# Skript 2 – Import Excel-Datei

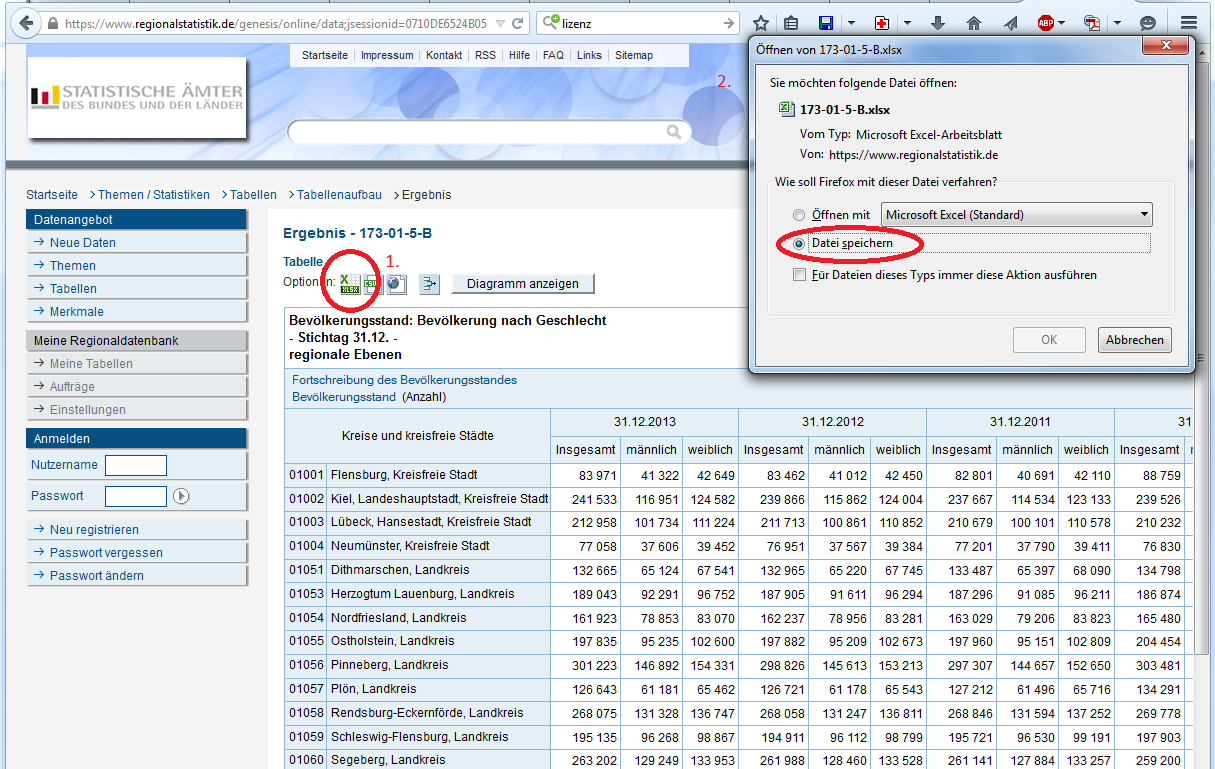
**Ziel:** Daten aus Excel-Dateien in SPSS öffnen

Statistische Daten werden häufig im Excel-Format (.xls oder .xlsx) bereitgestellt, zum Beispiel vom statistischen Bundesamt (<http://www-genesis.destatis.de/genesis/online> bzw. <http://www.regionalstatistik.de>). Dieses Skript beschreibt, wie solche Dateien in SPSS geöffnet werden können und wie sie dafür vorbearbeitet werden müssen.

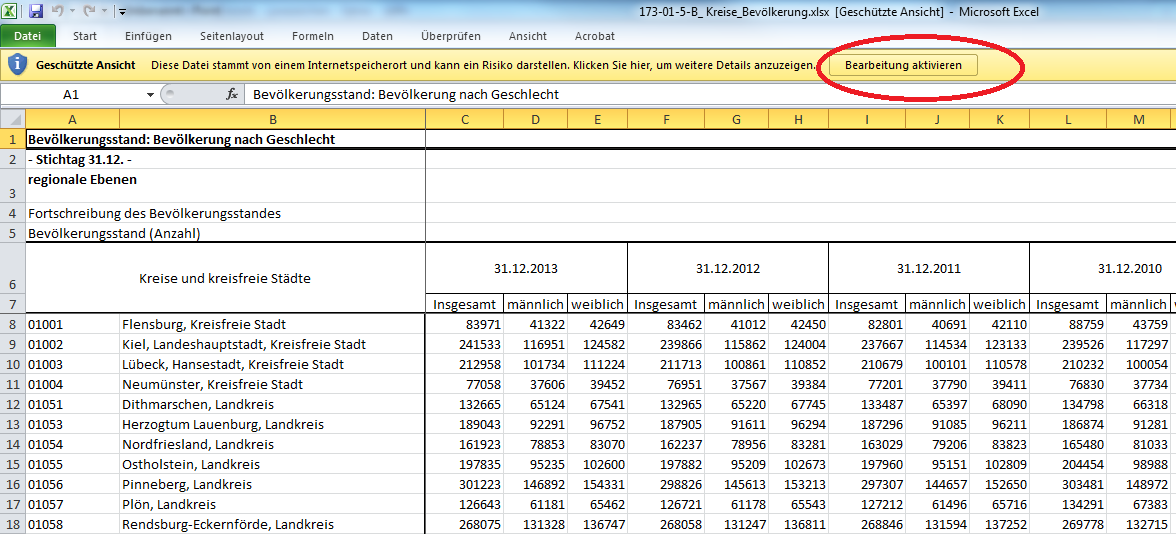
## Datei vorbereiten

Beispiel: Bevölkerungsentwicklung auf Ebene der Kreise und Kreisfreien Städte in Deutschland.  
*abrufbar unter:* [*http://www.regionalstatistik.de*](http://www.regionalstatistik.de) *|Themen | 12 Bevölkerung | 124 Bevölkerungsstand |* [*12411*](https://www.regionalstatistik.de:443/genesis/online/data;jsessionid=0710DE6524B05DF9AEA71C879497BB27?operation=statistikAbruftabellen&levelindex=0&levelid=1448378903790&index=2)[*Fortschreibung des Bevölkerungsstandes*](https://www.regionalstatistik.de:443/genesis/online/data;jsessionid=0710DE6524B05DF9AEA71C879497BB27?operation=statistikAbruftabellen&levelindex=0&levelid=1448378903790&index=2) *|* [*173-01-5-B*](https://www.regionalstatistik.de:443/genesis/online/data;jsessionid=0710DE6524B05DF9AEA71C879497BB27?operation=abruftabelleAbrufen&selectionname=173-01-5-B&levelindex=1&levelid=1448378938317&index=3)[*Bevölkerungsstand: Bevölkerung nach Geschlecht - Stichtag 31.12. - regionale Ebenen*](https://www.regionalstatistik.de:443/genesis/online/data;jsessionid=0710DE6524B05DF9AEA71C879497BB27?operation=abruftabelleAbrufen&selectionname=173-01-5-B&levelindex=1&levelid=1448378938317&index=3) *| Ebene(GEM): Kreise und Kreisfreie Städte.*

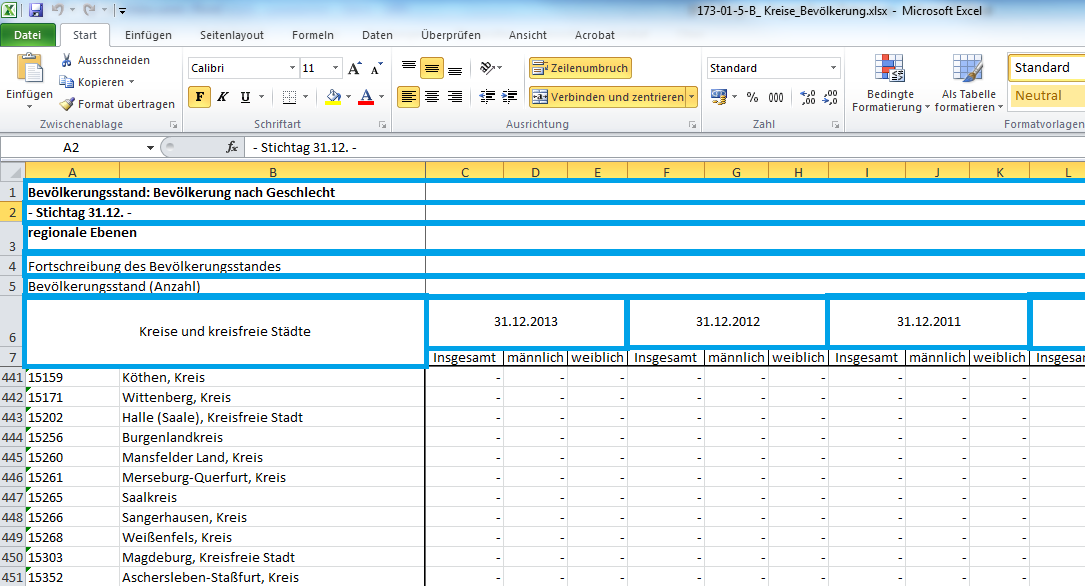
1. **Datei in einem lokalen Ordner abspeichern**



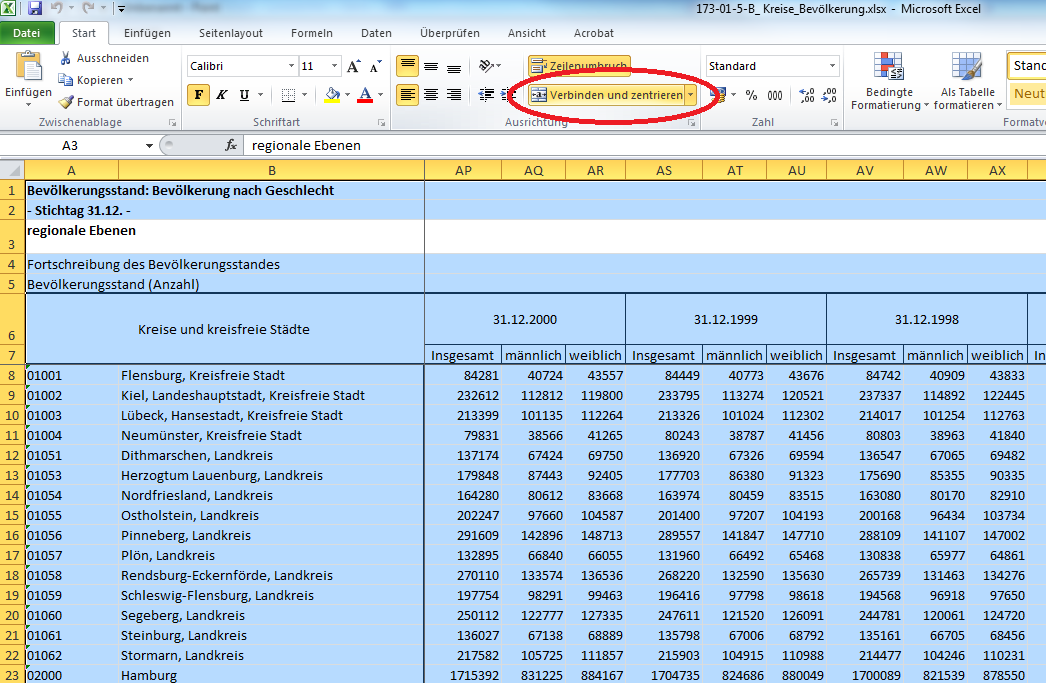
1. **Datei in Excel öffnen.** 
   1. Ggf. Bearbeitung aktivieren.



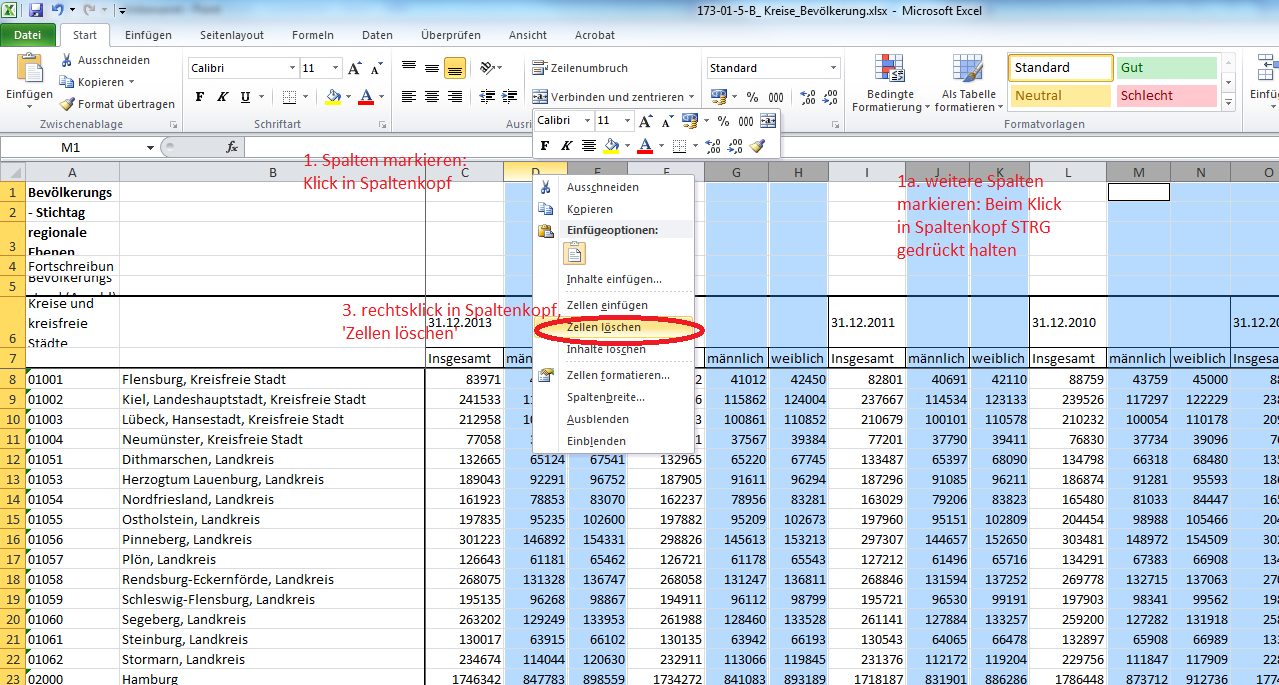
1. **Alle verbundenen Zellen auflösen**, so dass keine verbundenen Zellen mehr vorkommen.   
   *Verbundene Zellen kommen beispielsweise bei Dateien vom Statistischen Bundesamt am Anfang des Datenblattes und bei den mehrspaltigen Überschriften vor:*



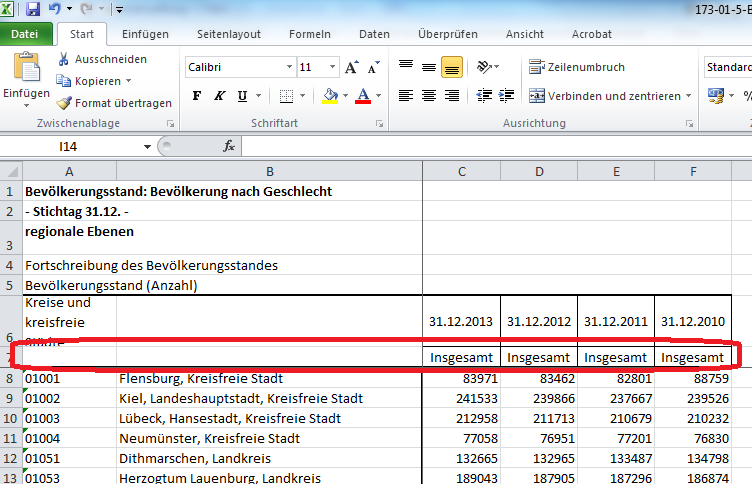
* 1. Mindestens alle verbundenen Zellen markieren.
  2. Besser jedoch: einfach das gesamte Arbeitsblatt markieren: Beliebige Zelle des Arbeitsblattes markieren, STRG+A drücken.
  3. Verbundene Zellen auflösen: einmal auf ‚Verbinden und Zentrieren‘ klicken, um diese Funktion bei verbundenen Zellen abzuschalten.



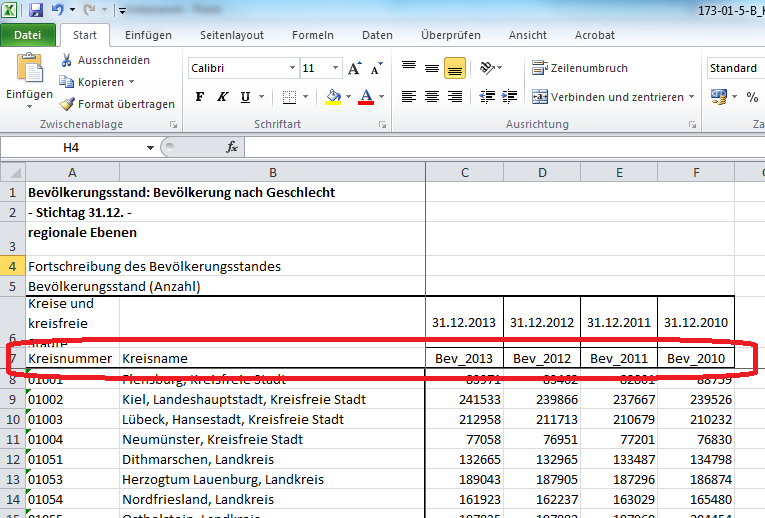
1. **Nicht benötigte Spalten löschen** 
   1. Spalte markieren: Klick in Spaltenkopf (‚Buchstabe‘)
   2. Weitere zu löschende Spalten markieren: Beim Klick in Spaltenkopf ‚STRG‘ gedrückt halten. Bereiche markieren, indem beim Klicken ‚Shift‘ (‚Großschreibung‘) gedrückt wird.
   3. Spalten löschen: Rechtsklick in einen den Kopf einer der zu löschenden Spalten, dabei ‚STRG‘ gedrückt halten. Klick auf ‚Zellen löschen‘



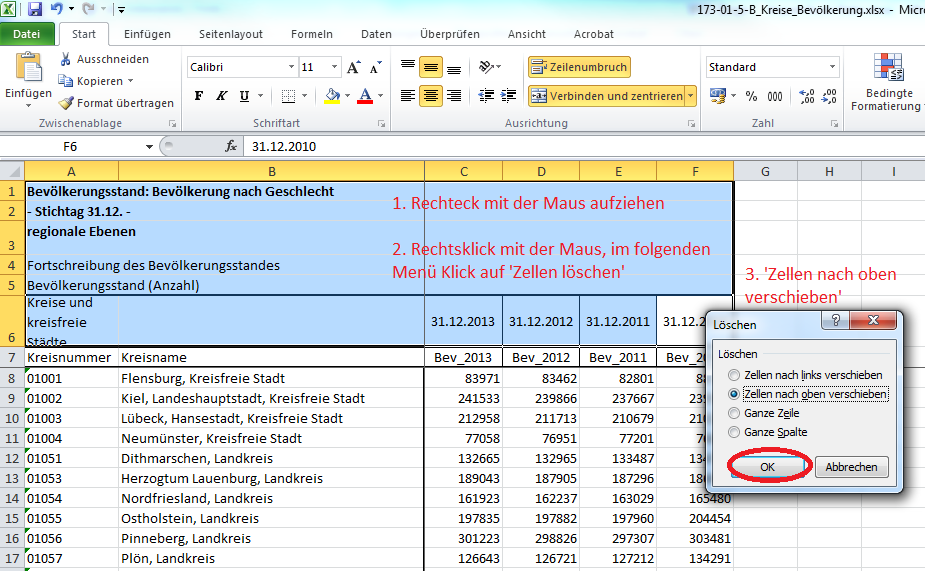
1. **Eindeutige Spaltenüberschriften vergeben** in die Zeile oberhalb der eigentlichen Daten.
   1. Vorher:



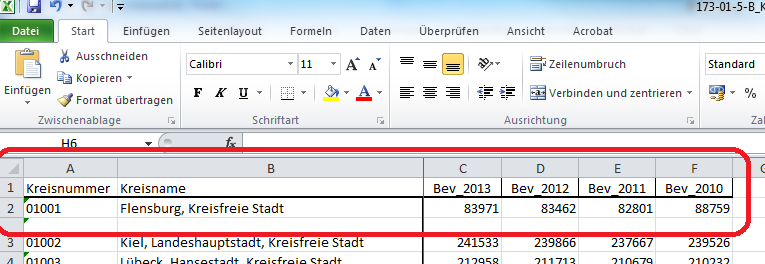
* 1. Nachher:



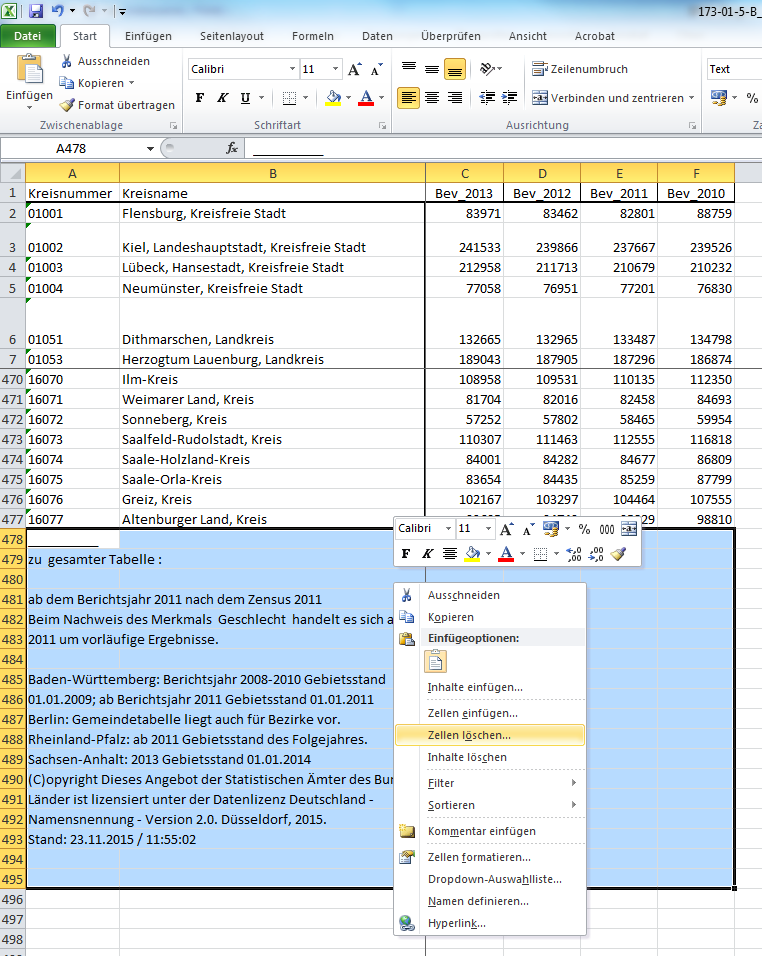
1. **Nicht benötigte Zeilen oberhalb und unterhalb der eigentlichen Werte löschen**. Es darf nur noch eine Zeile mit den Spaltenüberschriften außer den Werten vorhanden sein.
   1. Analog zu den Spalten:
      1. Zeile markieren: Klick in Zeilenkopf (‚Zahl‘)
      2. Weitere zu löschende Zeilen markieren: Beim Klick in Zeilenkopf ‚STRG‘ gedrückt halten. Bereiche markieren, indem beim Klicken ‚Shift‘ (‚Großschreibung‘) gedrückt wird.
      3. Zeilen löschen: Rechtsklick in einen den Kopf einer der zu löschenden Zeilen, dabei ‚STRG‘ gedrückt halten. Klick auf ‚Zellen löschen‘
   2. einfacher:
      1. Gesamtbereich oberhalb der Daten markieren, indem mit der Maus ein entsprechendes Rechteck aufgezogen wird. Rechtsklick in diesen Bereich, im folgenden Menü Klick auf 'Zellen löschen‘. Im folgenden Fenstern markieren: „Zellen nach oben verschieben“, Klick auf ‚OK‘



Ergebnis:



* + 1. Das gleiche unterhalb der Daten:   
       Im Datenblatt ganz nach unten Scrollen. Gesamtbereich unterhalb der Daten markieren, indem mit der Maus ein entsprechendes Rechteck aufgezogen wird. Rechtsklick in diesen Bereich, im folgenden Menü Klick auf 'Zellen löschen‘. Im folgenden Fenstern markieren: „Zellen nach oben verschieben“, Klick auf ‚OK‘



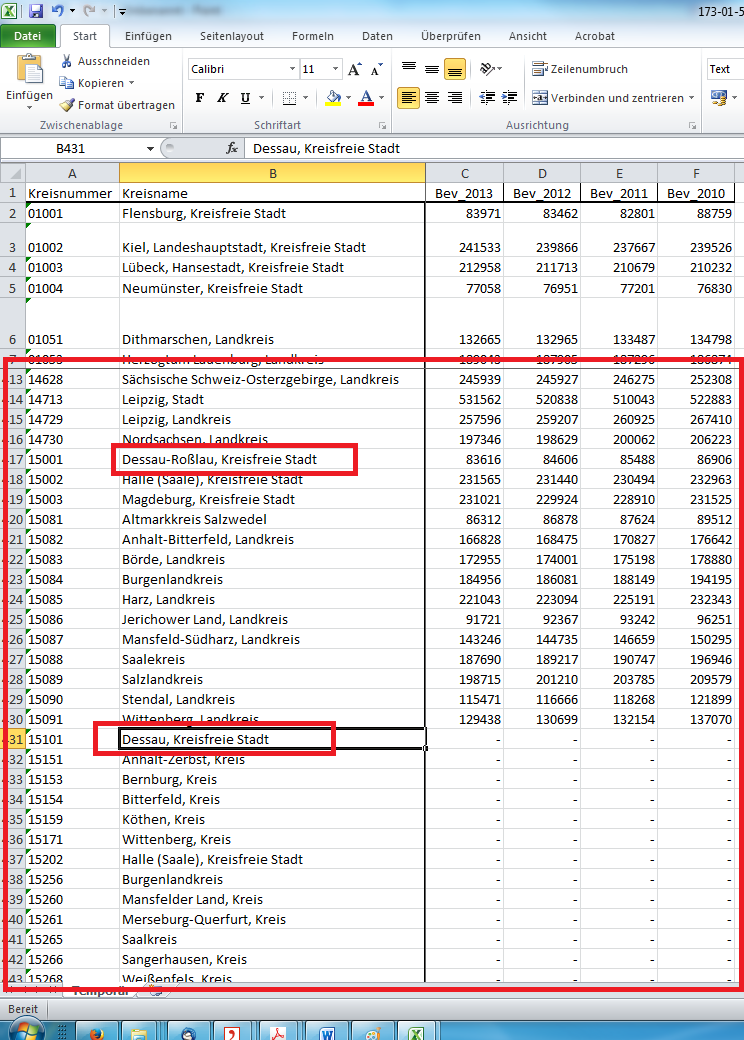
1. Daten auf ‚komische Werte‘ durchsehen –sich einen Überblick über die Daten verschaffen.

Wenn komische Werte auffallen, sollte deren Grund gesucht werden, und dann ein Umgang damit gefunden werden.

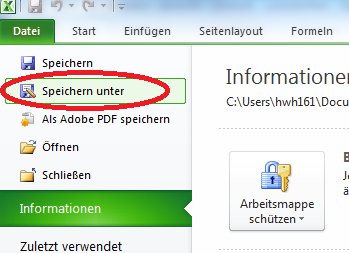
*Zum Beispiel fällt im vorliegenden Beispiel auf, dass bei der Hälfte der Einträge in Sachsen-Anhalt keine Werte zu finden sind, außerdem kommt z.B. Dessau zweimal vor. Ähnliches kommt auch z.B. für Werte in Mecklenburg-Vorpommern, Thüringen und Sachsen vor. Dies liegt an den Gebietsreformen, die z.B. in Sachsen-Anhalt 2007 durchgeführt wurde (vgl.* [*https://de.wikipedia.org/wiki/Kreisreform\_Sachsen-Anhalt\_2007*](https://de.wikipedia.org/wiki/Kreisreform_Sachsen-Anhalt_2007)*). Hier wurde z.B. die kreisfreie Stadt Dessau am südlichen Elbufer und die zum Landkreis Anhalt-Zerbst gehörende Stadt Roßlau (Elbe) am nördlichen Elbufer zur kreisfreien Stadt Dessau-Roßlau fusioniert.*

*In der Statistik werden neben den neuen auch noch die alten Einheiten geführt. Diese zu den alten Gebieten gehörenden Zeilen (ohne Werte) sollten in diesem Fall gelöscht werden.*

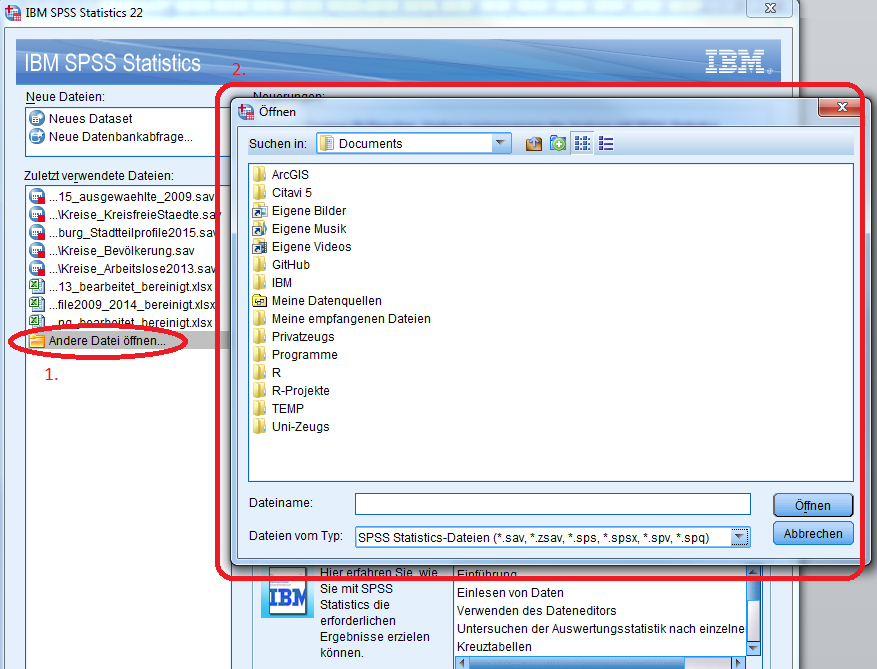
*Auch Hannover, Aachen und Saarbrücken kommen in der Tabelle zweimal vor, einmal als ‚Kommunalverband besonderen Typs‘ zusammen mit dem umgebenden Kreis, einmal nur als Kreisfreie Stadt bzw. dem sie umgebenden Kreis. Hier sollte sich überlegt werden, welche der Zeilen behalten werden sollen, weil sie für die Analyse gebraucht werden.*

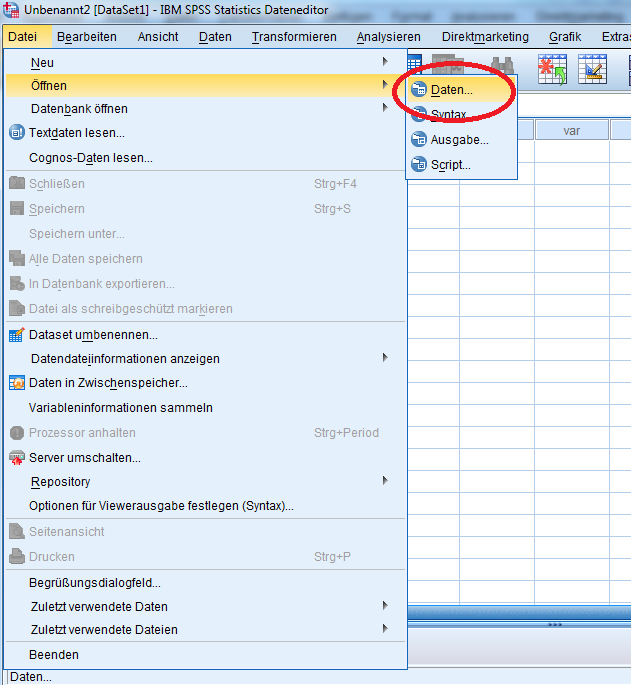


1. (Ggf. Weitere Variablen aus anderen Quellen hinzufügen, siehe ‚SPSS-Skript 3 - Variablen aus verschiedenen Quellen zusammenfügen‘)
2. (Ggf. per Hand weitere Werte hinzufügen, z.B. ‚West‘ ‚‘Ost‘ in einer neuen Spalte mit Überschrift ‚West-Ost‘).
3. **Datei (unter neuem Namen) abspeichern**.   
   Klick auf ‚Datei‘|‘Speichern unter‘. Excel kann anschließend geschlossen werden.

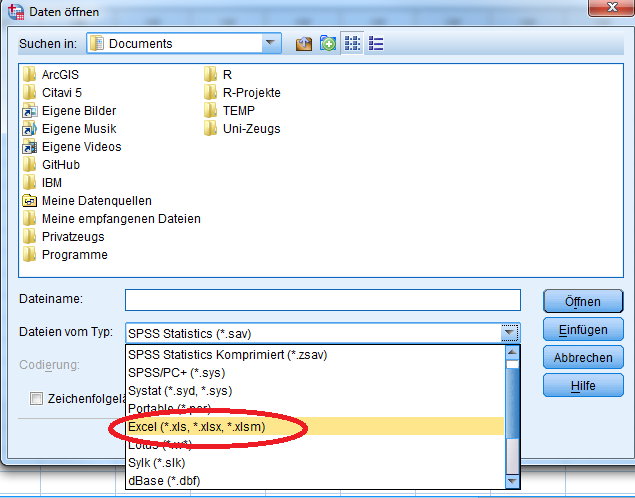


## Datei in SPSS öffnen

1. **SPSS öffnen**.
2. **Dialog für ‚Daten öffnen‘ aufrufen**:
   1. Direkt im Startfenster durch Klick auf ‚Andere Datei öffnen‘
   2. Über das Menü ‚Datei‘ | ‚Öffnen‘ | ‚Daten‘



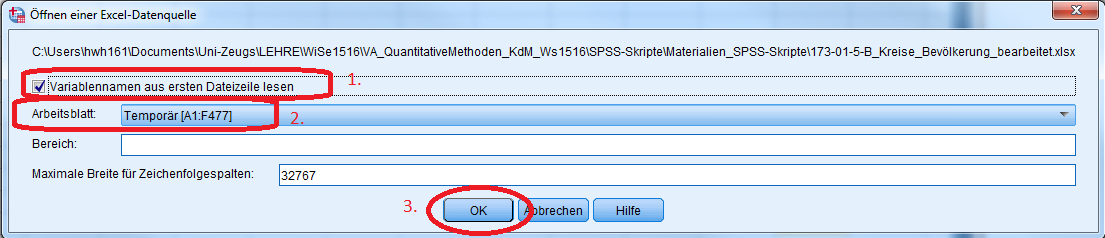
1. Im folgenden Dialogfenster **unter ‚Dateien vom Typ‘ aus dem Auswahlmenü ‚Excel (\*.xls, \*.xlsx, \*.xlsm)‘ auswählen.**



1. Den Ordner auswählen, wo die Excel-Datei zuvor abgespeichert wurde und **Datei** durch Klick auf ‚Öffnen‘ **öffnen**.



1. Im **folgenden Dialog ‚Öffnen einer Excel-Datenquelle**‘ sicherstellen, dass ‚Variablennamen aus der ersten Dateizeile lesen‘ aktiviert ist. Sicherstellen, dass das richtige Arbeitsblatt ausgewählt ist. Voreingestellt ist das erste Arbeitsblatt, was meistens das richtige sein sollte(im Beispiel mit dem Namen ’Temporär‘).   
   Abschließen durch Klick auf ‚OK‘

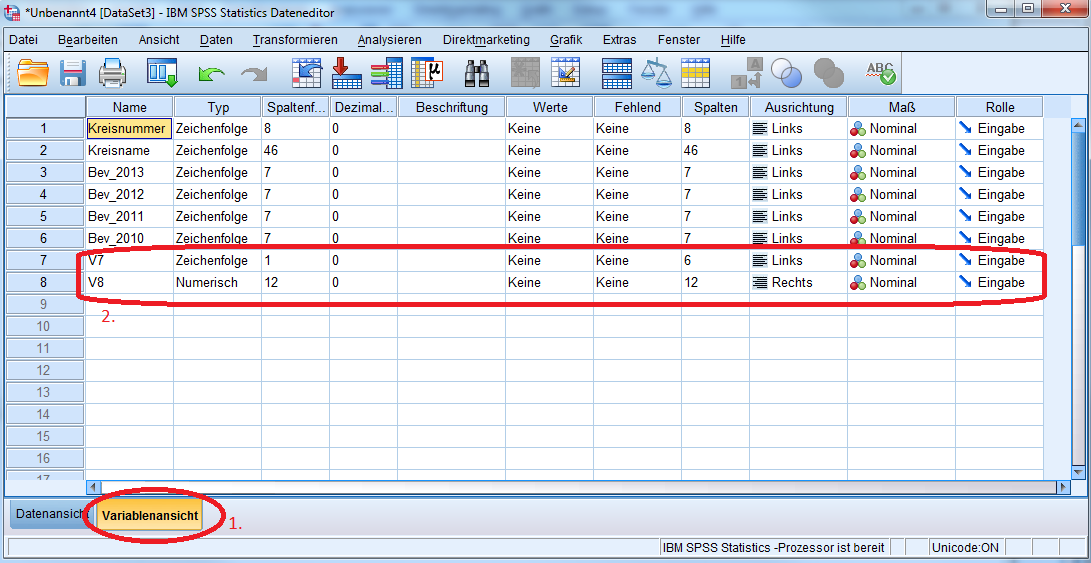


Jetzt sind die Daten in SPSS geöffnet. Jedoch sollten sie noch auf den richtigen Datentyp und evtl. Importfehler geprüft werden. Dies ist Gegenstand des nächsten Abschnitts.

## 

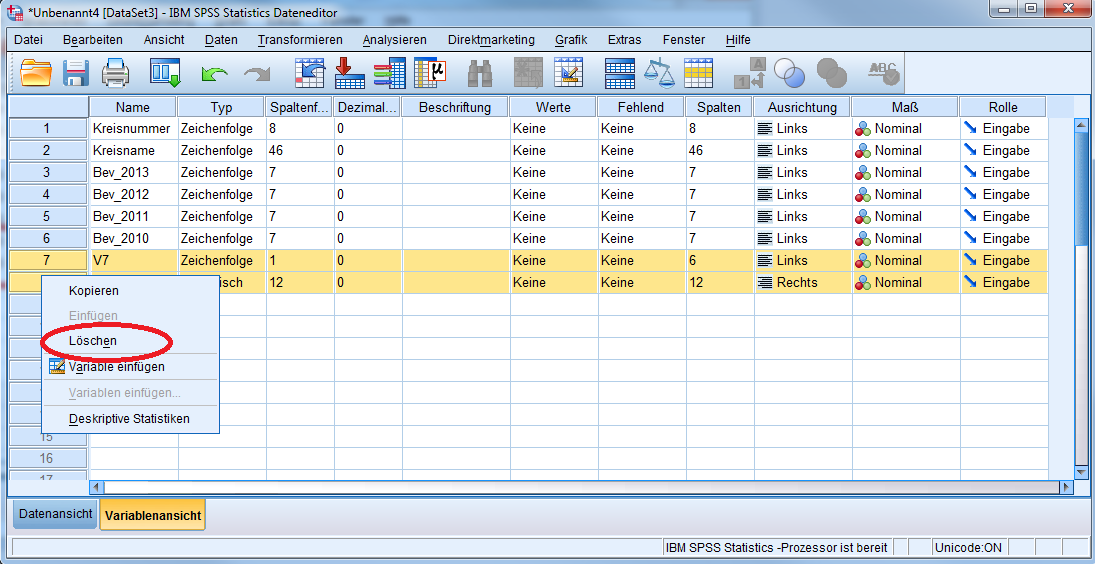
## Nachbearbeitung in SPSS

Nach dem Import müssen die Daten auf Korrektheit geprüft werden.

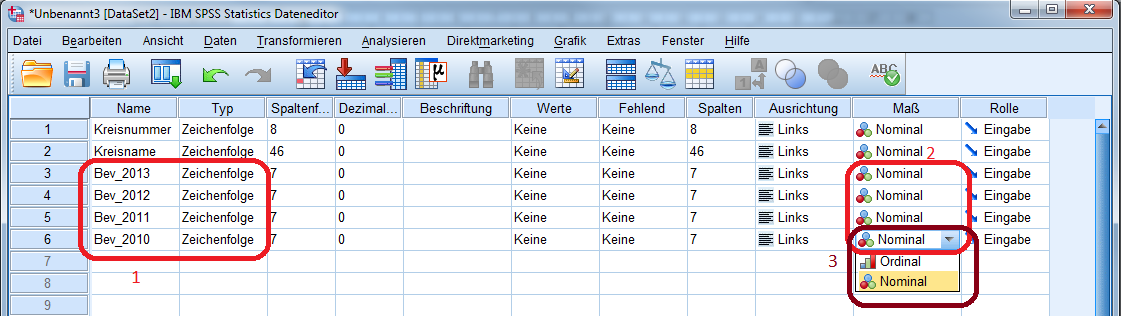
1. **Variablenansicht aufrufen**.
2. **Anzahl der Variablen prüfen**. 

Manchmal werden zu viele Spalten importiert und als Variablen interpretiert. Da diese dann keine Überschrift haben, werden hierfür die Namen ‚VXX‘ (mit XX=Zahl) vergeben. Diese Variablen sollten gelöscht werden.

* 1. Analog zur Vorgehensweise in Excel die entsprechenden Variablen (in der Variablenansicht: die Zeilen) durch Klick in den Zeilenkopf markieren.   
     Mehrere Zeilen markieren, indem beim Klicken ‚STRG‘ gedrückt gehalten wird.   
     Zeilenbereiche markieren, indem beim Klicken auf den Endpunkt ‚Shift‘ (‚Großschreibung‘) gedrückt wird.
  2. Variablen löschen: Drücken von ‚Entf‘ oder Rechtsklick in Zeilenkopf und auf ‚Löschen‘ klicken.

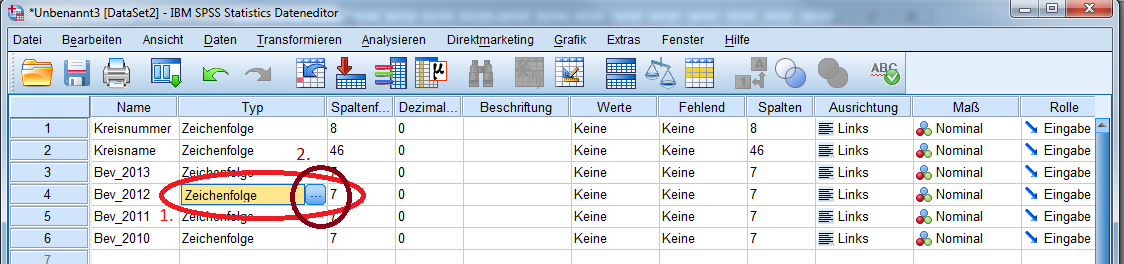


1. **Variablentypen und Variablenniveau(=‘Maß‘) auf Korrektheit prüfen**

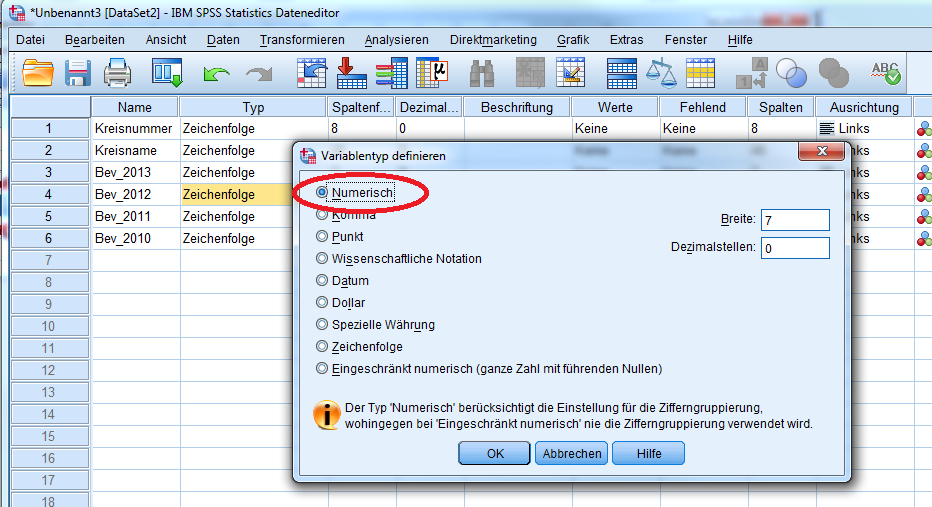
Im vorliegenden Fall sind Typ ‚Zeichenfolge‘ und Maß ‚nominal‘ für die Variablen   
‚Kreisnummer‘ und ‚Kreisname‘ korrekt. Die Variablen für Bevölkerung wurden jedoch nicht korrekt erkannt. Dies liegt daran, dass ein fehlender Wert durch ein Strich ‚-‘ symbolisiert wurde. Immer wenn mindestens ein Wert nicht numerisch ist, wird beim Import der Datentyp automatisch auf ‚Zeichenfolge‘ gesetzt (1). Deswegen ist auch das Maß ‚nominal‘ (2) und kann auch nicht auf ‚metrisch‘ geändert werden (3): 

Um dies zu ändern muss:

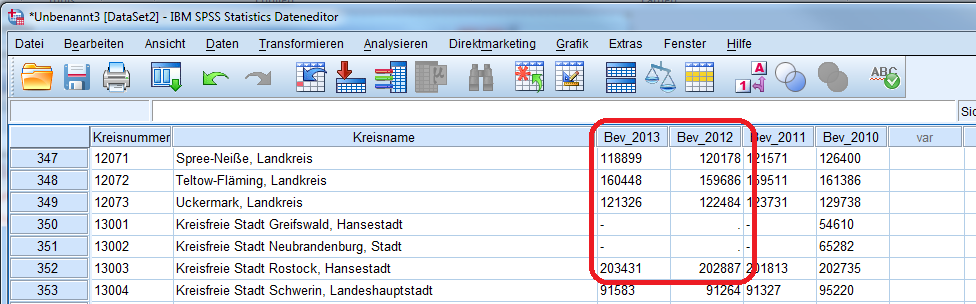
* 1. Der Variablentyp bei den betreffenden Variablen auf ‚Numerisch‘ gesetzt werden.
     1. Klick bei den entsprechenden Variablen auf das Feld ‚Typ‘. Dort Klick auf die drei Punkte ‚…‘



* + 1. Im folgenden Dialog den Datentyp von ‚Zeichenfolge‘ auf ‚Numerisch‘ ändern. Mit ‚OK‘ bestätigen.

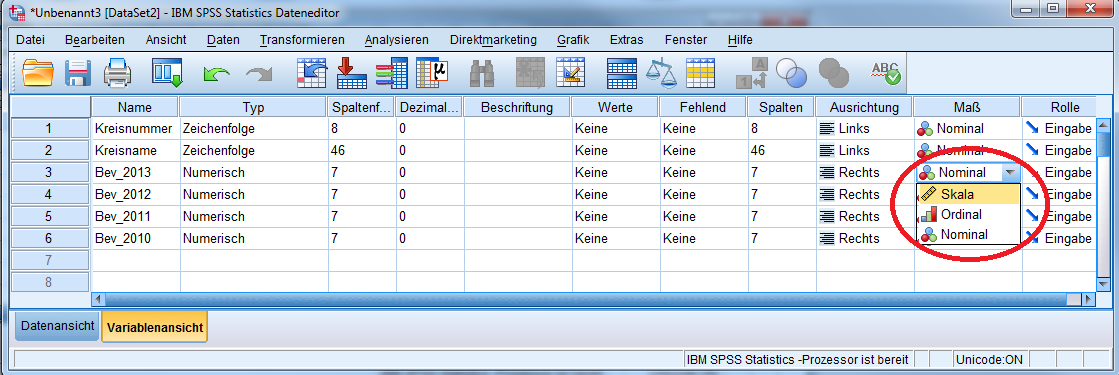


* + 1. Dabei werden die fehlenden Werte (‚-‘) automatisch durch den SPSS-eigenen fehlenden Wert (symbolisiert durch ‚.‘) ersetzt. Zum Vergleich die Datenansicht. Bei der Variable ‚Bev\_2012‘ ist der Typ bereits auf Numerisch gesetzt worden, bei den anderen noch nicht.

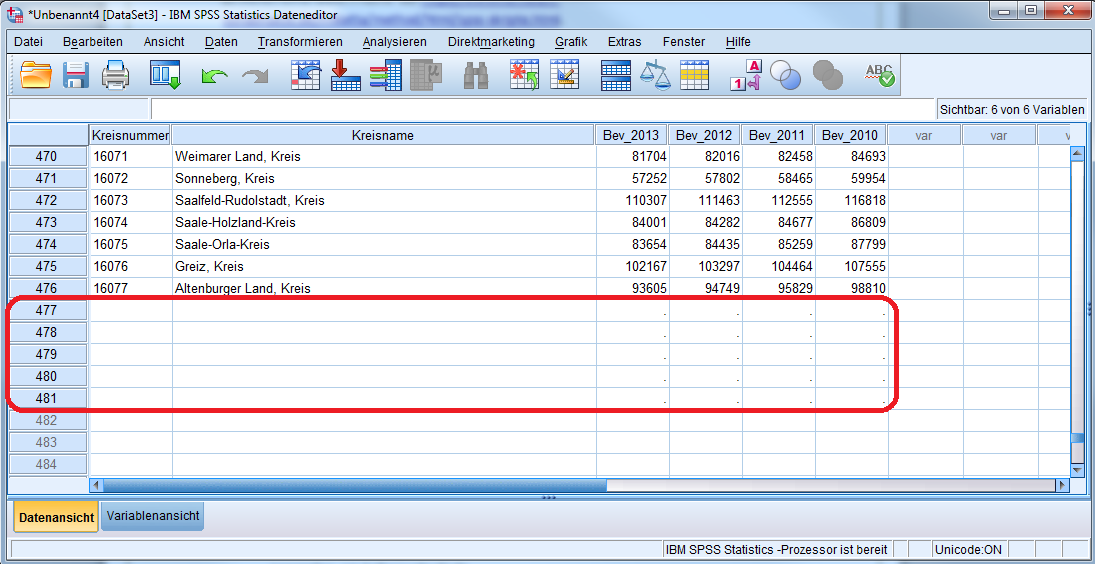


* 1. Nach dem Ändern des Datentyps kann das Maß auf ‚Skala‘ (=metrisch) gesetzt werden.

Hierzu muss in der Variablenansicht für jede betroffene Variable in das Feld ‚Maß‘ geklickt und aus dem Auswahlmenü ‚Skala‘ ausgewählt werden.



1. Wenn Typ und Maß aller Variablen korrekt sind, können ggf. noch eine **Beschriftung und Wertelabels hinzugefügt sowie fehlende Werte definiert werden.** Vergleiche hierzu das Skript ‚Erstellen einer Datenmaske‘ auf <http://www.home.uni-osnabrueck.de/rniketta/method/html/spss-skripte.html>.
2. **Daten auf Korrektheit prüfen.**Wenn die Variableneinstellungen korrekt sind, sollte noch ein Blick auf die Daten selbst geworfen werden. Insbesondere sollte überprüft werden, ob **nicht zu viele Zeilen** importiert worden sind (und jetzt als Leerzeilen erscheinen).
   1. In Datenansicht wechseln
   2. Ganz nach oben und ganz nach unten scrollen. Zuviel importierte Fälle erscheinen als Leerzeilen, am Anfang oder Ende.  
      Im Beispiel wurden die Zeilen 477 bis 481 zu viel importiert, zu erkennen an den schwarzen (nicht ausgegrauten) Zeilenköpfen bzw. den Punkten für die fehlenden Werte bei den numerischen Variablen.



* 1. Diese Fälle sollten analog zu den überflüssigen Variablen entfernt werden
     1. Analog zur Vorgehensweise in Excel die entsprechenden Fälle (in der Datenansicht: die Zeilen) durch Klick in den Zeilenkopf markieren.   
        Mehrere Zeilen markieren, indem beim Klicken ‚STRG‘ gedrückt gehalten wird.   
        Zeilenbereiche markieren, indem beim Klicken auf den Endpunkt ‚Shift‘ (‚Großschreibung‘) gedrückt wird.
     2. Fälle löschen: Drücken von ‚Entf‘ oder Rechtsklick in Zeilenkopf und Klick auf ‚Löschen‘

