

Opdracht 5a

-----  
Kruistabellen

Aan elk van 36 studenten werd gevraagd of zij alcohol drinken, en zo ja, welke soort alcoholische drank de voorkeur heeft. Tevens werd voor elke student de leeftijd en het geslacht geregistreerd. Op die manier werden de volgende gegevens verkregen:

leeftijd geslacht alcohol drank

19	1	1	3
26	1	1	3
23	1	1	3
19	2	2	3
24	2	1	1
29	2	1	1
26	1	1	3
23	1	1	3
19	1	2	1
24	2	1	1
22	1	1	3
23	1	1	3
22	2	1	3
23	2	1	1
24	1	1	2
36	2	1	1
26	1	2	
27	2	1	1
23	2	1	1
24	1	1	2
36	1	1	3
26	1	2	
23	1	1	3
24	1	1	2
18	1	1	3
21	2	1	1
20	2	1	3
31	2	1	1
26	1	1	2
32	1	1	3
21	2	1	1
20	2	1	3
31	2	1	1
26	1	1	2
24	2	1	1
20	2	1	3

Hierbij hebben de codes de volgende betekenis:

Geslacht: 1 = man  
          2 = vrouw

Alcohol : 1 = drinkt alcohol  
          2 = drinkt geen alcohol

Drank : 1 = wijn  
        2 = gedestilleerd  
        3 = bier

- a. Voor iemand die geen alcohol drinkt kan geen alcoholvoorkeur ingevuld worden. Als alcoholvoorkeur moet 'missing value' opgegeven worden. Wijzig de tabel zodanig dat dat inderdaad het geval is (niet handmatig !).
- b. Deel de leeftijd in in groepen: <-20, 21-25, 26-30, 31-> . Maak vervolgens een kruistabel van leeftijdsgroep en alcoholvoorkeur op basis van alleen die studenten die alcohol drinken. Zorg dat elke kolom correspondeert met een leeftijdsgroep en elke rij correspondeert met een alcoholvoorkeur. De cellen moeten de aantallen, de rijpercentages, de kolompercentages en de totaalpercentages weergeven.
- c. Maak een staafdiagram, waarbij voor elke leeftijdsgroep een staaf aanwezig is die het aantal studenten aangeeft dat binnen die leeftijdsgroep de voorkeur heeft voor bier.
- d. Maak een gesegmenteerd staafdiagram met voor elke leeftijdsgroep een staaf. Elke staaf visualiseert binnen de betreffende leeftijdsgroep voor elke alcoholvoorkeur het percentage studenten.
- e. Maak een taartdiagram met voor elke alcoholvoorkeur een punt. Hierbij wordt geen onderscheid gemaakt in leeftijdsgroepen.

#### Opdracht 5a - S-PLUS

-----

Voer de gegevens in in een tabel. Definieer de vier kolommen van de tabel en kies als kolomnamen 'leeftijd', 'geslacht', 'alcohol' en 'drank'.

- a. De bovenste toolbar is de Standard Toolbar. Klik in deze toolbar op het icoontje dat precies onder Help van de menubalk zit. Als je de cursor op dit icoontje plaatst, moet als bijschrift verschijnen: Commands Window. Na op dit icoontje te hebben geklikt verschijnt een window met de naam Commands. In dit window knippert achter de prompt (weergegeven door '>') de cursor (weergegeven door een verticaal streepje). Achter die cursor kunnen commando's ingetypt worden die uitgevoerd worden nadat je op Enter (of Return) hebt gedrukt.

Stel dat onze tabel 'data' heet. Dan moet we alleen die personen in de tabel selecteren geen alcohol drinken controleren. Dat zijn al die personen voor wie in de kolom ALCOHOL een 2 is ingevuld. De selectie zou dan worden:

```
data$ALCOHOL=='2'
```

Voor de personen uit deze selectie willen we nu dat als alcoholvoorkeur altijd 'missing value' wordt opgegeven. De code die in S-PLUS voor 'system missing' gebruikt moet worden is NA (Not Available). In de kolom DRANK moeten we voor de betreffende personen dus NA invullen. Het uiteindelijke commando dat we moeten geven wordt nu:

```
data$DRANK[data$ALCOHOL=='2']<-NA
```

Bekijk de tabel, speciaal de cases van personen die geen alcohol drinken.

- b. Kies >Data >Create Categories. Kies onder Data en achter Source Column de variabele 'leeftijd' en vul achter Target Column 'klasse' in. Kies onder Numeric Column >Cut points. Onderaan achter Cut Points geeft S-PLUS automatisch intervalgrenzen. De gegeven waarden geven steeds de eindpunten van de intervallen. De door S-PLUS voorgestelde verdeling in inter-

vallen is niet gelijk aan de verdeling die wij willen. De klassen die wij willen hebben zijn: <-20, 21-25, 26-30, 31-> . De eindpunten van de eerste drie intervallen zijn: 20, 25 en 30. Vul deze waarden, gescheiden door komma's, in achter Cut Points, ter vervanging van de door S-PLUS voorgestelde waarden. Het eindpunt van het laatste interval hoeft niet gegeven te worden. Klik op >OK.

Er is nu een nieuwe variabele 'klasse' gevormd die in de data-matrix kan worden bekeken. Voor iedere waarde in de kolom 'leeftijd' wordt in de kolom 'klasse' het interval gegeven waar die waarde invalt. Voor de leesbaarheid is het nodig de kolom 'klasse' iets breder te maken. Selecteer de kolom 'klasse' door op de kolomnaam te klikken. Ga vervolgens naar >Format en kies >Selected Object(s). Geef achter Width de waarde 20 en klik op >OK.

We moeten nu alleen die studenten selecteren die alcohol drinken. Kies >Data >Subset. Klik op >Create Subset. Selecteer onder Data en achter Column de variable 'alcohol' en klik op >Add. Klik onder Logical op ==. Klik onder Column Value op >Show Values. Selecteer de waarde 1, en klik op >Add. Beneden in het venster moet achter Expression deze expressie nu zijn vermeld:

ALCOHOL ==1

Klik op >OK. En klik nog eens op >OK. S-PLUS geeft nu een nieuwe tabel met daarin alleen de studenten die alcohol drinken. We gaan nu verder werken op basis van deze nieuwe tabel. Kies >Statistics >Data Summaries >Crosstabulations. Selecteer onder Data en achter Variables eerst de variabele 'klasse' door er een keer met de linker muisknop op te klikken. Selecteer vervolgens de variabele 'drank' door de Ctrl-toets ingedrukt te houden en met de linker muisknop een keer op de naam van die variabele te klikken. Klik nu bovenaan het venster op >Options. Zet de optie >Run Chi-Square Test uit. Vul achter Decimal Places een 3 in. Klik op >OK. In het Report-venster wordt de kruistabel nu gegeven.

In de cellen staan op de eerste regel de aantallen van de combinatie, op de tweede regel de rijproporties, op de derde regel de kolomproporties en op de vierde regel de proporties van de combinaties ten opzichte van het totale aantal in de tabel. Proporties worden omgerekend naar percentages door ze te vermenigvuldigen met 100.

- c. We moeten nu alleen die studenten selecteren die bier drinken. Kies >Data >Subset. Klik op >Create Subset. Selecteer onder Data en achter Column de variable 'drank' en klik op >Add. Klik onder Logical op ==. Klik onder Column Value op >Show Values. Selecteer de waarde 3, en klik op >Add. Beneden in het venster moet achter Expression deze expressie nu zijn vermeld:

DRANK ==3

Klik op >OK. En klik nog eens op >OK. S-PLUS geeft nu een nieuwe tabel met daarin alleen de studenten die bier drinken. We gaan nu verder werken op basis van deze nieuwe tabel.

Sorteer de tabel op basis van de klassen. Kies >Data >Restructure >Sort. Selecteer achter Sort By Columns de variabele 'klasse'. Klik op >OK. Kies >Graph >2D Plot. Kies onder Axes Type voor Linear, en onder Plot Type voor Bar with Base at Zero (x,y). Klik op >OK. Selecteer onder Data Columns en achter y Columns de variabele 'klasse'. Klik op >OK.

Ga naar het grafiek-venster. Zet de punt van de muispijl op de y-as van

de grafiek. Klik nu met de rechter muisknop op de y-as, en kies >Range. Vervang onder Axis Range en achter Axis Minimum Auto door 0. Klik op >OK.

- d. Gebruik nu weer de tabel die alle personen bevat, maar waarbij voor personen die geen alcohol drinken als alcoholvoorkeur 'missing value' is gegeven.

Kies >Data >Tabulate. Selecteer onder Data en achter Variables de variabelen 'klasse' en 'drank' als volgt: klik met de linker muisknop op de variabele 'klasse'. Houd vervolgens de Ctrl-toets ingedrukt en klik met de linker muisknop op de variabele 'drank'. Klik op >OK.

In het Report-venster wordt nu de kruistabel weergegeven. We moeten nu een echte gegevenstabel maken waarin we de gegevens van de kruistabel invoeren. Kies >File >New en selecteer >Data Set. We kunnen nu de gegevens uit de kruistabel kolom voor kolom invoeren. Het is van belang om per kolom steeds eerst de gegevens in te voeren, en pas daarna via >Format >Selected Object(s) de kolomnaam te geven door achter Name de door S-PLUS voorgestelde naam te vervangen door de gewenste naam.

De eerste kolom van de kruistabel bevat de klassen. In onze nieuwe gegevenstabel vullen we als gegevens van boven naar onder achtereenvolgens in: 18-20, 21-25, 26-30, 31-36. Daarna geven we als kolomnaam 'klasse'. Voor elk van de tweede tot en met de vierde kolom vullen we steeds eerst de frequenties in zoals gegeven in de kruistabel, en daarna geven we als kolomnaam respectievelijk 'wijn', 'gedest', 'bier'.

De tabel moet er als volgt uit komen te zien:

klasse	wijn	gedest	bier
18-20	0	0	5
21-25	7	3	6
26-30	2	2	2
31-36	3	0	2

Controleer na het intypen van de tabel de kolom 'klasse'. Selecteer de kolom 'klasse' door op de kolomnaam te klikken. Kies >Format >Selected Objects. Verwijder achter Factor Levels de klassen die je niet wilt hebben. Klik op >OK.

We maken nu het gesegmenteerde staafdiagram. Kies >Graph >2D Plot. Kies onder Axes Type voor Linear, en onder Plot Type voor Bar - Stacked (x, y1..yn). Klik op >OK. Selecteer onder Data Columns en achter x Columns de variabele 'klasse', en selecteer achter y Columns de variabelen 'wijn', 'gedest' en 'bier' als volgt: klik met de linker muisknop op de variabele 'wijn'. Houd vervolgens de Ctrl-toets ingedrukt en klik met de linker muisknop achtereenvolgens op de variabelen 'gedest' en 'bier'. Klik op >OK.

De onderste toolbar is de Graph Toolbar. Klik in deze toolbar op het icoontje dat precies onder Linear in de Standard Toolbar zit. Als je de cursor op dit icoontje plaatst, moet als bijschrift verschijnen: Auto Legend. Klik op dit icoontje. De legenda wordt nu in de grafiek geplaatst.

- e. Gebruik nu weer de tabel die alle personen bevat, maar waarbij voor personen die geen alcohol drinken als alcoholvoorkeur 'missing value' is gegeven.

Selecteer de kolom 'drank' door op de kolomnaam te klikken. Kies >Data

>Change Data Type. Onder From en achter Columns moet de variabele 'Drank' geselecteerd zijn. Selecteer onder Type en achter New Type het type factor. Klik op >OK. Kies >Graph >2D Plot. Kies onder Axes Type voor Pie, en onder Plot Type voor Pie Chart (x). Klik op >OK. Kies onder Data Columns en achter x Columns de variabele 'drank'. Klik op >OK.

De onderste toolbar is de Graph Toolbar. Klik in deze toolbar op het icoontje dat precies onder Linear in de Standard Toolbar zit. Als je de cursor op dit icoontje plaatst, moet als bijschrift verschijnen: Auto Legend. Klik op dit icoontje. De legenda wordt nu in de grafiek geplaatst. Klik binnen de legenda, maar zo dicht mogelijk tegen de rand, met de rechter muisknop. Kies >Position/Size. Vervang achter # of Items Auto door 3. Klik op >OK.

Klik met de rechter muisknop op Slice 1. Kies >Text. Vervang achter >Text @Auto door Wijn. Klik met de rechter muisknop op Slice 2. Kies >Fill/Border. Zet >Override auto legend item specs aan. Kies onder Fill en achter Fill Color voor Yellow. Klik op >OK. Klik met de rechter muisknop op Slice 2. Kies >Text. Vervang achter >Text @Auto door Gedestilleerd. Klik met de rechter muisknop op Slice 2. Kies >Fill/Border. Zet >Override auto legend item specs aan. Kies onder Fill en achter Fill Color voor Lt Blue. Klik op >OK. Klik met de rechter muisknop op Slice 3. Kies >Text. Vervang achter >Text @Auto door Bier. Klik met de rechter muisknop op Slice 2. Kies >Fill/Border. Zet >Override auto legend item specs aan. Kies onder Fill en achter Fill Color voor Lt Cyan. Klik op >OK.

#### Opdracht 5a - SPSS

-----  
Voer de gegevens in in een tabel. Definieer de vier kolommen van de tabel en kies als kolomnamen 'leeftijd', 'geslacht', 'alcohol' en 'drank'.

- a. Kies >Transform >Recode >IntoSameVariables. Breng in het window 'Recode into Same Variables' de variabele 'drank' met > (pijlte naar rechts) naar Variables.

Klik nu op >If. Zet >Include if case satisfies condition aan. Breng de variabele 'alcohol' met > (pijlte naar rechts) naar rechts. Klik op de '=' en daarna op de '2'. Klik nu op >Continue.

Klik op >OldandNewValues. De variabele 'drank' kan de waarden 1, 2 en 3 hebben. Geef bij OldValue als Value 1 en bij NewValue 'System missing' en klik op >Add. Geef bij OldValue als Value 2 en bij NewValue 'System missing' en klik op >Add. Geef bij OldValue Value 3 en bij NewValue 'System missing' en klik op >Add. Klik op >Continue.

Klik op >OK. Bekijk de tabel, speciaal de cases van personen die geen alcohol drinken.

- b. Deel de leeftijd in in groepen. Kies >Transform >Recode >Into Different Variables. Breng in het window 'Recode into Different Variables' de variabele 'leeftijd' met > (pijlte naar rechts) naar Input Variable -> Output Variable. Klik op het tekstvak 'Name' bij Output Variable, vul daar 'lftkl' in en klik op >Change. Klik vervolgens op >Old and New Values. Geef nu de vier leeftijdsgroepen op. Klik vervolgens op >Continue en in het windows 'Recode into Different Variables' op >OK. Er is nu een nieuwe variabele 'lftkl' gevormd die in de data-matrix kan worden bekeken.

Selecteer alleen die studenten die wel alcohol drinken. Kies >Data >Select Cases. Je komt nu in het window 'Select Cases'. Klik op het rondje voor >If condition is satisfied. Klik vervolgens op >If. Je komt nu in het window 'Select Cases: If'. Breng 'alcohol' naar rechts, klik op '=' en klik op '1'. Klik daarna op >Continue en tenslotte op >OK.

Klik achtereenvolgens op >Statistics >Summarize en >Crosstabs. Breng 'lftkl' over naar >Row en 'drank' naar >Column(s). Klik op >Cells. Onder Counts staat >Observed standaard aan. Zet onder Percentages >Row, >Column en >Total aan. Klik op >Continue. Klik op >OK. De tabel wordt nu gemaakt.

In de cellen staan op de eerste regel de aantallen van de combinatie, op de tweede regel de rijpercentages, op de derde regel de kolompercentages en op de vierde regel de percentages van de combinaties ten opzichte van het totale aantal in de tabel.

- c. Selecteer alleen die studenten die de voorkeur hebben voor bier. Kies >Data >Select Cases. Je komt nu in het window 'Select Cases'. Klik op >Reset. Klik daarna op het rondje voor >If condition is satisfied. Klik vervolgens op >If. Je komt nu in het window 'Select Cases: If'. Breng 'drank' naar rechts, klik op '=' en klik op '3'. Klik daarna op >Continue en tenslotte op >OK.

Kies nu >Graphs >Bar. Je komt in het window 'Bar Charts'. Klik op >Simple. Klik op het rondje voor 'Summaries for Groups of Cases'. Klik op >Define. Je komt in het window 'Define Simple Bar: Summaries for Groups of Cases'. Breng 'lftkl' naar >Category Axis. Kies onder 'Bars Represent' voor >% of cases. Klik nu op >OK.

- d. Maak eerst de selectie weer ongedaan. Klik op >Data >Select Cases. In het window 'Select Cases' moet je nu klikken op het rondje voor 'All Cases'. Klik dan op >OK.

Kies nu >Graphs >Bar. Je komt in het window 'Bar Charts'. Klik op >Stacked. Klik op het rondje voor 'Summaries for Groups of Cases'. Klik op >Define. Je komt in het window 'Define Stacked Bar: Summaries for Groups of Cases'. Breng 'lftkl' naar >Category Axis en 'drank' naar >Define Stacks by. Kies onder 'Bars Represent' voor >N of cases. Klik nu op >OK.

- e. Kies >Graphs >Pie. Je komt in het window 'Pie Charts'. Klik op het rondje voor 'Summaries for Groups of Cases'. Klik op >Define. Breng in het window 'Define Pie: Summaries for Groups of Cases' 'drank' naar 'Define Slices by'. Kies onder 'Slices Represent' voor >N of cases. Klik nu op >OK.

#### Opdracht 5a - verslag

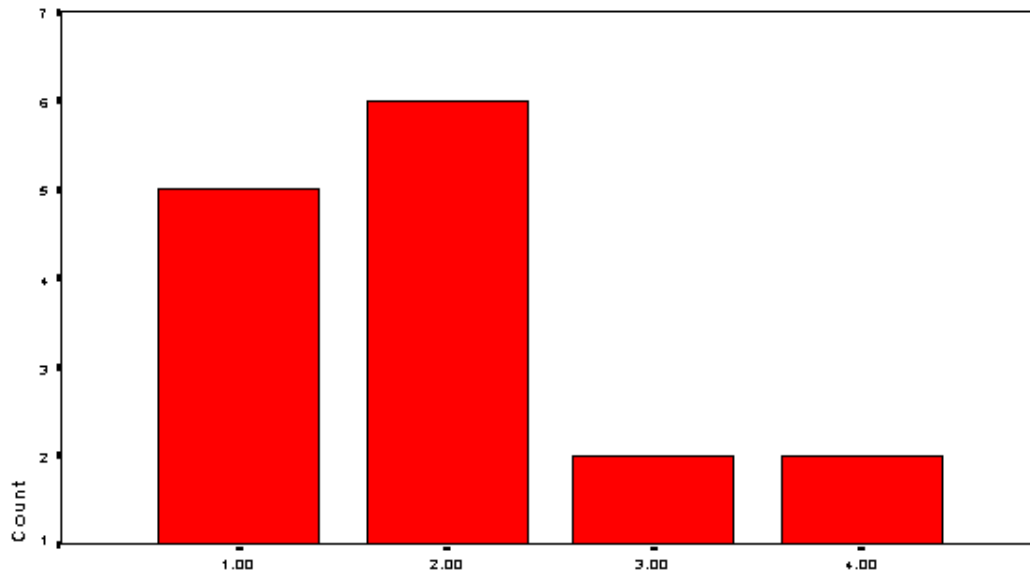
Aan elk van 36 studenten werd gevraagd of zij alcohol drinken, en zo ja, welke soort alcoholische drank de voorkeur heeft. Tevens werd voor elke student de leeftijd en het geslacht geregistreerd.

- a. Voor iemand die geen alcohol drinkt kan geen alcoholvoorkeur ingevuld worden. Als alcoholvoorkeur moet 'missing value' opgegeven worden. Wijzig de tabel zodanig dat dat inderdaad het geval is (niet handmatig !).
- b. Deel de leeftijd in in groepen: <-20, 21-25, 26-30, 31-> . Maak vervol-

gens een kruistabel van leeftijdsgroep en alcoholvoorkeur op basis van alleen die studenten die alcohol drinken. Zorg dat elke kolom correspondeert met een leeftijdsgroep en elke rij correspondeert met een alcoholvoorkeur. De cellen moeten de aantallen, de rijpercentages, de kolompercentages en de totaalpercentages weergeven.

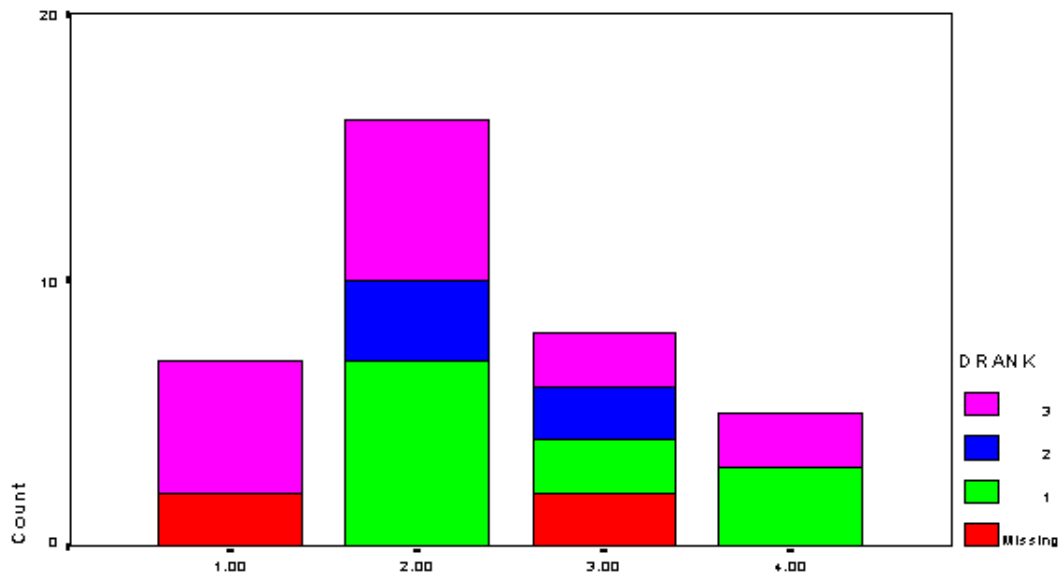
		DRANK			
	Count				
	Row Pct				
	Col Pct				Row
	Tot Pct	1	2	3	Total
LFTKL	1.00			5	5
				100.0	15.6
				33.3	
				15.6	
	2.00	7	3	6	16
		43.8	18.8	37.5	50.0
		58.3	60.0	40.0	
		21.9	9.4	18.8	
	3.00	2	2	2	6
		33.3	33.3	33.3	18.8
		16.7	40.0	13.3	
		6.3	6.3	6.3	
	4.00	3		2	5
		60.0		40.0	15.6
		25.0		13.3	
		9.4		6.3	
Column		12	5	15	32
Total		37.5	15.6	46.9	100.0

- c. Maak een staafdiagram, waarbij voor elke leeftijdsgroep een staaf aanwezig is die het aantal studenten aangeeft dat binnen die leeftijdsgroep de voorkeur heeft voor bier.



LFTKL

- d. Maak een gesegmenteerd staafdiagram met voor elke leeftijdsgroep een staaf. Elke staaf visualiseert binnen de betreffende leeftijdsgroep voor elke alcoholvoorkeur het percentage studenten.



LFTKL

- e. Maak een taartdiagram met voor elke alcoholvoorkeur een punt. Hierbij wordt geen onderscheid gemaakt in leeftijdsgroepen.



