

## Instructie omzetten van datasets naar vustat-format

Datasets die met een ander programma, bijvoorbeeld een spreadsheet of tekstverwerker, zijn gemaakt kun je omzetten naar een dataset in vustat format. De set kun je vervolgens opslaan. Aan de orde komt: dataset in excel-format, dataset in txt-format (ASCII file), dataset in een tekst

### Dataset in excel-format

Excel bestanden kun je direct importeren via het menu van Data analyse:

**Bestand > Import Excel bestand**

Het Excel bestand moet je eventueel aanpassen zodat de eerste regel de variabelen bevat.

Het Excel bestand moet een extensie xls hebben, de Excel 97-2003 Werkmap.

Een xlsx bestand moet je daarom via Excel eerst opslaan als een xls bestand.

### Dataset in txt-format (ASCII)

- Open in VUStat de module **Data analyse**.

- Kies in het menu **Opties > Importeren**

Het venster **Inlezen ASCII bestand** verschijnt.

In het txt-bestand kunnen op de eerste regel wel of niet de namen van de variabelen staan.

Het best is een variabele als één woord, zonder spaties of andere tekens.

- Je moet hier opgeven of er wel of geen variabelenrij is.

Als het niet bekend is moet je er door maar gewoon proberen achter komen.

Indien er in het txt-bestand geen variabelen zijn worden in de datatabel variabelen V1, V2 enzovoort geplaatst. Later kun je die dan aanpassen.

- Je kunt aanvinken met welk scheidingsteken rekening gehouden moet worden.

Als je niet weet wat voor scheidingsteken is gebruikt vink je gewoon alles aan.

- Open het txt-bestand - **OK**

- "Importeren gereed" dan zijn er geen problemen geconstateerd

- Pas zo nodig het type variabele aan.

- Klik rechts op de naam van een variabele

- Kies **Wijzigen naam variabele**

- Pas aan.

- Sla de dataset op

### Dataset in een tekst

- Kopieer de dataset.

- Plak de dataset in een leeg document.

- Sla op als txt-file (bv in Wordpad) of als tekst zonder opmaak (bv in Word).

- Open **VUStat > Data analyse > Opties > Importeren**

- Ga verder als bij Dataset in tekst-format

## Scheidingstekens

In een tekst-dataset staan soms geen nette kolommen, dat komt vanzelf wel goed als een scheidingsteken consequent is gebruikt.

Om de waarden in een tekstfile van elkaar te kunnen onderscheiden zijn scheidingstekens nodig.

Veel voorkomende scheidingstekens zijn bijvoorbeeld tabs, komma's of puntkomma's.

Een bron van fouten kan een tekstitem zijn met een spatie bv. de plaatsnaam Den Haag, waardoor Den en Haag bij twee opeenvolgende variabelen terecht komen. Je kunt dan beter de spatie weghalen (DenHaag) of opvullen (Den-Haag).

## Ontbrekende waarden

In datasets komen ook records voor die waarden missen. In tekstbestanden worden zulke ontbrekende waarden vaak gecodeerd aangegeven, bijvoorbeeld met 999. Deze code dient bekend te zijn opdat een ontbrekende waarde in VUStat ook correct als een ontbrekende waarde wordt aangegeven met \*\*.

## Categoriale variabelen (Kwalitatieve variabelen)

Categoriale variabelen (b.v. geslacht met labels man en vrouw) zijn interessant omdat de dataset op zo'n kenmerk kan worden gesplitst. In datasets zijn categoriale variabelen als een tekst variabele opgenomen of als een code. In VUStat kun je in beide gevallen een goede categoriale variabele met labels maken.

### Van tekst variabele naar categoriale variabele met labels.

- Open in **Data analyse** de dataset.
- Kies in menu **Opties > Tekstvariabele > Tekst naar labels**
- Selecteer de **Tekstvariabele** en de **Doelvariabele**  
De doelvariabele kan een bestaande variabele zijn of je kunt een nieuwe maken.
- **OK**. De categoriale variabele met labels is aangemaakt. De tekstvariabele is niet meer nodig.

### Van codes naar labels

- Open in **Data analyse** de dataset.
- Klik rechts op de variabele met de codes
- Selecteer **Wijzigen naam variabele**
- Verander type van de variabele in **Categorie**
- Klik op **Labels** en maak de labels passend bij de codes
- **OK**
- Sla de dataset op