

## 1st set SAS assignments

### 1. Horizontal Merging of DATA Sets

#### Creating a table with SQL code

```
Proc SQL;
create table name as select
variables
from dataset;
quit;
```

#### Merging two tables with SQL code

```
Proc SQL;
create table name as select
a.variables
,b.variables
from dataset1 a left join dataset2 b
on a.idvar1=b.idvar2;
quit;
```

#### Processing data

**Task:** Create an index for the different firms in the data set.

1.Step: Read data from text file;

2.Step: Produce table with distinct firm numbers;

3.Step: Use a data step to create an index number 1, 2, ... for each firm;

4.Step: Merge back the index number to the original dataset and write it permanently in the destination folder;

## 2. Horizontal Merging of DATA Sets

- 1.1 Make use of a data step to create the SAS data sets `stockprices07` and `stockprices03` in your working directory. The data set `stockprices07` and `stockprices03` contain three variables: a stock ticker, prices, and volume. The dataset `stockprices07` is from 2007 and `stockprices03` is from 2003.

The aim is to merge the data horizontally. For this use the `proc sql` statement:

```
proc sql;
create table tablename as select
a.*
, b.variablename1, variablename2
from stockprices07 a xxx join stockprices03 b
on a.variablename3=b.variablename3;
quit;

xxx join can be: left join, right join, inner join, full join
```

- 1.2 Create the data sets in Task a)-d) using `proc sql`.
- The resulting data set `name1` contains all information of 2007 data set and the price information of the 2003 data set for those stocks that are also available in the data set of 2007. Hint: The data set of 2007 dominates the one of 2003. For the stocks with no 2003 price information a missing value will be assigned.
  - The resulting data set `name2` contains all information of 2003 data set and the price information of the 2007 data set for those stocks that are also available in the data set of 2003. Hint: The data set of 2003 dominates the one of 2007. For the stocks with no 2007 price information a missing value will be assigned.
  - The resulting data set `name3` contains all information of 2003 and 2007 for all stocks. The variables of both data sets are preserved. Missing values are assigned for the stocks that are not in both data sets.
  - The resulting data set `name4` contains all information of 2003 and 2007 for those stock that are in both data sets.

## 2. Modifying and Vertical Merging of DATA Sets

- 2.1 Delete the variable `volume` from the `stockprices07` and name the data set `sp07`.
- 2.2 Rename the variable `p` in `stockprices03` as `prices` and keep the prices and the ticker symbol and name the data set `sp03`.
- 2.3 Merge the `sp03` and `sp07` by using the `set` statement in a data step. Name the data set `verticalmerge`.
- 2.4 Produce a table with the distinct stock tickers from `verticalmerge`.