Overzicht configuratie script tabellenboek

# Introductie

Dit script wordt gebruikt om de data van een gezondheidsmonitor om te zetten in tabellenboeken voor de verschillende gemeentes, scholen, regio’s, enz. In principe kan op iedere variabele een specifieke set tabellenboeken worden gemaakt, met daarbij ook uitsplitsing op een ‘hoger’ niveau. (Denk bijvoorbeeld aan tabellenboeken per gemeente, waarbij ook een kolom is meegenomen voor een subregio.)

Alle data wordt opgehaald uit de SPSS-bestanden. Het is van belang dat missende waardes en labels hierin goed zijn aangegeven: deze worden automatisch overgenomen.

Dit script is een herschreven versie van het tabellenboekscript dat al een paar jaar circuleert en telkens wordt uitgebreid. Deze situatie werd na het vijf keer doorgeven van het auteurschap onhoudbaar, waarop er vanuit GGData een werkgroep is opgezet om één universeel script op te zetten, welke voor alle monitors gebruikt kan worden. Het script zal daarom vanaf heden worden onderhouden door deze werkgroep. De laatste versie is altijd beschikbaar vanaf de Github van GGData: <https://github.com/ggdatascience/GGData_tabellenboek>

Werkgroepleden:

* Henk Broekhuizen (Twente)
* Aart Dijkstra (Noord- en Oost-Gelderland)
* Joanne Nies (Twente)
* Pieter Seinen (Gelderland-Zuid)

Voor foutmeldingen en functionaliteitsaanvragen verwijzen we in eerste instantie naar de ‘Issues’ pagina op Github: <https://github.com/ggdatascience/GGData_tabellenboek/issues> Problemen en foutmeldingen kunnen daar worden aangeleverd, inclusief bijbehorende reproductiestappen en bijlages.

# Werking

Het script bestaat uit een aantal R-bestanden (te herkennen aan de prefix “tbl\_”), deze handleiding, en een configuratiebestand in Excel. Het uitvoeren van het script gebeurt door tbl\_maken.R uit te voeren in R of Rstudio; de andere onderdelen worden vanuit dit script automatisch ingeladen.

Het script kan data gebruiken uit één of meerdere datasets. Als er meerdere datasets zijn worden deze gecombineerd tot één ‘superdataset’, waarin automatisch ‘superweegfactoren’ en ‘superstrata’ worden aangemaakt. Hierbij dient wel te worden aangegeven welke variabelen de strata en de weegfactoren bevatten. (Zie de naslag hieronder.)

Door deze combinatie tot een ‘superdataset’ is het mogelijk om chi2-testen uit te voeren *tussen* datasets, zoals het vergelijken van een regionale dataset met de landelijke dataset. Analyses worden uitgevoerd met het *survey* package van Thomas Lumley. Deze werkt volgens dezelfde principes als *Complex Samples* in SPSS, en zal gelijkwaardige resultaten geven.

Het uitvoeren van deze berekeningen is intensief en kan lang duren. Om dit werkbaar te houden met aanpassingen aan het tabellenboek is er een automatische opslag van resultaten ingebouwd. Deze analyseert of er in de configuratie veranderd zijn die een nieuwe berekening noodzakelijk maken. Veranderingen in opmaak of volgorde van variabelen zal daarom **niet** leiden tot een nieuwe berekening, waardoor het genereren van nieuwe tabellenboeken **na de eerste keer** beduidend sneller zal gaan.

## Plaatsing en configuratie

Het script is ontworpen om meerdere keren gebruikt te kunnen worden, voor meerdere monitors. Daarom wordt er ‘vanuit de configuratie’ gedacht; het script neemt aan dat de datasets in dezelfde map staan als het configuratiebestand, en dat de resultaten daar ook geplaatst dienen te worden. Pas, indien mogelijk, de configuratie niet naderhand aan, maar kopieer dit bestand voor een volgende monitor en maak in het nieuwe bestand de gewenste aanpassingen. Hierdoor kunnen oudere monitors ook simpel opnieuw gedraaid worden.

Het script zelf hoeft maar één keer op een makkelijke locatie te staan. Tijdens het uitvoeren wordt er gevraagd om een configuratiebestand te selecteren, waarna het script in die map aan de slag gaat. Tijdens het uitvoeren worden de mappen *output* en *resultaten\_csv* automatisch aangemaakt in de map waarin de configuratie zich bevindt. Tabellenboeken zullen worden geplaatst in de submap *output*. Berekende resultaten worden geplaatst in *resultaten\_csv*. Deze tabel **kan** gebruikt worden voor bijvoorbeeld het schrijven van de rapportage, aangezien alle aantallen, percentages en p-waarden hierin weergegeven zijn. (Al is dit niet heel aantrekkelijk gepresenteerd; de tabellen zijn bedoeld voor verwerking in R.)

# Stappenplan (LET OP: nog niet actueel voor het nieuwe script)

Het uitvoeren van het script bestaat globaal uit 3 stappen.

## Stap 1 – configuratie en plaatsing

* Verplaats de benodigde bestanden (tabellenboeken\_maken.R, functies\_tabellenboeken.R, configuratie.xlsx) naar een logische map. Dit kan bijvoorbeeld een NAS of gedeelde schijf zijn, of een projectmap.
* Maak in dezelfde map een submap aan met de naam “data”. Plaats hierin de relevante SPSS-bestanden.
* Vul de configuratie in (zie naslag hieronder). De bijgevoegde configuratie is puur ter voorbeeld, en geenszins volledig. Er is hieronder (bij [Snel van start](#_Snel_van_start_1)) een kort stappenplan toegevoegd om hiermee te beginnen.

## Stap 2 – uitvoeren script

* Open R of Rstudio en voer tabellenboeken\_maken.R uit. (functies\_tabellenboeken.R wordt automatisch ingeladen, hier hoeft niks voor gedaan te worden.)
* Selecteer het configuratiebestand dat in stap 1 is ingevuld.
* Ga een kopje koffie drinken. Dit duurt even. (Reken voor een volledige monitor op meer dan een uur.) Uiteraard kan tijdens het draaien de computer gebruikt worden, maar R zelf **absoluut niet**.

## Stap 3 – controle en herziening

* In de map waarin het configuratiebestand zich bevindt is nu een map aangemaakt, genaamd “output”. Hierin staan de tabellenboeken, gesplitst op de eerste variabele. (Zie tabblad *onderdelen*, hieronder.)
* Waarschijnlijk moeten er nog zaken worden aangepast. Denk aan kopjes, kleuren, lettertype, enz. Deze kunnen in het configuratiebestand worden aangepast.
* Als er een andere **dataset** nodig is of **meer variabelen** worden toegevoegd, moeten de CSV’s opnieuw berekend worden. Zet hiervoor *bestaande\_csv\_opnieuw\_berekenen* op tabblad *algemeen* op WAAR.
* Als er **geen** nieuwe dataset of variabele nodig is, kan het berekenen overgeslagen worden. Dit gebeurt automatisch in de standaardconfiguratie. Het script zal dan alleen nieuwe tabellenboeken maken, maar geen berekeningen meer uitvoeren. Dit scheelt aanzienlijk in tijd. (Reken nu op een minuut of 2-3.)
* Als alles naar wens is kunnen de tabellenboeken worden gepubliceerd.

## Snel van start

Het is makkelijker om vanuit een werkend prototype verder te werken aan verduidelijking of verdere toevoegingen dan om direct alles in te stellen. Om een eerste beeld te vormen is het daarom aan te raden om te beginnen met één dataset en een paar variabelen. Van daaruit wordt het vanzelf duidelijk wat iedere optie doet en kan er meer worden toegevoegd. Ter voorbeeld staat hieronder een stappenplan om vanuit de recente jongvolwassenenmonitor een snel tabellenboek te maken.

1. Plaats het script en alle bijbehorende bestanden op een logische plek en maak in dezelfde map een map aan met de naam “data”.
2. Plaats in de map “data” het regiobestand. In dit voorbeeld gaan we ervan uit dat het bestand “Regiobestand.sav” heet.
3. Open de voorbeeldconfiguratie en sla deze op met een duidelijke naam, bijvoorbeeld “configuratie JV 2022.xlsx”.
4. Ga naar tabblad *datasets* en verwijder de voorbeelden. Voeg het volgende toe:
   1. *naam\_dataset*: GGD <uw regio>
   2. *bestandsnaam:* Regiobestand.sav
   3. *jaarvariabele:* AGOJB401 (optioneel)
   4. *jaren\_voor\_analyse:* 2022 (optioneel)
   5. *waarden\_dichotoom:* 0,1,8|1,2,8 (zie op een later punt de toelichting hieronder)
5. Ga naar tabblad *onderdelen* en verwijder de voorbeelden. Voeg het volgende toe:
   1. *dataset:* GGD <uw regio> (moet gelijk zijn aan stap 4a)
   2. *subsetvariabele:* (leeg)
   3. *met\_crossing:* WAAR
   4. *weegfactor:* Standaardisatiefactor\_gemeente
   5. *stratum:* Stratum\_gemeente
6. Ga naar tabblad *indeling\_kolommen* en verwijder de voorbeelden. Voeg het volgende toe:
   1. *Tabblad:* Regio (bijvoorbeeld)
   2. *dataset:* GGD <uw regio> (moet gelijk zijn aan stap 4a)
   3. *subset:* zelf
   4. *geef\_crossing:* WAAR
   5. *jaar:* 2022 (moet gelijk zijn aan stap 4d)
   6. *sign\_toetsen\_totaal:* (leeg)
   7. *sign\_toetsen\_crossings:* intern
7. Ga naar tabblad *indeling\_rijen* en verwijder alles behalve de bovenste 7 regels. (Dus verwijderen vanaf “Mantelzorg”.)
8. Sla het bestand op en sluit dit in Excel. (Helaas kan maar één programma tegelijk het bestand gebruiken. Het script waarschuwt vanzelf als Excel nog draait.)
9. Voer het script uit. Er wordt gevraagd naar een configuratie; selecteer het bestand dat net is aangemaakt.
10. In de map van het script zijn extra mappen verschenen. In de map “output” staat nu een bestand met de naam “GGD <uw regio>.xlsx”. Controleer dit. (Voor een extra indicatie: pas *geef\_crossing* aan naar ONWAAR in tabblad *indeling\_kolommen*, voer het script uit, en kijk wat er gebeurt.)
11. Nu gaan we de nationale dataset toevoegen als referentieset. Open weer het configuratiebestand.
12. Ga naar tabblad *datasets* en voeg het volgende toe:
    1. *naam\_dataset*: NL totaal
    2. *bestandsnaam:* Totaalbestand.sav (let op dat deze naam klopt met het bestand)
    3. *jaarvariabele:* AGOJB401 (optioneel)
    4. *jaren\_voor\_analyse:* 2022 (optioneel)
    5. *waarden\_dichotoom:* 0,1,8|1,2,8 (zie op een later punt de toelichting hieronder)
13. Ga naar tabblad *onderdelen* en voeg het volgende toe:
    1. *dataset:* NL totaal (moet gelijk zijn aan stap 12a)
    2. *subsetvariabele:* (leeg)
    3. *met\_crossing:* ONWAAR
    4. *weegfactor:* Standaardisatiefactor
    5. *stratum:* Stratum
14. Ga naar tabblad *indeling\_kolommen* en voeg het volgende toe:
    1. *Tabblad:* Regio (bijvoorbeeld)
    2. *dataset:* NL totaal (moet gelijk zijn aan stap 12a)
    3. *subset:* (leeg)
    4. *geef\_crossing:* ONWAAR
    5. *jaar:* 2022 (moet gelijk zijn aan stap 12d)
    6. *sign\_toetsen\_totaal:* grote\_set
    7. *sign\_toetsen\_crossings:* (leeg)
15. Sluit wederom het bestand af en voer het script uit. Er is nu een extra kolom toegevoegd, waarin de totalen van Nederland zichtbaar zijn. Daarnaast zijn er nu een aantal getallen in de totaalkolom dikgedrukt; deze verschillen van de nationale data.
16. Zie voor verdere uitbreidingen de onderstaande naslag.

# Configuratie

Hieronder worden de verschillende tabbladen in het configuratiebestand toegelicht, op alfabetische volgorde. Variabelen die waarschijnlijk aangepast moeten worden zijn vetgedrukt, met uitzondering van het tabblad *opmaak* en *indeling\_rijen*; daar is alles belangrijk om aan te passen naar eigen huisstijl. Let erop dat alle instellingen hierin, inclusief de namen van de tabbladen en de gebruikte variabelen in SPSS, hoofdlettergevoelig zijn. Extra tabbladen toevoegen is geen probleem, tabbladen verwijderen zal leiden tot een fout in het script. Verplaatsen van tabbladen is eveneens geen probleem: het script zoekt op basis van naam.

**Statistische significantie van verschillen**

Er is in dit script vanwege technische redenen gekozen om een vergelijking met één kolom mogelijk te maken, of onderlinge vergelijkingen binnen kruisvariabelen, op basis van de betrouwbaarheidsinterval. Het is de vraag of dit binnen iedere GGD de gewenste methode is. Een uitbreiding naar meerdere soorten vergelijkingen was echter dermate ingewikkeld dat dit voor het huidige script niet is gedaan. Indien gewenst kan een interne data scientist of onderzoeker met kennis van R hier vrij makkelijk een aanpassing in maken. (Relevante stukken bevinden zich vanaf regel 634 in *tabellenboeken\_maken.R*.) Verder zijn er nog overwegingen rondom de validiteit van de gebruikte methode, zie ook de opmerkingen bij het tabblad *indeling\_kolommen*.

## Algemeen

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| bestaande\_csv\_opnieuw\_berekenen | WAAR / ONWAAR | Het berekenen van de survey designs en kruistabellen kost redelijk wat tijd. (Reken voor een volledige monitor met meerdere crossings op meer dan een uur.) Als er alleen geklust wordt aan de opmaak van de tabellenboeken is het aan te raden om deze waarde op ONWAAR te zetten; dan worden de eerder gegeneerde CSV’s opnieuw gebruikt. Bij het toevoegen van nieuwe variabelen dienen deze opnieuw berekend te worden. (Dit gaat **niet** automatisch.) Hierbij is het snelst om gewoon de map *resultaten\_csv* te legen, maar indien gewenst kan dit ook geforceerd worden middels deze optie. |
| min\_observaties\_per\_vraag | Getal | Minimaal aantal observaties per vraag. Vragen met minder observaties worden weergegeven als missend (met de tekst hieronder). |
| min\_observaties\_per\_antwoord | Getal | Minimaal aantal observaties per antwoord. Antwoorden met minder observaties worden weergegeven als missend (met de tekst hieronder). |
| tekst\_min\_vraag\_niet\_gehaald | Tekst | Tekst die moet worden weergegeven als het aantal observaties niet is gehaald. |
| tekst\_min\_antwoord\_niet\_gehaald | Tekst | Tekst die moet worden weergegeven als het aantal antwoorden per vraag niet is gehaald. |
| tekst\_missende\_data | Tekst | Tekst die moet worden weergegeven als de data volledig mist. (Dit kan bijvoorbeeld gebeuren bij het combineren van data uit voorgaande jaren.) |
| **sign\_toetsen** | WAAR / ONWAAR | Specificeert of significantie moet worden berekend a.d.h.v. betrouwbaarheidsintervallen. Het niveau kan worden aangepast middels de waarde in *confidence\_level*. De vergelijking die getoetst moet worden dient gespecificeerd te worden in *onderdelen*. Significante verschillen worden standaard dikgedrukt weergegeven. (Dit kan worden aangepast in *functies\_tabellenboeken.R* vanaf regel 1131.) |
| confidence\_level | Getal | Gewenste confidence interval voor significantie. 0.95 betekent een 95% confidence interval, 0.99 betekent 99%. (Zie ook *indeling\_kolommen*.) |
| subsets\_moeten\_voorkomen\_in\_onderwerp\_tbb | WAAR / ONWAAR | Let op; dit is GEEN filter. Als deze waarde op waar staat worden er alleen subsets weergeven die voorkomen in het onderwerp (= de eerste variabele in *onderdelen*). Als een school alleen havo-klassen heeft, wordt er bij de subset opleidingstype alleen met havo vergeleken (wel van de hele dataset). Hierdoor hoeft geen aparte dataset te worden gemaakt. |
| alleen\_tbb\_met\_actuele\_gegevens | WAAR / ONWAAR | Als dit op waar staat worden er geen tabellenboeken gemaakt voor onderwerpen waar recente data voor ontbreekt. Een school zonder data van 2021 wordt dus overgeslagen. |
| **missing\_weegfactoren** | “fout” / “verwijderen” / getal | Bepaalt wat er gebeurt als er missings zijn op de weegfactoren. “fout” stopt het script, “verwijderen” verwijdert observaties met missing weegfactoren. Een nummer vervangt alle missings met dat nummer. |
| weergave\_type | “n” / “percentage” / leeg | Het tabellenboek wordt standaard geëxporteerd met percentages. Het is (**voor intern gebruik!**) ook mogelijk om de onderliggende aantallen te exporteren, door *weergave\_type* op “n” te zetten. Dit verandert alleen de weergave in Excel, de onderliggende berekeningen en confidence intervallen worden nog steeds berekend op percentages. |

## Crossings

Een lijst met variabelen, zoals aangegeven in SPSS, waarop de dataset gesplitst moet worden. In de praktijk zal dit vaak geslacht, SES, of opleidingsniveau zijn. Meerdere crossings zijn mogelijk.

## Datasets

|  |  |  |  |
| --- | --- | --- | --- |
| Variabele | Waarde | Verplicht | Betekenis |
| **naam\_dataset** | Tekst | Ja | De (zelf gekozen) naam van de dataset. Deze wordt gebruikt in andere delen van de configuratie en weergegeven als kolomnaam in het tabellenboek. Deze hoeft niet overeen te komen met de bestandsnaam. |
| **bestandsnaam** | Tekst | Ja | Bestandsnaam van de dataset. Deze **moet** zich bevinden in de map *data*, dus één map dieper dan het configuratiebestand. |
| jaarvariabele | Variabelenaam / leeg | Nee | Indien er meerdere jaren in hetzelfde bestand aanwezig zijn, kan hier de naam van de gebruikte variabele ingevuld worden. |
| jaren\_voor\_analyse | Getal / leeg | Nee | Indien jaarvariabele is ingevuld kan hier gekozen worden voor een specifieke subset van de aanwezige jaren. Meerdere jaren kunnen worden gebruikt, gescheiden door komma’s. Alle jaren worden gebruikt als deze niet is ingevuld. |
| **waarden\_dichotoom** | Getal / leeg | Nee | Indien er bij dichotome variabelen een ander coderingsschema gebruikt wordt dan 0 (nee) en 1 (ja), kan dit hier worden aangegeven, gescheiden door komma’s. Meerdere opties zijn mogelijk, deze dienen te worden gescheiden door |. Bij de JV monitor in 2022 zijn er bijvoorbeeld variabelen die dichotoom zijn met waarden 0/1/8, maar ook met 1/2/8. Dit kan dan worden aangegeven met “0,1,8|1,2,8” Hierbij is het van belang dat waarde 1 altijd gebruikt wordt als waarde voor ja. De andere waardes worden aangenomen als nee. Dit kan bijvoorbeeld relevant zijn als er ook een waarde is voor ‘niet ingevuld’. |

## Dichotoom / niet dichotoom

Dichotome variabelen worden in theorie automatisch herkend door het script, en kunnen nog verder geborgd worden middels *waarden\_dichotoom* op het tabblad *algemeen*. Mocht er alsnog een variabele zijn die wel dichotoom is, maar niet zo gelezen wordt door het script, of vice versa, dan kunnen deze worden geplaatst in het bijbehorende tabblad. Hiermee wordt de detectie omzeilt en de status als wel/niet dichotoom geforceerd. Beide tabbladen bestaan simpelweg uit een lijst met variabelenamen.

## Headers afkortingen

Het kan voorkomen dat de volledige waarde van een categorie te groot is voor een cel. Voor dit geval is er een mogelijkheid om afkortingen op te geven, middels een simpele tekst/vervanging combinatie. In NOG wordt de regio bijvoorbeeld onderverdeeld in subregio’s, die te lange namen hebben voor het tabellenboek:

|  |  |
| --- | --- |
| tekst | vervanging |
| Achterhoek | AH |
| Noord-Veluwe | N-V |
| Midden-IJssel Oost-Veluwe | MIJ OV |

## Indeling kolommen

Er kunnen meerdere tabbladen gemaakt worden in het tabellenboek, met ieder meerdere kolommen. De kolom wordt gevuld met data gespecificeerd in de variabele *subset*, of de volledige dataset. Let op: de gebruikte data dient eveneens te worden gespecificeerd op het tabblad *onderdelen*. (Zie ook de inleidende beschrijving bovenaan dit document.) Verder *moet*

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| Tabblad | Naam van het tabblad | Spreekt voor zich. |
| dataset | naam\_dataset uit het tabblad *datasets* | De dataset waaruit data moet worden gehaald. |
| subset | “zelf” / variabelenaam / leeg | “zelf”: data van het onderwerp (de eerste rij in *onderdelen*) van het tabellenboek, oftewel het niveau waarop in eerste instantie uitgesplitst wordt.  Variabelenaam: data gesplitst op niveau van de subsetvariabele (dit moet overeenkomen met de naam in SPSS)  Leeg: data van de gehele dataset |
| geef\_crossing | WAAR / ONWAAR | Bepaalt of de variabelen in *crossings* moeten worden weergegeven binnen deze subset. Indien ONWAAR worden alleen totalen weergegeven. |
| jaar | Jaartal / leeg | Indien de dataset per jaar is gesplitst **moet** hier het gewenste jaar ingevuld worden. Indien er geen jaarvariabele is gebruikt **moet** deze leeg zijn. |
| sign\_toetsen\_totaal | “grote\_set” of leeg | Significantie wordt bepaald door controle van de betrouwbaarheidsinterval. (Let ook op de waarde hiervan op tabblad *algemeen*.) Hierbij wordt de data uit een kolom vergeleken met de ‘grote set’, die middels deze variabele wordt aangegeven. Als hierin geen overlap is, is de waarde ‘significant’. (Let op: dit is statistisch gezien discutabel, dus beoordeel zelf of dit een gewenste manier van vergelijken is.) In de praktijk is de totale dataset vaak de ‘grote set’ en worden andere waardes hiermee vergeleken. Er kan slechts één ‘grote set’ per tabblad zijn. Als in deze kolom “grote\_set” staat wordt die dataset gebruikt als referentie. |
| sign\_toetsen\_crossings | “intern” of leeg | Kruisvariabelen kunnen onderling vergeleken worden a.d.h.v. betrouwbaarheidsintervallen, gelijk aan de manier waarop de totaalkolommen vergeleken worden. Vul hier “intern” in indien deze met elkaar moeten worden vergeleken. In alle andere gevallen worden de waardes vergeleken met de kolom *grote\_set*, zoals de totaalkolommen. |

## Indeling rijen

De volgorde van variabelen in het tabellenboek kan op dit tabblad opgemaakt worden. Ieder element staat op een nieuwe rij, en wordt automatisch opgemaakt door het script. Dichotome variabelen krijgen **niet** automatisch een kop met kolommen; deze dient zelf geplaatst te worden middels de waarde “kop”. In het voorbeeld wordt de kop Financiën toegevoegd:

Afbeelding met tekst

Automatisch gegenereerde beschrijving

|  |  |
| --- | --- |
| kop | Financiën |
| var | FIIKA403 |
| var | FIIKA401 |
| var | FIIKA402 |

De kolomkoppen (Man/Vrouw) worden aangemaakt door de kop, niet door de variabelen. Het al dan niet zwaar vinden van het geven van mantelzorg (ZHMZGA406) is geen dichotome variabele, en krijgt daardoor zelf ook kolomkoppen:

|  |  |
| --- | --- |
| var | ZHMZGA406 |

Afbeelding met tekst

Automatisch gegenereerde beschrijving

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| type | “aantal” | Overzicht met de aantallen in de dataset per kolom. Accepteert geen tweede argument. (Tweede argument = kolom *inhoud*.) |
|  | “titel” | Een titel, vaak gebruikt als afscheiding tussen onderwerpen. |
|  | “tekst” | Tekst kan ingevoerd worden tussen variabelen voor meer uitleg. Een lege tekstregel kan gebruikt worden om een extra witregel in te voegen. Let op: introtekst bevindt zich op een aparte pagina: *tabblad\_intro\_tekst*. |
|  | “kop” | Een kop om een set dichotome variabelen samen te voegen. Zie bovenstaand voorbeeld. |
|  | “var” / “nvar” | Geeft de genoemde variabele weer. Bij “nvar” wordt een extra regel met n = … toegevoegd. |
| inhoud | Variabelenaam / tekst / leeg | In dit veld kan de bijbehorende tekst danwel variabele ingevuld worden. |

## Labelcorrectie

De labels in de dataset zijn soms onduidelijk of te lang voor een tabellenboek. Dit kan in principe gecorrigeerd worden in SPSS, aangezien het script de labels uit SPSS haalt, maar in veel gevallen zal de dataset niet voor niets zo opgezet zijn. Daarom is er in het script de mogelijkheid om labels aan te passen. Zowel variabelelabels als antwoordlabels kunnen worden aangepast. Hierbij is het van belang om te kijken welk type het is, aangezien het script hier zelf niet slim genoeg voor is. Bij een dichotome waarde wordt de **vraag** (= variabelelabel) als label weergegeven, bij een niet-dichotome waarde wordt het **antwoord** (= antwoordlabel) als label weergegeven.

|  |  |  |  |
| --- | --- | --- | --- |
| Variabele | Waarde | Verplicht | Betekenis |
| var | Variabelenaam / leeg | Alleen indien *var\_label* is ingevuld, optioneel voor antwoorden | Geeft de variabele aan waar de verandering moet plaatsvinden. |
| var\_label | Nieuw label / leeg | Nee | De vervangende waarde voor deze variabele. |
| antwoord\_oud | Oud antwoord / leeg | Alleen indien *antwoord\_nieuw* is ingevuld | Zoekwaarde om te vervangen. Deze dient precies identiek te zijn aan de waarde in SPSS. (Dit kan ook gekopieerd worden uit het tabellenboek.)  **Let op:** Als *var* niet is opgegeven in deze regel wordt het antwoord in alle variabelen vervangen. De waarde in *var* geldt daardoor als een filter. |
| antwoord\_nieuw | Nieuw antwoord / leeg | Alleen indien *antwoord\_oud* is ingevuld | Het vervangende antwoordlabel. |

## Logo’s

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| **tabblad** | Naam | Tabblad waarop het logo geplaatst moet worden. Deze naam moet overeenkomen met de naam van het tabblad zoals opgegeven in *indeling\_kolommen*. |
| **bestand** | Bestandsnaam | Indien er een logo toegevoegd moet worden aan het tabellenboek, kan hier de bestandsnaam worden gespecificeerd. Let erop dat het pad relatief vanaf het configuratiebestand moet zijn. (Dus stel dat de configuratie in Mijn documenten staat en het logo in Afbeeldingen, dan is het pad “../Afbeeldingen/plaatje.png”.) Voor het gemak is het aan te raden om het bestand in dezelfde map te zetten, dat voorkomt gedoe. |
| **rij** | Getal | De rij in het tabellenboek waarin de linkerbovenhoek van het logo zich moet bevinden. Om deze onderaan uit te lijnen mag een negatieve waarde worden ingevoerd. |
| **kolom** | Getal | Aantal kolommen vanaf rechts dat het logo moet worden ingevoegd in het tabellenboek. Om deze rechts uit te lijnen mag een negatieve waarde worden ingevoerd. (Bijvoorbeeld: -3 betekent 3 kolommen van rechts.) |
| **breedte** | Getal | Gewenste breedte van het logo in pixels. (Dit kan worden afgelezen door in Windows Verkenner het plaatje te selecteren, rechter muisknop > Eigenschappen > Tabblad *Details*.) |
| **hoogte** | Getal | Gewenste hoogte van het logo in pixels. |

## Onderdelen

Hierin wordt de opbouw van de dataset uitgewerkt. Op basis hiervan wordt de analyse uitgevoerd. Dit lijkt enigszins op *indeling\_kolommen*, maar loopt volgens een iets andere logica. Het huidige tabblad zorgt voor de analyse, *indeling\_kolommen* zorgt voor de weergave. Iedere nieuwe rij beschrijft een ‘set’ kruistabellen die gemaakt zullen worden. De eerste rij is leidend; hierop worden de tabellenboeken gesplitst. In de praktijk zal dit bijvoorbeeld gemeentecode of school zijn.

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| Dataset | naam\_dataset uit het tabblad *datasets* | De dataset waaruit data moet worden gehaald. |
| **subsetvariabele** | Variabelenaam / leeg | Variabele waarop gesplitst moet worden. Dit kan bijvoorbeeld op school, regio, of gemeente zijn. Indien leeg wordt de hele dataset berekend. |
| **met\_crossing** | WAAR / ONWAAR | Legt vast of dit onderdeel verder uitgesplitst moet worden met kruisvariabelen. (Zie ook tabblad *crossings*.) |
| weegfactor | Variabelenaam / leeg | Variabelenaam met informatie over de weging van het design. Als deze leeg is wordt er een ongewogen design gebruikt. |
| stratum | Variabelenaam / leeg | Variabelenaam met informatie over de indeling van het design. Als deze leeg is wordt er een ongewogen design gebruikt. |

## Opmaak

Voor alle waarden in dit tabblad geldt dat ze optioneel zijn. Indien een waarde mist wordt een standaardwaarde aangehouden.

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| (kop/titel)\_fontsize | Getal | Grootte van koppen/titels in pt (= lettergrootte in Word en Excel). |
| (kop/titel)\_fontcolor | Hexadecimale code of naam | De gewenste tekstkleur van koppen/titels. Excel accepteert hexadecimale codes (#FFFFFF) of namen (green/red). |
| (kop/titel)\_fontdecoration | “bold” / “strikeout” / “italic” / “underline” / “underline2” / “accounting” / “accounting2” | Letterstijl van de koppen/titels. Dit is vergelijkbaar met de knoppen bovenin Word en Excel. |
| (kop/titel)\_fill | Hexadecimale code of naam | De gewenste achtergrondkleur voor koppen/titels. Excel accepteert hexadecimale codes (#FFFFFF) of namen (green/red). |
| font\_color | Hexadecimale code of naam | De gewenste tekstkleur. Excel accepteert hexadecimale codes (#FFFFFF) of namen (green/red). |
| font\_type | Naam van lettertype | Lettertype dat gebruikt wordt voor alle tekst. Dit dient een geldige naam van een lettertype te zijn. Tip: gebruik de lijst bovenaan in Word als voorbeeld. |
| tekst\_fontsize | Getal | Grootte van de tekst in pt (= lettergrootte in Word en Excel). Dit geldt voor zowel introtekst als tekst tussen en bij variabelen. |
| rij\_hoogte | Getal | De rijen in Excel worden gelijkmatig verdeeld. Dit getal specificeert de hoogte van iedere rij in pt (= lettergrootte in Word en Excel). |
| rij\_hoogte\_kop | Getal | Het kan zijn dat de kolomkoppen een andere rijhoogte nodig hebben dan de normale rijen. Daarom is de hoogte van de koppen apart in te stellen. Wederom in pt. |
| kolombreedte\_antwoorden | Getal | Breedte van de kolom met antwoordteksten in pt. |
| kolombreedte | Getal | Breedte van de kolommen in pt. |
| border\_tussen\_gegevens | WAAR / ONWAAR | Opmaak van de variabelen. Moeten er zwarte lijnen (borders) zichtbaar zijn tussen de categorieën? |
| rijen\_afwisselend\_kleuren | WAAR / ONWAAR | Rijen afwisselend grijs/wit kleuren. |
| kolommen\_afwisselend\_kleuren | WAAR / ONWAAR | Kolommen afwisselend grijs/wit kleuren. |
| kolommen\_crossings\_kleuren | WAAR / ONWAAR | Kolommen afwisselend grijs/wit kleuren **per categorie**. Dit betekent dat iedere crossing en iedere totaalkolom een afwisselende kleur krijgt, waardoor het verschil goed zichtbaar is. |
| header\_stijl | “enkel” of “dubbel” | Om ruimte te besparen is het mogelijk om de kolomkoppen samen te voegen. Hierdoor wordt per kop slechts één rij gebruikt, i.p.v. twee. (Zie voorbeeld hieronder.) |
| crossing\_headers\_kleiner | WAAR / ONWAAR | Om onderscheid te maken tussen de kruisvariabelen en totaalkolommen kan ervoor gekozen worden om de kruisvariabelen een kleiner label te geven. Indien WAAR worden alleen de labels van totaalkolommen vet, bij ONWAAR worden alle labels vet. |
| label\_max\_lengte | Getal | Sommige variabelelabels zijn te lang om binnen de kolombreedte te passen. Excel kan automatisch tekstterugloop aanzetten, maar past niet automatisch de rijhoogte aan. Daarom dient handmatig te worden aangegeven na hoeveel karakters de rijhoogte voor een variabele groter moet worden. In de praktijk zal dit getal ongeveer overeenkomen met *kolombreedte\_antwoorden*. |

**Headerstijlen:**

*Enkel*



*Dubbel*

*Afbeelding met tekst

Automatisch gegenereerde beschrijving*

## Selectie hoofdvariabele

Als er meer gemeentes (of waar dan ook op geselecteerd wordt) in het databestand staan dan daadwerkelijk in de dataset aanwezig zijn, **kan** hier een lijst met de benodigde gemeentes worden geplaatst. (Als dit tabblad leeg is wordt de hele dataset gebruikt.) Deze dienen in een enkele kolom, genaamd *hoofdvariabele*, ingevoerd te worden met hun gemeente-/identificatiecode. Gewone namen zijn niet toegestaan. (Dit had ik graag toegevoegd, maar werd technisch ingewikkeld.) Indien er tevens gebruik gemaakt wordt van een subregio **kan** hier een tweede kolom met de bijbehorende waardes voor die variabele worden toegevoegd. Deze kolom moet de naam van de variabele in SPSS dragen. Ter voorbeeld, in de jeugdmonitor 2021/2019 werd een subregio toegevoegd aan de tabellenboeken. Dit is een deel van de configuratie:

|  |  |
| --- | --- |
| hoofdvariabele | subregio |
| 197 | 3 |
| 200 | 1 |
| 213 | 1 |

Hierbij droeg de variabele in SPSS de naam *subregio* en hoorde gemeente 197 in subregio 3 en gemeentes 200 en 213 in subregio 1.

## Tabblad intro tekst

Introductietekst kan per tabblad worden aangegeven. Dit is apart gezet van de algemene indeling in rijen, aangezien de introductietekst per tabblad kan verschillen, terwijl de rijen op alle tabbladen worden doorgevoerd.

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| **tabblad** | Naam van het tabblad | De naam van het tabblad waarop de introtekst dient te staan. Deze naam dient overeen te komen met de waarde in *indeling\_kolommen*. |
| **type** | “titel” / “kop” / “tekst” | Indeling van de tekst opgegeven in *inhoud*. Hierbij verschillen “titel” en “kop” in grootte van de normale tekst. |
| **inhoud** | Tekst / leeg | De weer te geven tekst. |