

(3)SORTプロシジャで重複データの削除

プログラム

◆NODUPKEYオプションを使用して重複データを削除します。
DUPOUT=オプションを使用して削除した重複データを出力します。

*サンプルデータの作成;

```
data sample;
  input id:$4. name:$10. date:yymmdd10.;
  format date yymmdds10.;
cards;
  1000      Yamada   2008/1/3
  2000      Tanaka   2008/1/3
  3000      Suzuki   2008/1/3
  4000      Kimura   2008/1/3
  1000      Yamada   2008/2/3
  2000      Tanaka   2008/2/3
  1000      Yamada   2008/3/3
  2000      Tanaka   2008/3/3
  5000      Sasaki   2008/3/3
  6000      Nakata   2008/3/3
;
run;

proc sort data = work.sample;
  by id descending date;
run;

proc sort data = work.sample nodupkey
  out = work.nodupsample /*重複削除後のデータセット*/
  dupout = work.dupsample; /*重複データとして削除したデータセット*/
  by id;
run;
```

結果

◆ログの出力結果

NOTE: データセット WORK.SAMPLE から 10 オブザベーションを読み込みました。
NOTE: キー値が重複した 4 オブザベーションは削除されました。
NOTE: データセット WORK.NODUPSAMPLE は 6 オブザベーション、3 変数です。
NOTE: データセット WORK.DUPSAMPLE は 4 オブザベーション、3 変数です。

◆WORK.NODUPSAMPLE(重複なしデータ)の出力結果

id	name	date
1000	Yamada	2008/03/03
2000	Tanaka	2008/03/03
3000	Suzuki	2008/01/03
4000	Kimura	2008/01/03
5000	Sasaki	2008/03/03
6000	Nakata	2008/03/03

◆WORK.DUPSAMPLE(削除した重複データ)の出力結果

id	name	date
1000	Yamada	2008/02/03
1000	Yamada	2008/01/03
2000	Tanaka	2008/02/03
2000	Tanaka	2008/01/03

※SASは、米国及び各国での米国SAS Institute Inc.の登録商標です。