

Opdracht 2a

Stamdiagrammen, histogrammen, tijdreeksgrafieken

De 'Old Farmers Almanac' vermeldt de groeiseizoenen voor de grote steden in de V.S., zoals gerapporteerd door het National Climatic Center. Het groeiseizoen wordt gedefinieerd als het gemiddeld aantal dagen tussen de laatste vorstdag in de lente en de eerste vorstdag in de herfst. De waarden zijn:

279	244	318	262	335	321	165	180	201	252
145	192	217	179	182	210	271	302	169	192
156	181	156	125	166	248	198	220	134	189
141	142	211	196	169	237	136	203	184	224
178	279	201	173	252	149	229	300	217	203
148	220	175	188	160	176	128			

- a. Maak een stamdiagram van deze gegevens. Rond elk aantal af door het laatste cijfer weg te laten, zodat de stammen honderdtallen en de bladen tientallen voorstellen. Beschrijf de vorm van de verdeling. Is zij ruwweg symmetrisch of duidelijk scheef? Zijn er uitschieters?
- b. Maak een histogram van de verdeling van het groeiseizoen. Kies eerst klassen en maak een geschikte frequentietabel.
- c. Maak een tijdreeksgrafiek. De opeenvolgende jaren verlopen langs de rijen van de tabel. Ga er vanuit dat de gegevens betrekking hebben op de jaren 1901 t/m 1957.

Opdracht 2a - berekening

- a. Wanneer we voor elke waarde het laatste cijfer weglaten krijgen we de volgende waarden:

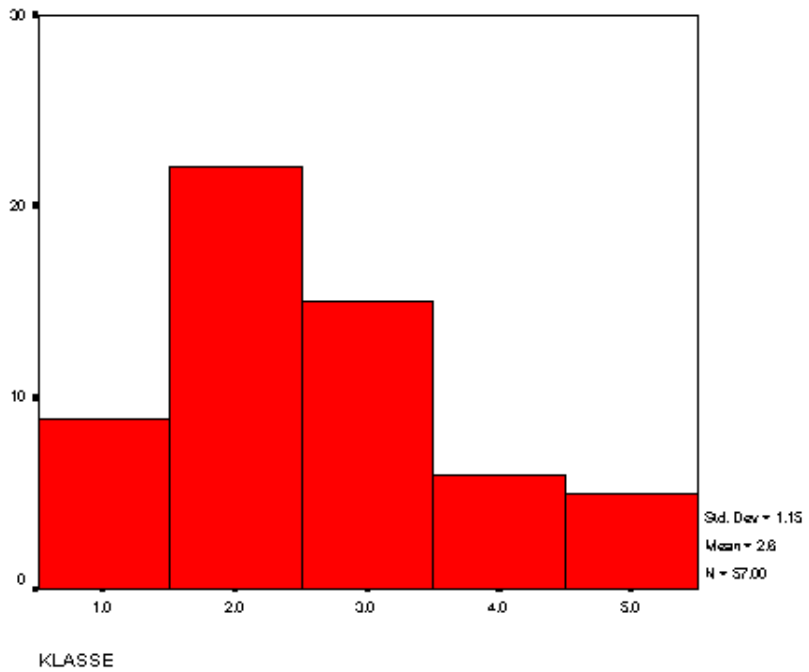
27	24	31	26	33	32	16	18	20	25	14	19	21
17	18	21	27	30	16	19	15	18	15	12	16	24
19	22	13	18	14	14	21	19	16	23	13	20	18
22	17	27	20	17	25	14	22	30	21	20	14	22
17	18	16	17	12								

Het stamdiagram komt er nu als volgt uit te zien:

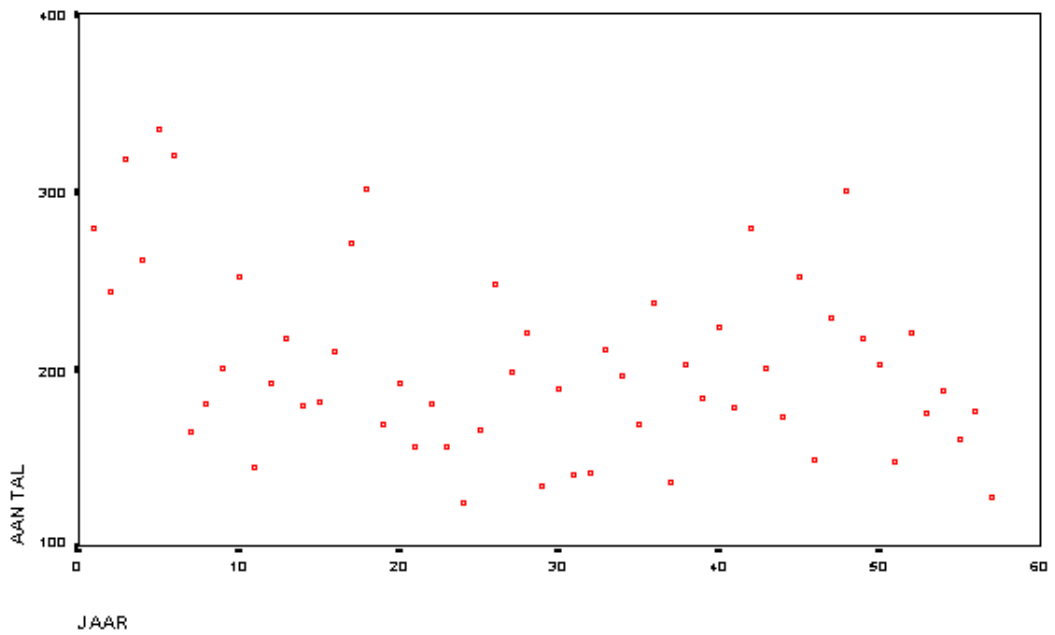
1		2	2	3	3	4	4	4	4	4	5	5	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	8	9	9	9	9
2		0	0	0	0	1	1	1	1	2	2	2	2	3	4	4	5	5	6	7	7	7										
3		0	0	1	2	3																										

- b. klasse | frequentie

1 - 149		9
150 - 199		22
200 - 249		15
250 - 299		6
300 - 349		5



C.



Opdracht 2a - S-PLUS

 Plaats de gegevens in een tabel. Kies >File >New. Kies voor >Data Set. Type de waarden in in de eerste kolom. S-PLUS geeft als kolomnaam 'V1'. We willen die veranderen in 'aantal'. Selecteer de kolom door te klikken op de kolom-

naam. Ga vervolgens naar >Format en kies >Selected Object(s). Vervang achter Name 'V1' door 'aantal'. Controleer of achter Format Type gekozen is voor Decimal. Achter Precision wordt het aantal decimalen opgegeven. Vul daar 0 in. Klik op >OK.

- 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.

Het commando dat gebruikt moet worden voor het maken van een stamdiagram heeft als algemene vorm:

```
stem(tabelnaam$kolomnaam, aantal bladeren per stam, scheidingspunt)
```

Hierbij is de tabelnaam de naam van de tabel waar we mee werken, en de kolomnaam de kolom op basis waarvan we het stamdiagram willen maken. Het aantal bladeren per stam betreft het aantal verschillende bladeren per stam. Hierbij kunnen we kiezen uit 2, 5 of 10. Met het scheidingspunt wordt bepaald wat de schaal van de stammen en de schaal van de bladeren wordt. Er geldt:

scheidingspunt	stammen	bladeren
1	tienden	honderdsten
0	eentallen	tienden
-1	tientallen	eentallen
-2	hondertallen	tientallen
-3	duizendtallen	honderdtallen

Stel de naam van de tabel die we gebruiken heet 'data', en de naam van de kolom op basis waarvan we het stamdiagram willen maken heet 'aantal'. Op basis daarvan willen we nu een stamdiagram maken met 10 bladeren per stam. Omdat we willen dat de stammen hondertallen en de bladeren tientallen representeren, kiezen we voor -2 als scheidingspunt. Het commando wordt nu:

```
stem(data$aantal,10,-2)
```

- b. Kies >Data >Create Categories. Kies onder Data en achter Source Column de variabele 'aantal' 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 intervallen is niet gelijk aan de verdeling die wij willen. De klassen die wij willen hebben zijn: 149 en lager, 150-199, 200-249, 250-299, 300 en hoger. De eindpunten van de eerste vier intervallen zijn: 149, 199, 249 en 299. 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 'aantal' 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.

Kies >Statistics >Data Summaries >Crosstabulations. Selecteer achter Variables de variabele 'klasse'. Klik nu bovenaan dit venster op >Options. Zet de opties >Show Cell Proportions, >Show Marginal Totals en >Run Chi-Square Test uit. Klik op >OK.

Het maken van het histogram gaat nu als volgt. Kies >Graph >2D Plot. Kies onder Axes Type voor Linear, en onder Plot Type voor Histogram(x). Klik op >OK. Selecteer onder Data Columns en achter x Columns de variabele 'klasse'. Klik op >OK.

- c. Kies Graph >2D Plot. Kies onder Axes voor Linear, en onder Plot Type voor Scatter Plot (x, y1, y2, ...). Klik op >OK. Selecteer onder Data Columns en achter y Columns de variabele 'aantal'. Klik op >OK.

Opdracht 2a - SPSS

Plaats de gegevens in een tabel. Kies >File >New >Data. Type de waarden in de eerste kolom. Definieer de kolom door te dubbelklikken op de tekst 'var' in de juiste kolom. Nu verschijnt het dialoogkader 'Define Variable'. Voer bij >Variable name de naam 'aantal' in. Kies vervolgens >Type. Kies als type 'Numeric'. Vul bij >With 3 (het aantal posities) en bij >Decimal places 0 in (aantal cijfers achter de komma). Klik vervolgens op >Continue en in 'Define Variable' op >OK.

- a. Kies >Statistics >Summarize >Explore. Plaats 'aantal' in Dependent List met > (pijlte naar rechts). Kies >Plots. Zet onder Descriptieve 'Stem-and-leaf' aan. Kies onder Boxplots voor >None. Klik op >Continue. Klik vervolgens op >OK. In het output-window vind je het stamdiagram.
- b. Kies >Window en kies vervolgens de SPSS Data Editor. Kies daarna >Transform >Recode >Into Different Variables. Breng in het window 'Recode into Different Variables' de variabele 'aantal' met > (pijlte naar rechts) naar Input Variable -> Output Variable. Klik op het tekstvak 'Name' bij Output Variable, vul daar 'klasse' in en klik op >Change. Klik vervolgens op >Old and New Values.

Stel we kiezen de volgende klassen: 149 en lager, 150-199, 200-249, 250-299, 300 en hoger.

Het opgeven van de eerste klasse gaat als volgt. Klik op het rondje voor >Range lowest through en type de waarde 49 in. Vul in het tekstvak >Value bij >New value de waarde 1 in en klik op >Add.

Het opgeven van de tweede klasse gaat als volgt. Klik op het rondje voor >Range through en type in het tekstvak voor through de waarde 150 en na through de waarde 199 in. Vul in het tekstvak >Value bij >New value de waarde 2 in en klik op >Add. Het opgeven van de derde en de vierde klasse gaat op analoge wijze. In het tekstvak >Value bij >New Value vul je dan respectievelijk de waarden 3 en 4 in.

Het opgeven van de vijfde klasse gaat als volgt. Klik op het rondje voor >Range Lowest through en type de waarde 300 in.

Klik nu op >Continue en in het windows 'Recode into Different Variables' op >OK. Er is nu een nieuwe variabele 'klasse' gevormd die in de data-matrix kan worden bekeken.

Het maken van een frequentietabel op basis van deze klassen gaat als volgt. Kies >Statistics >Summarize >Frequencies. Je komt dan in het window 'Frequencies'. Breng de variabele 'klasse' met > (pijlte naar rechts) naar Variable(s). Klik vervolgens op >Charts. Je komt dan in het window 'Frequencies: Charts'. Klik op het rondje voor Histograms. Klik op >Continue. Zorg dat Display frequency tables aan staat. Klik op >OK. In het output-window worden achtereenvolgens de frequentietabel en de histogram geplaatst.

- c. Kies >Window en kies vervolgens de SPSS Data Editor. Vul in de derde kolom de jaartallen 1901 t/m 1957 in. Definieer de kolom door te dubbelklikken op de tekst 'var' in de kolom. Nu verschijnt het dialoogkader 'Define Variable'. Voer bij >Variable name de naam 'jaar' in. Kies vervolgens >Type. Kies als type 'Numeric'. Vul bij >With 4 (het aantal posities) en bij >Decimal places 0 in (aantal cijfers achter de komma). Klik vervolgens op >Continue en in 'Define Variable' op >OK.

Kies nu >Graphs >Scatter. Kies in het window 'Scatterplot' voor 'Simple'. Klik vervolgens op 'Define'. Nu kom je in het window 'Simple Scatterplot'. Verplaats 'aantal' naar YAxis en 'jaar' naar XAxis. Klik nu op >OK. In het output-window verschijnt nu de tijdreeksgrafiek.

Opdracht 2a - verslag

De 'Old Farmers Almanac' vermeldt de groeiseizoenen voor de grote steden in de V.S., zoals gerapporteerd door het National Climatic Center. Het groeiseizoen wordt gedefinieerd als het gemiddeld aantal dagen tussen de laatste vorstdag in de lente en de eerste vorstdag in de herfst.

- a. Maak een stamdiagram van deze gegevens. Rond elk aantal af door het laatste cijfer weg te laten, zodat de stammen honderdtallen en de bladen tientallen voorstellen. Beschrijf de vorm van de verdeling. Is zij ruwweg symmetrisch of duidelijk scheef? Zijn er uitschieters?

Het stamdiagram ziet er als volgt uit:

```

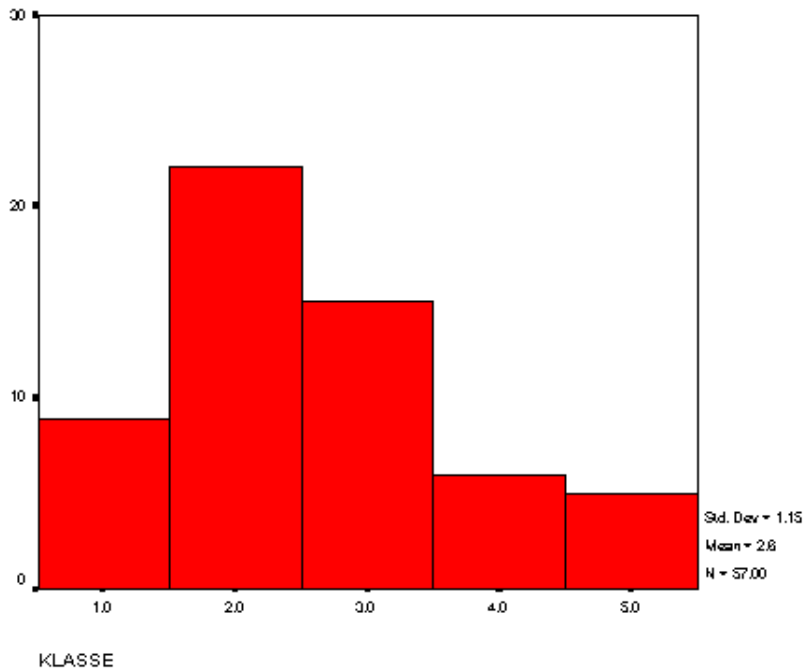
1 | 2 2 3 3 4 4 4 4 4 5 5 6 6 6 6 6 7 7 7 7 7 8 8 8 8 8 8 9 9 9 9
2 | 0 0 0 0 1 1 1 1 2 2 2 2 3 4 4 5 5 6 7 7 7
3 | 0 0 1 2 3

```

De vorm van de verdeling is duidelijk scheef.

- b. Maak een histogram van de verdeling van het groeiseizoen. Kies eerst klassen en maak een geschikte frequentietabel.

klasse	frequentie
1 - 149	9
150 - 199	22
200 - 249	15
250 - 299	6
300 - 349	5



c. Maak een tijdreeksgrafiek. De opeenvolgende jaren verlopen langs de rijen van de tabel. Ga er vanuit dat de gegevens betrekking hebben op de jaren 1901 t/m 1957.

