



Dear Readers,

When SAS opened for business in 1976, the company had seven employees selling one software program. Today, SAS employs more than 10,000 people in 424 offices worldwide, and our solutions are used at more than 40,000 sites.

As we celebrate our 30th anniversary as the leader in business intelligence – the hottest sector in the software market today – we anticipate an even greater future. SAS has been on the forefront of information technology for 30 years, maintaining an unbroken string of 29 years of growth and profitability. Revenues – totaling \$138,000 in 1976 – topped \$1.68 billion in 2005. And there's every indication our growth will continue.

We have you, our customers and our friends, to thank for our unprecedented success. We wouldn't be here without you.

Here's to another 30 years!

A handwritten signature in black ink that reads "Shelley Sessoms".

Shelley Sessoms

Editor, *Your SAS Technology Report*

FAQ # 4538

Q: What is the difference in using the SAS®9 function CATX to concatenate text versus using the TRIM and LEFT functions with the concatenation operator (||)?

A: The results of concatenating text are equivalent with both methods. However, the CATX function requires less code and processes the concatenation faster than the combination method, which uses multiple calls.

The following examples show the difference in coding for both methods. Notice that in these examples,

- 'Nina' is the user-supplied value for FIRSTNAME.
- 'Werner' is the user-supplied value for LASTNAME.
- 'Nina Werner' is the resulting concatenated string represented by MEMNAME.

Example 1: Concatenating First and Last Names by Using the Combination Method

```
memname=trim(left(firstname) || ' ' || left(lastname));
```

Example 2: Concatenating First and Last Names by Using the CATX Function

```
memname=catx(' ', firstname, lastname);
```

For environments that support variable lists, you can use the OF syntax in the CATX function to condense variable lists that you are concatenating (Example 4 below). Note: In the following examples, separator specifies the character string of your choice (comma, blank, and so on) that will be used to separate the concatenated strings.

Example 3: Concatenating a Variable List by Using the Combination Method

```
x=trim(left(x1)) || separator || trim(left(x2)) ||separator ||  
trim(left(x3)) || separator || trim(left(x4));
```

Example 4: Concatenating a Variable List by Using the CATX Function

```
x=catx(separator, of x1-x4);
```

For detailed information, see "CATX Function" in SAS Language Reference: Dictionary under Base SAS in [SAS OnlineDoc 9.1.3](#).

About the Author

Nina L. Werner is a SAS programmer with more years experience than she wants to say.

FASTats: Frequently Asked-for Statistics

A through J

Jump to the [K through Z](#) list

Alpha (Cronbach's)

ALPHA option in Base SAS procedure CORR.

Alternating Logistic Regression (ALR)

Beginning in Version 7, use the LOGOR= option in the REPEATED statement in the SAS/STAT procedure GENMOD. ALR is a form of the Generalized Estimating Equation (GEE) method for modeling repeated (longitudinal) binary-response data. Not available before Version 7.

[ANOVA on summary statistics](#)

Automatic Interaction Detector (AID)

See [CHAID](#) below.

Autoregressive Conditional Heteroscedasticity (ARCH)

SAS/ETS PROC AUTOREG.

[Area under a curve, estimation of](#)

Area under ROC curve

See [ROC curve](#).

Balanced incomplete block designs (BIBDs)

BLOCKS statement in the SAS/QC procedure OPTEX beginning in Release 6.10. Use the `structure=(b)k` block specification to construct block designs with b blocks of size k . See the "*Balanced incomplete block design*" example in the OPTEX chapter of the *SAS/QC User's Guide*.

Bartholomew test of ordering of proportions

Not available. But the [Cochran-Armitage](#) test of linear trend, which tests a stricter alternative hypothesis, is available.

[Bartlett test for variance homogeneity](#)

Berkson estimation

See [Minimum chi-square estimation \(Berkson\)](#)

Bhapkar's test

Use the REPEATED statement in the SAS/STAT procedure CATMOD to test marginal homogeneity as in the test of SIDE in the example "*Repeated Measures, 4 Response Levels, 1 Population*" in the CATMOD chapter of the *SAS/STAT User's Guide*. See Agresti (1990), *Categorical Data Analysis*, pp. 359, 499. Bhapkar's test is asymptotically equivalent to the [Stuart-Maxwell test](#).

[Binomial probability, test of or confidence interval for Binomial probabilities, test comparing two](#)

Biserial correlation

[BISERIAL macro](#). See also [point biserial correlation](#).

Bivariate normal data, Generating

Use the [MVN macro](#).

Bivariate Logit model

Not available.

Bivariate Probit model

SAS/ETS PROC QLIM beginning in SAS 9.

Bivariate Tobit model

SAS/ETS PROC QLIM beginning in SAS 9.

Bonferroni t-test

SAS/STAT procedures GLM, ANOVA, and MULTTEST.

[Bootstrapping](#)

Box and whisker plots

Low-resolution: PLOTS option in Base SAS PROC UNIVARIATE.

High- or low-resolution: SAS/QC PROC SHEWHART and SAS/STAT PROC BOXPLOT.

High-resolution: SAS/GRAPH PROC GPLOT (I=BOXXXX option in SYMBOL statement) and SAS/INSIGHT.

Box-Behnken Designs

SAS/QC ADX Interface gives them. There is no ADX macro for these.

Box-Cox Transformation

Beginning with Release 8.2, use the BOXCOX transformation in the MODEL statement in SAS/STAT PROC TRANSREG to transform the response variable in a model on independent observations (fit by procedures such as SAS/STAT procedures REG or GLM). For example:

```
model BoxCox(y) = identity(x1-x5);
```

Also beginning in Release 8.2, you can use the BOXCOX option in the MODEL statement of SAS/ETS PROC QLIM. This option allows transformation of both dependent and independent variables. Prior to Release 8.2, use the ADX Interface or ADXTRANS macro in SAS/QC. For an autoregressive model, use the BOXCOXAR macro in SAS/ETS.

Box-Tidwell Transformation

Used in logistic regression to test linearity in the logit, this can be done in SAS/STAT PROC LOGISTIC by adding a variable to your model of the form $x \cdot \log(x)$. Significance of the parameter estimate for this term indicates nonlinearity.

Bradley-Terry model

SAS/STAT PROC LOGISTIC. See this [example](#).

Breslow-Day test (of homogeneity of odds ratios)

Base SAS PROC FREQ, CMH option.

Brier score

Not computed by any procedure, but [Logistic Regression Examples Using the SAS System](#) gives a formula, and it could be computed easily using predicted values from LOGISTIC, PROBIT, or GENMOD.

CART (Classification and Regression Trees)

CART is part of the capability of the [SAS Enterprise Miner](#).

Censored regression

SAS/ETS PROC QLIM beginning in Release 8.2

CHAID (Chi-square Automatic Interaction Detector)

CHAID is part of the capability of the [SAS Enterprise Miner](#).

Chi-square goodness-of-fit test for One-way tables

CHISQ, TESTP=, or TESTF= options in base SAS PROC FREQ, beginning in Release 6.12, provide an asymptotic Pearson chi-square test. The CHISQ option provides a test of equal cell probabilities. (An exact test is available beginning in Version 7 by specifying the CHISQ option in the EXACT statement.) The TESTP= and TESTF= options allow you to specify expected cell probabilities or frequencies.

Chi-square (2-way tables)

CHISQ option in Base SAS PROC FREQ

Chi-square (corrected)

CHISQ option in Base SAS PROC FREQ

Chow test

Beginning in Release 6.11, use the CHOW= or PCHOW= options on MODEL statement in SAS/ETS PROC AUTOREG or, beginning in Release 6.12, on the FIT statement in SAS/ETS PROC MODEL. The PCHOW= option produces the predictive Chow test. See Example 11 of the book [Forecasting Examples for Business and Economics Using the SAS System](#) for information on testing forecasting models for break points using a Chow Test.

Cluster analysis, categorical data

See [DISTANCES](#).

Cochran-Armitage trend test

See [Trend test for ordered alternatives](#)

Cochran's Q

Beginning in Release 6.10, create a 2x2x...x2 table and use the AGREE option in Base SAS PROC FREQ. See the example titled "[Testing Marginal Homogeneity with](#)

Cochran's Q" in the FREQ chapter of the *SAS/STAT User's Guide*. Before Release 6.10, create a three-way table with a stratum variable identifying each subject (or matched group), a variable indicating each occasion (condition or individual within matched group), and a binary response variable. Then use the CMH option. For example, if each subject gives a binary response to each of several drugs, use the statement:

```
tables subject*drug*response/cmh2 noprint;
```

Cohen's kappa

See [Kappa](#).

Combinations and Permutations

You may either want to [compute the number of combinations or permutations](#), or produce a list enumerating the combinations or permutations. For the latter, see the COMB and PERM options in SAS/STAT PROC PLAN (beginning in Version 7) or see "*Generating Combinations and Permutations*" ([TS-498](#)).

Concordance index

Is the [area under the receiver operating characteristic \(ROC\) curve](#)

Concordance, Kendall's Coefficient

See [Kendall's Coefficient of Concordance](#).

[Conditional logistic model](#)

Conditional poisson model

See [Fixed effects poisson regression](#).

Confidence ellipse for mean or for prediction

In SAS/INSIGHT, click **Analyze : Multivariate**. After running the analysis, in the analysis results window click **Curves : Scatter Plot Conf. Ellipse** and select Mean or Prediction and the desired confidence level. Alternatively, use the [CONELIP macro](#).

Confidence intervals

On a mean: Use CLM, LCLM or UCLM options in Base SAS PROC MEANS. Beginning in Version 7, use the CIBASIC option in Base SAS PROC UNIVARIATE.

On a variance or standard deviation: Beginning in Version 7, use the CIBASIC option in Base SAS PROC UNIVARIATE. In Version 6, use the [VARTEST macro](#).

On a percentile (for example the median): Beginning in Version 7, use the CIPCTLDF or the CIPCTLNORMAL option in Base SAS PROC UNIVARIATE.

On a [binomial probability](#)

On a difference between two binomial probabilities: Arrange the [proportions as the two rows of a 2x2 table](#) for analysis by PROC FREQ and specify the RISKDIFF option in the TABLES statement.

On a [relative potency](#)

On odds ratios: Use the CLODDS= option in the MODEL statement of SAS/STAT PROC LOGISTIC. To output these to a data set, use the [ESTSTATS macro](#) in Version 6, or the ODS OUTPUT statement beginning in Version 7.

On hazard (or risk) ratios: Use the RISKLIMITS option in the MODEL statement of SAS/STAT PROC PHREG. To output these to a data set, use the [ESTSTATS macro](#) in Version 6, or the ODS OUTPUT statement beginning in Version 7.

On normal regression parameters: Beginning in Version 7, use the CLB option in the MODEL statement of SAS/STAT PROC REG or the CLPARM= option in the MODEL statement of SAS/STAT PROC GLM. Prior to Version 7, use the [ESTSTATS macro](#).

On logistic and probit regression parameters: Use the CLPARM= option in the MODEL statement of SAS/STAT PROC LOGISTIC. Use the LINK=PROBIT option to request a probit model. To output these to a data set, use the [ESTSTATS macro](#) in Version 6, or the ODS OUTPUT statement beginning in Version 7.

On proportional hazards or failure time (SAS/STAT PROC LIFEREG) model parameters: Use the [ESTSTATS macro](#).

Conjoint Analysis

SAS/STAT PROC TRANSREG and [Market Research Application](#). See [SAS Technical Report R-109, Conjoint Analysis Examples](#). Also see the examples titled "*Nonmetric Conjoint*"

Analysis of Tire Data" and *Metric Conjoint Analysis of Tire Data*" in the TRANSREG chapter of the *SAS/STAT User's Guide*.

Control Charts

SAS/QC procedures SHEWHART, MACONTROL, and CUSUM.

Cook's D

SAS/STAT procedures REG and GLM (COOKD= options in the OUTPUT statement), PROC RSREG (D option on MODEL statement), PROC LOGISTIC (C= option on the OUTPUT statement -- this value must be divided by the number of parameters in the model). A similar statistic can be computed for generalized linear models using the OBSTATS output from PROC GENMOD. See this [example](#).

Correlations

Biserial: [BISERIAL macro](#)

Hoeffding's D: Base SAS PROC CORR (HOEFFDING option)

Intraclass: See [Intraclass correlation](#)

Kendall's tau-a: SAS/STAT PROC LOGISTIC

Kendall's tau-b: Base SAS procedures CORR (KENDALL option) and FREQ (MEASURES option)

Partial: Base SAS PROC CORR (PARTIAL statement), SAS/STAT PROC REG (PCORR1 and PCORR2 options), SAS/STAT PROC CANCELL (SQPCORR option)

Pearson: Base SAS procedures CORR and FREQ (MEASURES option)

Point biserial: Base SAS PROC CORR (The point biserial correlation is equivalent to the Pearson product moment correlation between two variables where the dichotomous variable is given any two numeric values.); [BISERIAL macro](#)

Polychoric: Base SAS PROC FREQ (PLCORR option), [POLYCHOR macro](#)

Polyserial: Not available.

Rank Biserial: [BISERIAL macro](#)

Semipartial: SAS/STAT PROC REG (SCORR1, SCORR2 options), CANCELL (SPCORR and SQPCORR options)

Spearman: Base SAS procedures CORR (SPEARMAN option) and FREQ (MEASURES option)

Stuart's tau-c: Base SAS PROC FREQ (MEASURES option)

Tetrachoric: Base SAS PROC FREQ (PLCORR option), [POLYCHOR macro](#)

[Correlation, compare two populations using Fisher Z transformation](#)

[Correlation, confidence interval for, using Fisher Z transformation](#)

Correspondence Analysis

SAS/STAT PROC CORRESP performs simple and multiple correspondence analysis. Also, the [Market Research Application](#).

Correspondence Analysis, canonical

Not available.

Correspondence Analysis, detrended

Not available.

Covariance matrices, testing the equality of

SAS/STAT PROC DISCRIM with the POOL=TEST option provides Bartlett's test of the equality of two or more independent covariance matrices. You can input the raw data, or the covariance, correlation, or sums of squares and crossproducts (SSCP) matrices. The test assumes that the variables are normally distributed. [Here is an example](#).

Cox Regression

SAS/STAT PROC PHREG

CPK (process capability indices)

SAS/QC PROC CAPABILITY.

CPM (Critical Path Method)

SAS/OR PROC CPM.

[Crossvalidation](#)

d (Somers's)

See [Somers's d](#).

D-Optimal designs

SAS/QC PROC OPTEX.

Deciles

See [Quantiles](#)

Decision trees

SAS/OR PROC DTREE.

Deming regression

See [Errors-in-variables regression](#)

[Dendrogram](#)

Density estimation, Parametric and Nonparametric

Both [Parametric density estimation](#) (fitting theoretical distributions to data) and nonparametric, [kernel density estimation](#) are available.

Derivatives

The [DERIVS](#) sample program uses SAS/ETS PROC EXPAND to fit a cubic smoothing spline to paired (X,Y) data. The first and second derivatives of the spline are computed and output to a SAS data set. Finally, the first and second derivatives are plotted against X.

Design of experiments

SAS/QC procedures FACTEX, OPTEX and the ADX Interface. See the *SAS/QC User's Guide* and [Getting Started with the SAS ADX Interface for Design of Experiments](#)

DETMAX

SAS/QC PROC OPTEX.

Dickey-Fuller test for unit root

SAS/ETS DFTEST macro or STATIONARITY=ADF option in the IDENTIFY statement of SAS/ETS PROC ARIMA (6.12 or later) or the *White Noise and Stationarity Tests* window of the SAS/ETS Time Series Forecasting System (6.12 or later). If the test statistic has been computed but the p-value associated with the statistic is needed, use either the SAS/ETS DFPVALUE macro or the SAS/ETS PROBDF function.

Differential equations

See [ODE \(Ordinary Differential Equations\)](#)

[Discrete Choice model](#)

Distances

Beginning in SAS 9.1, use SAS/STAT PROC DISTANCE. Prior to SAS 9.1, use the [DISTANCE macro](#).

Distribution Fitting

See [Density estimation](#)

Distributions, Comparing

Use the EDF option in SAS/STAT PROC NPAR1WAY to compare the distributions of two or more samples.

Duncan multiple range test

SAS/STAT procedures GLM and ANOVA.

Dunnett's test

SAS/STAT procedures GLM and ANOVA.

Durbin-Watson statistic

SAS/STAT PROC GLM (CLI or CLM options), SAS/STAT PROC REG (DW option), SAS/ETS PROC AUTOREG (DW= option), SAS/ETS PROC MODEL (DW option in FIT statement).

ED50

See [LD50](#).

EM (Expectation Maximization) algorithm

Used in SAS/STAT PROC MI (see the EM statement) to compute the maximum likelihood estimate (MLE) of the data with missing values, assuming a multivariate normal distribution for the data. Used in SAS/Genetics PROC HAPLOTYPE to generate maximum likelihood estimates of haplotype frequencies.

Empirical distribution functions, comparison of

EDF option in SAS/STAT PROC NPAR1WAY.

Equality of Means

SAS/STAT procedures ANOVA, GLM, MULTTEST, and TTEST.

Equality of Variances

Errors-in-variables regression

This is a regression model that minimizes the perpendicular distances from the data points to the fitted line. Use SAS/OR PROC NLP with an appropriate minimization criterion. For example:

```
proc nlp;  
  min dist;  
  parms b1=1, b0=1;  
  dist=(y - (b0 + b1*x))**2 / (1 + b1*b1);  
run;
```

Alternatively, use SAS/STAT PROC CALIS (See "*Specifying Structural Equation Models*" in the "*Introduction to Structural Equations with Latent Variables*" chapter of the *SAS/STAT User's Guide*).

Exact and Monte Carlo methods

Exact confidence interval for binomial probability, p

See [Confidence interval for binomial probability, p](#)

Exact logistic regression

Beginning with Release 8.1, you can perform exact logistic regression by using the EXACT statement in SAS/STAT PROC LOGISTIC. See also the EXACTOPTIONS option in the PROC LOGISTIC statement.

Experimental Design

See [Design of Experiments](#).

Exponentially weighted moving average (EWMA) models

[EWMA models](#) can be fit for forecasting purposes. For information on using EWMA models in quality control, see SAS/QC PROC MACONTROL.

Factor analysis, Q-mode

Factorial

Beginning in Version 7, use the Base SAS FACT function. For example, FACT(7) computes 7! ("seven factorial"). Prior to Version 7, use the Base SAS GAMMA function. The factorial of an integer, x, is returned by GAMMA(x+1).

Fishbone diagrams

SAS/QC PROC ISHIKAWA.

Fisher's exact test

Base SAS PROC FREQ (FISHER option in the EXACT statement), SAS/STAT PROC MULTTEST (FISHER option in the TEST statement).

Fisher's least significant difference

SAS/STAT procedures ANOVA and GLM.

Fixed effects (conditional) logistic regression

SAS/STAT PROC PHREG using the STRATA statement. See the Conditional Logistic Regression example in the PHREG documentation. Beginning in SAS 9, SAS/STAT PROC LOGISTIC using the STRATA statement.

Fixed effects poisson regression

The conditional model which conditions out the strata parameters is not directly available, but can be fit using SAS/STAT PROC NLMIXED by specifying the conditional log likelihood function. However, the conditional and unconditional models yield identical estimates of the non-strata parameters (See [Fixed Effects Regression Methods for Longitudinal Data Using SAS](#)). The unconditional model can be fit in SAS/STAT PROC GENMOD by specifying the strata identifier variable in both the CLASS and MODEL statements.

Fold-over designs

Fold-over Plackett-Burman designs are directly available in the ADX Interface in SAS/QC. Fractional factorial designs, created by ADX or SAS/QC PROC FACTEX, can be folded over using the DATA step. See the example titled "*Fold-Over Design*" in *SAS/QC User's Guide*. These designs reduce the aliasing in the original design.

Fractional Factorial Designs

SAS/QC PROC FACTEX and the ADX Interface.

Freeman-Halton test

This is the extension of Fisher's exact test to tables larger than 2x2 and is available via the FISHER option in the EXACT statement in Base SAS PROC FREQ.

Friedman's test

Base SAS PROC FREQ, CMH2 option. The ANOVA or **Row Mean Scores Differ** CMH statistic is Friedman's test if there is only one response per treatment-block combination. See the example titled **Computing Friedman's Chi-Square Statistic** in the FREQ chapter of the *SAS Procedures Guide*. For more than one response per combination, this statistic is a generalization of Friedman's test. Alternatively, you can obtain an F-approximation to Friedman's test by using Base SAS PROC RANK to rank the data within blocks and then SAS/STAT PROC GLM to fit a two-way ANOVA model. For more information, see the article by Ipe (1987) in the *Proceedings of the Twelfth Annual SAS Users Group International Conference* (SUG12). By using the LSMEANS statement with the PDIFF option, you can also obtain approximations to the rank-sum multiple comparisons of treatment effects. For example:

```
proc rank data=in out=ranked;
  by block; var y; ranks ry;
run;
proc glm data=ranked;
  class block trt;
  model ry = block trt;
  lsmeans trt / pdiff;
run;
```

Full-information maximum likelihood (FIML)

SAS/ETS procedures SYSLIN and MODEL

Gage (or gauge) repeatability and reproducibility (R&R)

GAGE application in the SAS/QC [Sample Library](#). See the paper by LaBarr (1994) in the *Proceedings of the Nineteenth Annual SAS Users Group International Conference* (SUG19).

[GAM \(Generalized Additive Models\)](#)

GARCH (Generalized Autoregressive Conditional Heteroscedasticity)

SAS/ETS PROC AUTOREG, beginning in Release 6.08.

[GEE \(Generalized Estimating Equations\)](#)

Generalized Linear Models

SAS/STAT PROC GENMOD. Particular models in this class are also fit by other procedures. For example, logistic models can also be fit with SAS/STAT PROC LOGISTIC and probit models by SAS/STAT PROC PROBIT.

Genetic algorithms

Beginning in SAS 9.1 (experimental in that release), use SAS/OR PROC GA. Beginning in SAS 9.0 (experimental in that release), use the SAS/IML functions GAXXXXXX. See the *SAS/IML User's Guide* or the *SAS/OR User's Guide: Local Search Optimization* for discussion and examples. You can use these tools to optimize problems involving integer, continuous, binary, or combinatorial variables, especially for finding optima for problems where the objective function may have discontinuities or may not otherwise be suitable for optimization by traditional calculus-based methods.

Geometric means

To compute the geometric mean of values in a variable, use Base SAS [Sample Library](#) program [GEOMEAN](#). Beginning in SAS 9.1, you can compute the geometric mean of values in an observation using the Base SAS function GEOMEAN or GEOMEANZ.

Gini index (of diagnostic test performance)

This is another measure of logistic model fit. It is related to the [area under the ROC curve](#), c , by $\text{gini} = 2c - 1$.

Gini's mean difference

ROBUSTSCALE option in Base SAS PROC UNIVARIATE or SAS/QC PROC CAPABILITY. Also available in SAS/INSIGHT. This is a robust estimate of the population standard deviation.

Goodness-of-fit test

See [Distribution Fitting](#) and [Chi-square goodness-of-fit test for One-way tables](#)

Granger causality test

Beginning in Release 8.1, the CAUSAL statement in PROC VARMAX. Also, see [this example](#) which uses the autoregressive specification of a bivariate vector autoregression.

Guttman scaling

There was a procedure, PROC GUTTMAN, in the Version 5 supplemental library that handled up to twelve items. This procedure is not available after Version 5. Guttman recommended correspondence analysis as an alternative (see *Measurement and Prediction*, Stouffer and Guttman, Wiley 1966). SAS/STAT PROC CORRESP performs correspondence analysis. A similar method is the [Rasch model](#). See Andrich (1988), *Rasch Models for Measurement*, Sage Publication 07-068.

Harmonic means

Beginning in SAS 9.1, you can compute the harmonic mean of values in an observation using the Base SAS function HARMEAN or HARMEANZ.

Heckman model

See [Sample selection models](#).

[Heteroscedasticity or Homoscedasticity tests](#)**Hierarchical Linear Models (HLMs)**

Use SAS/STAT PROC MIXED with RANDOM statements. HLMs are commonly called *random coefficients models*. See [here](#) for more.

Hodrick-Prescott filter

Beginning in SAS System 9, the TRANSFORM= option in the CONVERT statement of SAS/ETS PROC EXPAND.

Hoeffding's D

See [Correlations:Hoeffding's D](#)

[Homogeneity of Variance, tests of](#)**Hotelling's T-square**

See [SAS System for Linear Models](#)

Impute missing values

See [Missing value Imputation](#)

Independent Component Analysis (ICA)

Not available. This is a multivariate variable reduction method related to principal components analysis which finds independent, not just uncorrelated, components for possibly nonnormal data.

Integration

SAS/IML (CALL QUAD); see also [Area under a curve, estimation of.](#)

Interquartile range

Base SAS PROC UNIVARIATE, SAS/QC PROC CAPABILITY, SAS/INSIGHT.

Intraclass correlation

Use the [INTRACC macro](#). SAS/STAT PROC NESTED can also compute an intraclass correlation. Using the second example in the INTRACC macro description, these statements produce the intraclass correlation in the PAIR row and **Percent of Total** column of the NESTED results:

```
proc sort data=table1 out=tt;
  by pair;
run;
proc nested data=tt;
  class pair;
  var score;
run;
```

For categorical ratings, the [kappa statistic](#) has the properties of an intraclass correlation coefficient and can be used for interrater reliability.

Inverse Mill's ratio

See [Mill's ratio](#).

Ishikawa diagrams

SAS/QC PROC ISHIKAWA and SAS/QC SQC Menu System

Item analysis

See the [ITEM macro](#).

Item Response Theory

Includes the [Rasch model](#).

[Jackknifing](#)**Jonckheere-Terpstra test**

See [Trend test for ordered alternatives](#).

Put SAS Training on your Desktop or Company Network

Self-paced training for basic to advanced programmers is available on CD-ROM. This award-winning interactive software, SAS OnlineTutor, combines lessons, interactive quizzes and exercises to deliver the premium-quality training you expect from SAS. The added benefit is that you can access it 24/7 and can focus your learning only on the topics you need, when you need them.

Read more <http://support.sas.com/training/elearn/catalog/olt/index.html>

SAS Talks Storage -- Intelligence Storage

And asks, why limit yourself to inferior operational RDBMS technology?

Most mid- to large-sized companies have historically maintained one or more operational relational database management systems (RDBMS) to store transactional data and to support their operational systems. As data environments have evolved, organizations have added storage to support data marts, data warehouses and desktop applications that support business analysts and users who want to work disconnected from source data.

Is the operational RDBMS still the only storage choice to support these endeavors?

As burgeoning data volumes continue to increase and a growing user audience demands access to information in a timely and ad hoc manner, the operational RDBMS has a hard time handling the workloads. Why? Because it was never designed to be applied to tasks outside the operational systems world.

Operational RDBMS vendors continue to push their solutions despite these limitations, suggesting to customers that they can overcome the limitations by throwing more hardware - including CPUs, disks and memory - at the problem. But this causes an escalation of hardware and software expenses that most organizations never planned or anticipated.

Fortunately, there's a better answer. With SAS Intelligence Storage, you can lower your costs and bring an end to the many problems that stem from using an operational RDBMS where it was never designed to be used.

Continue reading as Allan Russell, Senior Vice President of Strategy for SAS International, and Mark Torr, Technology Strategy Manager for SAS International, explain how SAS Intelligence Storage redefines the rules. And learn how SAS provides customers around the world with answers to the problems derived from a storage strategy based on the operational RDBMS.

Can you summarize [SAS® Intelligence Storage](#) and the issues it addresses?

Allan Russell: SAS Intelligence Storage is not an individual offering. It is a collection of interoperable specially designed intelligence stores ranging from the desktop to the warehouse. Each store has been built to suit the needs of the organizational area and the typical skill level of the person who needs to create, manage and use it. For example, SAS has storage on the desktop that enables business analysts to query data locally or allows data miners to work from their desktops to build models that can then be deployed against stores elsewhere and for many other purposes. In addition to desktop storage, SAS offers storage options designed to support subject-specific data marts and multidimensional storage for creating and delivering preaggregated data. At the very sharp end, SAS provides storage options to support the data warehouse itself. You could even say that SAS Intelligence Storage addresses everything except the operational systems themselves.

Why can't I use my operational RDBMS as my data storage platform for intelligence?

Mark Torr: Operational RDBMS systems play an important role in all organizations. They have been designed over many years to reliably support the transactional and operational systems in a business, such as airline reservations systems or inventory control systems. As a result, they have been built to handle many small transactions and record updates with low data volumes per transaction. They have been designed to ensure that this data gets updated in its entirety or rolled back to its previous state. To support quick, successive updates, operational RDBMS technology has produced many necessary mechanisms that add configuration and management complexity and massive additional disk overhead. The technology has evolved to deliver maximum speed for a constant number of small data updates from operational systems, which comes at the expense of the speed required for intelligence activities.

By contrast, in the intelligence world, it's more common to load large amounts of data and support large numbers of users who often issue long-running, complex queries. So the honest answer is that

you can use an operational RDBMS for things like business intelligence, but it means you're taking technology designed for one thing and applying it to something it was not designed to do. The operational RDBMS is the truck - a big, bulky, resource-guzzling machine constantly in need of help to keep it running - and it's out of place in the Formula One/Indy car race of intelligence. SAS Intelligence Storage is a highly tuned, lean race car.

So how does SAS® Intelligence Storage differ from operational RDBMS technology?

Russell: It differs because it has been designed to handle exactly the workload that the operational RDBMS cannot. It has been designed to support the loading and storage of vast amounts of data and regular long-running queries from large numbers of users. It is optimized to deal with many ad hoc queries without the baggage that comes with an operational RDBMS. It also differs because it is not an operational RDBMS trying to find new opportunities in the intelligence landscape. Rather, SAS Intelligence Storage is a collection of storage options that have been designed with the intelligence landscape in mind. For this very reason, SAS storage options are quickly replacing operational RDBMS, which have become overly expensive in many ways and are failing to deliver in this area.

What benefits should someone deploying SAS® Intelligence Storage expect to see?

Torr: First of all, they can expect to store data reliably on the desktop for analysts and others to use. Try having a business user set up, manage, maintain and use an operational RDBMS on the desktop without extensive IT support. That one set of costs can be offset straight away, not to mention the benefits of allowing IT to focus its manpower elsewhere. Next, when you get to the data mart and data warehouse level, you can expect to see disk space utilization decrease by a factor of four. It is not unusual to see a 40-terabyte warehouse stored in an operational RDBMS reduced to just 10 terabytes in SAS Intelligence Storage once the baggage is gone.

In addition, quicker performance times on queries can often be seen with even fewer CPUs than are commonly used today to get mediocre performance from operational RDBMS systems. We have seen performance increase by a factor of 40 in certain circumstances. Not only do you see faster performance, but you see hardware life extended where previously there was an impending cost. And, in many circumstances, you'll discover a newfound freedom to move platforms, which results in modern business requirements being met at a significantly lower cost than previously thought possible.

Finally, since the range of data stores included under the SAS Intelligence Storage banner is largely self-managing, the database administrator overhead incurred in organizations that try to serve their intelligence needs with operational RDBMS is removed from the equation. In such instances, it is not unusual to see the number of people required to support storage decreased by a factor of four, essentially freeing those people to either support the operational systems more effectively or simply be reassigned to other strategic projects.

Can you share any customer references?

Russell: Sure. When the [Belgian Railway](#) set up a single system to address management information needs for different types of users, it chose SAS Intelligence Storage to optimize its BI and analytical data stores. As a result, it delivered on-demand subsets of data quickly without straining its existing systems.

Torr: We have just had a recent case with a large financial institution in Denmark that should be released as a public success story soon. The company had its data warehouse stored in an operational RDBMS on the mainframe. Initially, this was not a problem, but, as daily data volumes have grown and with the overall data volume in the warehouse expanding, the company ran into some troubles. First, it had exhausted its batch window for loading the new data into the warehouse, and, since it was charged for usage of the mainframe, the costs had become astronomical.

One solution was to purchase a much bigger machine and throw more CPUs at the problem; the other solution was to look at SAS Intelligence Storage. In a one-week proof of concept, SAS was able to go to the site and demonstrate, with the company's own data and on a significantly smaller box, that we could load the data that was taking more than 10 hours to load (and in fact, this was not all

the data but simply the available window) in a little more than one hour and match the query performance on a box that was significantly smaller than its mainframe counterpart running the operational RDBMS. The company in question is currently transitioning to SAS and has already saved significantly on hardware costs.

Anything else you would like to add?

Torr: Yes. Organizations owe it to their shareholders and others to investigate alternative options. SAS is ready to come to your company to demonstrate the power of its intelligence storage options, and we're very confident of success. The question for readers is, why wouldn't you take a look and make your own decision? What do you have to lose?

F2006: World's Largest Forecasting Conference

Whether you're making automobiles, selling swimwear or recruiting college students, the ability to accurately predict demand for products or services represents an increasingly valuable tool for businesses, schools and government organizations.

Issues facing practitioners in the rapidly expanding field of forecasting were the subject of SAS' inaugural [F2006 Business Forecasting Conference](#), held June 5 and 6 at SAS world headquarters in Cary, NC.

The gathering brought together some of the most respected forecasting experts in the world, with 334 registered attendees from 35 states, five countries and 120 different business, government and academic organizations. SAS hosted this event – the world's largest forecasting conference of the year – to help organizations analyze and streamline their forecasting processes, and improve forecasting results.

"F2006 improved attendees' critical eye for both the statistical and the process side of forecasting," says Mike Gilliland, Product Marketing Manager at SAS. "The conference takeaways were powerful. Attendees learned which techniques and processes work best and returned to their offices with new ideas for more efficiently achieving objectives."

Sharing best practices

Networking opportunities abounded for participants, who included analysts, managers, directors and business executives, as well as scientists and researchers with Ph.D.s. Practitioners appreciated exposure to academic theory and the practical examples and applications of the latest research. Academics became more familiar with the real-life issues facing today's forecasting practitioners.

"The conference provided broad opportunities for idea-sharing," adds Gilliland. "For example, the call center staffing solution at a financial services institution could be applied directly to an airline's reservation center or a retailer's store staffing problem. The new product forecasting approach used at Warner Home Video can be tried by apparel or automotive manufacturers, electronics retailers or government agencies."

Forecasting experts led discussions of the latest software and statistical techniques, although many sessions were geared to people who work in nontechnical areas and focused on the forecasting process.

Keynote presentations

Among the outstanding presentations was a keynote address by Paul Goodwin, Ph.D., from the University of Bath (UK), who spoke on "Combining Management Judgment with Statistical Forecasts for Improved Accuracy." He described how statistical methods and expert management judgment can bring complementary benefits to the forecasting process. However, such judgment can be subject to both inconsistencies and biases. He then demonstrated a range of methods designed to allow judgment to play an effective role when used in combination with statistical forecasts.

In another keynote, Sven Crone, Lancaster University (UK), explained the best ways to incorporate the effects of external events into demand forecasts. He noted that management judgment must play a major role when the event history has not been properly recorded or the expected future event is not thought to be similar to those seen in the past. He concluded with a discussion of software enhancements that could help company forecasters improve their estimates of the impact of events.

Hands-on exhibits

The F2006 exhibit hall was full and busy. IBM and SAS staffed a unique exhibit where attendees took a "test drive" of [SAS® Forecast Server](#) on an IBM® p5™ 550 server. Attendees were invited to send their company data in advance of the conference and then view it running through SAS Forecast Server. The test drive enabled attendees to observe the functionality and speed of the SAS software

and IBM hardware combination, and it let them assess the level of forecast accuracy they were likely to achieve.

"The Forecast Server test drive was a big hit," says Jim Ferris, who leads the SAS Forecasting Practice. "We conducted test drives with 26 attendees – many who were from large global corporations. Attendees appreciated the chance for hands-on use of the solution accompanied by those with technical and business expertise. Based on attendee response, SAS and IBM plan to offer this exhibit in the future."

Many took advantage of the on-site forecasting training sessions following the conference. SAS is already planning its second annual business forecasting conference, F2007, to be held June 25-26, 2007, in Cary.

SAS® Forecast Server

SAS has been the recognized leader in forecasting software for more than a quarter-century, with more than 4,000 unique customers licensing SAS forecasting products worldwide.

In November 2005, SAS released SAS Forecast Server, which added the new SAS Forecast Studio GUI to the SAS High-Performance Forecasting engine. SAS Forecast Server automatically chooses the best forecasting model, optimizes the model parameters and produces the forecasts – even accounting for seasonality and intermittent data, both of which are common in many industries.

For example, graphical displays offer retailers an understanding of the effects of holidays, marketing events, sales promotions and unexpected events such as weather, improving the ability to plan effective sales promotions and marketing events. Already a favorite with users, SAS Forecast Server was named one of the "Trend-Setting Products of the Year for 2005" by KMWorld magazine and was also selected the "2005 Communications Solutions Product of the Year" by editors of Technology Marketing Corp.'s Communications Solutions magazine for excellence in technological advancement.

New SAS Training Centers – Opening Soon!

SAS Education is pleased to announce the opening of two new training centers in August 2006. Be sure to come and see our new state-of-the-art facilities in [Charlotte, NC](#) and [Cleveland, OH](#). To celebrate these grand openings, we're offering a 2-for-1 discount on all training classes in these locations through December 2006. Find out how to register for classes in [Charlotte](#) and [Cleveland](#).

FAQ # 4289

Q: In SAS Enterprise Miner 5.1, how can I update the data source metadata when a new variable is added to the underlying data set?

A: There are two ways to update the data source metadata:

1. Create a new data source. However, this is cumbersome and time-consuming, especially if you have made a lot of changes to the default metadata.
2. Use the %EmAdviseColumns and %EmAdviseTable macros to refresh the metadata.

Note: It's always a good idea to back up the data sources folder before updating the metadata.

In the SAS Enterprise Miner program editor, submit the following code:

```
%EmAdviseColumns(dataset=data-set-name, outdata=emds.datasource-id_CM,  
                newdataflag=N);  
%EmAdviseTable((dataset=data-set-name, outdata=emds.datasource-id_TM,  
                newdataflag=N);
```

The %EmAdviseColumns macro builds the column metadata (column names, variable roles, etc.), while the %EmAdviseTable macro builds the table metadata (number of rows, number of columns, etc.). The arguments of both macros are as follows:

dataset: name of the data set that is associated with the data source

outdata: name of the data set that contains the generated information

newdataflag: indicates whether a new metadata table should be created. When you set this argument to N, it will use the existing information as a base.

In order for the data source to be refreshed, make sure that the outdata is created in the EMDS library. The data source ID can be retrieved by viewing the ID property on the property sheet of that data source. For example, for the Fisher-Anderson Iris data source the ID is FISHERANDERSONIRIS. The outdata for the columns metadata is EMDS.FISHERANDERSONIRIS_CM, and the outdata for the table metadata is EMDS.FISHERANDERSONIRIS_TM. Here is the code to refresh that metadata:

```
%EmAdviseColumns(dataset=sampsio.dmairis,  
                outdata=emds.FISHERANDERSONIRIS_CM, newdataflag=N);  
%EmAdviseTable(dataset=sampsio.dmairis, outdata=emds.FISHERANDERSONIRIS_TM,  
                newdataflag=N);  
run;
```

Note: This refreshes the data source metadata but has no effect on existing data source nodes in the diagram. For the updated metadata to be reflected in the diagram, you must replace the existing data source node.

In SAS Enterprise Miner 5.2, it is no longer necessary to write code to refresh the metadata. The option to refresh metadata has been added to the choices of data source actions. Simply right-click the data source that you want to update, and select **Refresh Metadata**.

However, as in SAS Enterprise Miner 5.1, this refreshes the data source metadata and has no effect on existing data source nodes in the diagram. In order for the updated metadata to be reflected in the diagram, you must replace the existing data source node.

FAQ # 4537

Q: What is the difference in using the SAS®9 function CATS to concatenate text strings versus using the TRIM and LEFT functions with the concatenation operator (||)?

A: The results of concatenating text strings are equivalent with both methods. However, the CATS function uses less code and processes the concatenation faster than the combination method, which uses multiple calls.

The CATS function is similar to the CATX function in that both functions concatenate strings as well as remove leading and trailing blanks. However, unlike the CATX function, CATS does not insert separators into the string.

Example 1: Creating a SAS Date Constant by Using the Combination Method

The following example uses the combination method to create a SAS date constant in the format '05Jun2006'd:

```
data _null_;
  month="Jun ";
  day=" 05 ";
  year=" 2006 ";
  date=" " || trim(left(day)) || trim(left(month))
      || trim(left(year)) || "'d";
  put date= ;
run;
```

Example 2: Creating a SAS Date Constant by Using the CATS Function

The following code returns the same value ('05Jun2006'd) as that shown in Example 1. However, the code in this example consolidates the concatenation in one call (the CATS function), which saves time both in coding and in processing:

```
data _null_;
  month="Jun ";
  day=" 05 ";
  year=" 2006 ";
  date=cats(" ", day, month, year, "'d");
  put date= ;
run;
```

For detailed information, see "CATS Function" in SAS Language Reference: Dictionary under Base SAS in SAS [OnlineDoc 9.1.3](#).

About the Author

Nina L. Werner is a SAS programmer with more years experience than she wants to say.

Webcasts and Events

What Insurers Need to Know About Service Oriented Architecture

Aug. 1

12:00 p.m. ET

During this Web seminar, you'll hear how SOA is helping insurance companies reuse existing assets, allowing their IT departments to take the lead in business process innovation.

Inside SAS: Spotlight on Delivering Business Intelligence

Aug. 30

1:00 p.m. ET

Hear from BI visionary Claudia Imhoff as she answers the fundamental questions you must ask before embarking upon a successful BI implementation.

M2006

Oct. 23-24

Las Vegas

Don't miss this opportunity to network with colleagues from all over the world and get the most up-to-date information in the industry from data mining's top thought-leaders, visionaries and practitioners.

IT Management Summit on BI

Multi-City Tour

This series was created for busy IT professionals to acquire knowledge and insight from industry pundits as well as real-world experience on a variety of important topics concerning IT. Check the Web for dates and cities.