

A Macro to Assess the Key Drivers of Overall Product Acceptance Using the SAS® System

By Hurley, G. J. and Hengehold, D. A., Procter and Gamble Pharmaceuticals, Inc., Mason, OH

ABSTRACT

Factor Analysis is a multivariate technique that is useful for data reduction and eases interpretability of responses when the number of response variables is excessive. In particular, the Principal Components method for the Orthogonal Factor Model with m Common Factors (Johnson and Wichern, 2002) is used to obtain Principal Factors. Using these Principal Factors, a multiple or simple linear regression model can be fit to determine how these factors and/or the responses they "summarize" together or individually "drive" the growth of a particular overall response variable. Analogously logistic regression is used for binary responses. In the field of products research, this technique is often referred to as "Driver's Analysis" or factor regression. A SAS® macro has been developed to automate these analyses for most cases. This macro relies on the SAS/STAT® Procedures Factor, Reg, GLM, Logistic, and Genmod to perform these analyses and uses SAS® Output Delivery System (ODS) in conjunction with the Procedure Report to produce Rich Text File (RTF) output. Real data, from a products research study designed to investigate the extent to which key attributes drive overall acceptance of two beverages and to compare their overall acceptance, is analyzed using this macro.

INTRODUCTION

Factor Analysis, as defined by Johnson and Wichern, 2002, uses the construct of underlying "factors" within data. Conceptually speaking, the concept is to try to understand the covariance structure of a p -variate distribution in terms of $m < p$ underlying "factors." For instance, if a group of variables are highly correlated with each other, while not being as highly correlated with other groups of variables, it is conceivable to view that group as having an underlying "factor" responsible for the correlation. There are various methods at arriving at these constructs. However, this paper will focus specifically on the Principal Components Method for estimating the parameters of the Orthogonal Factor Model with m Common Factors (Johnson and Wichern, 2002).

The Orthogonal Factor Model with m Common Factors is as follows:

$$X_i = \mu_i + LF + \varepsilon_i$$

Where

X_i = the i^{th} component of the observable $p \times 1$ random vector X , where p is the number of variables observed

μ_i = the i^{th} component of the $p \times 1$ vector of means, μ , the mean of variable i

ε_i = the i^{th} component of the $p \times 1$ matrix of specific factors (error), ε , the i^{th} specific factor

F_j = the j^{th} component of the $m \times 1$ vector of common factors, F , the j^{th} common factor, where m is the number of factors that p is being reduced to

ℓ_{ij} = loading of the i^{th} variable on the j^{th} factor, a component of the $p \times m$ matrix L of factor loadings.

and the following are true for the unobservable, random factors F and ε

- (1) F and ε are independent,
- (2) $E(F)=0$, $\text{Cov}(F)=I$, $E(\varepsilon)=0$, $\text{Cov}(\varepsilon)=\Psi$, where Ψ is a

diagonal matrix.

The coefficients ℓ_{ij} determine the degree to which each factor is correlated to each component of X . It can further be shown

$\text{Cov}(X)=\Sigma=LL^T+\Psi$. It is noted that $\text{Var}(X_i)=h_i^2+\psi_i$, where

$$h_i^2 = \sum_{j=1}^m \ell_{ij}^2, \text{ and } i=1,2,\dots,p. \text{ Hence, } h_i^2 \text{ is the portion of the}$$

variance of the i^{th} variable explained by the $m < p$ common factors, known as the communality and ψ_i is known as the specific variance.

One method of estimation of these parameters is the Principal Component Method, beginning with the Sample Correlation Matrix, R (Johnson and Wichern, 2002). Eigenvalue/Eigenvector pairs of this matrix (denoted $(\hat{\lambda}_i, \hat{e}_i)$) are obtained and estimates of L , Ψ , and h_i^2 are made as follows:

$$\tilde{L} = \left[\sqrt{\hat{\lambda}_1} \hat{e}_1 \mid \sqrt{\hat{\lambda}_2} \hat{e}_2 \mid \dots \mid \sqrt{\hat{\lambda}_m} \hat{e}_m \right]$$

$$\tilde{\Psi} = \text{DIAG}(R - \tilde{L}\tilde{L}^T) \text{ and}$$

$$\tilde{h}_i^2 = \sum_{j=1}^m \tilde{\ell}_{ij}^2.$$

Alternative estimations of the parameters for the Orthogonal Factor Model can be determined by other techniques, including that of the Method of Maximum Likelihoods.

After obtaining the estimated factor loading matrix, $\tilde{L}$, it is often of interest to multiply it by some orthogonal matrix T , assuming independence of the factors. This is known as rotation and it can be shown to give the same representation as the original loadings. Often the "(normal) varimax" criterion is used to determine this matrix, as it often produces favorable interpretations in many applications (see Johnson and Wichern, 2002 for a more theoretical discussion of this criterion).

After factors are obtained, it is of interest to determine how they together "drive" a particular response. This is often done by estimating the slopes of multiple, linear or logistic regression models. From a marketing viewpoint, this is beneficial, as this method will determine how changes to a particular underlying "factor" within a given product may change a consumer's viewpoint of a particular product's overall acceptability ("liking").

Similarly, simple linear or logistic regression may be used to predict overall acceptability as a function of each individual predictor.

As these methods are utilized quite often in products research, it was determined to be useful to build a SAS® macro to automate these analyses.

APPLICATION / MACRO:

This macro has been built entirely of SAS® code. It predominantly relies on SAS/STAT® Procedures Factor, Reg, GLM, Logistic, and Genmod to perform the analyses described above. However, it utilizes a wide variety of Data Step Analysis, Base SAS® Procedures, PROC SQL, and, of course, SAS® Macro Language. Finally, the macro utilizes the SAS® Output Delivery System (ODS) in conjunction with the Procedure Report to produce Rich Text File (RTF) output.

A SAMPLE RUN

In a products research test, two beverages were rated by subjects. They rated the products for the response variable, "X1", representing "Overall Liking", and the following product attributes: "X2", "X3", "X4", "X5", "X6", "X7", "X8", representing "Aroma Liking", "Appearance Liking", "Thickness/Viscosity Liking", "Flavor Liking", "Sweetness Liking", "Mouth Feel Liking", and "Aftertaste Liking" respectively. The SAS® macro described above was used to analyze these data. Although all output from the macro is created via ODS and Proc Report at the end of the macro (unless otherwise specified), for clarity, output from the beverage study will be presented with discussion of the portion of the macro that created it. In addition, titles (footnotes) normally are headers (footers) in the RTF document when it is opened in Microsoft® Word. For convenience, and solely for the purposes of this paper, they have been attached as spanning titles to the tables. In addition, Table numbers and titles have been added above each table for easy reference. Finally, text size has been adjusted in Word for the same reasons. The macro call follows:

```
%factor(nfact=2, factornames=%nrstr(Factor1:
Appearance/Texture)\%nrstr(Factor2:
Taste/Flavor), outp=no, method=prin,
file1=H:\DATA\Macro\MWSUG,
outputname=Beverages,
lib=H:\DATA\Macro\MWSUG, priors=one, dsn=all,
varlist=x2 x3 x4 x5 x6 x7 x8, regvars=x1,
logvars=, mfprob=0.20, treatment=sample,
catcov=gender, intprob=0.15, capcut=0.05,
smallcut=0.10, bytrt=yes,
primarytitle=%str(A Comparison of Two
Beverages -- MWSUG),
suffix=MWSUG, catmain=gender)
```

The macro initially uses Proc Factor to perform the factor analysis using the method of principal components on the data set "all", located in "H:\DATA\Macro\MWSUG", as was specified by the dsn and lib parameters, respectively. This was dictated by the

method parameter being given value "prin." Another valid value for method is "ml", for the method of maximum likelihoods factor analysis. The nfact parameter specifies the number of factors desired. Although any integer value smaller than the number of attribute variables is acceptable from a theoretical standpoint, it is recommended that a Scree plot and examination of the Eigenvalues are considered in conjunction with interpretability of the factors to determine which value will provide the most desirable results (See Johnson and Wichern 2002 for interpreting Scree plots and Eigenvalues). In this instance, the desired number was determined to be 2. The priors parameter is to specify a method using the priors option on Proc Factor to specify a method for computing prior communality estimates. These methods are described in the SAS®/STAT User's Guide. The attribute variables are specified in the parameter varlist, any continuous variable is acceptable to use in the factor analysis and there is no limit on the number. As the macro uses Proc Factor without the cov option, the correlation matrix is used in the factor analysis. Hence, scale of the attributes is not a concern for the factor analysis; however, having attributes on different scales will greatly complicate the interpretation of the driver's analysis. Hence, this is not recommended. The two factors are titled "Factor1: Appearance/Texture" and "Factor2: Taste/Flavor" as is specified in the factornames parameter. These titles have been given based on interpretation of the loading scores. Furthermore, this analysis will be done overall and by treatment parameter, which is "sample", as the bytrt parameter is "yes." Specifically, sample contains the value 1 or 2 for each subject, with 1 corresponding to "Beverage A" and 2 to "Beverage B". These analyses are computed and then the factor loadings and communalities are output to SAS® data sets via ODS output statements. After some data set manipulation and organization, the following three tables are ready for output.

Positive factor loading scores with magnitudes greater than 40 are displayed in the factor analysis table. A warning is printed to the log when negative loading scores with magnitude greater than 40 are detected. It is recommended that for better interpretation these attributes have their scales reversed (generally by multiplying by -1) and their meanings negated to create positive loadings.

All tables, unless otherwise specified, are output to a file titled by the value of the parameter outputname followed by an .rtf extension. All Rich Text Files are output to the directory in the file1 parameter. The outp parameter controls whether or not SAS® output is listed. As the macro generates many pages of output, it is often more efficient to set outp=no.

Table 1 Factor Analysis – Both Beverages Together.

<i>A Comparison of Two Beverages -- MWSUG</i>			
<i>Principal Components Factor Analysis</i>			
Factor Definitions	Factor Loading Score		Final Communality
	F1	F2	
Factor1: Appearance/Texture EV=4.71 (of 7)	**	**	**
X4 Thickness/Viscosity Liking	88		0.82
X3 Appearance Liking	88		0.82
X7 Mouth Feel Liking	75	46	0.78
X2 Aroma Liking	63	44	0.59
Factor2: Taste/Flavor EV=0.83 (of 7)	**	**	**
X8 Aftertaste Liking		90	0.83
X5 Flavor Liking	51	78	0.88
X6 Sweetness Liking	50	76	0.82

Table 2 Factor Analysis – Beverage A.

A Comparison of Two Beverages – MWSUG Principal Components Factor Analysis Product: Beverage A			
Factor Definitions	Factor Loading Score		Final Communality
	F1	F2	
Factor1: Appearance/Texture EV=4.77 (of 7)	**	**	**
X4 Thickness/Viscosity Liking	91		0.87
X3 Appearance Liking	90		0.84
X7 Mouth Feel Liking	76	49	0.81
X2 Aroma Liking	66	48	0.66
Factor2: Taste/Flavor EV=0.94 (of 7)	**	**	**
X8 Aftertaste Liking		92	0.85
X6 Sweetness Liking	46	78	0.81
X5 Flavor Liking	54	76	0.87

Table 3 Factor Analysis – Beverage B.

A Comparison of Two Beverages – MWSUG Principal Components Factor Analysis Product: Beverage B			
Factor Definitions	Factor Loading Score		Final Communality
	F1	F2	
Factor1: Appearance/Texture EV=4.65 (of 7)	**	**	**
X7 Mouth Feel Liking	89		0.85
X4 Thickness/Viscosity Liking	83		0.74
X5 Flavor Liking	69	57	0.81
X8 Aftertaste Liking	69		0.55
X6 Sweetness Liking	65	63	0.82
Factor2: Taste/Flavor EV=0.73 (of 7)	**	**	**
X2 Aroma Liking		95	0.94
X3 Appearance Liking	56	60	0.67

As was aforementioned, the factor loading scores indicate the degree of correlation between the individual attributes and the factors (here, on a 0 to 100 scale). The final communality represents the degree of variability of the attribute that is explained by the two principal factors that were specified (0 to 1 scale). It is noted that greater final communalities are preferable and indicate that a correct number of factors was retained. Small communalities, as is the case with the "Aroma Liking" attribute for Table 1 often indicate that another factor may be necessary. As it has been determined from interpretability, Scree plots, and Eigenvalues, that a third factor is not appropriate for these data, this suggests that there may be "smell" factor that could be detected by asking about several "smell" attributes to better understand the product. It is also noted from the "Beverage B" analysis (Table 3) that the attributes did not fall out the same as in "Beverage A" (Table 2) or overall (Table 1). This considerable

difference may speak to some fundamental difference between the beverages in relation to how consumers view them.

Next, Proc Reg and/or Proc Logistic are used to perform the simple linear or logistic regression analyses (as appropriate) on each factor and on each individual variable with a given response variable. Response variables are specified in the **regvars** and **logvars** parameters for continuous and binary response, respectively. Here, only one continuous response variable, "X1", which is "Overall Liking" and no binary response variable has been specified. The parameter estimates, their associated p-Values (testing H_0 : parameter estimate = 0), and R^2 (for continuous variables) are output, in this case overall and for each treatment group (as determined by **bytrt** parameter). The following tables are produced by the macro:

Table 4 Individual Attribute Driver's Analysis – Both Beverages.

A Comparison of Two Beverages – MWSUG Simple Linear Logistic Regression for each Attribute			
Factor Definitions	X1		
	Slope	p-Value	R ²
Factor1: Appearance/Texture EV=4.71 (of 7)	*****	*****	****
X4 Thickness/Viscosity Liking	0.51	<0.0001	0.26
X3 Appearance Liking	0.63	<0.0001	0.21
X7 Mouth Feel Liking	0.77	<0.0001	0.48
X2 Aroma Liking	0.72	<0.0001	0.33
Factor2: Taste/Flavor EV=0.83 (of 7)	*****	*****	****
X8 Aftertaste Liking	0.72	<0.0001	0.51
X5 Flavor Liking	0.83	<0.0001	0.77
X6 Sweetness Liking	0.82	<0.0001	0.56

Table 5 Individual Attribute Driver's Analysis – Beverage A.

A Comparison of Two Beverages – MWSUG Simple Linear/Logistic Regression for each Attribute Product: Beverage A			
Factor Definitions	x1		
	Slope	p-Value	R ²
Factor1: Appearance/Texture EV=4.77 (of 7)	*****	*****	****
X4 Thickness/Viscosity Liking	0.47	0.0002	0.21
X3 Appearance Liking	0.54	0.0016	0.16
X7 Mouth Feel Liking	0.71	<0.0001	0.42
X2 Aroma Liking	0.70	<0.0001	0.29
Factor2: Taste/Flavor EV=0.94 (of 7)	*****	*****	****
X8 Aftertaste Liking	0.83	<0.0001	0.60
X6 Sweetness Liking	0.75	<0.0001	0.51
X5 Flavor Liking	0.75	<0.0001	0.69

Table 6 Individual Attribute Driver's Analysis – Beverage B.

A Comparison of Two Beverages – MWSUG Simple Linear/Logistic Regression for each Attribute Product: Beverage B			
Factor Definitions	x1		
	Slope	p-Value	R ²
Factor1: Appearance/Texture EV=4.65 (of 7)	*****	*****	****
X7 Mouth Feel Liking	0.84	<0.0001	0.58
X4 Thickness/Viscosity Liking	0.54	<0.0001	0.34
X5 Flavor Liking	0.93	<0.0001	0.90
X8 Aftertaste Liking	0.59	<0.0001	0.42
X6 Sweetness Liking	0.92	<0.0001	0.64
Factor2: Taste/Flavor EV=0.73 (of 7)	*****	*****	****
X2 Aroma Liking	0.74	<0.0001	0.40
X3 Appearance Liking	0.78	<0.0001	0.32

The slope in Tables 4 through 6 is indicative of what a one-unit increase in each attribute will equate to in terms of "Overall Liking" of the beverages. For instance for Both beverages together, an increase of one unit for "Flavor Liking" will on average provide a 0.83 unit increase in "Overall Liking," where "Thickness/Viscosity Liking" only will provide a 0.51 increase in "Overall Liking." "Flavor Liking" is therefore said to be more of a driver of "Overall Liking" than "Thickness/Viscosity Liking." It is noted that the attributes drive "Overall Liking" somewhat differently between "Beverage A" and "Beverage B." This could be indicative of consumers perceiving different outages or benefits between the two products. The p-Value column is testing whether or not these slopes are significantly different from zero. In this case, all attributes have slopes significantly different from zero. A slope that is not significantly different from zero indicates that no driving of the response was detected by the given attribute. The final column, R² indicates the percent of variability in overall liking that can be explained by the changes in the individual attribute. If more than one response variable had been specified, additional sets of three columns would be added to the table as appropriate to indicate these variables slopes, p-Values, and R² statistics. An analogous analysis using logistic regression would be done for binary response variables, however R² is not relevant to logistic regression and this column would be omitted.

Proc Reg and/or Proc Logistic are then repeated to perform multiple linear or logistic regression (as appropriate) using all

factors together and the analogous results are output. In addition, factor by independent variable ("categorical covariate") interaction is tested on each response variable, for each factor, by each "categorical covariate" specified in the parameter catcov for each response variable. This is done using Proc GLM or Proc Genmod as appropriate for each continuous or binary response variable. Here, "gender" and "sample" are chosen as independent variables and both "gender" by factor and "sample" by factor interactions are tested for each factor for the specified response variable, "x1" or "Overall Liking." Parameter estimates for each level of an interaction that is detected as significant at the level specified in the mfprob parameter are then output, along with their associated p-Value. If additional response variables were requested, columns would be added to the table as appropriate.

Table 7 Factor Driver's Analysis – Both Beverages.

A Comparison of Two Beverages – MWSUG Multiple Linear/Logistic Regression of Factors on Overall Variables (Driver's Analysis)		
Factors	x1 (R ² =0.75)	
	Slope	p-Value
Factor1: Appearance/Texture	0.60	<0.0001
Factor2: Taste/Flavor	1.23	<0.0001
Intercept	0.81	<0.0001

Table 8 Factor Driver's Analysis – Interactions Significant at the 0.20 Level.

A Comparison of Two Beverages – MWSUG Multiple Linear/Logistic Regression of Factors on Overall Variables Tests of Interactions – Interactions significant at the 0.20 level. (Driver's Analysis)				
Factors	Covariate	Level	X1	
Factor1: Appearance/Texture	SAMPLE	Beverage A	Slope	p-Value
		Beverage B	0.48	<0.0001
Factor2: Taste/Flavor	gender	Female	0.80	<0.0001
		Male	1.10	<0.0001
Of all calculable Logistic Regression models considered for the binary variables considered, none had a significant interaction term OR no Logistic Regression models were considered				

Both of these tables are interpreted similarly to tables 4 through 6 above. Similarly, analogous multiple logistic regression models would be used for binary response data.

Following these analyses, in preparation for the final model, comparing the two beverages' "Overall Liking" means, "treatment" by independent variable interactions are tested for each independent variable individually using the full model and Proc GLM, with each response variable and each factor as response

variables. These results, which include the tests for main effects in the presence of interactions are immediately output to a separate Rich Text File, titled interactions followed by the value of the parameter suffix and the .rtf extension. This is particularly useful to assess a reasonable level for the parameter *intprob*. This parameter determines at what level interactions are considered significant for reporting by-subgroup means and p-Values for the final model. Table 9 below enumerates this.

Table 9 Listing of Tests for Interactions.

A Comparison of Two Beverages – MWSUG Test to Compare Products Among Factors Overall									
Obs	MissPattern	Dependent	HypothesisType	Source	DF	SS	MS	FValue	ProbF
1	1	Factor1	3	SAMPLE	1	0.00383126	0.00383126	0.00	0.9513
2	1	Factor1	3	gender	1	0.23888988	0.23888988	0.23	0.6299
3	1	Factor1	3	SAMPLE*gender	1	0.05747927	0.05747927	0.06	0.8131
4	1	Factor2	3	SAMPLE	1	1.81415561	1.81415561	1.80	0.1821
5	1	Factor2	3	gender	1	0.03586458	0.03586458	0.04	0.8506
6	1	Factor2	3	SAMPLE*gender	1	0.28180598	0.28180598	0.28	0.5977
7	2	X1	3	SAMPLE	1	10.04775230	10.04775230	4.10	0.0451
8	2	X1	3	gender	1	1.74545795	1.74545795	0.71	0.4003
9	2	X1	3	SAMPLE*gender	1	0.44002986	0.44002986	0.18	0.6725

Next, a user specified model, using Proc GLM, is used to estimate Least Squares Means and p-Values for an *n*-Way ANOVA model containing Treatment and "main" independent variables ("categorical covariates") specified in the *catmain* parameter. Here gender was specified and hence the final model becomes *Y*= sample gender, where *Y* represents each of the response variables and each factor. Generally, ANCOVA models are not considered, however, a "beta" parameter, *contmain* allows the user to consider such models. Significance is assessed and capital or small letters are added to the results based on the value of the parameters *capcut* and *smallcut*, respectively. The value of *capcut* should be smaller than *smallcut*, hence indicating a greater level of significance. Already calculated, means and p-Values are also tabulated for those interactions that were considered significant at the level specified by *intprob*. Tables 10 through 13 illustrate these results.

Specifically, Table 10 compares means across the two Beverages. It is noted that for X1, the "Overall Liking" score, "Beverage B" had a significantly greater mean score than "Beverage A," at the $\alpha=0.05$ level.

Table 11 Main Model – Comparisons within Significant Interaction Subgroups.

A Comparison of Two Beverages – MWSUG Sub-Group Analysis – Tests for Product Differences No significant interactions were found at the alpha=0.15 level				
Covariate	Level	Beverage A	Beverage B	
		Product A	Product B	
--	--	--	--	Mean (SE)

The table above indicates that no significant interactions were found at the $p=0.15$ level. This is made quite clear by reviewing Table 9 above, which shows that of all tested interactions none were even detected at the $p=0.50$ level. Hence, it is determined that the *intprob* parameter being set to 0.15 is reasonable, as it is generally not customary to present breakouts by independent variable level when no interaction is detected at a reasonable level. However, if after reviewing Table 9, one would have detected an interaction significant at the 0.18 level and was interested in seeing it broken out in this table, one could rerun the macro with *intprob=0.20*

Table 10 Main Model – Tests for Product Differences.

A Comparison of Two Beverages – MWSUG Main Model – Tests for Product Differences		
	Beverage A	Beverage B
	Product A	Product B
Factor1: Appearance/Texture	-0.01 (0.1)	0.00 (0.1)
Factor2: Taste/Flavor	-0.13 (0.1)	0.13 (0.1)
X1	0.52 (0.2)	1.09 (0.2) A
Small and Large letters indicate a significant product comparison with the product group at the 0.10 and 0.05 levels, respectively		

Table 12 Main Model – Product Pairwise Comparisons.

A Comparison of Two Beverages – MWSUG Main Model – Tests for Product Differences p-Values	
	A vs. B
Factor1: Appearance/Texture	0.9654
Factor2: Taste/Flavor	0.1669
X1	0.0474
Product A: Beverage A	
Product B: Beverage B	

The table above enumerates the p-Values alluded to in the first final model table. Note that X1, "Overall Liking" is significant at the $\alpha=0.05$ level and in Table 10, the capital A is present next to Beverage B, indicating it has significantly greater "Overall Liking" at the $\alpha=0.05$ level (as determined by the capcut parameter.)

Table 13 Main Model – Sub Group Analysis.

A Comparison of Two Beverages – MWSUG Sub Group Analysis – Tests for Product Differences p-Values			
No significant interactions were found at the alpha=0.15 level			
	Covariate	Level	A vs. B
--	--	--	--

Again, no interactions were significant at the 0.15 level, so the next table produced contains no results. If however, a treatment by independent variable (covariate) interaction would have been detected as significant at this level, then this table would be populated with a breakout of the pairwise comparisons by independent variable ("categorical covariate") level.

Following these analyses is the section of the program where the main table, here titled *Beverages.rtf* (based on the parameter *outputname*) is actually output. This is done using a series of Proc Report calls with ODS RTF statements.

HOW IT WORKS

This macro is actually designed as four separate macros; however, the user calls one macro, which calls the other three as they are needed. The controlling macro, %factor, is the longest of the four. It has been designed to give the user maximum programming flexibility with minimum knowledge of SAS®. Hence, %factor has a wide variety of parameters, some of which were not discussed above. Six classes of parameters have been defined as follows:

TYPES OF PARAMETERS	
CLASS	DESCRIPTION
Analysis Variables	Variables in data to be used in analysis
Analysis Control	Controls certain aspects of analysis
Input	Parameter Dealing with data that enters macro
Output Control	Controls how results are output
Style	Controls certain aspects of how output looks
Beta	Untested Enhancements

Following are the entire list of %factor parameters and their explanations broken out by classification:

ANALYSIS CONTROL PARAMETERS	
PARAMETER	DESCRIPTION
Intprob	The alpha level for which treatment by categorical covariate interactions will be reported
Method	The method for factor analysis, the current default is prin (Principal Components), however ml (Maximum Likelihood) is another option
Mfprob	The significance level at which Main Effects or Interactions are reported in the additional Driver's Analysis Table
Nfact	Number of factors to be kept in factor analysis
Priors	Allows the priors option on Proc Factor, specifies what the priors statement reads.

ANALYSIS VARIABLE PARAMETERS	
PARAMETER	DESCRIPTION
Catcov	The names of the "long list" of categorical covariates
Catmain	The categorical covariates that go in the main model
Logvars	LOGVARS: Logistic Regression dependent variables (must be binary)
Regvars	REGVARS: Linear Regression dependent variables
Treatment	The name of the variable containing the treatment groups
Varlist	The set of variables for the factor analysis.

INPUT PARAMETERS	
PARAMETER	DESCRIPTION
Lib	The library where the permanent SAS® data set to be analyzed is stored
Dsn	The name of the permanent SAS® data set to be analyzed

OUTPUT CONTROL PARAMETERS	
PARAMETER	DESCRIPTION
Bytrt	Perform analyses by treatment group as well as overall
File1	Directory for .rtf output
Outp	Requests no .lst file
Outputname	Name of .rtf file
Rtfoutput	Request .rtf output
Suffix	Suffix for "Supplemental Outputs" and opportunity plots

STYLE PARAMETERS	
PARAMETER	DESCRIPTION
Capcut	The p-Value at which comparisons are considered statistically significant and receive capital letters
Factornames	The name of the NFACT factors
Primarytitle	The main title for all reports
Smallcut	The p-Value at which comparisons are considered marginally significant and receive small letters

BETA PARAMETERS	
PARAMETER	DESCRIPTION
Pages	Specifies which pages of the output you wish to construct
Contmain	The continuous covariates that go in the main model

CONCLUSION:

As factor and driver's analyses are common among products research testing, a tool to automate these analyses is time saving and cost effective. In the search for such a tool, this macro was created. It has been quite effective in analysis and has already saved much time. It is continually being improved, with more features being added to a strong core. This macro has proved to be a good investment of time and resources.

REFERENCES

Johnson, R.A. and Wichern, D.W. (2002), *Applied Multivariate Statistical Analysis*, Upper Saddle River, NJ: Prentice Hall

ACKNOWLEDGMENTS

We would like to thank the Biometrics and Statistical Sciences Department at Procter and Gamble for the optimism, encouragement and direction we have received in completing this macro and for the support and encouragement for this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.
Contact the author at:

George J. Hurley
Procter and Gamble Pharmaceuticals, Inc.
Health Care Research Center, Box 594
8700 Mason Montgomery Road
Mason, OH 45040
Phone: (513) 622-3020
Fax: (513) 622-0987
Email: hurlev.gj@pg.com

David A. Hengehold
Procter and Gamble Pharmaceuticals, Inc.
Health Care Research Center, Box 594
8700 Mason Montgomery Road
Mason, OH 45040
Phone: (513) 622-0976
Fax: (513) 622-0987
Email: hengehold.da@pg.com