

When analyzing exploration samples using geochemical techniques, some number of samples are split, with the two (or more) splits sent to separate laboratories. These duplicate samples can be used to help determine if any bias exists in the results reported from one lab or the other. The proper database setup can help to efficiently store the duplicate values and to maximize the information obtained from them. One advantage of storing this type of data in **TECHBASE** is that, as assay values are reported from the labs, data updates can be made and viewed at any time.

Procedure:

For the purpose of this **Technote**, assume that you have a FLAT table named **SAMPLES**. This contains the following fields:

samp_id	= unique id for each sample
samp_x, samp_y, samp_z	= coordinates of each sample
samp_Au, samp_Ag, etc.	= sample values

Method 1

One simple solution is to use the *DEFINE* program to add a second set of fields to the **SAMPLES** table to store the duplicate values. These would have names such as **samp_Au2, samp_Ag2**, etc. With both the original and duplicate values in the same record, you can add CALCULATED fields to calculate differences and averages:

Au_diff	"samp_Au2 samp_Au -"
Au_ave	"samp_Au2 samp_Au + 2 /"

Normally, this type of data is keyed on **samp_id**, to ensure that each sample is stored only once in the table. Because the table is keyed on **samp_id**, the *LOAD* program can be used with update-by-KEY to ensure that the duplicate values are merged into the proper records. This might produce a result such as the following:

*samp_id	samp_Au	samp_Ag	samp_Au2	samp_Ag2	Au_diff	Au_ave
1001	0.025	0.130				
1002	0.030	0.128	0.032	0.132	0.002	0.031
1003	0.028	0.134				
1004	0.018	0.230				
1005	0.032	0.208	0.028	0.204	-0.004	0.030
1006	0.026	0.184				

Note that if a given sample does not have a duplicate value, the CALCULATED fields are empty, so that those records will not be used in any statistical study of the duplicates. The advantage of this database setup is simplicity. A significant disadvantage is wasted disk space (which increases rapidly with the number of duplicate fields), when the number of duplicate samples is small with respect to the total number of samples.

Method 2

Instead of adding the duplicate fields to the **SAMPLES** table, create a second FLAT table named **DUPES**. To this table, add:

samp_id = same field as in the **SAMPLES** table
dupe_Au, dupe_Ag, etc. = duplicate sample values

The **DUPES** table should then be keyed on **samp_id**. Then a JOIN table should be created, called **SAMP_DUPE**. This joins the **SAMPLES** table to the **DUPES** table, using **samp_id** as a common key to match the records. (Note that it is not necessary to have the **SAMPLES** table keyed in order to create the JOIN table, but it may be keyed based on your preference.) Now the following CALCULATED fields can be added to the **SAMP_DUPE** table to allow comparison of the duplicate values with the original analysis:

Au_diff "dupe_Au samp_Au -"
Au_ave "dupe_Au samp_Au + 2 /"

The CALCULATED fields *must* be added to the JOIN table, since neither of the FLAT tables contains both the **samp_** and **dupe_** fields. Now the original sample values are loaded into the **SAMPLES** table, and the duplicate values are loaded into the **DUPES** table. Using the same data as before results in the following:

SAMPLES table				
samp_id	samp_Au	samp_Ag		
1001	0.025	0.130		
1002	0.030	0.128		
1003	0.028	0.134		
1004	0.018	0.230		
1005	0.032	0.208		
1006	0.026	0.184		

DUPES table with SAMP_DUPE				
*samp_id	dupe_Au	dupe_Ag	Au_diff	Au_ave
1002	0.032	0.132	0.002	0.031
1005	0.028	0.204	-0.004	0.030

As before, the CALCULATED fields are only defined for samples that have a matching duplicate value. This technique has the advantage of significantly reduced disk space requirements (so long as the number of duplicate records is comparatively small), with only a slight increase in complexity.

Viewing the Results

Now that the duplicate data has been calculated and stored, the information can be examined in a variety of ways. The **SUMMARY** program will display the basic statistics of the data and will perform hypothesis tests for data significance. **SCATTER** will calculate the basic data statistics and can also be used to determine Pearson correlation and least-squares regression.

A graphical comparison of the original and duplicate assay values can be created in the **SCATTER** program by plotting the **samp_** values as the X axis and the **dupe_** values as the Y axis. If you wish to examine the area distribution of the **Au_diff** data, you can use the **RANDCONT** program to contour those values with respect to the **samp_x** and **samp_y** locations.

See Also

[Define\(1tb\)](#), [Report\(1tb\)](#), [Load\(1tb\)](#)
[Summary\(1st\)](#), [Scatter\(1st\)](#)
[Randcont\(1gr\)](#)