Kenmerken van populatieverdelingen

<table>
<tr><th>Discrete veranderlijken / verdelingen</th><th>Continue veranderlijken / verdelingen</th></tr>
<tr><td>

Verwachtingswaarde:

$$\boxed{\mu_X = E[X] = \sum_{\text{alle } m_i} m_i p_X(m_i) = \sum_{i=1}^{k} m_i p_X(m_i)}$$

verwachtingswaarde: vermenigvuldig elke waarde met de kans dat die waarde optreedt en sommeer over alle mogelijke waarden.

$$E[g(X)] = \sum_{\text{alle } m_i} g(m_i) p_X(m_i) = \sum_{i=1}^{k} g(m_i) p_X(m_i)$$

Verwachtingswaarde van een functie van X: vermenigvuldig de functiewaarde voor elke waarde met de kans dat die waarde optreedt en sommeer over alle mogelijke waarden.

</td><td>

Verwachtingswaarde:

$$\boxed{\mu_X = E[X] = \int_{-\infty}^{+\infty} x\, f_X(x) dx}$$

verwachtingswaarde: vermenigvuldig elke waarde met de kansdichtheid in die waarde en integreer over alle mogelijke waarden.

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)\, f_X(x) dx$$

Verwachtingswaarde van een functie van X: vermenigvuldig de functiewaarde voor elke waarde met de kansdichtheid in die waarde en integreer over alle mogelijke waarden.

</td></tr>
<tr><td colspan="2">

<u>Eigenschappen</u>:

$$\boxed{E[aX+b] = aE[X]+b}$$

De verwachtingswaarde van een lineaire combinatie van de veranderlijke X is dezelfde lineaire combinatie van de verwachtingswaarde van de veranderlijke X.

<u>Gevolgen</u>:

$$E[b] = b$$

Een constante kan maar 1 waarde aannemen en dus is de verwachtingswaarde de constante zelf.

$$E[X - E[X]] = E[X - \mu_X] = 0$$

De verwachtingswaarde van de afwijking van de veranderlijke X t.o.v. zijn verwachtingswaarde is steeds gelijk aan 0. De veranderlijke $Y = X - E\left[X\right]$ is dus een symmetriesering/centralisering van X.

<u>Symmetrische veranderlijke</u>: De verwachtingswaarde van een veranderlijke die symmetrisch is t.o.v. de rechte $x = a$ is gelijk aan de symmetriewaarde a. (Kijk <u>hier</u> voor het bewijs)

</td></tr>
<tr><td colspan="2">

Variantie- standaardafwijking:

$$\boxed{\sigma_X^2 = Var(X) = E[(X - E[X])^2]}$$

De variantie is de verwachtingswaarde van de gekwadrateerde afwijking van de veranderlijke t.o.v. zijn verwachtingswaarde. Deze grootheid geeft een aanduiding van de spreiding van de waarden die de veranderlijke X kan aannemen.

$$\boxed{\sigma_X = \sqrt{Var(X)}}$$

De standaardafwijking is de vierkantswortel van de variantie. De standaardafwijking geeft ook een aanduiding van de spreiding van de waarden die de veranderlijke kan aannemen maar bovendien heeft ze dezelfde eenheid als de veranderlijke zelf en kan bijgevolg op eenzelfde as als X voorgesteld worden.

</td></tr>
<tr><td>

$$Var(X) = \sum_{i=1}^{k} (m_i - E[X])^2 p_X(m_i)$$

</td><td>

$$Var(X) = \int_{-\infty}^{+\infty} (x - E[X])^2 f_X(x) dx$$

</td></tr>
<tr><td colspan="2">

<u>Praktijkformule</u>:

$$\boxed{Var(X) = E[X^2] - E[X]^2}$$

</td></tr>
</table>

Eigenschappen:

$$\boxed{Var(aX+b)=a^2Var(X)}$$

Een verschuiving heeft geen invloed op de variantie en een factor komt gekwadrateerd voorop.

$$\boxed{\sigma_{aX+b}=|a|\sigma_X}$$

Een verschuiving heeft geen invloed op de standaardafwijking en een factor komt in absolute waarde voorop.

$$Var(b)=0,\sigma_b=0$$

Een constante kan maar 1 waarde aannemen en dus is er geen spreiding!

Standaardiseren: voor de praktijk is het soms nuttig om over te gaan op de gestandaardiseerde veranderlijke Z geassocieerd aan de veranderlijke X: $\boxed{Z=\frac{X-\mu_X}{\sigma_X}}$

Kenmerk: $\boxed{E[Z]=0 \quad , \quad Var(Z)=1}$

Ongelijkheid van Markov:

Voor een positieve kansveranderlijke X geldt:

$$\boxed{P(X\geq a)\leq\frac{E(X)}{a},\forall a>0}$$

De kans dat de veranderlijke een waarde aanneemt die groter of gelijk is aan het positieve getal *a* is naar boven begrensd door de verhouding $\frac{E(X)}{a}$. Het percentage van de waarden van X die groter of gelijk zijn aan *a* is dus kleiner of gelijk aan deze verhouding.

Bvb.: als de verhouding $\frac{E(X)}{a}$ gelijk is aan 0.25 dan zegt de ongelijkheid van Markov dat zeker niet meer dan 25% van de waarden die X kan aannemen groter of gelijk zijn aan het getal a, of m.a.w. dat minstens 75% van de waarden gelegen zijn tussen 0 en a.

Hoe groter a hoe kleiner deze kans is en hoe kleiner ook deze verhouding is.

Zonder de verdeling van X te kennen kan je via deze bovengrens dus een schatting geven van dit percentage.

Ongelijkheid van Chebyshev:

Voor willekeurige kansveranderlijke X geldt:

$$\boxed{P(|X-E(X)|\geq t)\leq\frac{Var(X)}{t^2},\forall t>0}$$

De kans dat de veranderlijke een waarde aanneemt die verder dan t gelegen is van *E(X)* is naar boven begrensd door de verhouding $\frac{Var(X)}{t^2}$. Het percentage van de waarden van X die ofwel groter of gelijk zijn aan $E(X)+t$, ofwel kleiner of gelijk aan $E(X)-t$ is dus kleiner of gelijk aan deze verhouding.

Bvb.: als de verhouding $\frac{Var(X)}{t^2}$ gelijk is aan 0.25 dan zegt de ongelijkheid van Chebyshev dat zeker niet meer dan 25% van de waarden die X kan aannemen rechts van $E(X)+t$ of links van $E(X)-t$ gelegen zijn of m.a.w. dat minstens 75% van de waarden gelegen zijn tussen $E(X)-t$ en $E(X)+t$.

Hoe groter t hoe kleiner de kans is en hoe kleiner ook deze verhouding is.

Zonder de verdeling van X te kennen kan je via deze bovengrens dus een schatting geven van dit percentage .

<table>
<tr>
<td>Kwantielen:
Het kwantiel van orde p is de kleinste waarde $x = Q_X(p)$ waarvoor $P(X \le x) = F_X(x) \ge p$
Het is m.a.w. de waarde van X waarvoor de cumulatieve verdelingsfunctie verspringt van minder dan p naar p of meer.</td>
<td>Kwantielen:
Het kwantiel van orde p is de waarde $x = Q_X(p)$ waarvoor $P(X \le x) = F_X(x) = p$
Of: $\int_{-\infty}^{x} f_X(t)dt = p$
Het is m.a.w. de waarde van X waarvoor de cumulatieve verdelingsfunctie gelijk is aan p.</td>
</tr>
<tr>
<td colspan="2">Eerste kwartiel = kwantiel van orde 0.25
Mediaan = Kwantiel van orde 0.5
Tweede kwartiel = kwantiel van orde 0.75</td>
</tr>
<tr>
<td>Modus
De modi zijn de lokale maxima van de kansfunctie p_X. Het zijn de mogelijke waarden m_i waarvoor de kans niet kleiner is dan in de waarden er onmiddellijk links of rechts van.</td>
<td>Modus
De modi zijn de lokale maxima van kansdichtheid functie f_X.</td>
</tr>
<tr>
<td colspan="2">Scheefheid
De scheefheid γ_X van een kansveranderlijke X is het volgende getal:
$$\gamma_X = \frac{E\left[(X-\mu_X)^3\right]}{\sigma_X^3}$$
<u>Scheefheid en een verschuiving</u>:
$$\gamma_{X+a} = \gamma_X$$
Een verschuiving verandert de scheefheid niet!

<u>Scheefheid van een symmetrische veranderlijke:</u>
Als X symmetrisch is t.o.v. $x = a$ dan is $\gamma_X = 0$
Een symmetrische veranderlijke heeft scheefheid 0. <u>Ziehier</u> voor het bewijs.

<u>Betekenis van scheefheid</u>:
De scheefheid geeft dus de afwijking aan t.o.v. een symmetrische verdeling. Liggen meer waarden links van μ_X dan is de veranderlijke links-scheef ($\gamma_X < 0$), in het andere geval rechts-scheef ($\gamma_X > 0$).</td>
</tr>
<tr>
<td colspan="2">Kurtosis
De kurtosis κ_X van een kansveranderlijke X is het volgende getal:
$$\kappa_X = \frac{E\left[(X-\mu_X)^4\right]}{\sigma_X^4}$$
De kurtosis geeft aan hoe dik de staarten van de verdeling zijn. (De kurtosis van de normale verdeling (zie verder) is 3)</td>
</tr>
</table>