Formuleblad Statistiek deel 2 (2024-2025)

Deel 1

Steekproefgemiddelde (gegeven een steekproef met n uitkomsten $x_1, x_2, \ldots, x_n$)

$$\overline{x} = \frac{\sum\limits_i x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Steekproefvariantie:

$$s^2 = \frac{\sum\limits_i (x_i - \overline{x})^2}{n-1} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1} \qquad \text{(optie 1)}$$

(1)

Rekenregels kansrekening:

$$P(A \text{ of } B) = P(A) + P(B) - P(A \text{ en } B) \qquad \text{(optelregel)}$$

$$P(B) = 1 - P(\text{niet } B) \qquad \text{(complementregel)}$$

$$P(A \mid B) = \frac{P(A \text{ en } B)}{P(B)} \qquad \text{(conditionele kansen)}$$

Discrete en continue kansverdelingen:

	Discrete kansvariabelen	**Continue kansvariabelen**
Uitkomstenruimte:	Eindig / aftelbaar oneindig	Overaftelbaar oneindig
Toepassingen:	Tellen / categoriseren	Meten
Kansbegrip:	Kansfunctie $p(k) = P(X = k)$	Kansdichtheidsfunctie $f(x)$
CDF:	$F(k) = P(X \leq k) = \sum_{\ell : \ell \leq k} p(\ell)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$
Verwachtingswaarde:	$E[X] = \sum_k k \cdot P(X = k)$	$E[X] = \int x \cdot f(x)\, dx$
Variantie:	$\text{Var}(X) = \sum_k (k - E[X])^2 \cdot P(X = k)$	$\text{Var}(X) = \int (x - E[X])^2 \cdot f(x)\, dx$
Standaardafwijking:	$\sigma(X) = \sqrt{\text{Var}(X)}$	$\sigma(X) = \sqrt{\text{Var}(X)}$

Speciale kansverdelingen:

- $X \sim \text{Binomiaal}(n, p)$: tellen van aantal successen bij onafhankelijke kansexperimenten met twee uitkomsten (Bernoulli-experimenten): succes / mislukking.
  - n: aantal Bernoulli-experimenten
  - p: succeskans per experiment
- $X \sim \text{Poisson}(\lambda \cdot t)$: tellen van aantal “gebeurtenissen” in een “interval” van tijd / ruimte.
  - λ: gemiddeld aantal gebeurtenissen per eenheid van tijd / ruimte.
  - t: aantal eenheden van tijd / ruimte van het interval $\rightarrow$ **Voorbeeld:** als “dag” de tijdseenheid is, dan bestaat “week” uit $t = 7$ tijdseenheden.
- $T \sim \text{Exponentieel}(\lambda)$: meten van de tijd / ruimte tot de volgende gebeurtenis.
  - λ: gemiddeld aantal gebeurtenissen per eenheid van tijd / ruimte.

Verwachtingswaarde en variantie van veelgebruikte kansverdelingen:

Verdeling	Kans(dichtheids)functie	CDF	$E(X)$	$\text{Var}(X)$
Discreet				
Uniform(a,b)	$p(k) = \frac{1}{b-a+1}$ $(k = a, a+1, \ldots, b)$	$F(k) = \begin{cases} 0 & x < a \\ \frac{k-a+1}{b-a+1} & a \leq k < b \\ 1 & k \geq b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Binomiaal(n,p)	$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$	$F(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$	np	$np(1-p)$
Poisson(λ)	$p(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$	$F(k) = \sum_{i=0}^{k} e^{-\lambda} \cdot \frac{\lambda^i}{i!}$	λ	λ
Continuous				
Uniform(a,b)	$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{elders.} \end{cases}$	$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponentieel(λ)	$f(x) = \lambda e^{-\lambda x}$, $x \geq 0$	$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Veelgebruikte functies op de grafische rekenmachine

Type vraag	TI-84 Plus	Casio
Continue kansverdeling (willekeurig)		
$P(a \leq X \leq b)$	$\int_a^b f(x)\,dx$	$\int_a^b f(x)\,dx$
$X \sim$ **Binomiaal**(n,p)		
$P(X = k)$ $P(X \leq k)$	binompdf(n,p,k) binomcdf(n,p,k)	BinomialPD(k,n,p) BinomialCD(k,n,p)
$X \sim N(\mu, \sigma)$		
$P(a \leq X \leq b)$ Grenswaarde g zodat $P(X \leq g) = p$?	normalcdf(a,b,μ,σ) invNorm(p,μ,σ)	NormalCD(a,b,σ,μ) InvNormCD(tail=left, p,σ,μ)
$X \sim$ **Poisson**(λ)		
$P(X = k)$ $P(X \leq k)$	poissonpdf(λ,k) poissoncdf(λ,k)	PoissonPD(k,λ) PoissonCD(k,λ)

z-score:

$$z = \frac{x - \mu}{\sigma}$$

Centrale limietstelling: Gegeven n kansvariabelen $X_1, X_2, \ldots, X_n$ die onderling onafhankelijk zijn en dezelfde kansverdeling hebben met een verwachtingswaarde μ en standaardafwijking σ, dan geldt (bij benadering) dat

- de som $\sum X = X_1 + X_2 + \ldots + X_n$ normaal verdeeld is met $E[\sum X] = n \cdot \mu$ en $\sigma(\overline{X}) = \sqrt{n} \cdot \sigma$.
- het gemiddelde $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ normaal verdeeld is met $E[\overline{X}] = \mu$ en $\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.

Deel 2:

$100 \cdot (1-\alpha)\%$**-betrouwbaarheidsinterval (BI) voor** μ **(wanneer** σ **bekend is):**

$$z_{\alpha/2} = \mathrm{InvNorm}(opp = 1 - \alpha/2; \mu = 0; \sigma = 1)$$

$$[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}]$$

Minimale steekproefomvang zodat het $100 \cdot (1-\alpha)\%$**-BI voor** μ **maximale afwijking** a **heeft:**

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{a}\right)^2$$

NB: soms kan in plaats van de afwijking a een gewenste intervalbreedte gegeven zijn, deze is gelijk aan $2a$.

$100 \cdot (1-\alpha)\%$**-betrouwbaarheidsinterval (BI) voor** μ **(wanneer** σ **NIET bekend is):**

$$t = \mathrm{InvT}(opp = 1 - \alpha/2; \mu = 0; \sigma = 1)$$

$$[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}]$$

Minimale steekproefomvang zodat het $100 \cdot (1-\alpha)\%$**-BI voor** μ **maximale afwijking** a **heeft:**

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{a}\right)^2$$

NB: soms kan in plaats van de afwijking a een gewenste intervalbreedte gegeven zijn, deze is gelijk aan $2a$.