

SAS DATA Step - Optionen vs. Anweisungen

Gigic, Biljana; Buck, Katharina; Ulrich, Cornelia

Abteilung Präventive Onkologie, Nationales Centrum für Tumorerkrankungen (NCT) und
Deutsches Krebsforschungszentrum (DKFZ)
biljana.gigic@nct-heidelberg.de

In diesem Vortrag soll die Verarbeitung der Datensätze während einer KEEP-, DROP-, IF-, WHERE- und RETAIN-Anweisung bzw. Option beleuchtet werden. Verschiedene Anwendungsmöglichkeiten von KEEP und DROP sowie die bedeutenden Unterschiede der IF- und WHERE -Anweisung werden dargestellt. Dabei sollen anhand von Beispielen die Datenverarbeitung innerhalb der Kompilierungs- und Ausführungsphase eines DATA Steps demonstriert und die daraus resultierenden Outputs gegenüber gestellt werden.

Das Verständnis dieser Prozesse erleichtert den richtigen Umgang bei der Erzeugung und Auswahl von Variablen und Datensätzen innerhalb eines DATA Steps.

Erläuterung: Ein SAS DATA Step wird in zwei aufeinanderfolgenden Phasen verarbeitet (Kompilierungs- und Ausführungsphase). Während der Kompilierungsphase werden der Program Data Vector (PDV) und der beschreibende Teil der neu zu erzeugenden Datei generiert. In der Ausführungsphase wird die Output-Datei erzeugt. Abhängig von den verwendeten Anweisungen und Optionen werden die einzelnen Beobachtungen im PDV gespeichert und nach jedem Durchgang des DATA Steps in die Output-Datei transferiert - oder doch nicht?

Literatur

Gigic B., Deckert A.: SAS Backstage. KSFE 2012 - Proceedings der 16. Konferenz der SAS®-Anwender in Forschung und Entwicklung. Shaker Verlag; 103-111