



SAS Proc Compare Rückgabewert – Alles drin!

30-Mar-2023

Sven Wichmann

sven.wichmann@mainanalytics.de



www.mainanalytics.de

Übersicht

- Rückgabewert – SYSINFO
- Erklärung des dezimalen Rückgabewertes
- “Entschlüsselung” des Rückgabewertes
- Auswertung mit SAS
- Kompakte Ausgabe

Automatische Makrovariable SYSINFO

- SYSINFO = globale system-makro Variable
- Rückgabewert wird in der automatischen Makrovariable SYSINFO gespeichert
- Der Wert des Rückgabewertes gibt Auskunft über das Ergebnis des Vergleichs
- Verarbeitung des Rückgabewertes vor dem Beginn eines anderen Schrittes
- Wert von SYSINFO wird von einem nächsten Schritt (Daten- oder Prozedur) überschrieben

Erklärung der dezimalen Rückgabewerte

- Proc Compare liefert 16 verschiedene Vergleichsergebnisse (kombinierbar)
- Jeder Resultatswert entspricht einer 1 in einem 16 Zeichen langen Bit-Wert
- Beispiele:
 - SYSINFO = 4096
 - Bitwert = 0001000000000000
 - 1 an Position 13

16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Erklärung des dezimalen Rückgabewertes

- Kombinationen der Vergleichsergebnisse
- Summe aller ergibt den Zurückgegeben SYSINFO Wert
 - $\text{SYSINFO} = 4105 (1 + 8 + 4096)$
 - Bitwert = 0001000000001001

16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1
32768	16384	8192	4096	2048	1024	512	256	128	64	32	16	8	4	2	1

- Position 1 = Datensatz Label unterscheidet sich (Dezimal = 1)
- Position 4 = Variablen haben unterschiedliche Formate (Dezimal = 8)
- Position 13 = Variablen unterscheiden sich Inhaltlich (Dezimal = 4096)

“Entschlüsselung” des Rückgabewertes

Bit	Code	Typ	Beschreibung
1	1	DSLABEL	Datensatz Label is unterschiedlich
2	2	DSTYPE	Datensatz Typ unterschiedlich
3	4	INFORMAT	Variablen mit unterschiedlichen Informat
4	8	FORMAT	Variablen mit unterschiedlichen Format
5	16	LENGTH	Variablen mit unterschiedlichen Längen
6	32	LABEL	Variablen mit unterschiedlichem Label
7	64	BASEOBS	BASE Datensatz hat Datenreihen die nicht in COMP enthalten sind
8	128	COMPOBS	COMP Datensatz hat Datenreihen die nicht in BASE enthalten sind

Bit	Code	Typ	Beschreibung
9	256	BASEBY	BASE enthält BY group welche nicht in COMP enthalten sind
10	512	COMPBY	COMP enthält BY group welche nicht in BASE enthalten sind
11	1024	BASEVAR	BASE enthält Variablen die nicht in COMP enthalten sind
12	2048	COMPVAR	COMP enthält Variablen die nicht in BASE enthalten sind
13	4096	VALUE	Werte/Inhalte sind unterschiedlich
14	8192	TYPE	Variablen haben unterschiedliche Typen
15	16384	BYVAR	BY-Variablen haben unterschiedliche Typen
16	32768	ERROR	Error, Compare wurde nicht durchgeführt

Auswertung mit SAS

- Zerlegen des SYSINFO Dezimalwertes
- Benutzung der BAND Funktion (Bit-weises logisches UND)
- Beispiel:
 - $\text{BAND}(4105, 1) = 1 \rightarrow$ „Match“
 - $\text{BAND}(4105, 2) = 0 \rightarrow$ kein „Match“
 - ...
 - $\text{BAND}(4105, 8) = 8 \rightarrow$ „Match“
 - $\text{BAND}(4105, 16) = 0 \rightarrow$ kein „Match“
 - ...
 - $\text{BAND}(4105, 4096) = 4096 \rightarrow$ „Match“

Auswertung mit SAS

```
%let compcode.=&sysinfo.;
data _null_;
  length MESSAGE $600. TEXT $60.;
  TESTVAL = &compcode.; * SYSINFO Wert *;
  if TESTVAL = 0 then MESSAGE="NO DIFFERENCE BETWEEN OLD & NEW"; * Wenn SYSINFO = 0 *;
  else do;
    MESSAGE=" "; * Nachricht leeren;
    do i = 1 to 16; * Loop über alle 16 potentiellen Ausgaben;
      BINVAL = 2**(i-1); * Konvertierung von i in einen Binärwert *;
      MATCH = band(BINVAL, TESTVAL); * Testen des Rückgabewertes *;
      KEY = sign(MATCH) * i; * Rückgabe von i wenn MATCH > 0, sonst 0 *;
      TEXT = msg(KEY + 1); * Rückgabe als Text aus dem Array holen *;
      MESSAGE = catx(", ", MESSAGE, TEXT); * Zusammenführen der gefunden Unterschiede;
    output;
  end;
end;
call symputx("message", MESSAGE); * Alle decodierten Ausgaben in eine Makrovariable *;
run;
```


Kompakte Ausgabe

- Erstellen eines “kleinen” SAS-Makros
- Aufbereitung der Rückgabewerte
- Zusammenfassende Darstellung

Base lib	Comp lib	Base DS	Comp DS	Sysinfo	Compare summary
WORK	WORK	a1	a2	0	NO DIFFERENCE BETWEEN a1 & a2
		a11	a12	16	Variable has different length
		a3	a4	640	a4 data set has observation not in a3, a4 data set has BY group not in a3
		a5	a6	4096	A value comparison was unequal
		a7	a8	1	Data set labels differ
		a9	a10	56	Variable has different format, Variable has different length, Variable has different label

The End Vielen Dank!

