

Opdracht 1a

Introductie

Bij het maken van deze opgave worden de volgende vragen beantwoord:

Hoe start ik S-PLUS op?

Hoe lees ik gegevens in vanuit een ASCII-bestand in een tabel?

Hoe kan ik berekeningen laten uitvoeren op de gegevens?

Hoe breng ik tekst over van S-PLUS naar Word?

Hoe voer ik gegevens in in een tabel?

Hoe maak ik een grafiek op basis van de gegevens?

Hoe breng ik een grafiek over van S-PLUS naar Word?

- a. Start S-PLUS op. Zet computer en beeldscherm aan. Geef username en password en log in. Kies in Windows voor SPlus2000.
- b. In een experiment leest een proefpersoon zinnen op het beeldscherm van een computer. Elke keer nadat de proefpersoon een woord heeft gelezen, drukt hij op een toets. Het woord verdwijnt, en het volgende woord verschijnt onmiddellijk rechts van waar het vorige woord stond. Wanneer de proefpersoon opnieuw op een toets drukt, verdwijnt ook dit woord en verschijnt het weer daaropvolgende woord. De tijd tussen twee toetsaanslagen is de tijd die de proefpersoon nodig had om het woord te lezen. Voor 24 woorden zijn de tijden hieronder achtereenvolgens gegeven in milliseconden (Bron: Edith Kaan en Laurie Stowe, *Developing an Experiment*, 1995. Techniques and Design, Klapper vakgroep Taalwetenschappen, Rijksuniversiteit Groningen):

450	390	467	654	30	542	334	432	421
357	497	493	550	549	467	575	578	342
446	547	534	495	979	479			

De data bevindt zich in z:\public\share\heeringa\woord.txt. Lees dit ASCII-bestand in. Definieer de kolom van de tabel en kies een goede kolomnaam.

- c. Bereken het gemiddelde, de mediaan en de standaardafwijking. Deze resultaten verschijnen in het output-venster.
- d. Breng de resultaten van de berekeningen over naar Word.
- e. In de Leidse lakenindustrie werd in de jaren 1651 - 1670 jaarlijks vastgelegd hoeveel exemplaren geproduceerd werden. Hieronder zijn de productiecijfers gegeven in duizenden stuks (Bron: N. W. Posthumus, *De geschiedenis van de Leidse lakenindustrie. II. De nieuwe tijd. De lakenindustrie en verwante industrieën ('s-Gravenhage 1939) 930-931*).

95	110	130	115
80	120	130	120
80	120	140	135
85	125	145	135
115	125	120	115

Maak een nieuwe tabel en voer de data in die tabel in.

- f. Verdeel de gegevens in vijf klassen: <- 90, 91-105, 106-120, 121-135, 136 ->. Maak vervolgens een histogram waarbij elke staaf correspondeert

met een klasse.

- g. Breng het histogram over naar Word.
- h. Het Word-document bevat nu zowel de resultaten van c. als van f. Bewerk het document zodanig dat het een goed verslag wordt. Print dit uit en lever het in (postvakje, W.J. Heeringa, 4e verdieping, centrale hal).

Opdracht 1a - S-PLUS

-
- a. Start S-PLUS op. Zet computer en beeldscherm aan. Geef username en password en log in. Kies in Windows voor SPlus2000.
 - b. Kies >File >Import Data >From File. Ga bij Look in eerst naar Sys on 'Lol' (Z:), dubbelklik hierop, een dubbelklik vervolgens op Public, Share, Heeringa. Kies nu bij Files of Type voor ASCII Text Files (*.prn, *.txt, *.asc). Klik nu op Woord, en kies daarna >Open. S-PLUS geeft als kolomnaam 'V1'. We willen die veranderen in 'tijd'. Selecteer de kolom door te klikken op de kolomnaam. Ga vervolgens naar >Format en kies >Selected Object(s). Vervang achter Name 'V1' door 'tijd'. Controleer of achter Format Type gekozen is voor Decimal. Klik op >OK.
 - c. Kies >Statistics >Data Summaries >Summaries Statistics. Selecteer onder Data en achter Variables de variabele 'tijd', en controleer of onder Summaries by Group en achter Group Variables voor (None) is gekozen. Klik nu bovenaan dit venster op >Statistics. Zorg dat alleen >Mean, >Median en >Std. Deviation geselecteerd zijn. Klik op >OK.
 - d. Breng de resultaten van de berekeningen over naar Word. Selecteer in S-PLUS met de linker muisknop de tekst in het Report-venster. Kies vervolgens >Edit >Copy. Ga daarna naar Word. Maak een nieuw document. Kies >File >New. Klik in het venster New op >OK. Kies nu >Edit >Paste. De tekst wordt nu in het nieuwe document geplaatst.
 - e. 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 'productie'. Selecteer de kolom door te klikken op de kolomnaam. Ga vervolgens naar >Format en kies >Selected Object(s). Vervang achter Name 'V1' door 'productie'. Controleer of achter Format Type gekozen is voor Decimal. Achter Precision wordt het aantal decimalen opgegeven. Vul daar 0 in. Klik op >OK.
 - f. Kies >Data >Create Categories. Kies onder Data en achter Source Column de variabele 'productie' 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: <- 90, 91-105, 106-120, 121-135, 136 ->. De eindpunten van de eerste vier intervallen zijn: 90, 105, 120 en 135. 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 'productie' 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.

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.

- g. Breng het histogram over naar Word. Kies >Edit >Copy Graph Sheet Page. Ga vervolgens naar Word. Kies >Edit >Paste Special. Kies onder As: voor Picture. Klik op >OK. De grafiek wordt nu in het document geplaatst.
- h. Het Word-document bevat nu zowel de resultaten van c. als van f. Bewerk het document zodanig dat het een goed verslag wordt.

Opdracht 1a - SPSS

- a. Start SPSS op. Zet computer en beeldscherm aan. Geef username en password en log in. Kies in Windows voor SPSS_8.
- b. Kies >File >Read Ascii Data >Free Field. Klik op >Browse. Ga bij Look in eerst naar Sys on 'Lol' (Z:), dubbelklik hierop, een dubbelklik vervolgens op Public, Share, Heeringa. Kies nu bij Files of Type voor All files (*.*). Klik nu op Woord.txt, en kies daarna >Open. Vul nu in het vak achter Name een kolomnaam in: tijd. Kies als Data Type voor Numeric. Klik op >Add en daarna op >OK.
- c. Kies >Statistics >Summarize >Frequencies. Plaats de variabele 'tijd' onder Variable(s). Klik op >Statistics. Klik op >Mean, >Median en >Std. Deviation, en vervolgens op >Continue. Zorg dat Display frequency tables uitgeschakeld is. Klik op >OK.
- d. Breng de resultaten van de berekeningen over naar Word.
- e. Kies >File >New >Data. De oude gegevens hoeft je niet te bewaren. Type de waarden in 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 'product' 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.
- f. Kies daarna >Transform >Recode >Into Different Variables. Breng in het window 'Recode into Different Variables' de variabele 'product' met > (pijlje 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.

Verdeel de gegevens in vijf klassen: <- 90, 91-105, 106-120, 121-135, 136 ->.

Het opgeven van de eerste klasse gaat als volgt. Klik op het rondje voor >Range lowest through en type de waarde 90 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 91 en na through de waarde 105 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 klas-

se 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 136 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 de histogram gaat nu als volgt. Kies >Graphs >Histogram. Verplaats 'klasse' naar rechts onder Variable door op > (pijlje naar rechts) te klikken. Klik vervolgens op >OK. In het output-window verschijnt nu de histogram.

- g. Breng het histogram over naar Word.
- h. Het Word-document bevat nu zowel de resultaten van c. als van f. Bewerk het document zodanig dat het een goed verslag wordt.

Opdracht 1a - verslag

- b. In een experiment leest een proefpersoon zinnen op het beeldscherm van een computer. Elke keer nadat de proefpersoon een woord heeft gelezen, drukt hij op een toets. Het woord verdwijnt, en het volgende woord verschijnt onmiddellijk rechts van waar het vorige woord stond. Wanneer de proefpersoon opnieuw op een toets drukt, verdwijnt ook dit woord en verschijnt het weer daaropvolgende woord. De tijd tussen twee toetsaanslagen is de tijd die de proefpersoon nodig had om het woord te lezen. Voor 24 woorden zijn de tijden hieronder achtereenvolgens gegeven in milliseconden (Bron: Edith Kaan en Laurie Stowe, Developing an Experiment, 1995. Techniques and Design, Klapper vakgroep Taalwetenschappen, Rijksuniversiteit Groningen):

450	390	467	654	30	542	334	432	421
357	497	493	550	549	467	575	578	342
446	547	534	495	979	479			

- c. Bereken het gemiddelde, de mediaan en de standaardafwijking. Deze resultaten verschijnen in het output-venster.

gemiddelde = 483.6667
mediaan = 486.0000
standaardafwijking = 160.6299

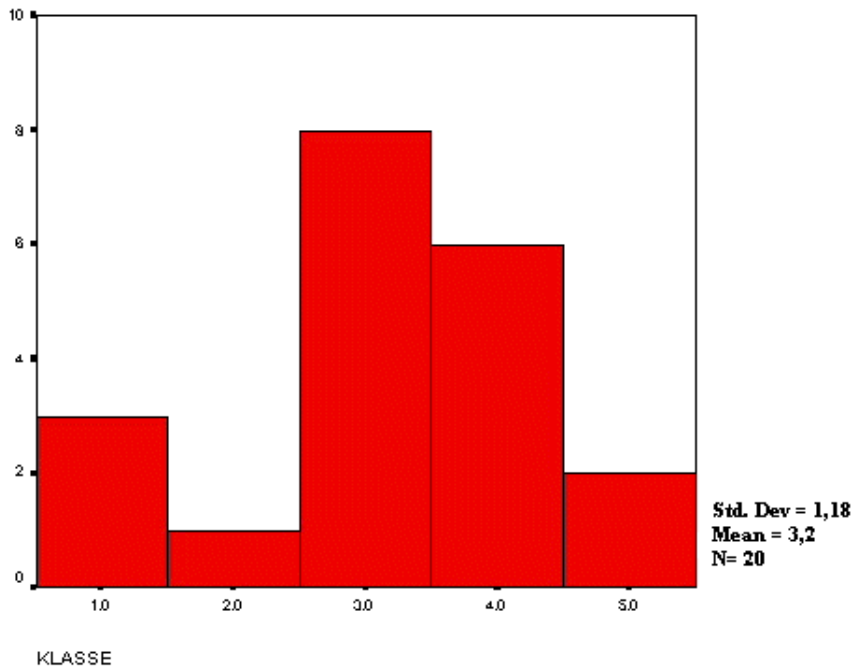
Het gemiddelde en de mediaan liggen dicht bij elkaar. De standaardafwijking is vrij fors. De tijd die nodig is voor het lezen van een woord is dus niet erg constant.

- e. In de Leidse lakenindustrie werd in de jaren 1651 - 1670 jaarlijks vastgelegd hoeveel exemplaren geproduceerd werden. Hieronder zijn de productiecijfers gegeven in duizenden stuks (Bron: N. W. Posthumus, De geschiedenis van de Leidse lakenindustrie. II. De nieuwe tijd. De lakenindustrie en verwante industrieën ('s-Gravenhage 1939) 930-931).

95	110	130	115
80	120	130	120
80	120	140	135
85	125	145	135

115 125 120 115

f. Verdeel de gegevens in vijf klassen: <- 90, 91-105, 106-120, 121-135, 136 ->. Maak vervolgens een histogram waarbij elke staaf correspondeert met een klasse.



We zien dat de derde klasse het grootste aantal metingen bevat, gevolgd door de vierde klasse.