information scattered across paper, PCs, network drives, DMS, LIMS, databases



File Systems and DMS

SigmaCERF enables researchers to manage experimental data, informatics, documents, databases – and capture their work



No. 17352- 80

Project No. Protocol / Experiment No. Date

Subject Purpose

Cont'd from page 79

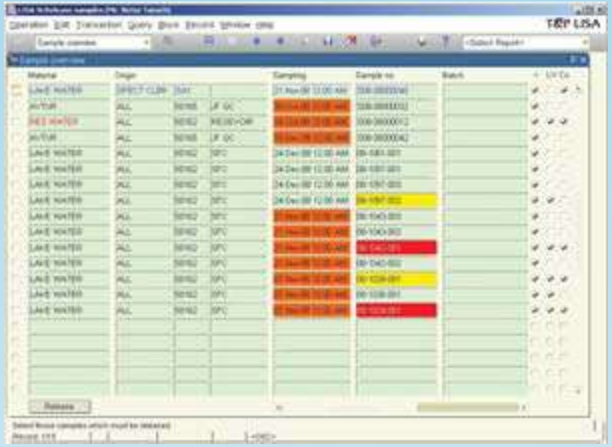
Cont'd on page 81 Investigator Date

Read and Understood K. Simmer Date

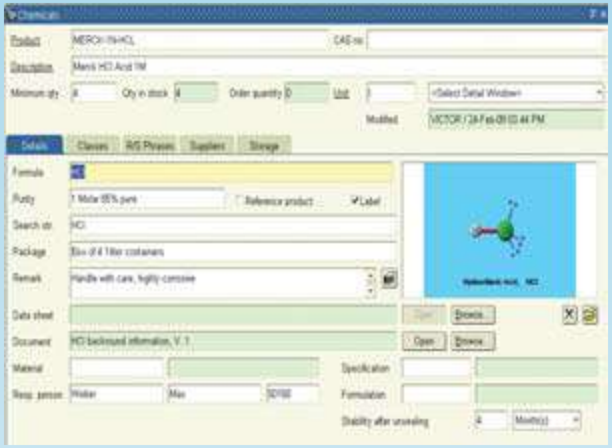
LISA.lims, the Premier Laboratory Information Management System

SigmaPlot
Exact Graphs and Data Analysis

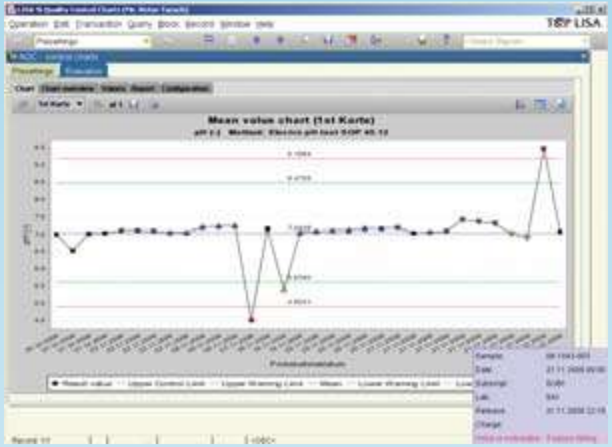
- Lets automate, organize, and standardize complex workflows for tracking crucial information and creating sophisticate reports



The Dashboard lets you quickly identify limit violations and samples ready for release



Complete chemical and formulation management

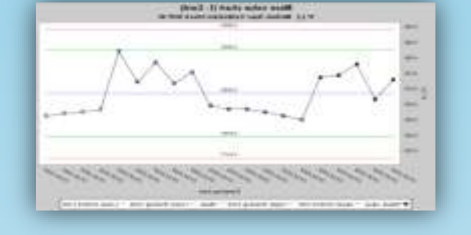


Identify trends quickly with automatically generated quality control charts

Module Based Design to Grow With Your Business and Accommodate Your Budget

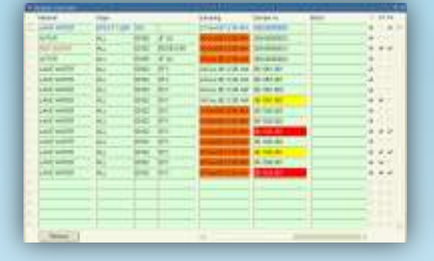
SAP integration

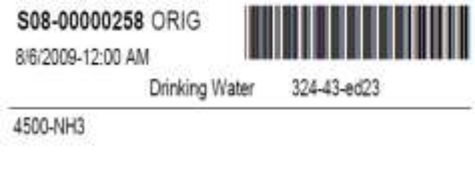
Automatic Data



Chemical Management

Laboratory Basics





S08-00000258 ORIG
8/6/2009-12:00 AM
Drinking Water 324-43-ed23
4500-NH3

Web-Services

Document

Stability Testing

Capacity Planning

Mobile Data Recording

Raw Data Archiving

Data Management Offers

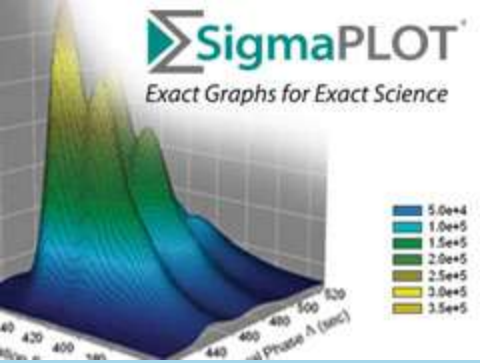
SigmaPlot
Exact Graphs and Data Analysis

- Free Trial Sign up here: http://www.sigmaplot.com/products/SIGMA_CERF/demo.php
- Webinar Signup: https://www1.gotomeeting.com/register/977570744
- PPT Download: http://www.sigmaplot.com/products/SIGMA_CERF/download.php

SigmaCERF

- Demonstration Sign up http://www.sigmaplot.com/products/lisa/demo.php
- PPT Download http://www.sigmaplot.com/products/lisa/ppt.php

LISA.lims



SIGMAPLOT
Exact Graphs and Data Analysis

Special Offer For Registered Attendees!!!

$50 off & Free Shipping of Single Box Purchases.

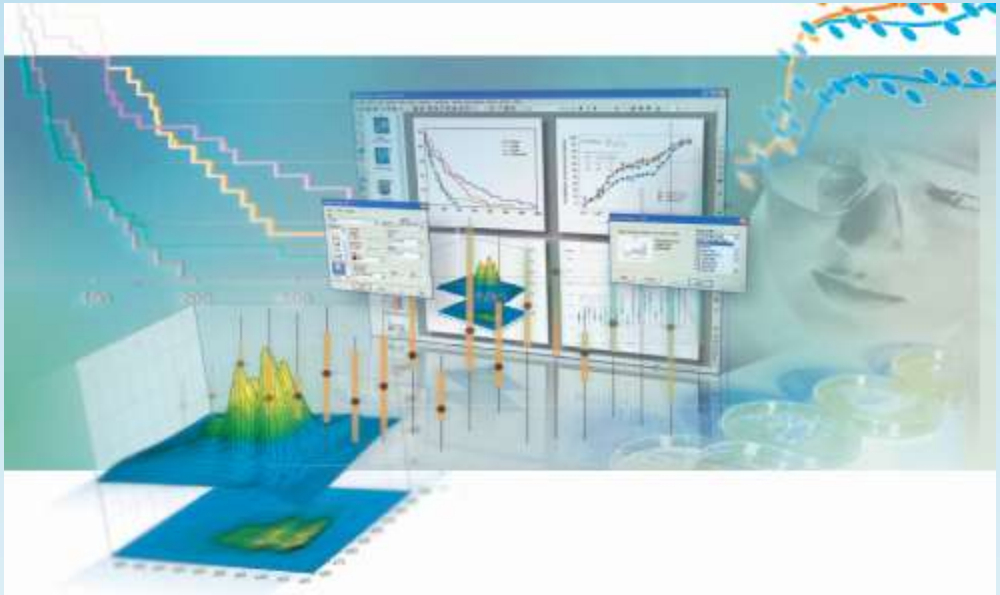
20% of multiuser licenses

USE PROMO CODE:

SigmaWBNR7

Expires on 8/15/2010

Not available for internet purchases



Contact Information

Systat Software Inc

Contact the Sales/Marketing Department (USA & Canada):

Email:	vtaruch@systat.com
Phone:	+1 (800) 797-7401
Fax:	+1 (800) 797-7406

Contact Technical Support (USA & Canada):

Email:	techsupport@systat.com
Phone:	+1 (408)452-9010

Thank You !