Handleiding 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. Labels kunnen in het tabellenboek anders worden weergegeven indien gewenst, maar dit is een handmatig en foutgevoelig proces. (Zie ook de naslag van het tabblad *labelcorrectie* hieronder.)

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. Hierdoor kan iedereen zien welke problemen al bekend zijn en kunnen mensen met dezelfde problemen makkelijk gebundeld worden.

# 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 verwerken 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 zaken 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.

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. Door gebruik van tbl\_onderzoeksexport.R worden hiervan leesbare tabellen gemaakt die eveneens in de map *output* zullen verschijnen.

# Stappenplan

Het uitvoeren van het script bestaat globaal uit 3 stappen.

## Stap 1 – configuratie en plaatsing

* Verplaats de benodigde bestanden naar een logische map. Dit kan bijvoorbeeld een NAS of gedeelde schijf zijn, of een projectmap.
* Vul de configuratie in (zie naslag hieronder). De bijgevoegde configuratie is puur ter voorbeeld, en geenszins volledig.

## Stap 2 – uitvoeren script

* Open R of Rstudio en voer tbl\_maken.R uit. (Andere bestanden worden automatisch geladen.)
* 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**.
* Het kan voorkomen dat er foutmeldingen optreden. Deze worden zoveel mogelijk vóór de berekening getoond, zodat er geen kostbare tijd verloren gaat. Daarnaast worden meldingen automatisch opgeslagen in een logbestand: log.txt.

## Stap 3 – controle en herziening

* In de map waarin het configuratiebestand zich bevindt is nu een map aangemaakt, genaamd “output”. Hierin staan de tabellenboeken.
* Waarschijnlijk moeten er nog zaken worden aangepast. Denk aan kopjes, kleuren, lettertype, enz. Deze kunnen in het configuratiebestand worden aangepast.
* 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.
* De resultaten kunnen ook gebruikt worden voor verdere analyse of het maken van een rapportage. Voer tbl\_onderzoeksexport.R uit om hier een leesbare tabel van te maken.

# Configuratie

Hieronder worden de verschillende tabbladen in het configuratiebestand toegelicht, op alfabetische volgorde. Variabelen die waarschijnlijk aangepast moeten worden zijn vetgedrukt, met uitzondering van de tabbladen *onderdelen,* *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 tussen totaalkolommen of onderlinge vergelijkingen binnen kruisvariabelen mogelijk te maken. Deze worden per antwoord berekend middels een chi2-test. Indien er meer dan twee categorieën zijn in een kruisvariabele worden deze als geheel in de chi2-test gestopt, waardoor bij een significant verschil alle kolommen binnen die kruisvariabele als significant worden weergegeven. Een post-hoc test om te bepalen welke groep afwijkt is niet ingebouwd. **Let op:** De chi2-test kan alleen omgaan met één lege groep. Ter voorbeeld: binnen de monitor VO was er een vraag die alleen aan 65+’ers werd gesteld. Als de leeftijd dan gesplitst is op 65+ en 65- kan er een chi2-test worden uitgevoerd; deze is vanzelfsprekend significant. Als er gesplitst wordt op 18-44, 44-65, en 65+, dan zijn er twee volledige groepen leeg, en zal de functie vastlopen. In zo’n geval zal het script een waarschuwing weergeven en wordt de chi2-test overgeslagen. Hierdoor zullen de cellen **niet** aangegeven worden als significant, ook als er tussen de wel aanwezige groepen wel verschillen bestaan.

Het vergelijken gebeurt op basis van de kolomselectie. Vanuit de configuratie wordt de onderliggende data voor iedere kolom bepaald, bestaande uit een dataset en optioneel een jaar of subset. Vanuit deze data worden dummies aangemaakt, welke aangeven of een rij bij een bepaalde kolom hoort. Wanneer een test wordt uitgevoerd wordt data geselecteerd op WAAR in beide bijbehorende dummies, en vervolgens getest op WAAR voor de eerste dummy versus ONWAAR voor de eerste dummy. Dit klinkt ingewikkelder dan het is. Stel bijvoorbeeld dat we een splitsing hebben op gemeente (dus subset = Gemeentecode) en het totaal van deze gemeente willen vergelijken met het regionale totaal, zie onderstaande tabel. Data in de gemeentekolom (kolom a) bevindt zich per definitie ook in kolom b; gemeentedata behoort immers ook in de gehele regio, en dit bevindt zich in dezelfde dataset. De chi2-test wordt uitgevoerd op alle rijen die in één van deze kolommen voorkomen en gesplitst op “is onderdeel van kolom a”. In dit geval wordt de gemeente daarom vergeleken met alle regionale data **behalve** die van de gemeente. Deze data zit immers **wel** in kolom a, dus in de vraag “bevindt de data zich wel of niet in kolom a” is het antwoord ja.

|  |  |  |
| --- | --- | --- |
|  | Kolom a | Kolom b |
| Dataset | Regionaal | Regionaal |
| Jaar | 2022 | 2022 |
| Subset | Gemeentecode |  |

Vergelijk dit met de situatie waarin een totaalkolom van een gemeente uit een regionale dataset wordt vergeleken met het totaal van de landelijke dataset, zie onderstaande tabel. Data uit kolom c kan voorkomen in kolom d, maar dit is op basis van een universeel script niet te herleiden. Daarom zal een test op deze vergelijking mogelijk ook data bevatten van de gemeente in kolom c. De vergelijking vindt daarom plaats tussen data van de gemeente in de regionale dataset versus **alle** data in de landelijke dataset. De gemeentelijke data is in de frequentietabel dus mogelijk aanwezig aan beide kanten, aangezien hier niet op gesplitst kan worden.

|  |  |  |
| --- | --- | --- |
|  | Kolom c | Kolom d |
| Dataset | Regionaal | Landelijk |
| Jaar | 2022 | 2022 |
| Subset | Gemeentecode |  |

Om te vergelijken tussen kolommen met wederzijdse uitsluiting is het noodzakelijk om vanuit één dataset te werken, waarin de splitsing op basis van één variabele (optioneel binnen een specifiek jaar) gemaakt kan worden. Dit is per definitie het geval bij het vergelijken binnen kruisvariabelen.

**Weegfactoren** (relevant voor V&O)

Het script is ontworpen om met één weegfactor per dataset te werken (gelijk aan SPSS Complex Samples). Desondanks kan het voorkomen dat er verschillende weegfactoren per variabele of zelfs per variabele per jaar nodig zijn. Deze mogelijkheid is toegevoegd, maar op eigen risico: de controles die uitgevoerd worden over de weegfactoren vinden voor de analyse plaats. Wisselen midden in de analyse kan daarom fouten opleveren. Deze dienen opgelost te worden in de dataset.

Het aangeven van een afwijkende weegfactor kan op het tabblad *indeling\_rijen­,* door het toevoegen van een kolom met de naam “weegfactor” (voor een andere weegfactor per variabele) of een kolom met de naam “weegfactor.d\_(datasetnaam)” (voor een weegfactor per variabele per jaar). Het is enkel mogelijk een andere weegfactor **per dataset** of per gehele variabele aan te geven. Als er voor jaar x een andere weegfactor geldt dan voor jaar y dienen deze jaren dus ingevoerd te worden als aparte datasets. Hiervoor kunnen dan twee kolommen worden toegevoegd: “weegfactor.d\_datasetx” en “weegfactor.d\_datasety”.

## Algemeen

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| 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. 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 desgewenst worden aangepast in *tbl\_MakeExcel.R*, zoek op “style.sign”.) |
| confidence\_level | Getal | Gewenste p-waarde voor significantie. (Zie ook *indeling\_kolommen*.) |
| **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. |
| multithreading | WAAR / ONWAAR | Om de rekentijd korter te maken kan er gebruik worden gemaakt van parallelle berekeningen. Hierbij worden meerdere processorkernen gebruikt om de berekeningen uit te voeren, waardoor de rekenkracht beter benut wordt. Het kan echter zijn dat deze mogelijkheid door de afdeling IT is begrensd. Verder moet R voor deze mogelijkheid gedeblokkeerd worden in de firewall. Deze melding verschijnt automatisch, maar accepteren vereist meestal administratorprivileges. |
| **waarden\_dichotoom** | Getal / leeg | 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’. |
| vergelijk\_variabelelabels | WAAR / ONWAAR | Het script controleert over variabelen overeenkomen aan de hand van de labels in de dataset. Het kan bij het gebruik van oudere sets voorkomen dat deze niet overal gelijk zijn. Daarom is het mogelijk om deze controle uit te schakelen. **Let op:** het is aan te raden het script eerst te draaien met deze instelling op WAAR, zodat het inzichtelijk wordt waar mogelijke problemen voorkomen. Deze dienen handmatig gecontroleerd te worden. |

## Crossings

Een lijst met variabelen, zoals aangegeven in SPSS, waarop de dataset gesplitst moet worden. In de praktijk zal dit bijvoorbeeld 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. Het script neemt aan dat dit bestand in dezelfde map staat als de configuratie. Een relatief pad (bijv. “../2021/data.sav”) is mogelijk, maar niet aan te raden vanwege de onleesbaarheid. N.B.: Wachtwoordbeveiliging van de databestanden is **niet** beschikbaar binnen R. Voor verwerking binnen dit script moet een mogelijk wachtwoord daarom eerst verwijderd worden. |
| jaarvariabele | Variabelenaam / leeg | Nee | Indien er meerdere jaren in hetzelfde bestand aanwezig zijn, kan hier de naam van de gebruikte variabele ingevuld worden. |
| weegfactor | Variabelenaam / leeg | Nee | Variabelenaam met informatie over de weging van het design. Als deze leeg is wordt er een ongewogen design gebruikt. **Let op:** als er per variabele een andere weegfactor wordt gebruikt **moet** deze leeg zijn. |
| stratum | Variabelenaam / leeg | Nee | Variabelenaam met informatie over de indeling van het design. Als deze leeg is wordt er een ongewogen design gebruikt. |

## 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 omzeild en de status als wel/niet dichotoom geforceerd. Beide tabbladen bestaan simpelweg uit een lijst met variabelenamen.

## Forceer datatypen

De datatypen worden overgenomen uit het databestand van SPSS. Het kan voorkomen dat hier om moverende redenen een ‘onhandig’ datatype is gekozen. Zo is er bijvoorbeeld bij de VO monitor 2022 bij Gemeentecode gekozen om hier een string van te maken i.v.m. de combinatie met CBS-data. Dit is anders dan bij eerdere monitors, waardoor een combinatie tussen deze datasets niet meer mogelijk is. Om dit mogelijk te maken kan het datatype geforceerd worden. Het script zal dan de gewenste variabele omzetten naar het gewenste datatype.

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| variabele | Variabelenaam | Spreekt voor zich |
| type | “numeric” / “character” | Gewenst datatype. Er zijn momenteel twee opties ingebouwd: getallen (“numeric”) en strings (“character”). De variabele wordt binnen R omgezet naar het gewenste type. **Let op:** als er ongeldige waarden bestaan worden deze verwijderd. Bij een conversie naar getallen kunnen bijvoorbeeld geen letters bestaan. |

## 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 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. Van iedere variabele wordt uit elke opgegeven dataset data gezocht. Indien deze afwezig is wordt er een vervangende tekst weergegeven. (Zie *algemeen*.) 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 kolomkoppen:

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

Afbeelding met tekst

Automatisch gegenereerde beschrijving

Er kunnen afwijkende weegfactoren per variabele worden aangegeven door het toevoegen van een kolom met de naam “weegfactor” (voor de hele variabele) of “weegfactor.d\_(naam)” (voor een splitsing per dataset). (Specifiek voor VO: als er gebruik wordt gemaakt van een combinatiebestand met meerdere jaren kan er dus gebruik gemaakt worden van een kolom met de naam “weegfactor”, als er gebruik wordt gemaakt van een bestand per jaar dient dit per dataset te worden aangegeven.) Als deze kolommen ontbreken wordt de weegfactor van het tabblad *datasets* gebruikt.

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| type | “aantallen” | 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: *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. |
| weegfactor | Variabelenaam / leeg | Indien er per variabele een andere weegfactor gebruikt moet worden kan dit hier ingevuld worden. Standaard wordt de weegfactor op het tabblad *datasets* gebruikt. **Let op:** controles op weegfactoren worden niet uitgevoerd . **Let op:** indien deze functie gebruikt wordt **moet** de weegfactor op het tabblad *datasets* leeg zijn. |
| weegfactor.d(getal) | Variabelenaam / leeg | Indien deze kolom aanwezig is wordt de hierin aangegeven variabele gebruikt als weegfactor. Hiermee kan een splitsing worden gemaakt tussen bijvoorbeeld verschillende jaren. (getal) duidt op het volggetal van de dataset in *datasets*, waarbij 1 de eerste dataset is, 2 de tweede, enz. |
| weegfactor.d\_(naam) | Variabelenaam / leeg | Identiek aan de kolom hierboven, maar dan op basis van de naam van een dataset i.p.v. het volggetal. N.B.: Het script differentieert tussen deze twee mogelijkheden door het zoeken naar een underscore (\_). Deze is daarom verplicht. |

## Introtekst

Introductietekst staat op een apart tabblad voor de overzichtelijkheid en om de opmaak makkelijker te maken in het script.

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| **type** | “titel” / “kop” / “tekst” / leeg | Indeling van de tekst opgegeven in *inhoud*. Hierbij verschillen “titel” en “kop” in grootte van de normale tekst. “tekst” en een lege cel zijn equivalent. |
| **inhoud** | Tekst / leeg | De weer te geven tekst. Een witregel kan gemaakt worden door deze leeg te laten. “[naam]” wordt vervangen door de naam van het tabellenboek. (Bijvoorbeeld de gemeentenaam.) |

## 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. Dit is zeer foutgevoelig, dus een dubbel- of triplecheck is aan te raden. 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* of *antwoord\_waarde* is ingevuld, optioneel voor antwoorden met zoektekst. | Geeft de variabele aan waar de verandering moet plaatsvinden. |
| var\_label | Nieuw label / leeg | Nee | De vervangende waarde voor deze variabele. |
| antwoord\_waarde | Antwoordoptie / leeg | Alleen indien *antwoord\_oud* leeg is. | De waarde van het te vervangen antwoordlabel. Dit is de numerieke waarde die in SPSS staat. |
| antwoord\_oud | Oud antwoord / leeg | Alleen indien *antwoord\_nieuw* is ingevuld en *antwoord\_waarde* **niet** 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 | Nee | Het vervangende antwoordlabel. |

## Logo’s

|  |  |  |
| --- | --- | --- |
| Variabele | Waarde | Betekenis |
| **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. (Bijvoorbeeld: -2 betekent 2 rijen van onderen.) |
| **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. Iedere nieuwe rij beschrijft een ‘set’ kruistabellen die gemaakt zullen worden. Zo kunnen meerdere datasets naast elkaar worden weergegeven, bijvoorbeeld een regionale en een landelijke dataset, of een huidige en een voorgaande monitor, door op een volgende rij een bijpassend ‘onderdeel’ toe te voegen. De eerste rij met een subset 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. |
| subset | Variabelenaam / leeg | Variabele waarop gesplitst moet worden. Dit kan bijvoorbeeld op school, regio, of gemeente zijn. Hier kan tevens een tweede niveau worden aangegeven, welke op basis van de eerste (bovenste) subset opgezocht zal worden. Als de eerste regel bijvoorbeeld *Gemeentecode* als subset heeft, en de tweede regel *subregio*, dan wordt de subregio waarin de gemeente zich bevind meegenomen. Indien leeg wordt de hele dataset berekend. |
| jaar | Jaartal / leeg | Indien de dataset per jaar is gesplitst kan hier het gewenste jaar ingevuld worden. |
| met\_crossing | WAAR / ONWAAR | Legt vast of dit onderdeel verder uitgesplitst moet worden met kruisvariabelen. (Zie ook tabblad *crossings*.) |
| sign\_crossing | “intern” / leeg | Indien onderlinge verschillen getest moeten worden kan hier worden aangegeven in welke data een chi2-test moet worden uitgevoerd. **Let op:** als *sign\_toetsen* op het tabblad *algemeen* niet op WAAR staat wordt de test alsnog niet uitgevoerd. |
| sign\_totaal | Naam van een dataset / nummer van een kolom / leeg | Indien totalen getest moeten worden kan hier worden aangegeven met welke dataset/kolom vergeleken dient te worden. Hier kan de naam van een dataset worden ingevuld, of een kolomnummer. **Let op:** indien een naam wordt opgegeven neemt het script de **laatste** kolom van die dataset. Dit kan relevant zijn als een dataset bijvoorbeeld met en zonder subset wordt weergegeven. Het kolomnummer betreft de kolom in het tabellenboek. Kruisvariabelen dienen dus meegeteld te worden. (Bijv.: 2 datasets: regionaal en nationaal. Regionaal splitst op 2 categorieën met twee levels: totaal 2 + 2 + 1 = 5 kolommen. Nationaal zit dan in kolom 6.)  **Let op:** vanuit een kolom **zonder** subset kan niet worden vergeleken met een kolom **met** subset, maar wel andersom. Het is bijvoorbeeld wel mogelijk om vanuit de totaalkolom van een gemeente te vergelijken met het totaal van Nederland, maar niet vice versa. De reden is wat lastig uit te leggen; voor nu gewoon niet doen. (Vraag het een keer op een fysieke bijeenkomst, dan doe ik een poging. - AD) Van een kolom **met** subset naar een kolom **met** subset is ook geen probleem. |

## 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)\_size | Getal | Grootte van koppen/titels in pt (= lettergrootte in Word en Excel). |
| (kop/titel)\_color | Hexadecimale code of naam | De gewenste tekstkleur van koppen/titels. Excel accepteert hexadecimale codes (#FFFFFF) of namen (green/red). |
| (kop/titel)\_decoration | “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. |
| font\_size | 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. |
| 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*. |
| 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.) |
| header\_template | Tekst | Titel die weergegeven moet worden in de kop van de totaalkolom. “[naam]” en “[jaar]” worden vervangen door de naam van de dataset en het onderzochte jaar, respectievelijk. |
| 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. |

**Headerstijlen:**

*Enkel*



*Dubbel*

*Afbeelding met tekst

Automatisch gegenereerde beschrijving*