

STATISTIEK: INLEIDING EN HERHALING*

Free High School Science Texts Project

This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License 3.0[†]

1 Inleiding

In die wêreld rondom ons word inligting dikwels in die vorm van syfers, grafieke en tabelle gegee. Ons sien dit op die televisie, op die radio en in die koerante. Ons word blootgestel aan misdaadsyfers, sportuitslae, reënval, die uitgawes van die regering, die tempo van HIV/VIGS infeksie, bevolkingsgroei en ekonomiese groei.

Hierdie hoofstuk demonstreer hoe Wiskunde gebruik kan word om data te manipuleer, om data en tendense voor te stel of wan voor te stel en om oplossings te bied wat direk betrekking het op die wêreld rondom ons.

Vaardighede wat in vorige grade verwerf is en verband hou met die versameling, organisering, uitbeelding, analise en interpretasie van inligting, word hier verder ontwikkel.

2 Hersiening van vorige Werk

Dataversameling is in vorige grade bekendgestel as 'n manier om antwoorde te kry vir vrae wat te make het met die wêreld rondom ons.

2.1 Data en Datainsameling

2.1.1 Data

Definition 1: Data

Data verwys na inligting wat waargeneem of opgeneem is as deel van 'n eksperiment of 'n meningspeiling. Daar is twee tipes data: primêre en sekondêre data. Die woord "data" is die meervoud van die woord "datum".

Data kan geklassifiseer word as *primêr* of *sekondêr*, en primêre en sekondêre data kan verder geklassifiseer word as *kwalitatief* of *kwantitatief*. Figure 1 som die klassifikasie van data op.

*Version 1.1: Aug 4, 2011 7:21 am +0000

[†]<http://creativecommons.org/licenses/by/3.0/>

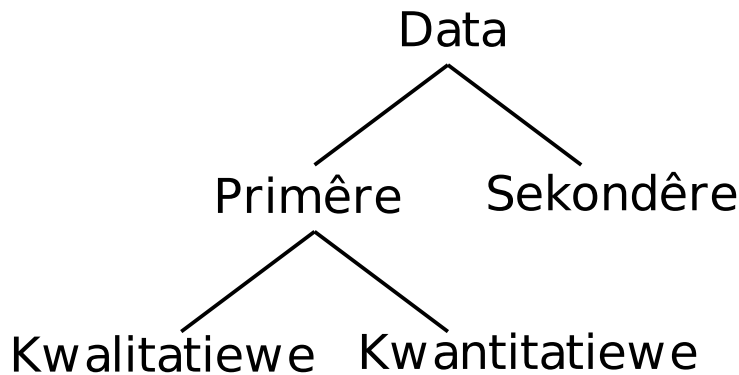


Figure 1: Klassifikasie van data

Primêre data: beskryf die oorspronklike data wat versamel is. Hierdie tipe data staan ook bekend as *rou* data. Dikwels is die primêre datastel baie groot en moet dit opgesom of verwerk word om betekenisvolle inligting uit te lig.

Kwalitatiewe data: is inligting wat nie met getalle beskryf kan word nie, byvoorbeeld as jy data insamel oor hoe mense voel of wat hul gunstelingkleur is.

Kwantitatiewe data: is inligting wat geskryf kan word as getalle, byvoorbeeld as jy data insamel oor mense se lengte of massa.

Sekondêre data: is primêre data wat opgesom of verwerk is. Die stel kleure wat mense as hul gunstelingkleure aangedui het, is 'n voorbeeld van sekondêre data omdat dit 'n opsomming is van mense se antwoorde.

Die proses om primêre data om te skakel na sekondêre data deur analise, groepering of organisering is die proses waardeur informasie geskep word.

2.1.2 Doelwit van die Insameling van Primêre Data

Data word versamel om antwoorde te kry wat help om 'n sekere situasie beter te verstaan. Hier is 'n paar voorbeelde van dataversameling uit die regte wêreld wat kwalitatiewe en kwantitatiewe data illustreer.

2.1.3 Kwalitatiewe Data

- Die plaaslike regering sal wil weet hoeveel inwoners elektrisiteit het en mag dan aan hulle vra: "Het jou huis 'n veilige, ononderbroke toevoer van elektrisiteit vanaf die nasionale kragvoorsiener (Eskom)?"
- 'n Maatskappy wat medisyne vervaardig mag vra: "Hoe effektief verlig ons pil hoofpyn?" Daar kan dan aan mense wat die pil gebruik om hoofpyn te verlig, gevra word: "Hoe baie help ons pil om jou hoofpyn weg te neem?" Die maatskappy kan dan op grond van mense se antwoorde bepaal hoe effektief hul produk is.
- 'n Motormaatskappy wil hulle kliëntediens verbeter en kan dan aan kliënte vra: "Hoe kan ons ons kliëntediens verbeter?"
- 'n Supermarkbestuurder mag vra: "Watter tipes gaskoeldrank moet ek in my supermark in voorraad hê?" Daar mag dan miskien aan klante gevra word: "Wat is jou gunsteling gaskoeldrank?" Die klante se keuse of voorkeur is 'n voorbeeld van kwalitatiewe data.

2.1.4 Kwantitatiewe Data

- 'n Maatskappy wat selfone vervaardig mag moontlik data versamel oor hoe dikwels mense nuwe selfone koop en watter faktore hul besluit beïnvloed. Sodoende kan die selfoonmaatskappy bepaal op watter fasiliteite en eienskappe hulle moet fokus om hul produk meer aantreklik te maak vir kopers.
- 'n Lid van die stadsraad wil moontlik weet hoeveel ongelukke by 'n spesifieke padkruising gebeur om te besluit waar verkeersligte opgerig moet word. Die raadslid kan dan die plaaslike polisiekantoor besoek en hul rekords nagaan om die nodige data te bekom.
- 'n Supermarkbestuurder mag vra: "Watter tipes gaskoeldrank moet ek in my supermark in voorraad hê?" Daar mag dan aan klante gevra word: "Wat is jou gunsteling gaskoeldrank?" Op grond van hoeveel mense 'n spesifieke gaskoeldrank verkies, kan die bestuur dan 'n ingeligte besluit neem van watter gaskoeldrank om aan te hou.

Dit is belangrik om daarop te let dat verskillende vrae, verskillende fasette van die 'n situasie na vore bring en dat dit die begrip van die situasie sal beïnvloed. Byvoorbeeld, die eerste vraag op die lys kan geformuleer word om te vra: "Het jou huis elektrisiteit?". Indien die gebruiker "Ja" antwoord hierop, terwyl hy krag van sy buurman af herlei, sal dit die verkeerde indruk gee dat die betrokke persoon nie krag nodig het van die amptelike verskaffer nie.

2.2 Metodes van Dataversameling

Die metode of strategie van dataversameling moet ooreenstem met die vrae wat gevra word. 'n Paar voorbeelde van dataversamelingsmetodes is:

1. Vraelyste, meningspeilings en onderhoude
2. Eksperimente
3. Ander bronne (vriende, familie, koerante, boeke, tydskrifte en die Internet)

Die belangrikste aspek van elke metode van dataversameling is om die vrae wat beantwoord moet word duidelik te formuleer. Die tipe dataversamelingsmetode moet gekies word om by jou vrae te pas.

Byvoorbeeld vraelyste, meningspeilings en onderhoude sal die beste pas by die vrae in die voorbeelde in "Die Doelwit van die Versameling van Primêre Data" (Section 2.1.2: Doelwit van die Insameling van Primêre Data).

2.3 Steekproef en Bevolking (Populasie)

Voordat daar begin word met die datainsameling is dit belangrik om te besluit hoeveel data nodig is sodat die resultate 'n verteenwoordigende aanduiding sal gee van die antwoord op 'n sekere vraag. In die ideale geval sal die studie so ontwerp word dat die maksimum hoeveelheid inligting verkry word met die minimum hoeveelheid moeite. Die konsepte *bevolking* en *steekproef* is baie belangrik ten einde die energie- en koste-uitset so min as moontlik te hou.

Die volgende terme behoort bekend te wees:

Bevolking: beskryf die hele groep wat in die studie in ag geneem word. Byvoorbeeld, as jy wil weet hoeveel leerders in jou skool in die winter verkoue gekry het, dan sal jou bevolking al die leerders in jou skool wees.

Steekproef: beskryf 'n groep wat gekies word om die bevolking te verteenwoordig. Byvoorbeeld, vir die opname oor verkoue in die skool kan jy miskien net 'n paar leerders kies - byvoorbeeld een uit elke klas.

Ewekansige steekproef: beskryf 'n steekproef wat op so manier uit die bevolking gekies word dat elke lid van die bevolking 'n gelyke kans het om gekies te word.

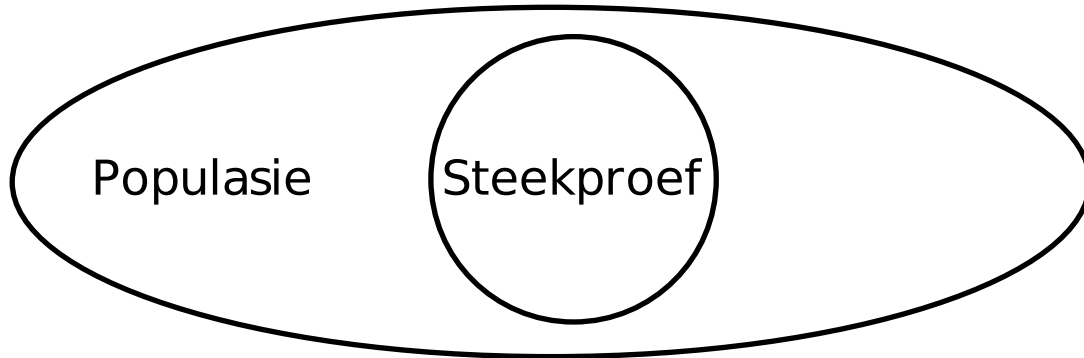


Figure 2: 'n Illustrasie van hoe 'n steekproef uit die bevolking gekies word

Om resultate te verkry wat verteenwoordigend is, is die van kritieke belang om 'n verteenwoordigende steekproef te kry. Byvoorbeeld, as ons wou bepaal hoe groepsdruk die besluit om te begin rook, beïnvloed - dan sou die resultate baie anders gewees het as slegs seuns ondervra is, in vergelyking met 'n studie waar onderhoude met beide seuns en meisies gevoer is.

Daarom moet vrae soos: "Hoeveel onderhoude word benodig?" en "Hoe kies ons kandidate vir onderhoude?" tydens die ontwerpfasie van die steekproefproses gevra word.

Die mees akkurate resultate word verkry indien die hele bevolking gebruik word as steekproef vir 'n opname, maar dit kan baie duur wees en/of baie lank neem. Die tweede beste metode is om 'n steekproef ewekansig te kies. Dit beteken dat elke lid van die bevolking 'n gelyke kans het om geselekteer te word, onafhanklik van hoe die lede gekies word. Daar is verskeie metodes om lede op hierdie manier te kies, byvoorbeeld, name kan uit 'n hoed getrek word. Meeste moderne wetenskaplike sakrekenaars het 'n sleutel wat mens kan druk om ewekansige getalle te genereer wat mens kan gebruik om 'n steekproef te kies. Sigbladpakkette op 'n rekenaar het ook gewoonlik so 'n funksie.

So, as jy byvoorbeeld 'n bevolking van 1 000 leerders in jou skool het, kan jy moontlik 100 leerders ewekansig kies en dit sal dan die steekproef wees wat jy vir die opname gebruik.

3 Voorbeelde van Datastelle

Die res van hierdie hoofstuk handel oor die wiskundige besonderhede wat nodig is om data wat versamel is, te analiseer.

Hier volg nou 'n paar voorbeelde van datastelle wat gebruik kan word om die metodes wat toegepas word, te verduidelik.

3.1 Datastel 1: Gooi van 'n Muntstuk

'n Ewekansige muntstuk is 100 keer gegooi en die kant waarop die muntstuk land, is opgeneem (kop of stert). Die data is opgeneem in "Datastel 1: Gooi van 'n muntstuk" (Table 1).

K	S	S	K	K	S	K	K	K	K
K	K	K	K	S	K	K	S	S	S
S	S	K	S	S	K	S	K	S	K
K	K	S	S	K	S	S	K	S	S
S	K	K	K	S	S	K	S	S	K
K	S	S	S	S	K	S	S	K	K
S	S	K	S	S	K	S	S	K	S
K	S	S	K	S	S	S	S	K	S
S	K	S	S	K	K	K	S	K	S
S	S	S	K	K	S	S	S	K	S

Table 1: Resultate wanneer 'n ewekansige muntstuk 100 keer gegooi word. K beteken dat die muntstuk met sy kop na bo geland het terwyl S beteken dat die muntstuk met sy stert na bo geland het.

3.2 Datastel 2: Gooi van 'n Dobbelsteen

'n Ewekansige dobbelsteen is 200 keer gegooi en die waardes waarop die dobbelsteen land, is opgeneem. Die data is opgeneem in "Datastel 2: Gooi van 'n Dobbelsteen" (Section 3.2: Datastel 2: Gooi van 'n Dobbelsteen).

3	5	3	6	2	6	6	5	5	6	6	4	2	1	5	3	2	4	5	4
1	4	3	2	6	6	4	6	2	6	5	1	5	1	2	4	4	2	4	4
4	2	6	4	5	4	3	5	5	4	6	1	1	4	6	6	4	5	3	5
2	6	3	2	4	5	3	2	2	6	3	4	3	2	6	4	5	2	1	5
5	4	1	3	1	3	5	1	3	6	5	3	4	3	4	5	1	2	1	2
1	3	2	3	6	3	1	6	3	6	6	1	4	5	2	2	6	3	5	3
1	1	6	4	5	1	6	5	3	2	6	2	3	2	5	6	3	5	5	6
2	6	6	3	5	4	1	4	5	1	4	1	3	4	3	6	2	4	3	6
6	1	1	2	4	5	2	5	3	4	3	4	5	3	3	3	1	1	4	3
5	2	1	4	2	5	2	2	1	5	4	5	1	5	3	2	2	5	1	1

Table 2: Resultate wanneer 'n ewekansige dobbelsteen 200 keer gegooi word

3.3 Datastel 3: Massa van 'n Brood

In Suid-Afrika is daar regulasies oor die vervaardiging van brood om verbruikers te beskerm. Hier is 'n uittreksel uit 'n verslag oor die wetgewing:

"Wetgewing vereis dat 'n brood 800g moet weeg indien dit nie gemerk is nie, met 'n speling van 5 persent bo en 10 persent onder hierdie massa. Die gemiddelde massa van 10 van hierdie brode moet egter presies die aangeduide massa wees." - Sunday Tribune op 10 Oktober 2004, bladsy 10.

Ons kan die massa van brode bepaal en dit gebruik om vas te stel of verbruikers waarde vir hulle geld kry. 'n Ongemerkte brood moet 800g weeg. Vir een week is 10 verskillende brode by 'n sekere winkel elke dag geweeg. Die data word getoon in Table 3.

Maandag	Dinsdag	Woensdag	Donderdag	Vrydag	Saterdag	Sondag
802.39	787.78	815.74	807.41	801.48	786.59	799.01
796.76	798.93	809.68	798.72	818.26	789.08	805.99
802.50	793.63	785.37	809.30	787.65	801.45	799.35
819.59	812.62	809.05	791.13	805.28	817.76	801.01
801.21	795.86	795.21	820.39	806.64	819.54	796.67
789.00	796.33	787.87	799.84	789.45	802.05	802.20
788.99	797.72	776.71	790.69	803.16	801.24	807.32
808.80	780.38	812.61	801.82	784.68	792.19	809.80
802.37	790.83	792.43	789.24	815.63	799.35	791.23
796.20	817.57	799.05	825.96	807.89	806.65	780.23

Table 3: Massas (in g) van 10 verskillende brode, vanaf dieselfde vervaardiger, bepaal by dieselfde winkel oor 'n tydperk van een week

3.4 Datastel 4: Temperature Wêreldwyd

Die wêreldwye gemiddelde temperature van 1861 tot 1996 word in Table 4 getoon. Die data, verkry by <http://www.cgd.ucar.edu/stats/Data/Climate/>¹, is na temperatuur in grade Celsius omgeskakel.

Jaar	Temperatuur	Jaar	Temperatuur	Jaar	Temperatuur	Jaar	Temperatuur
1861	12.66	1901	12.871	1941	13.152	1981	13.228
1862	12.58	1902	12.726	1942	13.147	1982	13.145
1863	12.799	1903	12.647	1943	13.156	1983	13.332
1864	12.619	1904	12.601	1944	13.31	1984	13.107
1865	12.825	1905	12.719	1945	13.153	1985	13.09
1866	12.881	1906	12.79	1946	13.015	1986	13.183
1867	12.781	1907	12.594	1947	13.006	1987	13.323
1868	12.853	1908	12.575	1948	13.015	1988	13.34
1869	12.787	1909	12.596	1949	13.005	1989	13.269
1870	12.752	1910	12.635	1950	12.898	1990	13.437
<i>continued on next page</i>							

¹<http://www.cgd.ucar.edu/stats/Data/Climate/>

1871	12.733	1911	12.611	1951	13.044	1991	13.385
1872	12.857	1912	12.678	1952	13.113	1992	13.237
1873	12.802	1913	12.671	1953	13.192	1993	13.28
1874	12.68	1914	12.85	1954	12.944	1994	13.355
1875	12.669	1915	12.962	1955	12.935	1995	13.483
1876	12.687	1916	12.727	1956	12.836	1996	13.314
1877	12.957	1917	12.584	1957	13.139		
1878	13.092	1918	12.7	1958	13.208		
1879	12.796	1919	12.792	1959	13.133		
1880	12.811	1920	12.857	1960	13.094		
1881	12.845	1921	12.902	1961	13.124		
1882	12.864	1922	12.787	1962	13.129		
1883	12.783	1923	12.821	1963	13.16		
1884	12.73	1924	12.764	1964	12.868		
1885	12.754	1925	12.868	1965	12.935		
1886	12.826	1926	13.014	1966	13.035		
1887	12.723	1927	12.904	1967	13.031		
1888	12.783	1928	12.871	1968	13.004		
1889	12.922	1929	12.718	1969	13.117		
1890	12.703	1930	12.964	1970	13.064		
1891	12.767	1931	13.041	1971	12.903		
1892	12.671	1932	12.992	1972	13.031		
1893	12.631	1933	12.857	1973	13.175		
1894	12.709	1934	12.982	1974	12.912		
1895	12.728	1935	12.943	1975	12.975		
1896	12.93	1936	12.993	1976	12.869		
1897	12.936	1937	13.092	1977	13.148		
1898	12.759	1938	13.187	1978	13.057		
<i>continued on next page</i>							

1899	12.874	1939	13.111	1979	13.154		
1900	12.959	1940	13.055	1980	13.195		

Table 4: Wêreldwye gemiddelde temperature van 1861 tot 1996. Tans is daar 'n groot bespreking oor die verandering in weerpatrone en die moontlike verwantskap met besoedeling en kweekhuisgasse.

3.5 Datastel 5: Prys van Petrol

Die prys van petrol in Suid-Afrika vanaf Augustus 1998 tot Julie 2000 word in Table 5 getoon.

Datum	Prys (R/l)
Augustus 1998	R 2.37
September 1998	R 2.38
Oktober 1998	R 2.35
November 1998	R 2.29
Desember 1998	R 2.31
Januarie 1999	R 2.25
Februarie 1999	R 2.22
Maart 1999	R 2.25
April 1999	R 2.31
Mei 1999	R 2.49
Junie 1999	R 2.61
Julie 1999	R 2.61
Augustus 1999	R 2.62
September 1999	R 2.75
Oktober 1999	R 2.81
November 1999	R 2.86
Desember 1999	R 2.85
Januarie 2000	R 2.86
Februarie 2000	R 2.81
Maart 2000	R 2.89
April 2000	R 3.03
Mei 2000	R 3.18
Junie 2000	R 3.22
Julie 2000	R 3.36

Table 5: Petrolpryse in Suid-Afrika vanaf Augustus 1998 tot Julie 2000

4 Groepering van Data

Een van die eerste stappe in die verwerking van 'n groot stel rou data is om die datawaardes te rangskik in 'n kleiner aantal groepe en dan te tel hoeveel daar van elke datawaarde in elke groep is. Die groepe is gewoonlik gebaseer op een of ander interval van datawaardes, sodat datawaardes wat binne 'n sekere interval val, saamgegroeper word. Die gegroepeerde data word dikwels grafies of in 'n frekwensietabel uitgebeeld. Frekwensie beteken "hoeveel keer kom iets voor".

Exercise 1: Groepering van Data

(Solution on p. 11.)

Groepeer die elemente van Datastel 1 (Table 1) om te bepaal hoeveel keer die muntstuk op kop land en hoeveel keer die muntstuk op stert land.

4.1 Oefeninge: Groepering van Data

- Die lengtes van 30 leerders word hier aangetoon. Groepeer die data in die gegewe tabel (onder). (Telmerke is 'n gerieflike manier om in 5'e te tel. Ons gebruik die vierstreep-hek metode: 1111 om 4 aan te dui en 1111 met 'n horisontale streep deur die 4 vertikale strepies om 5 aan te dui.)

142	163	169	132	139	140	152	168	139	150
161	132	162	172	146	152	150	132	157	133
141	170	156	155	169	138	142	160	164	168

Table 6

Groep	Telmerke	Frekwensie
$130 \leq h < 140$		
$140 \leq h < 150$		
$150 \leq h < 160$		
$160 \leq h < 170$		
$170 \leq h < 180$		

Table 7

Klik hier vir die oplossing.²

- 'n Eksperiment is uitgevoer in 'n klas en 50 leerders is gevra om te raai hoeveel lekkertjies daar in 'n gegewe fles is. Die volgende raaiskote is opgeneem:

56	49	40	11	33	33	37	29	30	59
21	16	38	44	38	52	22	24	30	34
42	15	48	33	51	44	33	17	19	44
47	23	27	47	13	25	53	57	28	23
36	35	40	23	45	39	32	58	22	40

Table 8

²<http://www.fhsst.org/lab>

Trek 'n gegroepeerde frekwensietabel op vir die intervale 11 tot 20, 21 tot 30, 31 tot 40, ens.
Kliek hier vir die oplossing.³

³<http://www.fhsst.org/laj>

Solutions to Exercises in this Module

Solution to Exercise (p. 9)

Step 1. Daar is twee unieke datawaardes: K en S. Daarom is daar twee groepe: een vir die K-datawaardes en nog een vir die S-datawaardes.

Step 2.

Datawaardes	Frekwensie
K	44
S	56

Table 9: Frekwensie van datawaardes in Datastel 1

Step 3. Daar is 100 datawaardes en die total van die frekwensiekolum is $44 + 56 = 100$.