DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 17(2), 165-181 (1991)

STATISTICAL ANALYSIS OF STABILITY DATA BY MEANS OF GENERAL LINEAR MODELS (GLM)

F. Langenbucher

CIBA-GEIGY AG, CH-4002 Basel

ABSTRACT

Application of a General Linear Model (GLM, "Analysis of Covariance") to the statistical interpretation of stability data combines the methods of regression and analysis of variance in one common model. Expanding the well accepted method of linear regression upon time, the GLM model permits one to include supportive factors which may be either continuous regressors (temperature, humidity, etc.) or class effects (batch number, formulation type, manufacturer, etc.). Using the GLM procedure of SAS as a convenient software tool, the technique is illustrated by several examples. It is concluded that GLM provides a most suitable approach for the interpretation of stability data.

INTRODUCTION

Since long, scientific techniques are applied to the stability assessment of pharmaceutical products, including principles of reaction kinetics (reaction order, effect of temperature) and statistical techniques such as regression and confidence intervals, e.g. [1-4]. They have been summarized by Hartmann et al.[3] and are fixed in recent FDA guidelines [4]: Individual regression lines are calculated for each experimental condition (temperature, humidity, package); Analysis of Variance is applied to test differences between various batches; confidence limits and their intersection with specification limits define the expiration dating period (shelf life).

Obviously it would be attractive to combine all storage conditions and batches into a common statistical analysis. For continuous factors such as temperature or humidity this is easily achieved by Multiple Regression. Class effects like batch number or package are treated by means of ANOVA, and in connection with regressors they are handled by a General Linear Model (GLM), also known as Analysis of Covariance [5, 6]: "The technique of analysis of covariance is designed for circumstances in which the data are in such a form that they must be analyzed by a combination of the methods of regression and analysis of variance... If... interest centers in the regression of y upon x, but the data are obtained from a number of sources or may be otherwise classified, our estimate of the regression coefficient may be influenced by class effects" (p. 441 of [5]).

THEORETICAL BACKGROUND

Starting point of the statistical analysis is regression of a changing variable upon TIME as main regressor. In most cases regression will be linear, since stability tests are rarely continued to such an extent of degradation that the experimental error permits one to distinguish reaction orders. If the data indicate a curvature according to first-order degradation, the dependent variable could be used in logarithmized form.

The most logical form of the dependent variable is one which directly reflects the change of a property from its initial set value, expressed on either an absolute or a relative scale. For chemical degradation, the appearance of a decomposition product (DECOMP) is such an effect which is zero at the beginning and develops in some proportion with time; the regression intercept for TIME=0 is zero ideally and any set-off indicates an "instability" resulting from the manufacturing process. In contrast, the actual drug content (CONTENT) does not meet this criterion and should be replaced by 100-CONTENT, the sum of all degradation products. This becomes advantageous if loss of potency and appearance of degradation products are to be treated by a common analysis [8].

A full stability test protocol involves a number of additional factors which fall into one of these groups, see Figure 1: Storage factors such as temperature (TEMP), humidity (HUMI), and package (PACK) are absent in the beginning and develop in proportion with TIME, i.e. they affect the slope of the time regression (1b). They are typically continuous regressors or may converted to regressors. Batch factors such as batch number (BATCH), size or scale-up level (SIZE), dose strength (DOSE), formulation type (FORM), or manufacturing place (MANUF) are already present at the beginning. They are usually class effects and contribute a constant shift to the regression line, i.e. the intercept of the curve (2a). Interactive factors may affect both, intercept and slope (2b).

STORAGE FACTORS

The storage conditions TEMP, HUMI, and PACK have (or may be easily converted into) a reasonable continuous scale. This permits one to combine them with TIME in a common multiple regression model. When setting-up the model, it is important to not include the storage factors as main effects, since this would correspond to a constant contribution already present at TIME zero. The only meaningful model for two factors TIME and TEMP is presented by case (1b) of Figure 1, with TIME and TIME*TEMP as regressors. A generalized model for all storage factors would have a form which also takes into account that PACK is effective only in combination with TEMP and/or HUMI:

$$\text{DECOMP} = \text{TIME}\ \text{TIME*TEMP}\ \text{TIME*HUMI}\ \text{TIME*TEMP*PACK}\ \text{TIME*HUMI*PACK} \quad (1)$$

TEMPERATURE - In a model of this structure the coefficient of TIME estimates the slope of a TEMP=0 sample, independent of whether this condition exists in the data or not. The coefficient of TIME*TEMP characterizes the "spreading" of the higher temperatures from the TEMP=0 reference line; this spreading of the slopes is linear, i.e. at a given TIME value, the single TEMP's are equally spaced. If a reference temperature, e.g. 25°C, is subtracted from the TEMP values, then the same results are obtained. However, the coefficient of TIME now refers to that reference temperature

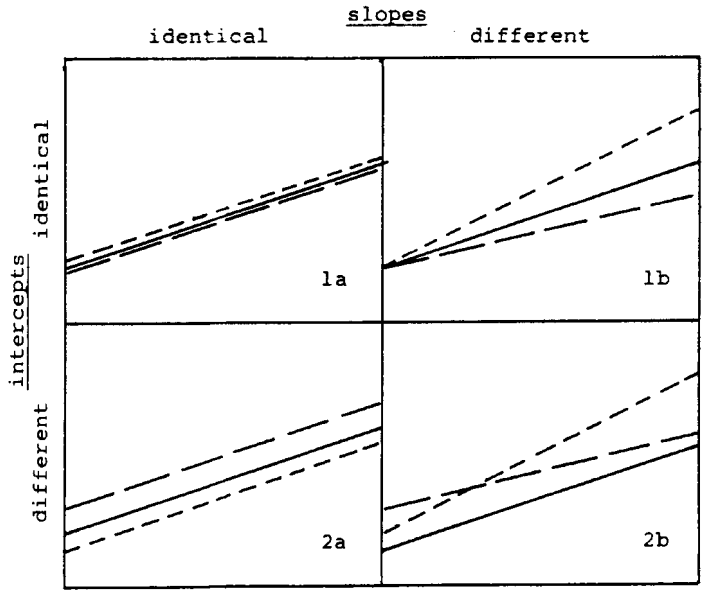


Figure 1. Influence of Storage and Batch Factors on Regression Model: Storage factors affect the slope {1b}, Batch factors the intercept {2a}, Interactive factors both {2b}

and the interaction represents the deviation from that reference. In this form the interpretation of the effects is more easy.

If the data cover a wide range of temperatures, a linear effect of TEMP may be found inadequate. A quadratic or higher polynomial may be used, but it is also possible to introduce a non-linear temperature dependence by calculating a transformed variable

$$TARR = \frac{EXP(-A/(R*(TEMP+273.15)))}{EXP(-A/(R*273.15))} \qquad (2)$$

where R = 8.3144 J/(mole K) is the universal gas constant. This represents a scale according to the Arrhenius equation, with an activation energy A (in J/mole). Dividing by the exponential for 0°C results in convenient values which are normalized to 1 for this reference temperature and all activation energies. This is illustrated in Figure 2 for two activation energies.

A model of the form TIME*TARR forces -273.15°C to be absolutely stable (which makes sense), since the absence of TIME as main effect forces the zero level of TARR to be a horizontal line. A model TIME TIME*TARR permits -273.15°C to have a positive or negative slope, and seems to be the most generalized model for non-linear TEMP effects. An optimum value of the activation energy can be found by iterative "search" technique, minimizing the residual sum of squares, but a value of 60 kJ/mole will be found satisfactory in most cases.

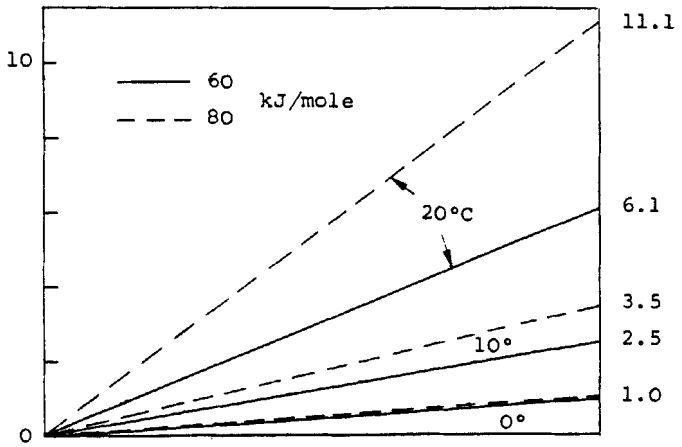


Figure 2. Slopes of Regression Lines with Arrhenius Transformation: Three equidistant temperatures and activation energies of 60 and 80 kJ/mole

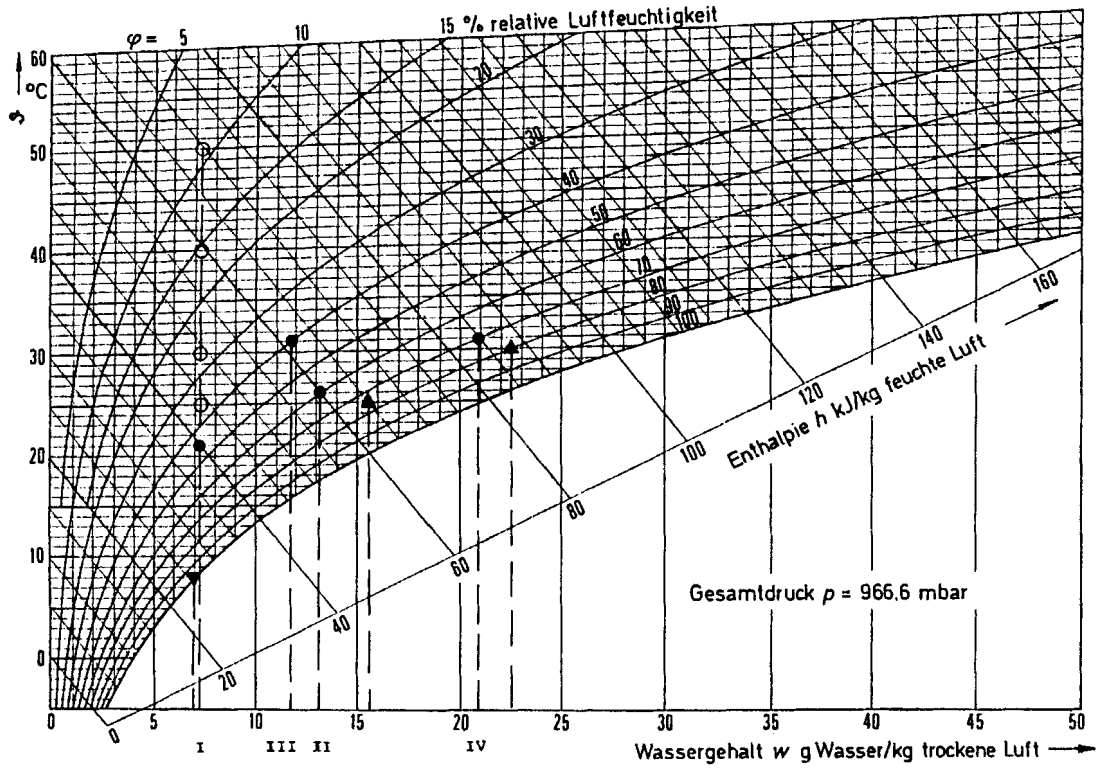
An alternative to the GLM procedure is to combine the Arrhenius equation directly with the linear regression on TIME, which results in a nonlinear model [9]. In this approach, however, there is no simple way to include class effects in the model.

HUMIDITY - Humidity is usually expressed as "relative humidity", but this is not useful as regressor: Most data are obtained without humidity control, and temperature and relative humidity are highly correlated. The absolute water content w of the air is more suitable to clearly separate the effect of humidity from that of temperature. It may be obtained from the Mollier diagram which interrelates the three variables, see Figure 3.

The two controlled stress conditions are read immediately from the diagram. Ambient (room) storage may be identified with the four international climatic zones, their kinetic mean temperatures and mean annual relative humidities [10]. Since temperature cabinets without humidity control are assumed to contain air with the same water content as the environment, the same w values apply likewise to all elevated temperatures, as shown for zone I. For storage in a refrigerator, the excess moisture will condense rapidly and the long-term water content is that corresponding to 100 percent at the relevant refrigerator temperature.

PACKAGE - Although packages represent a class effect primarily ("GLASS", "PE"), they may be handled as a continuous regressor by using their water vapor permeability (PERM). Values (in $mL/m^2/MPa/day$) are reported for 25°/90% and a film thickness of 25 um [11]:

glass, Alu/Alu	0
polychloro-trifluoroethylene	19
polyvinylidene chloride (Saran)	94
polypropylene	4500
polyamide (Nylon 6)	47000
cellulose acetate (P912)	500000



condition	temperature °C	rel. humidity %	w g/kg
refrigerator	8	100	7.0
zone I (temperate)	21	45	7.2
zone II (sub-tropical)	26	60	13.2
zone III (hot/dry)	31	40	11.7
zone IV (hot/humid)	31	70	20.9
stress 25°/75%	25	75	15.6
stress 30°/80%	30	80	22.5

Figure 3. Conversion of Relative Humidity into Absolute Water Content for Multiple Regression, using the Mollier Diagram for Humid Air [see textbooks on drying technology]

For application, e.g. with blister packs, one would use permeabilities of the finished containers instead. For open storage one could add a class effect PACK with levels YES and NO.

Increasing values of PERM represent increasing stress, same as TIME, TEMP, and HUMI. A full model treats PERM in the form of Eq.(1): "TIME*HUMI*PERM" estimates the increase of the HUMI effect due to permeability of the package, i.e. increasing stress with more permeable packaging. "TIME*TEMP*PERM" estimates the

interaction of package permeability with increasing TEMP. This term is expected with negative sign in cases where initial moisture of the product may escape at higher temperatures and higher permeability.

BATCH FACTORS

In a simple case a batch effect contributes a constant additive value: Each batch starts with a different initial content and this difference remains constant with TIME, see the "homogeneity of slopes model" in case {2a} of Figure 1. The "separate slopes model", case {2b}, permits each BATCH to have in addition different values of the slopes, i.e. different stability.

Batch factors represent class effects primarily. If two or more class effects are present, one has to distinguish between crossed and nested effects. If two batches are present, each in three different packages, then BATCH has a distinct meaning throughout and is crossed with PACK. If, however, three formulations are present, each with two batches, then the levels of BATCH have no common meaning, even if they have the same label; in this case BATCH is nested within FORM.

BATCH SIZE - The study may comprise batches of various size, from laboratory to production scale. If there are enough batches present, it is possible to arrange them on a continuous SIZE scale acoording to their actual size, or with the values "1" (research), "2" (pilot), and "3" (production). Regression upon this factor shows any systematic trends during scale-up.

DOSE STRENGTH - Several dose strengths, based on the same formulation principle, may be analyzed in common, with DOSE as a class effect or as a regressor. When comparing content data of various dose strengths, it is necessary to use a common scale for the content. The most reasonable way is to define a "normalized" content, RELCONT, in terms of the actual set value, including the nominal ("label") strength and any manufacture overage or bias of the assay.

MANUFACTURING SITE - Another possible factor is the manufacturing site, coded as a class effect MANUF. The levels represent development and production sites of one company in one or more countries, or even various companies in the case of a licenced product.

FORMULATION TYPE - Formulation types of the same product may be combined by a class effect FORM. According to FDA guidelines, this could not only involve formulations on the market but also "summary of information on previous formulations obtained during product development" [4].

GLM PROCEDURE EXECUTION

SAS is a software package well suited for GLM analysis, available on mainframe computers as well as PC's [7]. It provides file handling and graphical facilities, and a large set of statistical procedures for linear regression, ANOVA, nonlinear regression, and multivariate analysis. PROC GLM is the appropriate procedure for combined regression and ANOVA analysis. Denoting class effects by A, B, and C, continuous regressors by X1 and X2, and the observed dependent variables by Y, Y1, and Y2, various models may be handled by the procedure [7]:

REGRESSION:

```
Y       = X1              simple regression
Y       = X1 X2           multiple regression
Y       = X1 X1*X1        polynomial regression
Y1 Y2   = X1 X2           multivariate regression
```

ANOVA:

```
Y       = A               one-way ANOVA
Y       = A B C           main-effects model
Y       = A B A*B         factorial model (with interaction)
Y       = A B(A) C(B A)   nested model
Y1 Y2   = A B             multivariate analysis of variance
```

GLM:

```
Y       = A X1            analysis-of-covariance model
Y       = A X1(A)         separate-slopes model
Y       = A X1 X1*A       homogeneity-of-slopes model
```

The application of PROC GLM to stability data is illustrated in Figure 4, showing typical data input and model specification.

DATA INPUT - The raw data shown on the top form a "flat table" of all independent and dependent variables for each observation: BATCH, TIME, TEMP, and CONTENT. Arbitrary zero values are coded for the regressor TEMP at time zero, in order to avoid missing conditions which would automatically be excluded from the GLM analysis. A DATA step copies the relevant data from permanent storage into a file TRANS, and creates temporarily new variables to be used in the analysis: DECOMP as the percent decomposition and TDIF as the difference of TEMP from the standard storage condition.

MODEL DEFINITION - The SAS program, consisting of DATA step and PROC GLM is shown in the center part. The PROC GLM statement starts the GLM procedure on the data set TRANS. In the MODEL statement, the statistical model is specified in terms of all con-

Data Input

BATCH	TIME	TEMP	CONTENT	TDIF	DECOMP
A	0	0	99.8	-20	0.2
A	6	20	95.6	0	4.4
A	6	30	93.6	10	6.4
A	6	40	94.7	20	5.3
A	12	20	94.7	0	5.3
A	12	30	92.0	10	8.0
A	12	40	90.1	20	9.9
B	0	0	98.2	-20	1.8
B	6	20	95.4	0	4.6
B	6	30	93.0	10	7.0
B	6	40	91.3	20	8.7
B	12	20	90.9	0	9.1
B	12	30	90.2	10	9.8
B	12	40	86.5	20	13.5

Model Specification

```
DATA TRANS; SET CASE0; DECOMP=100-CONTENT; TDIF=TEMP-20;
PROC GLM DATA=TRANS; CLASS BATCH;
     MODEL DECOMP=BATCH TIME TIME*TDIF
           /NOINT SOLUTION P CLM ALPHA=0.10;
     ESTIMATE 'BATCH A-B' BATCH 1 -1;
     OUTPUT OUT=GLMOUT P=PRED R=RESI;
```

Figure 4. Data Input and Model Specification of a GLM Analysis by means of SAS: For explanation, see text

tributing main effects and interactions. If class variables are present, these are separately mentioned in the CLASS statement; all variables not mentioned there are treated as regressors. The INT/NOINT option of the MODEL statement specifies whether the model is to contain an overall intercept or not.

Regarding any factor present in the raw data we have the choice between three different treatments: i. the factor is not mentioned at all in the specification, i.e. "ignored" (or "pooled"), and the model cannot account for this factor; ii. the factor is not mentioned in the MODEL but in a BY statement, and the analysis is performed "separately" for each level of this factor; iii. the variable is "included" in the MODEL, as either regressor or class effect, and each level is permitted to have an individual response.

Crossed interaction terms are specified in the form "TIME*TDIF", and nested effects in the form "BATCH(DOSE)". A term "DOSE*BATCH" is also interpreted as nested, if the nested factor (BATCH) does not appear as a main effect. In addition, the nesting effect (DOSE) must be a class effect, whereas the nested effect (BATCH) may be either class or continuous regressor.

OUTPUT - A first output table lists the overall breakdown of the sum of squares into the contributions of the model and the error. It gives the overall F value of the model, together with its significance level (PR > F). In addition the table gives the estimates for: overall MEAN, ROOT MSE, R-SQUARE, and C.V. It then shows the breakdown according to the effects of the model, together with their significance levels (PR>F). For this PROC GLM uses four types of estimable functions and decomposition of the sums of squares, designated I through IV. The default print shows: Type I representing the sequential sums of squares depending on the order in which the effects are entered in the MODEL statement; Type III representing the partial sums of squares in which each effect is treated as if it had been entered last.

A second table is generated by the SOLUTION option of the MODEL statement. It lists the estimated values of the specified parameters, together with the standard error and the significance level for rejecting the null hypothesis that the factor is zero. Additional estimates may be specified by ESTIMATE statements, e.g. the contrast of the intercepts of the two batches.

A third table consists of the input data, augmented by the predicted (PRED) and residual (RESI) values, the confidence limits for the mean (CLM) or the individual observation (CLI), on the chosen significance level (ALPHA). This table is printed by the "P" option of the MODEL statement; the OUTPUT statement creates a similar data set GLMOUT for further processing.

CASE I: "STABLE" DATA

These data represent drug content for two packages and five storage conditions up to 12 months:

TIME	PACK	temperature/humidity 23	35	50	23/80	30/80
0		100.0/97.3				
3	glass	98.0	98.7	99.3	.	.
	PVC	98.7	98.7	100.0	.	.
6	glass	98.7	98.7	98.0	.	.
	PVC	100.0	99.3	100.0	99.3	98.7
12	glass	95.3/100.0	96.0/100.7	.	.	.
	PVC	100.7	99.3	.	99.3	96.7

They are reported as single determinations, with the exception of duplicates in the beginning and at 12 months in glass. The 24 observations range from 95.3 to 100.7, with mean 98.8 and standard deviation 1.4 (CV=1.4%), thus indicating a stable product with error of usual size. No effect of any of the factors involved can be detected from the table or a plot, comparing humidity stress with corresponding ambient storage, looking across the temperatures, or comparing glass and PVC data pairwise. But all comparisons suffer from the small number of data points and the unbalanced structure.

For the statistical analysis it is, therefore, attractive to combine data from various factors in a common model. Disregarding HUMI for simplicity, the two factors PACK and TEMP lead already to a total of nine possible GLM models, depending on whether the levels of the factors are treated separately, combined within the model, or simply ignored ("pooled"):

no.	PACK	TEMP	inter-cept	slope	separate "BY"
{1.1}	separate	separate	INT	TIME	PACK TEMP
{1.2}		combined	INT	TIME TIME*TEMP	PACK
{1.3}		ignored	INT	TIME	PACK
{2.1}	combined	separate	PACK	TIME	TEMP
{2.2}		combined	PACK	TIME TIME*TEMP	
{2.3}		ignored	PACK	TIME	
{3.1}	ignored	separate	INT	TIME	TEMP
{3.2}		combined	INT	TIME TIME*TEMP	
{3.3}		ignored	INT	TIME	

Separate analyses according to case {1.1} are exemplified by the first two plots in Figure 5. The number of data points is extremely small: Even for room temperature we have only six values for glass and five for PVC. Accordingly, regression lines are highly erratic and confidence limits extremely wide. It is interesting to note that the large difference between glass and PVC results from one single low value at 12 months for glass.

The two plots labeled {1.3} illustrate the gain in reliability achieved by combining the three TEMP levels for each of the two packages. A similar improvement is achieved in case {3.1} for the 23°C data, with the two packages combined. Best model performance is found for case {3.3} which handles all 24 observations in one single model, assuming a common error term for all conditions and no significant effect of any factor.

The justification of this model is seen from the analysis according to case {2.2}, with the full form

$$\text{CONTENT=TIME TIME*TEMP TIME*PACK TIME*HUMI/INT} \qquad (3)$$

PACK is used as regressor with values 0 (glass) and 1 (PVC), HUMI is used as regressor with values 0 and 80. The resulting ANOVA table is given below:

SOURCE	DF	type I SS	type I PR>F	type III SS	type III PR>F	ESTIMATE
MODEL	5	38.77	0.015			
INTERCEPT	1	34.08	0.001	5.69	0.113	0.9977
TIME	1	0.52	0.618	1.05	0.483	0.1160
TIME*TEMP	1	0.01	0.945	0.08	0.848	-0.0010
TIME*PACK	1	4.12	0.173	4.08	0.175	-0.1074
TIME*HUMI	1	0.02	0.921	0.02	0.921	0.0002
ERROR	19	39.07				
TOTAL	24	77.84				

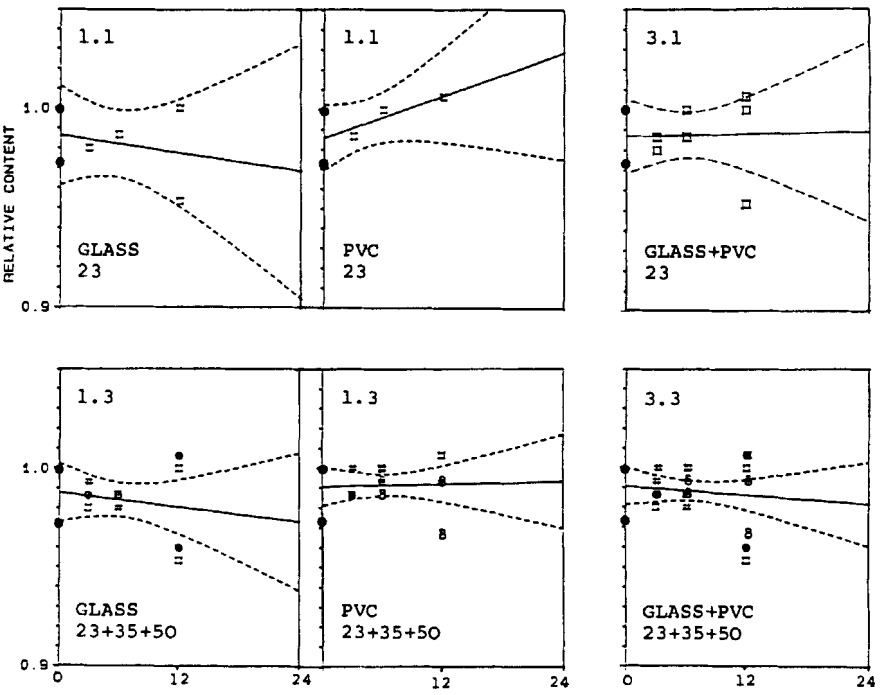


Figure 5. Various Possibilities for Combining Case I Data: {1.1} keeps all effects separate; {3.3} combines and ignores all storage conditions; {1.3} and {3.1} are intermediate

The only significant effect is the overall intercept, if entered first. All other terms are far from being significant. Hence, model {3.3} with INT and TIME as only parameters is sufficient to fit the data.

CASE II: STRONG TEMPERATURE EFFECT

Data of an unstable preparation with high temperature effect are shown in Figure 6. Initial values include a 5 percent overdose; the two values represent two batches A and B. Inspection of the data points indicates a strong and non-linear temperature effect which requires Arrhenius transformation. There is also a systematic difference between the two batches: In 14 of 16 paired comparisons B is lower than A, and a paired t-test shows the mean difference of -2.38 as highly significant (alpha=0.0006).

SEPARATE-TEMPERATURE MODELS - A separate regression analysis of the four temperatures gives the dashed regression lines, with these parameters:

	8°	15°	25°	30°
A	-5.58	-6.31	-5.11	-5.19
B	-4.36	-2.71	-4.19	-3.14
b (%/month)	0.45	0.99	2.03	3.53
RMSE	1.96	2.19	2.20	1.34

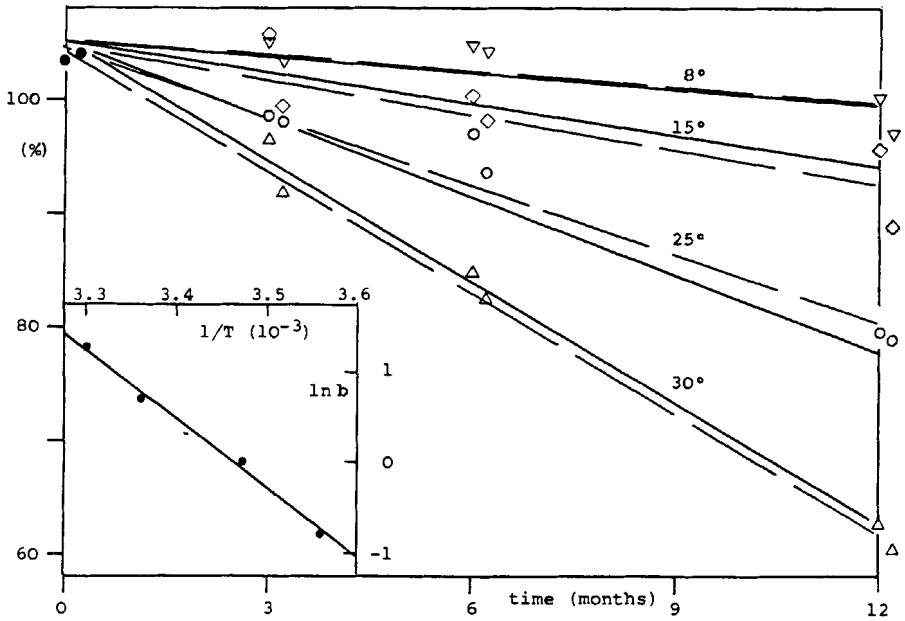


Figure 6. Decomposition Data with Strong Temperature Effect (Case II): The staggered points represent two batches "A" and "B"; dashed lines are individual regressions for 8°(▽), 15°(◇), 25°(○), and 30°(△); the inset shows the Arrhenius plot with activation energy estimated as 65 kJ/mole; solid lines represent the corresponding common Arrhenius model

For all temperatures we obtain a satisfactory linear model with TIME, with residual RMSE in the range from 1.3 to 2.2 percent. As expected all temperatures reveal a constant difference B-A between 1.2 and 2.6. These differences reach a weak significance (alpha=0.07) for 15° and 30° only, for the other two they are not significant at all. Plotting ln(b) vs. 1/T we obtain a straight line corresponding to a activation energy of 65 kJ/mole, see the inset in Figure 6.

"ARRHENIUS" MODELS - Based on these findings we define an Arrhenius model for all temperatures in common. Two forms are used, according to whether the factor BATCH is included or not, and the RMSE estimates as a function of the activation energy are found as follows:

A (kJ/mole)	BATCH TIME*TARR/NOINT	TIME*TARR/INT
50	2.44	2.74
60	2.04	2.41
65	1.98	2.36
70	1.99	2.37
80	2.15	2.50

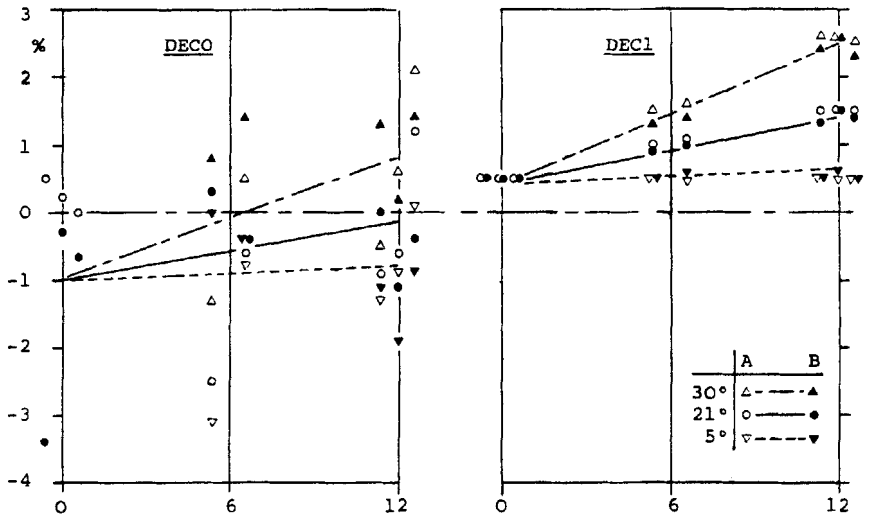


Figure 7. Input and GLM Analysis of Decomposition Data (Case III): Ordinate is decomposition in %, either determined directly (DEC1) or calculated from drug content (DEC0); staggered data points represent doses 0.1, 0.2, and 0.5 mg; each dose is represented by two batches "A" and "B"; regression lines according to model (1) with regressor "TIME*TARR" and activation energy of 60 kJ/mole

The best fit is obtained for 65 (kJ/mole), as expected from the Arrhenius plot. It is worth to keep the BATCH factor in the model, and the difference B-A is found as 2.58, which is significant with alpha=0.003. Although the fit and the estimate of the activation energy is reasonable, this does not imply a uniform reaction mechanism but is interpreted as a formal description of the temperature effect on a physicochemical process. Storage above 30°C showed degradation much faster than expected from the low-temperature model, thus indicating a break in the relevant processes. Since these data do not contribute to the interpretation of ambient storage, they are dropped from the statistical analysis.

CASE III: VARIOUS DOSES

The data in Figure 7 represent chemical degradation in percent of the nominal dose, but obtained by two different analytical methods: DEC0 = 100 - CONTENT is obtained from the unchanged drug; DEC1 is determined directly as decomposition product.

The two storage factors, TIME and TEMP, are continuous regressors. Visual inspection shows that the effect of TIME may well be interpreted as linear; however, the slopes are certainly not in a linear proportion with TEMP. Hence, we use the transformed TARR scale generated by Eq.(2), with 60 kJ/mole as a default value for the activation energy (TARR values are 1.61 for 5°, 6.59 for 21°, and 13.66 for 30°). Although "TIME*TARR" appears as one single regressor in all models, it actually represents all three temperature values together.

The DOSE of active ingredient is clearly a main class effect, present on three levels. It appears advantageous to combine the three doses in one common model rather than to analyze separately "BY DOSE": The increase of degrees of freedom will improve the power of detecting systematic effects. One could also think to use DOSE as a regressor rather than a class effect. The present data, however, exhibit no regular trend with DOSE; actually the performance of a regressor is found inferior to a class effect.

Within each DOSE, two batches are labeled by a batch identifier ID with values "A" and "B". The variable ID is nested as subgroup factor within DOSE as main group factor and is specified as "ID(DOSE)". (We might likewise use the original six BATCH numbers; this causes different labels in the output but has no influence on the results).

FULL MODEL - A full model of the form

TIME*TARR DOSE TIME*TARR*DOSE ID(DOSE) TIME*TARR*ID(DOSE)/INT

provides six regression lines with individual slopes and intercepts, i.e. one line for each batch. It contains all factors in the sequence of a "forward selection" procedure: Starting with the elementary one-line model (INTERCEPT and TIME*TARR), DOSE is added as crossed effect and in interaction with TIME*TARR; finally ID(DOSE) is added as nested effect, again as contribution to intercept and slope.

The model, however, is "over-parameterized" in a sense that it names more parameters than there are actually in the model. E.g., ten intercept terms are named: 1 INTERCEPT + 3 DOSE + 6 ID(DOSE), whereas a maximum number of six may be estimated independently from the model; the same is true for the slopes. PROC GLM does not remove this redundancy by itself, and parameters may not be estimated from this model.

SPECIFIC MODELS - Without redundancy we may specify reasonable models on these levels:

	model		DEC0		DEC1	
	slope	intercept	RMSE	R^2	RMSE	R^2
{6}	TIME*TARR*ID(DOSE)	ID(DOSE)/NOINT	0.94	0.61	0.10	0.99
{3}	TIME*TARR*DOSE	DOSE /NOINT	0.97	0.49	0.11	0.99
{1}	TIME*TARR	/INT	1.04	0.28	0.11	0.98

where the performance may be judged from their error terms, expressed as either residual standard deviation (RMSE) or correlation coefficient (R^2). With six regression lines {6}, each batch is permitted to have its own intercept and slope; all factors in the study design are accounted for. Three regression lines {3} permit the three DOSE levels to have different intercepts and/or slopes; the two batches within each dose are assumed as identical. A single regression line for all six batches in common {1} represents the effects of time and temperature only; no effect is provided for the three doses or the batches within the doses.

For DEC1 the total sum of squares of 66.8 is broken down into 66.6 for the model and 0.3 for the error. Thus the fit is excellent with RMSE=0.10, R^2=0.99, and CV=8.9%. The first two terms (INT and TIME*TARR) together represent more than 99 percent of the variance; hence a one-line model {1} is completely adequate.

For DEC0 the performance is poorer. The total sum of squares of 55.1 is separated into 33.8 for the model and 21.2 for the error. The fit is characterized by RMSE=0.94, R^2=0.58, and CV=271%. Again the one-line model represents most of the variance, about 54 percent of the variance explained by the six-lines model. DOSE is

significant when added as contribution to the intercept, but not as slope term (TIME*TARR*DOSE). The ID term reaches no significance at all. A three-lines model, either with uniform or differing slopes, seems to represent the data best.

We find that the overall fit of the models is much better for DEC1 than for DEC0. RMSE is not much affected by the choice of the model: 1.0 for DEC0 and 0.1 for DEC1. R^2 is constantly high for all models and DEC1 as response; for DEC0 there is a significant influence of the model.

CASE IV: DISSOLUTION DATA

Dissolution testing provides a good example how to reduce excessive experimental workload by statistical data evaluation via GLM. In simplified form, the standard testing plan for the first batch of a solid dosage form may have this structure:

TIME	25			30			40			50			25/75			30/80			N	
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	x	x+o
0									x										1	1
3	x	x	x	o	o	o	o	o	o	x	x	x	o	x	x	o	x	x	11	19
6	x	x	x	o	o	o	o	o	o	x	x	x	x	x	x	x	x	x	23	37
12	x	x	x	o	o	o	x	x	x				x	x	x	x	x	x	35	52
24	x	x	x	x	x	x	x	x	x				x	x	x				47	64
36	x	x	x	x	x	x													53	70
48	o	o	o	o	o	o													53	76
60	x	x	x	x	x	x													59	82

It comprises three packages: PE bottle ("A"), PVC/PE/PVDC blister ("B"), and PVC blister ("C"). At three months, the 25/75 and 30/80 conditions are also tested under open storage, but since this stress is extreme and the condition not continued to later time values, the two values are not included in the GLM analysis. Denoting compulsory tests by "x" and optional tests by "o", the column "N" in the list shows the cumulative number of tests required. This is for the first batch only and has to be repeated for two additional batches.

A number of six replicates is the basis of sampling in quality control according to USP [12]. There is a tendency to require this number of replicates even for stability testing: "At least 5 tablets should be subjected individually to the dissolution test at each stability testing station" [13]. The resulting number of about 1000 tests is felt excessive in most cases.

SEQUENTIAL SAMPLING - Since all packages have a common starting point, they are due at the same time and may be sampled together as illustrated in Figure 8. At each time point a sample of 6 or 12 units is composed from the compulsory conditions due at that time. If the GLM analysis of this first step gives a clear picture, e.g. because no changes are detected at all, then one may stay with this stage. If the analysis, however, gives doubtful indication of an effect, then further steps of 6 units are added sequentially. These may repeat the most important ambient room condition, repeat those conditions which had shown some effect in the previous step, or include optional conditions.

NUMERICAL EXAMPLE - The procedure is illustrated by an example, assuming effects of temperature, humidity, and package, in strict proportion with TIME. The changes at six months are:

PACK	25	30	40	50	25/75	30/80
A	2 (2.8)	4 (3.8)	8 (7.3)	16(14.0)	2 (4.6)	4 (7.0)
B	2 (3.2)	4 (4.2)	8 (7.6)	16(14.4)	3 (4.9)	6 (7.3)
C	2 (6.4)	4 (7.4)	8(10.9)	16(17.6)	7 (8.2)	14(10.6)

(values in parentheses are the GLM predictions at TIME=6). After adding random noise with a standard deviation of +/- 0.5, the simulated data are shown in Figure 8.

At TIME=0, step {0/1} suggests a small scatter but the presence of a doubtful "outlier". Adding another sample of 6 units in step {0/2}, the analysis of the N=12 data gives a consistent estimation: Standard deviation 5.3, with a 95 % confidence interval from 3.8 to 9.1, and mean -0.4 +/- 1.5.

The two "open" humidity stress conditions at TIME=3 indicate a strong effect of humidity, hence the model includes all factors in the form "INTERCEPT TIME*TARR TIME*WATER TIME*PERM". Here, TARR is the Arrhenius' temperature scale with levels of 9.2 (25°), 13.7 (30°), 29.2 (40°), and 59.6 (50°). WATER is the absolute air water content with levels of 7.4 (ambient), 15.6 (25/75), and 22.3 (30/80). PERM is the permeability of the package with levels 0.01 (A), 0.12 (B), 1.20 (C).

The 10 values of step {3/1} show a slight instability at higher stress conditions, and this is substantiated by adding another 12 observations with 50° and 30°/80% in step {3/2}; similarly we add a step {6/2} at TIME=6. The final parameter estimates and their

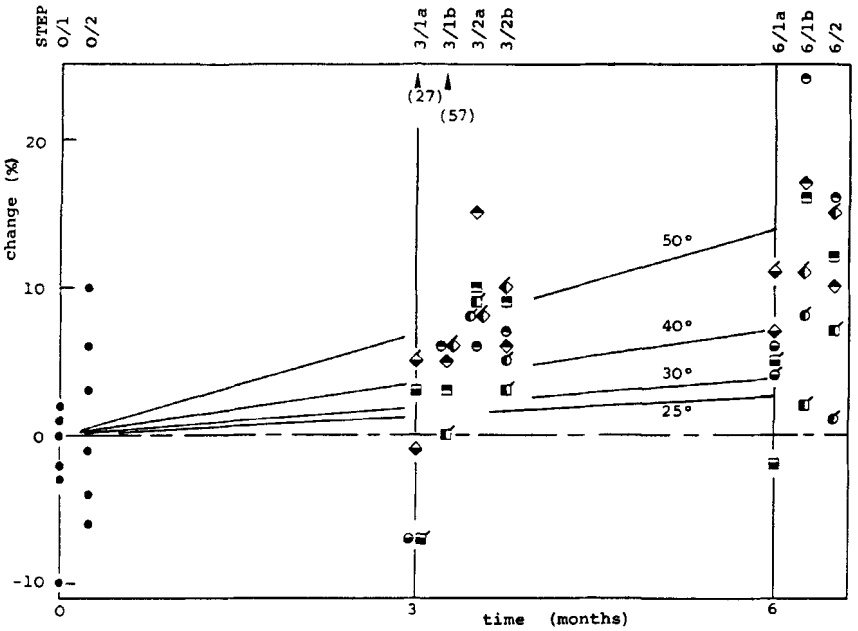


Figure 8. Sequential Sampling Plan for Dissolution Tests: Arrows indicate two high values for open storage at 25/75 and 30/80; regression lines are for package A and ambient humidity

significance levels (PR>F) at 3 and 6 months are:

TIME =	3	6
INTERCEPT	-1.146 (0.3811)	-0.779 (0.4656)
TIME*TARR	0.035 (0.0047)	0.037 (0.0001)
TIME*WATER	0.059 (0.0929)	0.035 (0.0305)
TIME*PERM	0.539 (0.3759)	0.503 (0.0803)
N/RMSE/R^2	34/4.7/0.386	52/4.5/0.572

The effect of TARR is significant already at three months, WATER is just below significance at three and significant at six months, and even PERM becomes obvious at six months. The predicted values for TIME=6, estimated at six months and listed above, illustrate the consistency attainable also with less than 6 replicates.

DISCUSSION AND CONCLUSIONS

From the nature of stability testing, linear regression upon time and assessment of shelf-life in terms of confidence limits is most appropriate. Although main interest focusses on the "primary" data at ambient conditions, additional information is gained by combining data of various batches and stress conditions. Such a synopsis of relevant data is encouraged by the FDA [4]: "If batch-to-batch variability is small ..., it would be advantageous to combine the data into one overall estimate"; "pooled estimates may be submitted if individual data points are provided". GLM is the most suitable tool for such common treatment since it includes continuous regressors as well as class effects. The choice of main effects and/or interaction terms, together with the possibility to transform variables, permits one to adjust the model to any features suggested by theory or by the data themselves.

The most simple model involves one single regression line with common intercept and slope according to the TIME factor. Class effects or regressors as main effects split the regression lines into parallel groups with intercepts according to the levels of the factors (batch factors). In interaction with TIME, the effects add further lines with the same intercept but differing slopes (storage factors).

The "power" of the effects increases in this sequence: Nested effects without any distinct "meaning", crossed effects with uniform physical meaning, and regressors with physical ranking on a continuous scale. Whereas class effects are restricted to a few distinct levels, continuous effects may have differing values any time they occur in the study. Independent of the actual number of levels, a regression effect "consumes" only one degree of freedom, whereas a class effect with N levels uses (N-1) degrees of freedom. From these reasons and the easier visualization it is advisable to treat an effect as continuous rather than categorical whenever a physical background exists for such scaling. This may be beneficial even for effects with only two levels, although the benefit increases with the number levels.

The choice of factors to be included in a model depends on the design of the study as well as the data themselves. Combining data of more effects increases the degree of freedom for the error estimate, sharpens the detection of effective factors, and narrows the confidence intervals. On the other hand, the intelligibility of a model decreases rapidly with the number of factors involved. Strategies such as "backward elimination", "forward selection", "stepwise regression", or "stagewise regression" apply to GLM, same as to regression or ANOVA [6]. F tests on the ratio of effect-to-error mean squares and t tests on contrasts show which factors may be ignored in a next step and which must be kept in the model.

GLM, same as regression and ANOVA, requires that the independent variables be exact and free of experimental error. This seems most crucial for TIME itself, but proper organization can easily handle the problem: Either samples are removed from storage at the exact time and then frozen at very low temperature until measurement, or they are kept on storage until the measurement is possible, and the correct TIME value is used for the regression.

A more series problem with stability tests is their lack of randomization: They are separated by months or years during which the testing methodology itself may develop considerably. This contamination by artificial trends or jumps cannot be handled by statistics but must be avoided a priori: Either one has to stay rigorously with the initial method even if a better one becomes available, or one has to conserve a sample at a sufficiently deep temperature to prevent any changes and to "restore" initial values with the improved method [3, 14].

In concluding, it may be stated that combining results from several "sources" can only improve the confidence of a stability assessment, and that GLM is the most appropriate statistical tool for this purpose.

ACKNOWLEDGMENTS

I thank Dr. G. Mueller for "stable" discussions of stability problems over more than twenty years, and Dr. W. Kremers for valuable statistical assistance. The discussions in the working group "Statistische Qualitätskontrolle" with members of Ciba-Geigy, Hoffmann-La Roche, and Sandoz are appreciated gratefully.

REFERENCES

[1] J.T. Carstensen and E. Nelson, J. Pharm. Sci., 65, 311 (1976)
[2] O.L. Davies and H.E. Hudson, in "Statistics in the Pharmaceutical Industry", by C.R. Buncher and J.Y. Tsay, Marcel Dekker, New York (1981)
[3] V. Hartmann, K. Krummen, G. Schnabel, and H. Bethke, Pharm. Ind., 44, 71 (1982)
[4] "Guidelines for Submitting Documentation for the Stability of Human Drugs and Biologics", Center for Drugs and Biologics, Food and Drug Administration. Rockville, Md (1987)
[5] C.A. Bennet and N.L. Franklin, "Statistical Analysis in Chemistry and the Chemical Industry", John Wiley & Sons, Inc., New York (1963)
[6] N.R. Draper and H. Smith, "Applied Regression Analysis", John Wiley and Sons, New York (1966)
[7] "SAS User's Guide: Basics", "SAS User's Guide: Statistics". 5th Edition. SAS Institute Inc., Cary, NC, USA (1985),
[8] K. Kowalski, M. Beno, C. Bergstrom, and H. Gaud, Drug Dev. Ind. Pharm., 13, 2823 (1987)
[9] S.P. King, M. Kung, and H. Fung, J. Pharm. Sci., 73, 657 (1984)
[10] European Federation of Pharmaceutical Industries' Associations, "Stability Tests on Active Substances and Finished Products", document III/66/87 rev. 4 (1987)
[11] J.A. Cairns, C.R. Oswin, and F.A. Paine, "Packaging for Climatic Protection", London (1974)
[12] M.J. Choo, I.H. Pitman, and M. Pernarowski, J. Pharm. Sci., 62, 174 (1973)
[13] Commonwealth Dept. of Health (Australia), Letter to CIBA-GEIGY, March 13 (1987)
[14] J.T. Carstensen and C.T. Rhodes, Drug. Dev. Ind. Pharm., 14, 1785 (1988)