

## Konzept zur Anonymisierung der Landwirtschaftszählung 1999 zur Erstellung eines CAMPUS-Files

### 1. Einleitung

Das CAMPUS-File ist ein absolut anonymes Public-Use-File, das speziell für Lehrende und Studierende erstellt wird. Seine Funktion besteht darin, die praktische Statistikausbildung mit amtlichen Einzeldaten anzureichern und damit den Hochschulen ein effektives Werkzeug für eine qualitativ hochwertige Lehre zu liefern.

Das Basismaterial des vorliegenden CAMPUS-Files bildet die Landwirtschaftszählung 1999. Diese umfasst Einzeldaten mit 425 Merkmalen von 498.369 land- und forstwirtschaftlichen Betrieben.

Im nachfolgenden Abschnitt werden die Maßnahmen beschrieben, die zur Einhaltung der Maßgabe der absoluten Anonymität der Einzeldaten eingesetzt wurden.

### 2. Anonymisierungsmaßnahmen

Zur Anonymisierung des CAMPUS-Files aus der Landwirtschaftszählung 1999 wurden verschiedene gängige Kriterien überprüft bzw. Maßnahmen durchgeführt, die im Folgenden erläutert werden:

- **Das Datenmaterial ist nur eine Stichprobe des Originalmaterials**

Für das CAMPUS-File wird aus den in der Totalerhebung der Landwirtschaftszählung enthaltenen landwirtschaftlichen Betrieben (471.960 Betriebe) eine 1%-Zufallsstichprobe gezogen = 4.720 Betriebe. Die Ziehung der Stichprobe erfolgt mit der sample-Prozedur der Statistiksoftware SPSS.

- **Das Datenmaterial hat einen stark reduzierten Merkmalsumfang.**

Es werden von den im Basismaterial enthaltenen 425 Merkmalen nur 43 aus Sicht der Geheimhaltung unkritische Merkmale für das CAMPUS File ausgewählt. Die ausgewählten Merkmale können der Merkmalsliste entnommen werden.

- **Das Datenmaterial weist ein bestimmtes Mindestalter auf**

Das Alter der Erhebung (1999) bietet eine weitere Schutzwirkung für die enthaltenen Einzeldaten. Die Werte sind durch die nachfolgenden totalen Agrarstrukturerhebungen der Jahre 2003 und 2007 bereits überholt.

- **Regionalangaben werden nur als Typisierungsangaben weitergegeben**

Als regionale Gliederungsebene wird lediglich die Unterscheidung „alte“ und „neue“ Bundesländer (Merkmal Region) aufgenommen. Die Berechnung erfolgt aus dem Merkmal EF2U1 (Bundesland). Dieses Merkmal wird nicht in das CAMPUS-File übernommen.

- **Sensible und identifizierende Merkmale werden klassifiziert übermittelt**

Des Weiteren wird das Standardbetriebseinkommen nur als kategoriale Variable (Merkmal EF7) zur Verfügung gestellt. Gleiches gilt für das möglicherweise identifizierende Merkmal des Alters des Betriebsleiters (EF8). Die Größenklassen in die die Variablen EF7 und EF8 eingeteilt wurden, können der Merkmalsliste zum CAMPUS File entnommen werden.

- **Die Datensätze sind systemfrei angeordnet**

Aus der Anordnung der Datensätze im Originalmaterial wäre es möglich, indirekt Regionalinformationen abzuleiten. Daher wurde das CAMPUS-File systemfrei nach dem Zufallsprinzip sortiert.

Weitere sensible oder identifizierende Merkmale sind im Datenmaterial nicht enthalten. Eine Mindestbesetzung aller Merkmale ist ebenfalls gegeben. Aufgrund der fehlenden Regionalangaben und des reduzierten Merkmalsumfangs sind weitere Anonymisierungsmaßnahmen nicht erforderlich.

Eine Überprüfung des CAMPUS-Files auf auffällige Betriebe (Maximalwerte, Merkmalskombinationen) ergab keine Ansatzpunkte für mögliche Reidentifikationen.

### 3. Fazit

Die durchgeführten Maßnahmen stellen die absolute Anonymität der im CAMPUS-File enthaltenen Einzeldaten sicher. Zum Einsatz in der Methodenlehre bietet das vorliegende CAMPUS-File trotz der durchgeführten Maßnahmen vielfältige Analysemöglichkeiten.

gez. Dr. Hendrik Tietje