3 Verwachtingswaarde, variantie, standaard deviatie en covariantie

Het populatiegemiddelde μ geeft het gewogen gemiddelde aan van alle waarden in een populatie. De variantie σ^2 geeft de spreiding aan van de waarden binnen de populatie. Hieronder zijn de formules voor beide termen te vinden:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Mocht men bijvoorbeeld een onderzoek doen naar het salaris van docenten in Nederland dan zou x_1 het salaris zijn van docent 1, x_2 het salaris van docent 2, enzovoorts. In dat geval is N het aantal docenten waaruit de populatie uit bestaat.

Het populatiegemiddelde noemt men ook wel de verwachtingswaarde van *X* en noteert men als *E(X)*. De formule voor de verwachtingswaarde van *X* is

$$E(X) = \mu = \sum_{allex} xp(x)$$

Op dezelfde manier berekent men de populatievariantie *V(X)* wat het gewogen gemiddelde is van de gekwadrateerde standaarddeviaties van het gemiddelde.

$$V(X) = \sigma^2 = \sum_{allex} (x - \mu)^2 p(x)$$

Deze formule kan enigszins vergemakkelijkt worden:

$$V(X) = \sigma^2 = \sum_{allex} x^2 p(x) - \mu^2$$

De standaarddeviatie van de populatie krijgt men simpel door de wortel te trekken van de populatievariantie:

$$\sigma = \sqrt{\sigma^2}$$

3.1 Regels voor verwachtingswaarde

Er zijn drie regels van toepassing op de verwachtingswaarde:

1. $E(c) = c$
2. $E(X + c) = E(X) + c$
3. $E(cX) = cE(X)$

Waarbij *X* de stochastische variabele is en *c* een constante.

3.2 Regels voor variantie

Tevens zijn er drie regels van toepassing op de variantie:

1. $V(c) = 0$
2. $V(X + c) = V(X)$
3. $V(cX) = c^2 V(X)$

Voorbeeld 3[a]

Een boekwinkel op een Universiteit heeft een maandelijkse omzet met een gemiddelde van €15.000 en een standaarddeviatie van €1.200. Winst wordt berekend door de omzet te vermenigvuldigen met 25% en te verminderen met €3.000 vaste kosten. Wat is het gemiddelde en de standaarddeviatie van de winst?

Uit het bovenstaande kan afgeleid worden dat

$$Winst = 0{,}25(Omzet) - 3000$$

De verwachte of gemiddelde winst is

$$E(Winst) = E[0{,}25(Omzet) - 3000]$$

Als men de tweede regel van de verwachtingswaarde toepast, krijgt men

$$E(Winst) = E[0,25(Omzet)] - 3000$$

Past men de derde regel van de verwachtingswaarde toe, dan

$$E(Winst) = 0,25E(Omzet) - 3000 = 0,25(15000) - 3000 = 750$$

De gemiddelde maandelijkse winst is dus €750.

De variantie is

$$V(Winst) = V[0,25(Omzet) - 3000]$$

Als men de tweede regel van de variantie toepast, krijgt men

$$V(Winst) = V[0,25(Omzet)]$$

En de derde regel van variantie houdt in dat

$$V(Winst) = (0,25)^2 V(Omzet) = 0,0625(1200)^2 = 90000$$

De standaarddeviatie van de winst is dan

$$\sigma_{winst} = \sqrt{90000} = 300$$

3.3 Covariantie

Indien twee stochastische variabelen een lineaire relatie met elkaar hebben zijn er twee begrippen van belang, de covariantie en de correlatiecoëfficiënt. De covariantie *COV* geeft informatie over de richting en de grootte van de lineaire relatie tussen de variabelen. De formule voor de covariante tussen variabelen X en *Y* is als volgt:

$$COV(X,Y) = \sum_{allex} \sum_{alley} (x - \mu_x)(y - \mu_y) p(x,y)$$

Wat opvalt is dat men de deviaties van het gemiddelde van zowel *X* als *Y* vermenigvuldigen en dan vermenigvuldigen met de gemeenschappelijke kans. Daardoor kan men de formule ook herschrijven als:

$$COV(X,Y) = \sum_{allex} \sum_{alley} xyp(x,y) - \mu_x \mu_y$$

De correlatiecoëfficiënt ρ geeft de daadwerkelijk sterkte en grootte weer van de lineaire relatie. De formule voor de correlatiecoëfficiënt is

$$\rho = \frac{COV(X,Y)}{\sigma_x \sigma_y}$$

Als de correlatiecoëfficiënt tussen -1 en 0 ligt is er sprake van een negatieve lineaire relatie. Omgekeerd geldt dat als de correlatiecoëfficiënt tussen de 0 en 1 ligt er een positieve lineaire relatie geconstateerd kan worden. Is de correlatiecoëfficiënt gelijk aan 0 dan is er geen lineaire relatie.