Paper 207

Use of Multivariate Techniques to Solve an Urgent, Real-Life Problem in Pharmaceutical Manufacturing

Karl Glaser, Roche Colorado Corporation Technology Center, Boulder, Colorado

Keywords

Canonical discrimination, multiple discriminant analysis, manufacturing
PROC CANDISC

Abstract

Batches of raw material were, with increasing frequency, yielding finished product which was unacceptable. Investigation revealed that all of the raw material met all existing specifications, and the processing parameters had not changed. Univariate methods were used to try to predict batches which would give rise to failing finished product, with little success. Multiple Discriminant Analysis was then used to sort the raw material into bad and good batches.

Background

Laidlomycin, Sodium Salt (LSS) is a biologically derived raw material used in the manufacture of Laidlomycin Propionate, Potassium Salt (LPPS). On occasion, batches of LSS would, upon processing, give rise to LPPS which did not meet final product specifications, for no known reason. Further, the reason for the failure could be one or both of two different and unrelated causes:

1. a "UV Impurity" seen in the reaction completion check. This impurity subsequently disappeared, but batches in which this problem was seen gave rise to LPPS with poor yield and low assay.

2. brown or amber lumps in the final product.

To make matters worse, this was occurring with increasing frequency, approaching fifty percent of the batches manufactured.

Results of chemical analyses on the raw material LSS were checked, and all batches were seen to have met all specifications. The processing conditions in the batch records were likewise checked, and no changes in these conditions had been recorded.

With no test available to determine which batches of LSS would yield good batches of LPPS, the production facility was faced with a shutdown.

A two-pronged approach was decided upon in order to solve the problem. The first was to develop a process which was more robust, so that it would be able to make final product which met specification from any batch of raw material. The second was to try to determine if any analysis could be performed which would sort the existing raw material into "good" and "bad" batches so that production could continue using the current process. The latter approach is the subject of this paper.

A new High Pressure Liquid Chromatographic (HPLC) assay was developed; this procedure showed the presence of two compounds not seen with the existing method. It was determined that these two compounds were the same as the brown or amber lumps in the final product, and further, that they existed to a greater or less degree in all raw material batches tested. However, the amount of these two compounds in a batch of LSS was not, by itself, sufficient to determine which batches would give rise to LPPS which would fail specifications based on the "Brown Lump" problem.

Attempts using Univariate Methods

Since visual examination of the chromatograms of the LSS raw material did not yield any further clues as to what might distinguish good batches from bad, it was decided to sort them into groups, to see if any existing assay gave results which clustered in the groups to a greater or less extent.

The results from chemical analyses of the raw material were entered into a Microsoft EXCEL spreadsheet, and the spreadsheet was used as input to the SAS system using Access and View Descriptors.

The data, consisting of the numerical results from ten different chemical assays, as well as the batch number and the outcome of processing (i.e., did the finished product meet specifications, and if not, why not), were then sorted using the SORT procedure. The results of the sort were examined to determine if any clusters of good or bad batches occurred. A total of ten sorts were performed (one for each result of chemical testing), and while some analyses gave better discrimination than others, none was able, by inspection, to unequivocally separate "good" batches of raw material from "bad".

Next, Student's t-tests (using the TTEST procedure) were performed on the data, again with a view to determining which, if any, of the results of the chemical testing could be used to identify which batches would give rise to acceptable finished product and which would not.

The batches were separated into "good" and "bad" categories, and the t-tests were performed on the numerical results of the chemical assays. Again, while some assays provided better discrimination power than did others, no single chemical assay was able to identify which batches would give acceptable final product and which would not; certain of the t-tests gave statistically significant results (i.e., the mean values of the assay results were different for good and bad batches) but there was too much overlap of the individual results for any useful screening.

The experimental classification of batches as good or bad has four possible results:

1. a good batch is identified as "good" (correct)
2. a bad batch is identified as "bad" (correct)
3. a good batch is misidentified as "bad" (incorrect, Type I error)
4. a bad batch is misidentified as "good" (incorrect, Type II error)

If there is a danger of misclassification, it is most important to identify all the bad batches as "bad", since the cost of calling a bad batch "good" and using it in production is much greater than that of calling a good batch "bad" and not using it at all. Thus, any classification scheme must correctly identify all bad batches, even if some good batches may be called "bad" incorrectly. At the same time, it is important to identify as many as possible of the good batches, so that production can continue (i.e., the power of the classification scheme must be as high as possible). The chemical analyses which were in place at the time could not distinguish between good and bad lots; therefore, in order to eliminate a Type II error, all incoming batches would have to be classified as "bad" and production would cease.

It was determined that a recently developed absorbance assay, in which a known amount of the LSS raw material was dissolved in a known amount of methanol and the absorbance at 270 nM was determined, was the single best method to identify bad lots; it would identify about 60% of the good lots as good without misclassifying any of the bad lots as "good". This method was put into place as an initial screen to keep production going while further analyses were developed.

Bi- and Trivariate Analyses

Since no single assay result could distinguish between good and bad lots, an analysis using two results at the same time was tried. A SAS program was written to plot (using the GPLOT procdeure) pairs of results for samples, with the plotted point designated by a color or symbol to identify it as a good or bad batch. Results which on inspection had the best chance of correctly characterizing a batch were used for this analysis . This at first appeared promising, with a larger percentage of good batches of LSS being classified

as "good"; however, beginning with batches made in 1995, the process by which the LSS was made appeared to have changed (either feedstock, or fermentation conditions, or strain of bacteria, etc.) and batches which were in reality good would not be identified as such upon analysis. (These batches also had a somewhat different "impurity profile" upon analysis by HPLC.) Thus, while the probability of accepting a bad lot remained quite low, the probability of rejecting a good lot increased markedly.

In order to obtain more accuracy in the prediction (i.e., increase the power of the classification scheme), trivariate analysis (three variables at the same time) was tried. Triples of assay results were plotted (using the G3D procedure) with the lots being identified as good or bad by different symbols or colors. This did in fact give somewhat greater discriminating power, but did not appear to be useful enough to serve as a screening tool.

Multiple Discriminant Analysis

Results from the bi- and trivariate analyses were encouraging; therefore, a procedure using all the available data at the same time was attempted.

Multiple Discriminant Analysis, or in SAS terminology, Canonical Discrimination, is a dimension-reducing technique which is useful in cases where data have a non-numeric classification variable (e.g., bad vs. good, high vs. middle vs. low) and several quantitative variables. It involves deriving a linear combination of the independent (quantitative) variables which will best discriminate between the groups defined by the classification (qualitative) variable.

In this case, the classification (also variously known as the categorical, qualitative, nominal or taxonomic) variable was one of three types:

1. the result of the final product as either BAD or GOOD
2. the result of the final product as GOOD or type of bad (i.e., LUMPS, UV Impurity, or BOTH)
3. the type of BAD result (i.e., LUMPS, UV Impurity, or BOTH)

The results from the canonical discrimination analysis gives a set of weights (known variously as canonical coefficients, canonical weights, or discriminant weights) which are then multiplied by the corresponding independent variables, then summed to give a canonical variable (also known as a Z-score), as follows:

$$Z = W_1X_1 + W_2X_2 + ... + W_nX_n$$

where

Z is the canonical variable (Z-score)
W_i are the canonical weights
X_i are the independent variables (e.g., results of chemical assays)

The canonical variables were calculated from the results of the ten chemical assays using the CANDISC procedure and output to a data set using the OUT= option. The output data set was then sorted by the value of the canonical variable, and the results of the sort printed, including the designation of "bad" or "good" associated with each batch.

The results clearly indicated that the canonical discrimination possessed sufficient power to predict fairly accurately which batches of raw material could be processed into final product.

Determination of the Cutting Score

In this type of analysis, it is necessary to determine a value of the canonical variable (Z-score) below which, for example, a batch will be classified as "bad", and above which it will be classified as "good". This value is called the cutting score, the determination of which depends on the application.

Normally, it is desirable for the classification scheme to determine both good and bad lots as accurately as possible, such that the probability of calling a good lot "good" and a bad lot "bad" are both as high as possible. This would be the case if the cost of misclassification is the same in either instance. However, in the present situation, since the cost of misclassifying a bad lot as "good" is much higher than the cost of the reverse, it was desired to reduce this probability to a minimum, even at the risk of misclassifying some good lots as "bad".

The means and standard errors were calculated (using the MEANS procedure) for the canonical variables in the output data set from PROC CANDISC, with the data separated by the result of processing (either good or bad). The cutting score was then determined by using the 95% upper confidence level of the "bad" scores, which had the effect of minimizing the Type II error (above). The cutting score for this analysis was estimated as -0.42; any batch with a canonical variable (Z-score) greater than -0.42 was classified as "bad".

Results

This analysis identifies as "good" approximately 75% of the batches known to be good and which had failed the absorbance test (above). Thus, using the absorbance test as an initial screen (accepting about 60% of the batches known to be good) and the canonical discrimination procedure (accepting about 75% of the remaining 40%), approximately 90% of good batches will be identified as "good" while rejecting all (or nearly all) bad batches.

In order to validate this scheme, 12 batches of raw material which were known to be good, and which had not been used to develop the canonical variables or cutting score, were subjected to the analysis. Ten of the batches (83%) were identified as "good", in line with the prediction of the ability of the discrimination procedure.

The following procedure was put into place to determine the status of raw material which was received into the plant:

1. If the material passed all specifications, and also passed the absorbance screening test, the material was approved for use.

2. If the material passed all specifications and failed the absorbance screening test, the numerical results of the chemical tests were multiplied by the corresponding canonical weights, then summed to give a canonical variable which was then compared to the cutting score. If the material passed the cutting score test, it was approved for use.

3. If the material passed all specifications, but failed both the absorbance screening test and the canonical discrimination test, it was placed on "hold" status, to await approval and installation of the new process.

While this scheme was in place, no bad batch of raw material was misidentified, and enough good material was identified as such, that production could continue.

Additional Work

Batches known to be bad were placed in a separate data set and subjected to the canonical discrimination analysis, with the number of canonical coefficients set to two. This gives two sets of canonical variables or Z-scores, which may be used for additional discriminating power. Again using PROC GPLOT, the data were plotted with the type of failure (i.e., Lumps, UV Impurity, or Both) designated by different symbols or colors.

Not only could the classification scheme differentiate between good and bad lots, but once a batch was identified as "bad", the type of failure in the processing could be predicted with a fair amount of certainty.

Conclusion

The evidence indicates that the cause of the failures of some of the batches of LSS to give acceptable final product is the interaction of two or more impurities in the raw material. This interaction might be synergistic (i.e., Impurity A at a higher level requires that Impurity B be at a low level), or it might be antagonistic (i.e., if Impurity A is high, then the level of Impurity B must be high also) in order to yield good material. The interaction of impurities is something which may be profitably monitored (by means of results of different chemical analyses) in manufacturing processes.

References

Hair, Joseph H., Anderson, Rolph E., Tatham, Ronald L. "Multivariate Data Analysis with Readings", 2nd Edition (Macmillan Publishing Company)

Jellum, Harboe, Bjune, Wold "Interpreting Complicated Chromatographic Patterns" Journal of Pharmaceutical and Biomedical Analysis, vol 9 no 8 (1991)

SAS/STAT Users Guide, Version 6, Fourth Edition

SAS/GRAPH Software (Reference), Version 6, First Edition

SAS Companion for the Microsoft Windows Environment, Version 6, Second Editon

Karl Glaser
Roche Colorado Corporation Technology Center
2075 N. 55th Street
Boulder, Colorado 80501

karl.glaser@roche.com