

1.3 Eigenschappen van variabelen

Op basis van hoe ze gemeten zijn, kan je verschillende soorten variabelen onderscheiden. In de volgende subparagrafen gaan we hier dieper op in.

1.3.1 Schaalfamilies

Getallen kunnen een verschillende betekenis hebben in verschillende situaties. Het getal 12 bijvoorbeeld: het kan een leeftijd voorstellen van een kind dat 12 jaar oud is, het rugnummer van een voetballer, de twaalfde plaats in de marathon van Londen, de temperatuur in graden Celsius, etc. Afhankelijk van de context verschilt de betekenis van het getal 12. De informatie die het getal 12 bevat, hangt af van de *meetschaal* die werd gebruikt om het getal te bekomen. In deze cursus onderscheiden we vier meetschalen: nominaal, ordinaal, interval en ratio^j. Om de informatie in data ten volle te kunnen benutten, moeten we het verschil tussen deze meetschalen goed begrijpen. De wiskundige bewerkingen die we kunnen toepassen op data, zullen ook afhangen van de meetschaal.

Nominale schaal

Bij de nominale schaal worden de waarden van de variabele gebruikt voor identificatie zonder dat ze een hoeveelheid aanduiden.

Enkele voorbeelden:

- *Variabele*: Geslacht. *Waarden*: man, vrouw of andere.
- *Variabele*: Land van herkomst. *Waarden*: België, Frankrijk, Spanje, etc.
- *Variabele*: Politieke voorkeur in de VS. *Waarden*: Democraat, Republikein of Andere.

Bij deze voorbeelden is het duidelijk dat de waarde van de variabele geen numerieke betekenis heeft, het zijn zelfs geen getallen (man, vrouw, België, etc.). Er bestaan

^jNiettegenstaande deze vier meetschalen algemeen worden aanvaard, staan ze soms toch ter discussie. Er zijn ook onderzoekers die de meetschalen onderverdelen in bijvoorbeeld 5 of 10 schalen. In ‘Statistiek II’ wordt hier dieper op ingegaan.

echter ook nominale variabelen die numerieke waarden aannemen zonder dat het getal zelf een specifieke betekenis heeft (het wordt louter gebruikt ter identificatie).

Enkele voorbeelden:

- *Variabele*: Rekeningnummer. *Waarden*: 37 0000 0000 2828 (Oxfam), 73 0000 0000 6060 (Artsen Zonder Grenzen), etc.
- *Variabele*: Rugnummer in het voetbal. *Waarden*: 1, 2, 10, 31, etc.
- *Variabele*: Het nummer van de tram in Gent. *Waarden*: 1, 2 en 4.

Het kan ook zijn dat een onderzoeker de waarden van een nominale variabele, zoals geslacht, numeriek codeert. Een 1 kan dan bijvoorbeeld overeenkomen met de mannen, een 2 met de vrouwen en een 3 met personen van het derde geslacht. Deze getallen hebben op zich geen betekenis, maar ze kunnen de verwerking van de gegevens eenvoudiger maken. Het coderen van variabelen wijzigt de schaal niet: als geslacht de waarden ‘man’, ‘vrouw’ of ‘andere’ aanneemt of ‘1’, ‘2’ of ‘3’, het blijft een variabele gemeten op nominale schaal. In plaats van ‘1’, ‘2’ of ‘3’ kunnen ook andere getallen gekozen worden. Het maakt niet uit welke numerieke waarden er worden gebruikt. Bij een nominale schaal worden dus alleen nummers toegekend aan de waarden van een variabele om ze van elkaar te kunnen onderscheiden.

Variabelen gemeten op nominale schaal worden ook nominale variabelen genoemd. Nominaal drukt uit dat de waarden van de variabelen slechts ‘namen’ zijn.

Ordinale schaal

De ordinale schaal erft alle eigenschappen van de nominale schaal samen met een extra eigenschap: de waarden duiden een volgorde aan. Behalve om een volgorde weer te geven, is de waarde van de variabele niet van belang.

Enkele voorbeelden:

- *Variabele*: Uitslag wedstrijd. *Waarden*: goud, zilver of brons.
- *Variabele*: Officiersgraad. *Waarden*: Onderluitenant, Luitenant, Kapitein, Kapitein-commandant, etc.
- *Variabele*: Mate van instemming met een bepaalde uitspraak. *Waarden*: volledig mee oneens, mee oneens, neutraal, mee eens, volledig mee eens.

Net zoals bij de nominale schaal kunnen we de waarden van de variabelen vervangen door getallen. Bij de uitslag van een wedstrijd bijvoorbeeld: 1 = goud, 2 = zilver en 3 = brons. Een kleinere waarde wil zeggen dat de atleet beter is. Het verschil tussen de getallen speelt echter geen rol: het verschil in prestatie tussen de eerste en tweede is niet noodzakelijk gelijk aan het verschil in prestatie tussen de tweede en derde.

Een andere keuze van numerieke waarden is: 1 = goud, 10 = zilver en 100 = brons. Nog steeds drukt een kleiner getal een beter resultaat uit. Opnieuw kunnen dus verschillende coderingen gebruikt worden, maar met de restrictie dat de codering de volgorde moet behouden. Een voorbeeld van een incorrecte codering zou zijn: 1 = goud, 3 = zilver en 2 = brons.

Variabelen gemeten op ordinale schaal worden ook ordinale variabelen genoemd. Ordinaal drukt uit dat de waarden van de variabele geordend kunnen worden. Bij een nominale variabele is dit niet het geval: we kunnen niet zeggen dat voetballers met een kleiner rugnummer slechter zijn dan die met een groter nummer (of omgekeerd).

Intervalschaal

De intervalschaal erft alle eigenschappen van de ordinale schaal met de extra eigenschap dat verschillen tussen waarden een betekenis hebben. Er is echter geen absoluut nulpunt^k.

Enkele voorbeelden:

- *Variabele*: Temperatuur in graden Celsius. *Waarden*: 0, 10, -30, etc.
- *Variabele*: IQ^l. *Waarden*: 96, 100, 130, etc.

Niettegenstaande de temperatuur de waarde 0 kan aannemen, is het geen absoluut nulpunt omdat 0 °C niet willen zeggen dat er geen temperatuur aanwezig is. Het is de aanduiding van een bepaalde temperatuur^m. Een andere manier om te zien dat het nulpunt bij temperatuur niet absoluut is, is door een andere meeteenheid te gebruiken. Nul graden Celsius komt overeen met 32 graden Fahrenheit. Dit geeft aan dat het nulpunt hier relatief is: ze hangt af van de gebruikte meeteenheid (Celsius of Fahrenheit).

^kMet absoluut nulpunt bedoelen we een getal dat aangeeft dat niets van de variabele aanwezig is.

^lHet IQ is een variabele waarover nog steeds discussie bestaat tot welke klasse ze behoort. Sommige wetenschappers argumenteren dat ze ordinaal is. We gaan hier echter niet dieper op in en zullen, zoals de meeste handboeken, veronderstellen dat het IQ gemeten is op intervalschaal.

^mNul graden Celsius duidt de temperatuur aan waarbij water bevriest bij een luchtdruk van 1 bar.

Variabelen gemeten op intervalschaal worden ook intervalvariabelen genoemd. De benaming interval drukt uit dat gelijke verschillen op de meetschaal (intervallen) duiden op gelijke verschillen in de variabele. Een stijging van $10\text{ }^{\circ}\text{C}$ naar $20\text{ }^{\circ}\text{C}$ is evenveel als een stijging van $20\text{ }^{\circ}\text{C}$ naar $30\text{ }^{\circ}\text{C}$ in termen van het warmteverschil. Dit blijft behouden bij omzetting naar Fahrenheit ($10\text{ }^{\circ}\text{C} = 50\text{ }^{\circ}\text{F}$, $20\text{ }^{\circ}\text{C} = 68\text{ }^{\circ}\text{F}$ en $30\text{ }^{\circ}\text{C} = 86\text{ }^{\circ}\text{F}$): het verschil tussen $68\text{ }^{\circ}\text{F}$ en $50\text{ }^{\circ}\text{F}$ is gelijk aan het verschil tussen $86\text{ }^{\circ}\text{F}$ en $68\text{ }^{\circ}\text{F}$. Bij een ordinale variabele is dit niet het geval: bv. het verschil in uitslag tussen de eerste (gouden medaille) en de tweede (zilveren medaille) hoeft niet gelijk te zijn aan het verschil in uitslag tussen de tweede en de derde (bronzen medaille).

Ratioschaal

De ratioschaal erft alle eigenschappen van de intervalschaal én heeft daarbij ook een absoluut nulpunt.

Enkele voorbeelden:

- *Variabele*: Lengte in cm. *Waarden*: 0, 1, 354, etc.
- *Variabele*: Geldbedrag in euro. *Waarden*: 0, 5, 2400, etc.
- *Variabele*: Reactietijd tussen stimulus en gedrag in seconden. *Waarden*: 0, 5, 2, 58, etc.

Hier is het nulpunt absoluut: als je 0 euro hebt, wil dit zeggen dat je geen geld hebt. Een lengte van 0 centimeter blijft 0, ook als je de schaal omzet naar meter. Een reactietijd van 0 seconden wil zeggen dat er geen tijd zit tussen stimulus en gedrag.

Variabelen gemeten op ratioschaal worden ook ratiovariabelen genoemd. De benaming ratio drukt uit dat verhoudingen (ratio's) een betekenis hebben: 10 euro is bijvoorbeeld dubbel zoveel als 5 euro. Bij intervalvariabelen zijn verhoudingen niet zinnig: een temperatuur van $10\text{ }^{\circ}\text{C}$ is niet dubbel zo warm $5\text{ }^{\circ}\text{C}$ omdat deze uitspraak bij een omzetting naar Fahrenheit ($5\text{ }^{\circ}\text{C} = 41\text{ }^{\circ}\text{F}$ en $10\text{ }^{\circ}\text{C} = 50\text{ }^{\circ}\text{F}$) niet meer opgaat: 50 is niet dubbel zoveel als 41.

Voor dit vak (en in de praktijk) is het vooral van belang om het onderscheid te kennen tussen de nominale, ordinale en interval/ratioschaal. Het onderscheid tussen de interval- en ratioschaal is minder van belang.

Als we terugkeren naar de verschillende situaties van het nummer 12, kunnen we ze als volgt indelen in de verschillende schaaftamilies:

- Nominaal: rugnummer van een voetballer.
- Ordinaal: de twaalfde plaats in de marathon van Londen.
- Interval: de temperatuur in graden Celsius.
- Ratio: de leeftijd van een kind dat 12 jaar oud is.

Niettegenstaande het getal 12 steeds dezelfde blijft, verschilt de informatie die het getal weergeeft van situatie tot situatie.

1.3.2 Discrete en continue variabelen

Naast de vier schaaftamilies kunnen we ook een onderscheid maken tussen continue en discrete variabelen.

Continue variabelen kunnen tussenwaarden aannemen. Met tussenwaarde bedoelen we dat er tussen elke twee willekeurige waarden een derde waarde ligt. Dit impliceert dat er tussen twee waarden (theoretisch gezien) oneindig veel andere waarden kunnen liggen. We zeggen ook dat variabelen gemeten op continue schaal continu variëren.

Enkele voorbeelden:

- Lengte in cm: tussen 2 en 3 cm liggen nog vele andere waarden: vb. 2.5 cm, 2.34 cm, 2.323548 cm, etcⁿ.
- Temperatuur in °C: tussen twee willekeurige temperaturen ligt altijd een derde. Tussen 25.8 °C en 25.9 °C ligt bijvoorbeeld 25.85 °C, 25.82147 °C, etc.
- De tijd in seconden: tussen twee willekeurige tijdstippen ligt altijd een derde tijdstip.

Bij *discrete variabelen* bestaan steeds twee waarden waar geen derde waarde kan tussen liggen. Dit impliceert dat de variabele maar een eindig aantal waarden kan aannemen^o.

Enkele voorbeelden:

ⁿVoor kommagetallen wordt een punt gebruikt, dus 2.5 lees je als 2 komma 5.

^oStrikt genomen kan het aantal ook aftelbaar oneindig zijn, maar dit behoort niet tot de inhoud van deze cursus.