

A simple macro to identify samples for reanalysis.

Robert H. Gallavan, Jr. and James M. Mckim, Jr. Dow Corning Corporation, Midland, MI 48686-0994.

Abstract

It is a common practice to perform two or more determinations of each sample endpoint during an analytical procedure to protect against measurement error. When cost or time constraints prohibit multiple measurements, a check for internal consistency can partially protect against measurement error. Automating the evaluation process is particularly important in a regulated environment where decisions to reanalyze individual samples must be unbiased and the decision process must be documented. In the demonstration case, stained nuclei from liver sections were counted using a blinded plan of analysis. The data were then partially unmasked to allow sorting into undisclosed, randomly ordered treatment groups. A macro ordered the observations in each group by absolute distance from the group median and sequentially tested the extreme values against the remaining observations in that group. Potential outliers were reanalyzed and tested until two consistent results were obtained, indicating that the initial result was valid and reflected the inherent variation in the method or that a measurement error was made. In the demonstration case, 19 of 90 samples were flagged and 4 were found to be the result of measurement error. The program provided a rapid, unbiased and well documented method to select samples for reanalysis.

Introduction

Under normal circumstances, assays of biological endpoints are often performed two or more times for each sample to reduce the probability of measurement error. If the results exceed some quality guideline, such as maximum allowable coefficient of variation, they are reanalyzed until they pass the test. There are times, however, when either the cost of the analysis or the time required to perform the test prohibit multiple determinations on all samples. In these cases, each sample is analyzed once and the investigator must identify and reanalyze suspect samples. This decision is usually based upon internal consistency, i.e., all the experimental units in a given treatment group came from a common population and have received the same treatment, therefore, the values obtained using a given analytical method should be similar. It is not always easy to draw this line in an objective manner, especially when the inherent variation in the method is high or the investigator is aware of the treatment each group has received. The purpose of this paper is to describe a simple program which automatically identifies and orders extreme values within a group and then sequentially tests them to determine if it is likely that they belong to the same population as the remaining observations. Samples which fail this test are then flagged for reanalysis.

Background

One analytical procedure used in our laboratory involves determining the percent of nuclei in a liver section which stain positively to an immunohistochemical agent following treatment with a test article. This is determined by counting the number of positively stained cells in a series of microscopic fields during a random search of the entire tissue section. The technician performing the counts is blinded to the treatment each liver section has received. This is a time consuming process and the results show a high level of variability. It is not uncommon to see several extreme values in a given treatment group which might be the result of measurement error or which might reflect the

inherent variation of the method. Time constraints prohibit recounting all the slides and yet the investigator needs some protection against measurement error. Because we operate in a regulated environment, the selection process must be unbiased and well documented.

Methods

The decision was made to select samples for reanalysis within a group based on internal consistency, i.e., the extreme values within a treatment group must fall outside some confidence interval based on the remaining observations in that group in order to be flagged for reanalysis. There are a number of standard tests to identify 'outliers', however, the decision being considered in those cases is to eliminate the observation from analysis and the tests are extremely conservative. Because we are only deciding which samples should be reanalyzed to confirm the original analysis, we used a more liberal test.

The procedure we used has two steps. First, the identity of the slides was unmasked only enough to group them into treatment groups without revealing the actual treatment received. The groups were then presented in random order for further analysis and a review of the results by the investigator. The analysis involved the application of a macro to the data from each treatment group. The first step of this procedure was to order observations based on their absolute distance from the group median. The median was selected as the measure of central tendency because it is less sensitive to extreme values than the mean. The most extreme value was removed from the data and the mean and standard deviation of the remaining observations in the group was calculated and used to construct a 90 % confidence interval for individual observations. If the removed observation fell outside of that confidence interval, it was flagged for reanalysis. The process was repeated with sequential elimination and testing of the ordered observations.

Two general patterns emerged from this type of analysis and are shown in the table below. In the pattern indicated as A, the first few extreme values are flagged and then there is a break. In that case, all samples before the break are reanalyzed. Samples flagged after the break are not reanalyzed. In the pattern indicated as B in the table below, the first few samples are not flagged but a subsequent sample is flagged for reanalysis. This is taken as an indication that there are two distinct populations in the group and the flagged sample and all preceding samples are reanalyzed.

Pattern Type	Results	Action taken
A	Flagged	Reanalyzed
	Flagged	Reanalyzed
	Not flagged	
	Flagged	
B	Not Flagged	Reanalyzed
	Flagged	Reanalyzed
	Not Flagged	
	Flagged	

Results

A data set consisting of the results of the analysis of 90 slides derived from nine treatment groups with ten animals per group was tested using the attached macro. The results indicated that the percent of labeled nuclei in 19 slides were outside the specified bounds and should be reanalyzed. In each case the slides were recounted and re-tested until two consistent results were obtained. If both the original result and the retest were flagged, it was concluded that the data was valid and reflected the inherent variation of the system and the mean of the counts was used. If the result after reanalysis was not flagged, the slide was counted a third time and the mean of the two consistent results was used. Based on this decision rule it was concluded that four of the nineteen slides originally flagged were the result of measurement error.

Conclusion

The method presented here provides a well documented, unbiased method to identify samples for reanalysis based on internal consistency. The decision rule involved is flexible in that the size of the confidence interval can be fixed *a priori* based on whatever considerations the investigator thinks are pertinent.

Code

```
*****;
*The following steps randomize the order in which the groups are
to be analyzed. A new variable 'Order' is created and the data set
is sorted again by group in order to merge it with the data set;
*****;
```

```
data a;
input group;
ran=ranuni(0);
cards;
1
2
3
4
5
6
7
8
9
proc sort;
by ran;
data b;
set a;
order=_N_;
proc sort;
by group;
```

```
*****;
*The data set containing the order of analysis is merged with the
data set containing the analytical results;
*****;
```

```
data c;
merge b libname.filename;
by group;
proc sort;
by order;
run;
```

```
%macro flag05(v1);
```

```
*****;
*This step calculates the median for the variable 'li' in the group
under analysis and prepares data set 'd' in order to merge the value
of the median to every observation. 'Slideno' is the slide number
of the tissue section;
*****;
```

```
data mac;
set c (keep=order slideno li);
where order=&v1;
mer=1;
proc univariate noprint;
var li;
output out=out1 median=median;
data d;
set out1;
mer=1;
```

```
*****;
*This step calculates the absolute difference between each
observation and the median for the group and then sorts the data
set by descending absolute difference. This identifies the most
extreme values and sorts them into descending order;
*****;
```

```
data e;
merge mac d;
by mer;
drop mer;
absdiff=abs(li-median);
proc sort data=e;
by descending absdiff;
proc print;
```

```
*****;
*These steps sequentially eliminate the most extreme values and
after each elimination the number of remaining observations and
their mean and standard deviation are calculated;
*****;
```

```
data all;
set d(keep=li);
data min1;
set all(firstobs=2);
proc means mean std noprint;
output out=omin1 n=n mean=mean std=std;
data min2;
set min1(firstobs=2);
proc means mean std noprint;
output out=omin2 n=n mean=mean std=std;
data min3;
set min2(firstobs=2);
proc means mean std noprint;
output out=omin3 n=n mean=mean std=std;
data min4;
set min3(firstobs=2);
proc means mean std noprint;
output out=omin4 n=n mean=mean std=std;
data min5;
set min4(firstobs=2);
proc means mean std noprint;
output out=omin5 n=n mean=mean std=std;
data min6;
set min5(firstobs=2);
proc means mean std noprint;
output out=omin6 n=n mean=mean std=std;
```

```

data min7;
set min6(firstobs=2);
proc means mean std noprint;
output out=omin7 n=n mean=mean std=std;
data min8;
set min7(firstobs=2);
proc means mean std noprint;
output out=omin8 n=n mean=mean std=std;

```

```

*****;
*This step creates a data set that will be merged with the ordered
data for this group so that each observation is paired with the
mean, standard deviation and number of observations of the data
remaining after that observation was deleted. It also calculates
the degrees of freedom (df) for the t-test;
*****;

```

```

data means;
set omin1 omin2 omin3 omin4 omin5 omin6 omin7 omin8;
drop _TYPE_ _FREQ_;
df=n-1;

```

```

*****;
*This step performs a t-test to determine if the deleted observation
(li) falls within the confidence interval of the remaining
observations, if not it is flagged. 'Lowtest' represents the lower
bound of the 90% confidence interval. 'Hightest' represents the
upper bound of the 90% confidence interval. 'Test' is the
difference between the observation and the upper or lower bound.
'Test' is calculated to yield a positive result when 'li' exceeds
either bound;
*****;

```

```

data e ;
merge d means;
t=inv(.95,df);
lowtest=mean-t*std;
hightest=mean+t*std;
if li lt median then do;
test=lowtest-li;
end;
if li gt median then do;
test=li-hightest;
end;
if test gt 0 then flag=1;

```

```

*****;
*This step generates the final report containing each observation,
all the data used to calculate the t-test and the results of the test in
the form of flag=0 (not flagged) or flag=1 (flagged).
*****;

```

```

proc print data=e;
title3 "Flagging for 90% outliers following sequential elimination
of extreme values";
title4 "Order=&v1";
run;

```

```

%amend;

```

```

*****;
*The macro is applied to each group in random order as dictated
by the variable 'order';
*****;

```

```

%flag10(1)
%flag10(2)

```

```

%flag10(3)
%flag10(4)
%flag10(5)
%flag10(6)
%flag10(7)
%flag10(8)
%flag10(9)

```

Contact

Dr. Robert H. Gallavan, Jr.
 Biostatistician
 Dow Corning Corporation C03101
 Midland, MI 48686-0994