INTRODUCTION

The requirement to compute the standard deviation from a geometric (rather than arithmetic) mean has been prompted by the desire to model a portion of the Federal government’s methodology for reimbursing hospitals for services provided to Medicare recipients. The Health Care Financing Administration (HCFA) has used this approach as the mechanism to identify unusually high-cost cases, known as outliers. The geometric mean charge for each diagnosis-related group (DRG) serves as a benchmark to which multiples of the DRG’s standard deviation are added. The resulting threshold must be exceeded for an individual case to achieve "outlier" status. These observations are deleted from the database prior to computing the average charge for each DRG.

Procedures available within the SAS* System for computing averages are based primarily on the arithmetic mean. The arithmetic mean is distinguished from other seldom-used statistics such as the geometric mean, which averages the logarithms of numbers, and the harmonic mean, which averages reciprocals. Fortunately, the SAS* System provides the flexibility and data manipulation power to satisfy this computing requirement.

APPROACH

The data set displayed in Exhibit 1 provides a simplified example of the raw data which were encountered.

Exhibit 1
Sample Raw Data

<table>
<thead>
<tr>
<th>OBS</th>
<th>DRG</th>
<th>CHARGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>105</td>
<td>$18,200.00</td>
</tr>
<tr>
<td>2</td>
<td>105</td>
<td>16,750.00</td>
</tr>
<tr>
<td>3</td>
<td>105</td>
<td>17,345.00</td>
</tr>
<tr>
<td>4</td>
<td>105</td>
<td>22,750.00</td>
</tr>
<tr>
<td>5</td>
<td>201</td>
<td>5,950.00</td>
</tr>
<tr>
<td>6</td>
<td>201</td>
<td>5,538.00</td>
</tr>
<tr>
<td>7</td>
<td>379</td>
<td>7,103.00</td>
</tr>
<tr>
<td>8</td>
<td>379</td>
<td>1,238.00</td>
</tr>
<tr>
<td>9</td>
<td>379</td>
<td>1,310.00</td>
</tr>
<tr>
<td>10</td>
<td>379</td>
<td>1,540.00</td>
</tr>
</tbody>
</table>

These data simulate a sample of ten cases and associated charges for three DRGs. The objective was to develop the code necessary to compute for each DRG: 1) the geometric mean charge and 2) the standard deviation of charges from the geometric mean. The seven-step process outlined below was used for these purposes:

1. Compute the natural logarithm of the charge variable for each observation.
2. Compute the mean of the log values for each diagnostic group.
3. Merge each observation with the appropriate group mean.
4. Compute the antilog of the group mean.
5. Compute the squared deviation of each observation from the antilog of its group mean.
6. Sum the squared deviations for each group.
7. Divide the sum of squared deviations by the number of observations (less one) in each group. The square root of this value yields the standard deviation.

The statements used to accomplish these calculations and print the results are displayed in Exhibit 2 on the following page.

RESULTS

Exhibit 3 displays several of the statistics associated with each DRG. The outlier thresholds in this example are computed by adding three multiples of the standard deviation to the geometric mean. Therefore, any case exceeding its threshold would qualify as an outlier and be deleted from the database.

* SAS is a registered trademark of SAS Institute, Inc., Cary, North Carolina, U.S.A.
options nodate:
libname in 'd:sas';

* compute logarithm for each value *:
data one:
set in.raw:
logchg=log(charge);

proc sort data=one:
by drg:

* compute mean of log values for each group *:
proc means data=one mean noprint:
var logchg;
by drg;
output out=stats mean=mchg;

* merge each observation with its group mean *:
data two:
merge one stats:
by drg:
if first.drg then do:
sum_sd=0;
end:

* compute geometric mean *:
geo_mean = exp(mchg);
* accumulate squared deviations *:
squardev = (charge - geo_mean)**2:
sum_sd + squardev;

* compute the standard deviation and threshold *:
if last.drg and _freq_ > 1 then do:
std_dev = sqrt(sum_sd/(_freq_ - 1));
thes = geo_mean + (std_dev * 3);
output:
end:

proc print data=two label:
var drg _freq geo_mean std_dev thres:
format geo_mean std_dev thres dollar10.2:
label _freq  'GEO. MEAN';
geo_mean  'GEO. MEAN';
std_dev  'STD. DEV.';
thes  'THRESHOLD';
title1 'OUTLIER THRESHOLDS BASED ON':
title2 'THREE STANDARD DEVIATIONS FROM THE GEOMETRIC MEAN':
run:
Exhibit 3

Sample Output

OUTLIER THRESHOLDS BASED ON THREE STANDARD DEVIATIONS FROM THE GEOMETRIC MEAN

<table>
<thead>
<tr>
<th>OBS</th>
<th>DRG</th>
<th>N</th>
<th>GEO. MEAN</th>
<th>STD. DEV.</th>
<th>THRESHOLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>105</td>
<td>4</td>
<td>$18,623.46</td>
<td>$2,729.59</td>
<td>$26,812.23</td>
</tr>
<tr>
<td>2</td>
<td>201</td>
<td>3</td>
<td>6,162.69</td>
<td>812.30</td>
<td>8,599.59</td>
</tr>
<tr>
<td>3</td>
<td>379</td>
<td>3</td>
<td>1,356.76</td>
<td>157.90</td>
<td>1,830.48</td>
</tr>
</tbody>
</table>