Combinatoriek

Volgorde van belang?	herhaling toegestaan?
✓	✗
✓	✗
✗	✗
✓	✓
✗	✓

Variatie: $V_n^p = \frac{n!}{(n-p)!}$
↳ ketting met parels

Permutatie: $P_n = V_n^n = n!$
↳ n elementen ordenen

Combinatie: $C_n^p = \binom{n}{p} = \frac{n!}{p!(n-p)!}$
↳ "p kiezen uit n"

Herhalingsvariatie: $\overline{V}_n^p = n^p$
↳ nummerplaten

Herhalingscombinatie: $\overline{C}_n^p = \binom{n+p-1}{p}$
↳ "stipjes & streepjes"

Herhalingspermutatie: $\overline{P}_n^{p,q,\dots,r} = (p\ q^n \dots\ r) = \frac{n!}{p!q!\dots r!}$
↳ anagrammen

Het aantal deelverzamelingen van een verzameling met n elementen = $2^n = \#D(A)$

Binomium van Newton

$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} \cdot b^i$

Multinomiale ontwikkeling

$(x_1 + x_2 + \dots + x_u)^n = \sum (n_1\ n_2^n \dots n_u)\ x_1^{n_1} x_2^{n_2} \dots x_u^{n_u} \quad (n_1 + n_2 + \dots + n_u = n)$

1 Kansruimten

Basisbegrippen

- universum Ω: alle mogelijke uitkomsten
- gebeurtenissen: deelverz. v.h. universum
- sigma-algebra's van deelverzamelingen:
  een klasse $\mathcal{A}$ van deelverz. van Ω heet een σ-algebra als het aan de volgende axioma's voldoet:
  1. $\Omega \in \mathcal{A}$
  2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
  3. $\forall n \in \mathbb{N}: A_n \in \mathcal{A} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$

  ↳ $(\Omega, \mathcal{A})$ is een meetbare ruimte
  - ↳ $\forall n \in \mathbb{N}: A_n \in \mathcal{A} \Rightarrow \bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$
  - ↳ $\forall A,B \in \mathcal{A}: A \Delta B = (A \cup B) \setminus (A \cap B) \in \mathcal{A}$
- discrete σ-algebra: machtsverz. $= D(\Omega) = P(\Omega) = 2^\Omega$
- $P: \mathcal{A} \to \mathbb{R}$ is een kansmaat indien:
  1. $P(\Omega) = 1$
  2. $\forall A \in \mathcal{A}: P(A) \geq 0$
  3. $\forall n \in \mathbb{N}: A_n \in \mathcal{A} \vee \forall i \neq j: A_i \cap A_j = \emptyset$ (paarsgewijs disjunct)
     $\Rightarrow P(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$
     - ↳ axioma van σ-additiviteit
- kansruimte $(\Omega, \mathcal{A}, P)$: Ω = universum, $\mathcal{A}$ = σ-algebra, P = kansmaat over Ω
- Stijgende/dalende rij van verzamelingen:
  $$\lim_{n\to\infty} A_n = \begin{cases} \bigcup_{n=1}^{\infty} A_n & \text{als } A_n \uparrow \\ \bigcap_{n=2}^{\infty} A_n & \text{als } A_n \downarrow \end{cases}$$

st. Zij $(\Omega, \mathcal{A}, P)$ een kansruimte:

1. Eindige additiviteit: Als $\{A_n \mid n \in \{1,\dots,N\}\}$ paarsgewijs disjunct en $\forall n \in \{1,\dots,N\}. A_n \in \mathcal{A}$, dan $P(\bigcup_{n=1}^{N} A_n) = \sum_{n=1}^{N} P(A_n)$
2. $\forall A \in \mathcal{A}: P(A^c) = 1 - P(A)$
3. Als $\forall n \in \mathbb{N}_0: A_n \in \mathcal{A}$ en (A_n) monotoon, dan $P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$

↳ $\forall A,B \in \mathcal{A}$:
- $P(B) = P(B \cap A) + P(B \cap A^c)$
- $A \subset B \Rightarrow P(A) \leq P(B)$
- $P(A) \leq 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

niet-aftelbaar universum

↳ gebeurtenissen in vorm $]-\infty, a]$

↳ σ-algebra voortgebracht door een collectie verzamelingen $\mathcal{C} \subset 2^\Omega$ = de kleinste σ-algebra die $\mathcal{C}$ bevat:
- $\sigma(\mathcal{C})$ is een σ-algebra
- $\mathcal{C} \subset \sigma(\mathcal{C})$
- $\forall$ σ-algebra $\mathcal{A}$ met $\mathcal{C} \subset \mathcal{A}$: $\sigma(\mathcal{C}) \subset \mathcal{A}$
  uniekheid van $\sigma(\mathcal{C})$

eig. Voor elke $\mathcal{C} \subset 2^\Omega$ bestaat $\sigma(\mathcal{C})$

bewijs p. 13

Voorwaardelijke kans en onafhankelijkheid

def. kans op A, gegeven B met $P(B) \neq 0$:
$P(A|B) = P(A \cap B)/P(B)$
↳ $P(A|A) = 1$, $P(A|\Omega) = P(A)$

st. Zij $(\Omega, \mathcal{A}, P)$ en $B \in \mathcal{A}$ met $P(B) > 0$, dan definieert $P_B : \mathcal{A} \to \mathbb{R} : A \mapsto P(A|B)$ een kansmaat op $\mathcal{A}$.

st. • kettingregel: Zij $(\Omega, \mathcal{A}, P)$, $(A_n)_{n=1}^{k} \in \mathcal{A}$ met $2 \leq k \in \mathbb{N}$, dan:
$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \dots P(A_k|A_1 \cap A_2 \cap \dots \cap A_{k-1})$
bewijs via inductie p. 18

St. • wet van de totale kans: Zij (Ω, A, P) en $\{A_n\}_{n \in \mathbb{N}}$ een partitie van Ω met alle $P(A_n) \neq 0$, dan geldt:
$\forall B \in \mathcal{A} : P(B) = \Sigma_{n \in \mathbb{N}} P(A_n) P(B|A_n)$
↳ een partitie is een opdeling van een verzameling in niet-lege, paarsgewijs disjuncte deelverzamelingen

St. • stelling van Bayes: zie voorwaarden vorige stelling
$P(A_n|B) = P(A_n) P(B|A_n) / \Sigma_{k \in \mathbb{N}} P(A_k) P(B|A_k)$

Onafhankelijkheid

def. $A, B \in \mathcal{A}$ zijn onafhankelijk $\Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

def. paarsgewijs onafhankelijk $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ $\forall i \neq j$

def. onderling onafhankelijk: $\forall m \in \{2, \dots, n\}$, $i_1 \neq i_2 \neq \dots \neq i_m$
$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_m})$
↳ aantal voorwaarden: paarsgewijze: $\binom{n}{2}$
onderling: $2^n - n - 1$

Systeembetrouwbaarheid

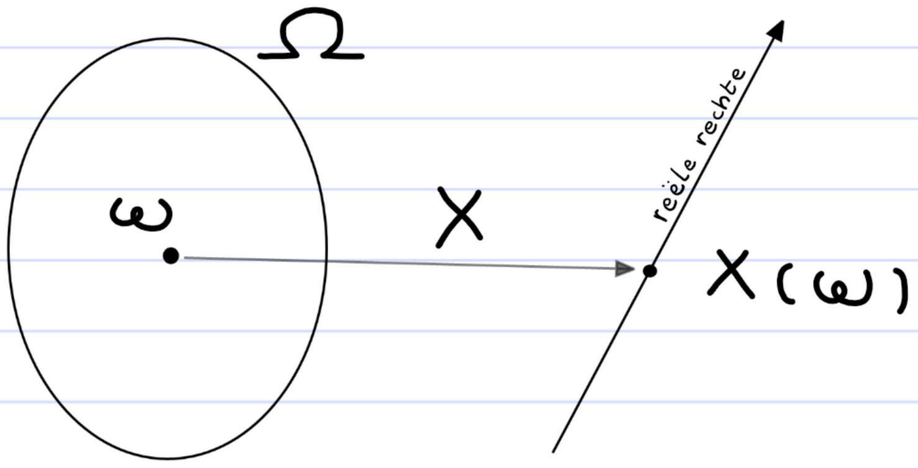
in serie: $P(S \text{ werkt}) = P(A \text{ werkt}) \cdot P(B \text{ werkt})$
in parallel: $P(S \text{ faalt}) = P(A \text{ faalt}) \cdot P(B \text{ faalt})$

② Stochastische verand.

Inleiding & definitie

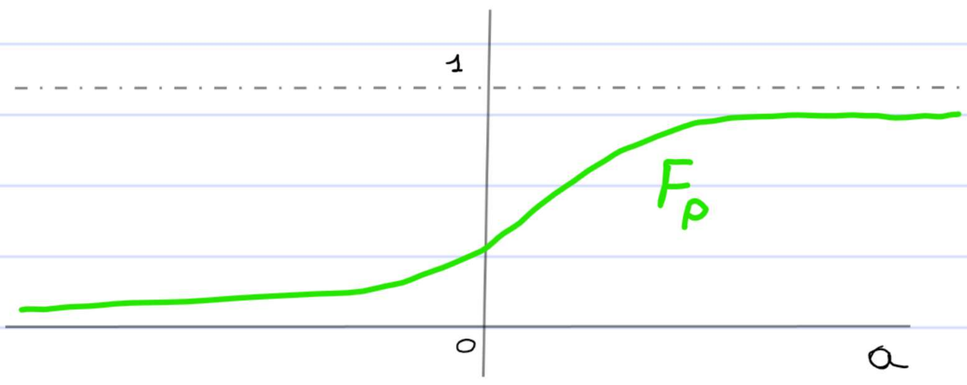
def. **stochastische veranderlijke**: een reële functie die aan elke uithomst $\omega \in \Omega$ van een kansruimte (Ω, A, P) een reëel getal $X(\omega)$ toekent met:
$\forall B \in \mathcal{B}(\mathbb{R}) : X^{-1}(B) = \{\omega \mid X(\omega) \in B \in A$
↳ A-meetbare afbeelding
$\mathcal{B}(\mathbb{R})$ is de σ-algebra voortgebracht door $\mathcal{C} = \{]-\infty, a] \mid -\infty < a < +\infty\}$

st. Zij $X : (\Omega, A) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X is een s.v. a.s.a.:
X onafh. van P $\forall a \in \mathbb{R} : X^{-1}(]-\infty, a]) = \{\omega \mid X(\omega) \leq a\} \in A$. (= def.)
↳ als $\Omega = \mathbb{R}$ en $A = \mathcal{B}(\mathbb{R})$, dan is deze Borel-meetbaar



st. X induceert een kansmaat $P_X(B)$ of $P(X \in B)$ op $\mathcal{B}(\mathbb{R})$:
$P_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega \mid X(\omega) \in B\})$.
Noteer: $(\Omega, A, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ kansruimte
zie bewijs p. 29

def **verdelingsfunctie** F_X: $F_X(a) = P_X(]-\infty, a]) = P(\{\omega \mid X(\omega) \leq a\})$
$\Rightarrow F_X(a) = P(X \leq a)$ $a \in \mathbb{R}$



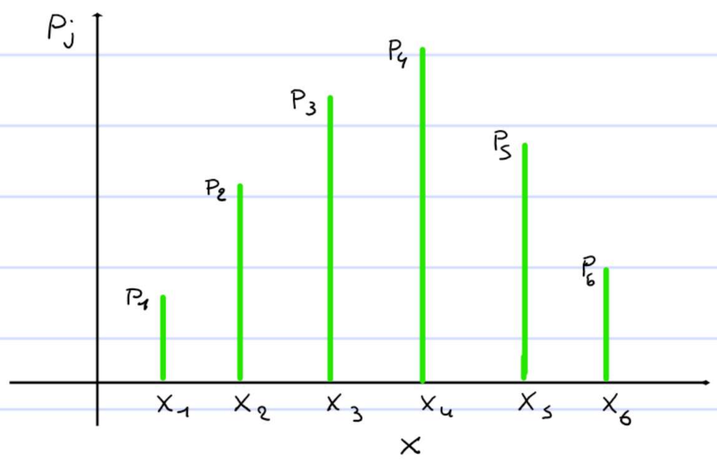
- monotoon stijgend:
  $\forall a \leq b : F_X(a) \leq F_X(b)$
- $\lim_{a \to +\infty} F_X(a) = 1$, $\lim_{a \to -\infty} F_X(a) = 0$
- rechtscontinu:
  $\forall a \in \mathbb{R} : \lim_{h \to 0^+} F_X(a+h) = F_X(a)$

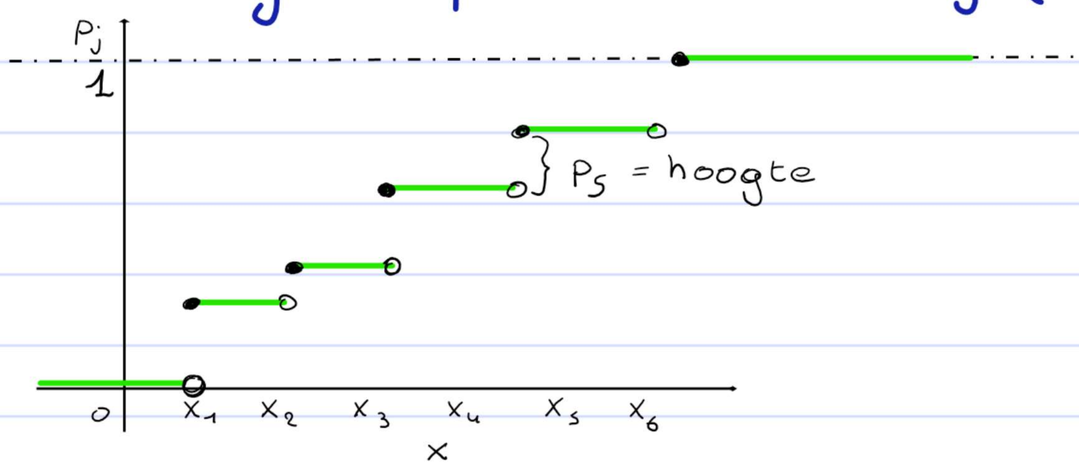
def. **kwantielfunctie** Q_X: de inverse van de verdelingsfunctie F_X. $Q_X(p)$ is de kleinste a met $F_X(a) \geq p$ met $0 < p \leq 1$. Als F_X inverteerbaar: $Q_X = F_X^{-1}$
25%, 50%, 75% → 1e, 2e en 3e kwantiel

Types stochastische veranderlijken

Discrete stochastische veranderlijken

↳ Ω wordt afgebeeld op een eindig of aftelbaar aantal verschillende punten x_j
$\Rightarrow p_j = P(\{\omega \in \Omega \mid X(\omega) = x_j\}) = P(X = x_j)$ met $p_j \geq 0$ en $\sum p_j = 1$
de rij $\{p_j \mid j \in \mathbb{N}\}$ is de **discrete verdeling** van de **discrete s.v.** X
↑ x_j zijn de waarden horende bij p_j
$\Rightarrow$ een discrete verdeling bepaald een rij $\{(x_j, p_j) \mid j \in \mathbb{N}\}$





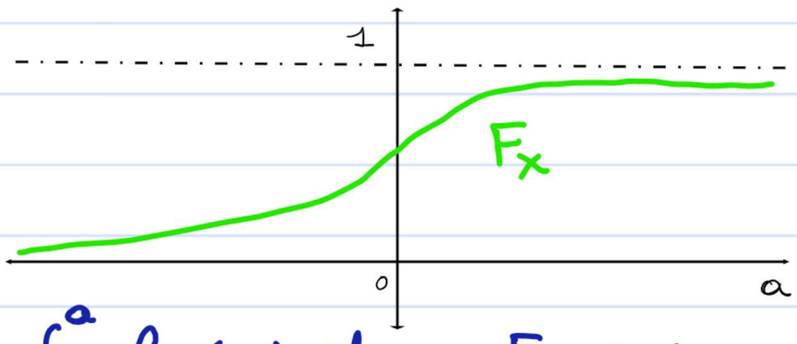
Continue stochastische veranderlijken

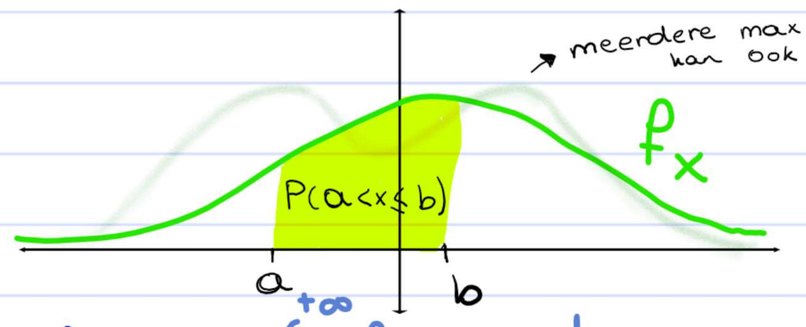
↳ $P_X(\{x\}) = 0$, geen sprongen in grafiek

Als F_X continu afleidbaar:

dichtheidsfunctie: $f_X(x) = \frac{d F_X(x)}{dx} = F_X'(x)$

↳ $f_X(x) \geq 0$, want F_X zwak stijgend





$\int_{-\infty}^{a} f_X(x)\,dx = F_X(a) = P(X \leq a)$ → $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$

$\int_{a}^{b} f_X(x)\,dx = F_X(b) - F_X(a) = P(a < X \leq b) = P(a \leq x \leq b) = P(a < x < b)$

→ F_X heeft geen eenheden en f_X die van $\frac{1}{x}$

Opm.

Er zijn s.v. die niet discreet en niet continu zijn

Momenten van een s.v.

Verwachtingswaarde & variantie

• verwachtingswaarde $E[X]$ (bestaat enkel als $E[X] < \infty$ volgens formule)

$$E[X] = \int_{-\infty}^{+\infty} x\,dF_X(x) = \begin{cases} \sum_j x_j P_j & \text{(discreet)} \\ \int_{-\infty}^{+\infty} x \cdot f_X(x)\,dx & \text{(continu)} \end{cases}$$

→ eigenschappen:

- $E[aX] = a\,E[X] \quad (a \neq 0)$
- $E[X+b] = E[X] + b$
- $E[b] = b$
- $|E[X]| = E[|X|]$

$E[X]$ is het eerste moment van X en $E[|X|]$ het eerste absolute moment.

• variantie $Var[X]$

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

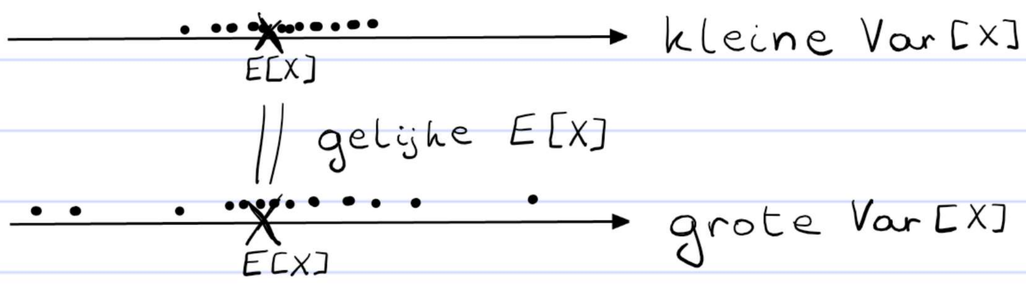
$$Var[X] = \begin{cases} \sum_j (x_j - E[X])^2 P_j & \text{(discreet)} \\ \int_{-\infty}^{+\infty} (x - E[X])^2 f(x)\,dx & \text{(continu)} \end{cases}$$

→ eigenschappen:

- $Var[aX] = a^2 Var[X]$
  - ↳ $Var[-X] = Var[X]$
- $Var[X+b] = Var[X]$
- $Var[b] = 0$

$Var[X]$ in eenheid van X^2 => Standaardafwijking σ

$\sigma = \sqrt{Var[X]}$



Som en product van s.v.

bewijzen vanaf p. 44

- $(X+Y)(\omega) = X(\omega) + Y(\omega)$
  - ↳ $E[X+Y] = E[X] + E[Y]$
- $(XY)(\omega) = X(\omega)\,Y(\omega)$

→ X en Y zijn onafhankelijk a.s.a

$P((X \in A) \cap (Y \in B)) = P(X \in A) \cdot P(Y \in B)$,

$P((X \leq a) \cap (Y \leq b)) = P(X \leq a) \cdot P(Y \leq b)$

↳ $E[XY] = E[X]\,E[Y]$

$Var[X+Y] = Var[X] + Var[Y]$

$Var[XY] = Var[X]\,Var[Y] + E[X]^2\,Var[Y] + Var[X]\,E[Y]^2 \geq Var[X]\,Var[Y]$

$Var[X-Y] = Var[X] + Var[Y]$

Ongelijkheden & grenzen

- $E[X] < \infty$ als $E[|X|] < \infty$ als $\sum_{n=1}^{\infty} P(|X| \geq n)$ convergeert

bew p. 46 $\sum_{n=1}^{\infty} P(|X| \geq n) \leq E[|X|] \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n)$

p. 48 - indien $E[|X|^n] < \infty$, dan $E[|X|^k] < \infty$ voor $0 \leq k \leq n$

$\Rightarrow$ Als X positieve s.v. enkel gehele waarden kan aannemen:
$E[X] = \sum_{n=1}^{\infty} P(X \geq n)$

p. 49 - ongelijkheid van Chebyshev: X s.v., $\phi: \mathbb{R} \to \mathbb{R}$, $\phi(X)$ s.v., $E[\phi(X)] < \infty$:

① $\phi \geq 0$, even en niet-dalend voor $x \geq 0$, dan:
$\forall a \geq 0: P(|X| \geq a) \leq \frac{1}{\phi(a)} E[\phi(X)]$

② $\phi \geq 0$ en niet-dalend voor $-\infty < x < +\infty$, dan:
$\forall a \in \mathbb{R}: P(X \geq a) \leq \frac{1}{\phi(a)} E[\phi(X)]$

$\Rightarrow$ X s.v. en $E[|X|^n] < \infty$, $n > 0$, dan:
$\forall a > 0: P(|X| \geq a) \leq a^{-n} E[|X|^n]$

$\Rightarrow$ X s.v. en $E[X] = \mu$, $Var[X] = \sigma^2 < \infty$, dan:
$\forall a > 0: P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$

Hogere momenten en momentgenererende functie

- ruw moment van orde $k \geq 1$: $\alpha_k(X) = E[X^k]$
- centraal moment van orde $k \geq 1$: $\mu_k(X) = E[(X - E[X])^k]$
  $\mu_1(X) = 0$ $\mu_2(X) = Var[X]$
- momentgenererende functie van X: $M_X: \mathbb{R} \to \mathbb{R}^+$, $M_X(t) > 0$

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_j e^{tx_j} p_j & \text{(discreet)} \\ \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx & \text{(continu)} \end{cases}$$

$e^a = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \dots$

$e^{tX} = 1 + (tX) + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots$

$E[e^{tX}] = 1 + E[X]t + E[X^2]\frac{t^2}{2!} + E[X^3]\frac{t^3}{3!} + \dots$

$\Rightarrow M_X(t) = 1 + \alpha_1 t + \alpha_2 \frac{t^2}{2!} + \alpha_3 \frac{t^3}{3!} + \dots$

$\Rightarrow \alpha_k = \frac{d^k}{dt^k}[M_X(t)]_{t=0}$ $(k \geq 1)$

Stel $Y = aX + b$, dan $M_Y(t) = e^{bt} M_X(at)$

bew p. 52 Als X en Y onafhankelijk zijn, dan:
$M_{X+Y}(t) = M_X(t) M_Y(t)$

Kentallen

van Locatie

- gemiddelde: $\mu_X = E[X]$
- mediaan: $Med(X) = Q_X(0,5)$
- modus: meest voorkomende waarde → piek in $f(x)$
  ↳ 1 modus: unimodinaal, 2 modi: bimodaal, multimodaal

van schaal

- variantie (en standaardafwijking): $\sigma^2 = Var[X] = E[X^2] - E[X]^2$
- interkwartielafstand: $IQR = Q_X(\frac{3}{4}) - Q_X(\frac{1}{2})$
- median absolute deviation: $MAD = Med|X - Med(X)|$
  ↳ 50% van massa tussen $Med(X) - MAD(X)$ en $Med(X) + MAD(X)$
- variatiebreedte: $Q_X(1) - Q_X(0)$

$V(aX) = V(X)$ - variatiecoëfficiënt: $(X \geq 0)$: $v(X) = \frac{\sigma_X}{\mu_X}$ → dimensieloos

van scheefheid

- centrale derde moment: $\mu_3(X) = E[(X - E[X])^3]$
  ↳ 0  , >0  , <0 
- scheefheidscoëfficiënt: $\gamma_1 = \frac{\mu_3}{\sigma^3}$ → dimensieloos
- $\frac{3(\mu - Med(X))}{\sigma}$
- $\frac{(Q(\frac{3}{4}) - Q(\frac{1}{2})) - (Q(\frac{1}{2}) - Q(\frac{1}{4}))}{Q(\frac{3}{4}) - Q(\frac{1}{4})}$

Belangrijke verdelingen

Discrete verdelingen

- **discrete uniforme verdeling**: elke uitkomst even waarschijnlijk → gedefinieerd op eindig universum
  $\Omega = \{x_1, \dots, x_n\}$, $p_j = P(\{x_j\}) = \frac{1}{n}$
- **Bernoulli verdeling**: slechts 2 uitkomsten, $X \to \mathcal{B}(1,p)$
  $\Omega = \{0,1\}$ $\begin{cases} P(X=1) = p & \text{(kans op succes)} \quad (0<p<1) \\ P(X=0) = q = 1-p & \text{(kans op mislukking)} \end{cases}$
- **binomiaalverdeling**: aantal successen bij n keer herhalen van het experiment, $Y \sim \mathcal{B}(n,p)$
  $\Omega = \{0,1,\dots,n\}$, $P(Y=k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$
  ▽ als $X_i \sim \mathcal{B}(1,p)$, $i = 1,\dots,n$ onderling onafhankelijk ($Y = X_1 + \dots + X_n$)
- **geometrische verdeling**: aantal mislukkingen tot eerste succes bij Bernoulli experiment. $X \sim Geom(p)$
  $\Omega = \mathbb{N}$, $P(X=j) = q^j p = q^j (1-q)$
- **negatief binomiaalverdeling**: aantal mislukkingen tot r-de succes bij Bernoulli experiment. $X \sim NB(r,q)$
  $\Omega = \mathbb{N}$, $P(X=j) = \binom{j+r-1}{j} (1-q)^r q^j = \binom{-r}{j} p^r (-q)^j$
- **hypergeometrische verdeling**: s goed uit N opties, n trekken zonder terugleggen → # goede. $Y \sim \mathcal{H}(N,s,n)$
  $\Omega = \{0,1,\dots,n\}$, $\max(0, n-(N-s)) \le j \le \min(s,n)$, $P(Y=j) = \frac{\binom{s}{j}\binom{N-s}{n-j}}{\binom{N}{n}}$
  ↳ met teruglegging: $Y \sim \mathcal{B}(n, s/N)$
- **poissonverdeling**: aantal gebeurtenissen met rate α, $0<\alpha<\infty$, $\mathcal{P}(\alpha)$, $\Omega = \mathbb{N}$, $P(X=j) = \frac{\alpha^j}{j!} \cdot e^{-\alpha}$
  ↳ aantal gebeurtenissen in disjuncte intervallen is onafhankelijk van elkaar
  ↳ $X_1 \sim \mathcal{P}(\alpha_1)$ en $X_2 \sim \mathcal{P}(\alpha_2)$ onafhankelijk, dan: $X_1 + X_2 \sim \mathcal{P}(\alpha_1 + \alpha_2)$

Continue verdelingen

- **continu uniform**: verdeling op $[a,b]$, $-\infty < a < b < +\infty$,
  $X \sim \mathcal{U}[a,b]$, $f_{a,b}(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$. $\Omega = [a,b]$
- **exponentieel**: Levensduur (continue versie van geometrische)
  $X \sim \mathcal{E}(\alpha)$, $0<\alpha<\infty$, $\Omega = [0,+\infty[$, $f_\alpha(t) = \begin{cases} \alpha \cdot e^{-\alpha t} & t \ge 0 \\ 0 & t < 0 \end{cases}$
  ↳ geheugenloos: $P(X > s+t \mid X > t) = P(X > s)$
- **univariate normale verdeling**:
  * **standaard normaal**: $Z \sim \mathcal{N}(0,1)$, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ $\forall x \in \mathbb{R}$
    met ϕ de dichtheidsfunctie, ϕ is even, verdeling:
    $\Phi(x) = \int_{-\infty}^{x} \phi(t)\,dt$ ← tabellen, $P(-z \le Z \le z) = 2\Phi(z) - 1$
  * **algemeen normaal**: $X \sim \mathcal{N}(\mu, \sigma^2)$, $X = \sigma Z + \mu$, $Z \sim \mathcal{N}(0,1)$
    $f_{\mu,\sigma}(x) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
    $F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
    $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ en $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ onafhankelijk:
    $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- χ^2_n: $Z_1, \dots Z_n$, $Z_i \sim \mathcal{N}(0,1)$. $X = Z_1^2 + \dots + Z_n^2$, chi-kwadraat verdeling met n vrijheidsgraden $X \sim \chi^2_n$
  $f_{\chi^2_n}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} & (x>0) \\ 0 & (x \le 0) \end{cases}$ $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\,dx$
  → $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
  ↳ $\forall t > 0: \Gamma(t+1) = t\,\Gamma(t)$
  ↳ $\forall n \in \mathbb{N}: \Gamma(n+1) = n!$

Formule van Stirling: $\Gamma(t) \sim \left(\frac{t-1}{e}\right)^{t-1} \sqrt{2\pi(t-1)}$ *

$\lim_{n\to\infty} \frac{\Gamma(t)}{*} = 1$

$\Rightarrow t = n+1 \to n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$

- **gammaverdeling**: moddeleren van wachttijden, $X \sim \Gamma_{\gamma,\beta}$, $\gamma,\beta > 0$, $f_{\gamma,\beta}(x) = \frac{x^{\gamma-1} e^{-\frac{x}{\beta}}}{\beta^\gamma \Gamma(\gamma)} 1_{\{x>0\}}$
- t_n: $Z \sim \mathcal{N}(0,1)$, $X \sim \chi^2_n$ $(n \geq 1)$ onafhankelijk, $T \sim t_n$ een studentverdeling met n vrijheidsgraden $T = \frac{Z}{\sqrt{X/n}}$
  ↳ $Med(T) = 0$, $E[t] = 0$, even, $Var[T] = \frac{n}{n-2}$ $(n>2)$
- **Cauchy**: t_n voor $n=1$, als $X \sim \mathcal{N}(0,1)$ en $Y \sim \mathcal{N}(0,1)$ onafhankelijk, dan $\frac{X}{Y}$ Cauchy verdeeld
  $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ ← zwaardere staart dan ϕ
  $F(x) = \frac{1}{\pi} \text{bgtan}(x) + \frac{1}{2}$
- **F**: als $W \sim \chi^2_m$ en $V \sim \chi^2_n$ onafhankelijk, $X = \frac{W/m}{V/n}$ de F-verdeling met m vrijheidsgraden in de teller en n in de noemer, $X \sim F_{m,n}$, α_k bestaat als $k < \frac{n}{2}$,
  $F_{m,n}(x) = 1 - F_{n,m}(x)$
- **Lognormaal**: $Y > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$, $X = \ln(Y) \sim \mathcal{N}(\mu, \sigma^2)$
  $F_Y(y) = \phi\left(\frac{\ln(y) - \mu}{\sigma}\right)$

Transformatie van s.v.

Monotone transformatie

Zij X een continue s.v. met $f_X(x)$. $x \notin S: f_X(x) = 0$, neem $h: \mathbb{R} \to \mathbb{R}$, zodat $U = h(X)$ een s.v. Als h differentieerbaar en strikt stijgend/dalend op S, dan:

bew p. 72
$$f_U(u) \begin{cases} f_X(h^{-1}(u)) \left|\frac{d\,h^{-1}(u)}{du}\right| & u \in h(S) \\ 0 & u \notin h(S) \end{cases}$$

Integraaltransformatie

bew p. 74 Zij X een continue s.v. met F_X strikt stijgend op $F_X^{-1}(]0,1[)$, stel $U = F_X(X)$, dan geldt $U \sim \mathcal{U}[0,1]$

Genereren van s.v.

bew p.74 Zij F strikt stijgend en $U \sim \mathcal{U}[0,1]$, $X = F^{-1}(U)$ met F de verdelingsfunctie van X.

3 Bivariate verd.

Verdeling van een stochastisch koppel

Stochastisch koppel (X,Y) ← stochastische vector van lengte 2
↳ versch. verdelingen.

Gezamelijke verdeling

↳ X en Y samen beschouwen

def. Zij X en Y s.V. op (Ω, A, P). dan: $(X,Y): \Omega \to \mathbb{R}^2 : \omega \mapsto (X(\omega), Y(\omega))$
met (X,Y) de stochastische vector $(\Omega, A, P) \xrightarrow{(X,Y)} (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), P_{X,Y})$
$\mathcal{B}(\mathbb{R}^2) \to]-\infty, a_1] \times]-\infty, a_2] \quad a_1, a_2 \in \mathbb{R}$
$P_{X,Y}(]-\infty, x] \times]-\infty, y]) = P(X^{-1}(]-\infty, x]) \cap Y^{-1}(]-\infty, y]))$
⇒ gezamelijke verdelingsfunctie $F_{X,Y}$ van X en Y:
$F_{X,Y}(x,y) = P(X \le x, Y \le y) = P_{X,Y}(]-\infty, x] \times]-\infty, y])$

- stijgend in alle argumenten
- rechtscontinu in alle argumenten
- $F_{X,Y}(x,y) \to 0$ als $x \to -\infty$ en/of $y \to -\infty$
- $F_{X,Y}(x,y) \to 1$ als $x \to +\infty$ en $y \to +\infty$

↑ elke bivariate f die hieraan voldoet is een verdelingsf.

$F_{X,Y}$ bepaald volledig de verdeling van (X,Y):
$P((X,Y) \in]a_1, b_1] \times]a_2, b_2])$
$= F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, a_2)$

→ continue bivariate verd.:
$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, du\, dv$
met $f_{X,Y}(x,y)$ de gezamelijke dichtheidsfunctie
- $f_{X,Y}(x,y) \ge 0 \quad \forall x, y \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(u,v)\, du\, dv = 1$

$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x\, \partial y}$

Marginale verdeling

↳ verdeling van X of verdeling van Y
$F_X(x) = P(X \le x) = P(X \le x, Y \le +\infty) = F_{X,Y}(x, +\infty)$ (idem voor $F_Y(y)$)
→ marginale dichtheidsfunctie van X: $f_X(x) = \int_{-\infty} f_{X,Y}(x,y)\, dy$

Onafhankelijkheid

def. De s.v. X en Y zijn onafhankelijk als voor alle $A, B \in \mathcal{B}(\mathbb{R})$
$\{X \in A\}$ en $\{Y \in B\}$ (onderling) onafhankelijk zijn.
↳ $F_{X,Y}(x,y) = F_X(x) F_Y(y)$
⇒ $E[XY] = E[X]E[Y]$, $Var[X+Y] = Var[X] + Var[Y]$ +zie s.v.

Voorwaardelijke verdeling

↳ verdeling van X, gegeven Y
$P(X = x \mid Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}$ als $P(Y=y) > 0$
voorwaardelijke dichtheidsfunctie $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ als $f_Y(y) > 0$

- $f_{X|Y}(x|y) \ge 0 \quad \forall x$
- $\int_{-\infty}^{+\infty} f_{X|Y}(x|y)\, dx = \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$

discreet / continu: 2 s.v. zijn onafh. a.s.a. $P(X = x \mid Y = y) = P(X = x) \quad \forall x, y$ + omg
$f_{X|Y}(x|y) = f_X(x)$

Karakteristieken

Momenten & MGF

def. Zij (X,Y) stochastisch koppel en $g: \mathbb{R}^2 \to \mathbb{R}$ Borelmeetbaar

dan $E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)\, P(X=x, Y=y) & \text{als } < \infty \text{ (discreet)} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy & \text{als } < \infty \text{ (continu)} \end{cases}$

- (ruwe) momenten: $g(x,y) = x^{r_1} y^{r_2}$, $E[X^{r_1} Y^{r_2}]$
- centrale momenten: $g(x,y) = (x-\mu_x)^{r_1} (y-\mu_y)^{r_2}$ met $\mu_i = E[i]$

  $E[(X-\mu_x)^{r_1} (Y-\mu_y)^{r_2}]$
- gezamelijke momentgenererende functie

Covariantie en correlatie

def. $Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$

↳ $Cov(X,X) = E[(X-E[X])^2] = Var[X]$

$Cov(X,Y) = Cov(Y,X)$

zie p. 86 $Cov(X,Y) = E[XY] - E[X]E[Y]$

def. correlatiecoëfficiënt $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X] Var[Y]}}$

→ beschrijven Lineair verb. tussen X en Y

cov: mate van verband (pos / neg \)

corr: maat voor sterkte verband $\in [-1,1]$ → ± 1 → perfect

Eigenschappen

st. Zij X en Y onafhankelijk en $g: \mathbb{R}^2 \to \mathbb{R}$ Borelmeetbaar, zodat

bew. p. 87 $\forall x,y \in \mathbb{R}: g(x,y) = g_1(x)\, g_2(y)$ en $E[g_1(x)], E[g_2(y)] < \infty$, dan:

$E[g(x,y)] = E[g_1(x)]\, E[g_2(y)]$

$\Rightarrow M_{X,Y}(t_1,t_2) = M_X(t_1)\, M_Y(t_2)$

st. Zij X, Y s.v., dan:

bew. p. 87
- $|Cov(X,Y)| \leq \sqrt{Var[X] \quad [Y]}$
- Zij X, Y onafh., dan $E[XY] = E[X]E[Y]$ en $Cov(X,Y) = 0$

  ↳ omgekeerd niet altijd waar

$\Rightarrow$ Voor X, Y s.v. geldt:

bew. p. 88
- $-1 \leq Corr(X,Y) \leq 1$
- Zij $X_2 = aX_1 + b$ met $a,b \in \mathbb{R}$, dan $Corr(X,Y) = \frac{a}{|a|} = sgn(a)$

st. Zij X en Y s.v. en $a_1, a_2 \in \mathbb{R}$, dan:

bew. p. 89
- $E[a_1 X + a_2 Y] = a_1 E[X] + a_2 E[Y]$
- $Var[a_1 X + a_2 Y] = a_1^2 Var[X] + a_2^2 Var[Y] + 2 a_1 a_2 Cov(X,Y)$

$\Rightarrow$ Zij X en Y onafh. en $a_1, a_2 \in \mathbb{R}$, dan:

$Var[a_1 X + a_2 Y] = a_1^2 Var[X] + a_2^2 Var[Y]$

Bivariate normale verdeling

def. (X,Y) heeft een bivariate normale verdeling als de gezamelijke verdelingsfunctie wordt gegeven door:

$$f_{X,Y}(x,y) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\frac{1}{1-\rho^2} \cdot \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$$

$\forall (x,y) \in \mathbb{R}^2$ met $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_0^+$ en $\rho \in]-1,1[$.

Hierin is: $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ met $\mu_1 = E[X]$ en $\sigma_1^2 = Var[X]$

en $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ met " en $\rho = Cor(X,Y)$

$\Rightarrow Cov(X,Y) = \rho\, \sigma_1 \sigma_2$

→ standaard bivariate normale verdeling

↳ $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0$

$\Rightarrow f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2+y^2)\right)$

4 Benaderingen

Limietstellingen

def. Zij $X_1, X_2, \dots$ s.v. met verdelingsfuncties $F_1(x), F_2(x), \dots$
Dan convergeert X_n in verdeling naar X als er een $F(x)$ bestaat met $\lim_{n\to\infty} F_n(x) = F(x)$ $\forall x$ waar F continu is

- notatie: $F_n \xrightarrow{D} F$ of $X_n \xrightarrow{D} X$

↑ puntsgewijze conv. van F_n naar F in continuïteitspunten van F

- steekproefgemiddelde: $X_1 \dots X_n$ i.i.d., dan $\bar{X} = \frac{X_1 + \dots + X_n}{n}$

Dan is: • $E[\bar{X}] = E[X_1]$
• $\text{Var}[\bar{X}] = \frac{1}{n}\text{Var}[X_1]$
$\Rightarrow X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ en onafh. $\Rightarrow \bar{X} \sim (\mu, \frac{\sigma^2}{n})$

Centrale Limietstelling

st. Zij (X_n) een rij i.d.d. s.v. met $E[X_1] = \mu$ en $\text{Var}[X_n] = \sigma^2 < \infty$.
Stel $\bar{X} = (X_1 + \dots + X_n)/n$, dan geldt:
$\forall x \in \mathbb{R}: \lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du$ ofwel
$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z$ met $Z \sim \mathcal{N}(0,1)$

- $E[\bar{X}] = \frac{1}{n} E[X_1 + \dots + X_n] = \frac{1}{n} \cdot n \cdot E[X_1] = E[X_1]$
- $\text{Var}[\bar{X}] = \text{Var}[X_1 + \dots + X_n]/n^2 = \frac{n}{n^2}\text{Var}[X_1] = \text{Var}[X_1]/n$

Herschrijving CLS:
$S_n = n\bar{X} = X_1 + \dots + X_n \Rightarrow E[S_n] = n\mu$, $\text{Var}[S_n] = n\sigma^2$
voor $n \to \infty$: $P\left[\left(\frac{S_n - n\mu}{\sqrt{n}\,\sigma}\right) \le x\right] \to \Phi(x)$
→ in praktijk: Zij (X_n) i.i.d. en n voldoende groot $(\mu, \sigma < \infty)$:
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})$
$S_n = \sum_{i=1}^{n} X_i \approx \mathcal{N}(n\mu, n\sigma^2)$

Opm.:
- benadering
- verd. v. X_i moet niet continu zijn
- X_i moet niet symmetrisch zijn (hoe schever, hoe groter n)

Stelling van De Moivre-Laplace

st. Bernouilli verd.: $S_n = X_1 + \dots + X_n$ met $X_i \sim B(1,p)$, dan $S_n \sim B(n,p)$
↳ $E[X_1] = \mu = p$, $\text{Var}[X_1] = \sigma^2 = p(1-p)$
$E[S_n] = np$, $\text{Var}[S_n] = np(1-p)$
$\frac{S_n - np}{\sqrt{npq}} \xrightarrow{D} Z$ met $Z \sim \mathcal{N}(0,1)$ $\Rightarrow B(n,p) \approx \mathcal{N}(np, np(1-p))$
↑ $n \ge 30$, $np > 5$ en $n(1-p) > 5$

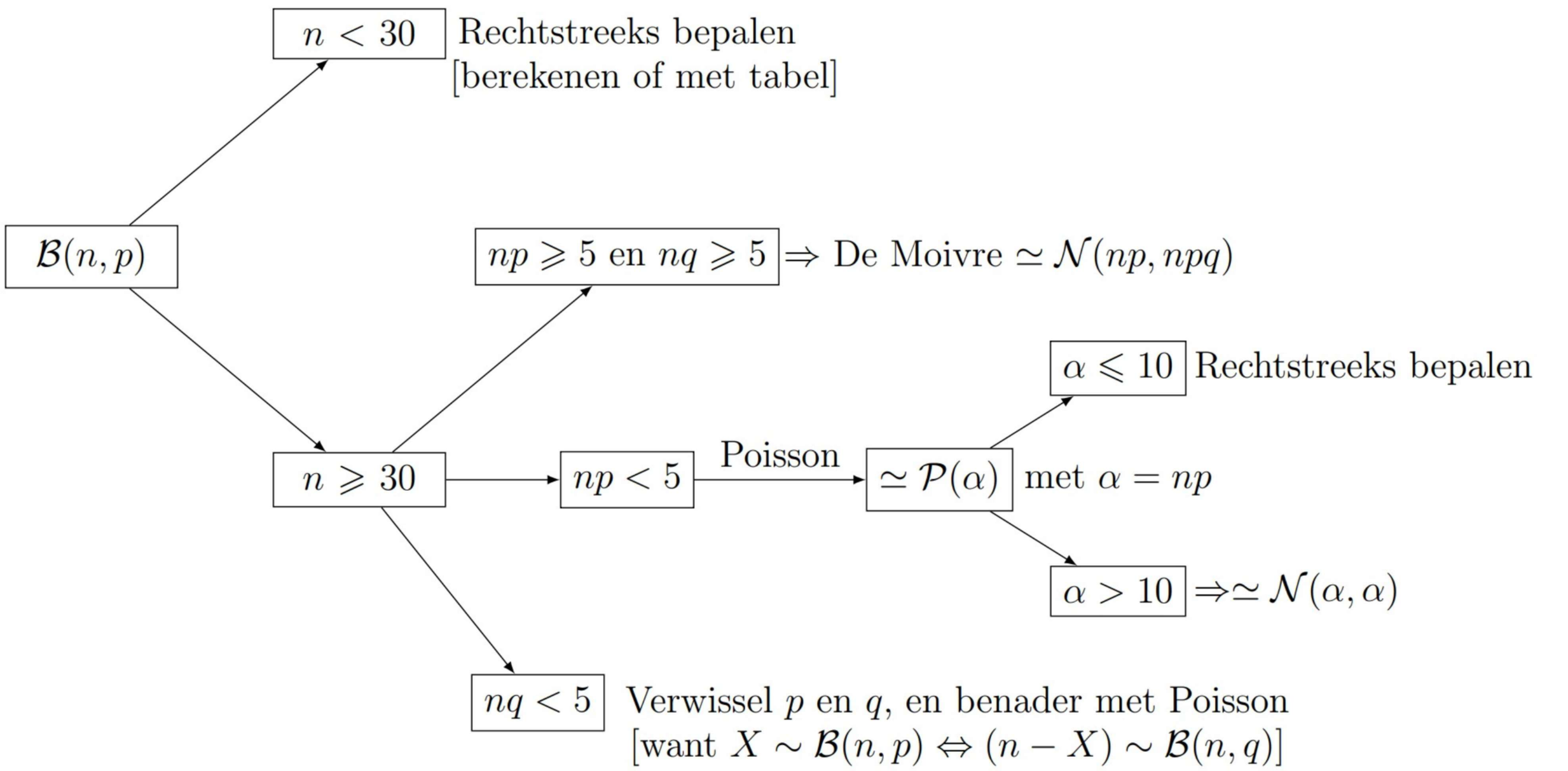
Continuïteitscorrectie

$P(Y \ge a) \approx 1 - \phi\left(\frac{a - 0{,}5 - np}{\sqrt{npq}}\right)$
$P(Y \le b) \approx \phi\left(\frac{b + 0{,}5 - np}{\sqrt{npq}}\right)$
$\Big\} P(a \le Y \le b) \approx \phi\left(\frac{b + 0{,}5 - np}{\sqrt{npq}}\right) - \phi\left(\frac{a - 0{,}5 - np}{\sqrt{npq}}\right)$

Limietstelling van Poisson

st. Zij (X_n) een rij s.v. met $X_n \sim B(n, p_n)$ met als $n \to \infty$, dan $p_n \to 0$,
bew. p.94 + slides zodat $np_n \to \alpha$ met $0 < \alpha < \infty$, dan $X_n \xrightarrow{D} X$ met $X \sim P(\alpha)$
$\Rightarrow B(n,p) \approx P(np)$
↳ $n \ge 30$ en $np < 5$ of $nq < 5$ → p en q van plaats wisselen

→ Normale benadering voor Poissonverdeling:
st. Zij $X \sim P(\alpha)$, dan geldt voor $\alpha \to \infty$ dat $\frac{X - \alpha}{\sqrt{\alpha}} \xrightarrow{D} \mathcal{N}(0,1)$
bew slides $\Rightarrow P(\alpha) \approx \mathcal{N}(\alpha, \alpha)$
↳ $\alpha > 10$


$n < 30$
Rechtstreeks bepalen
[berekenen of met tabel]
$\mathcal{B}(n,p)$
$np \geqslant 5$ en $nq \geqslant 5$
$\Rightarrow$ De Moivre $\simeq \mathcal{N}(np, npq)$
$\alpha \leqslant 10$
Rechtstreeks bepalen
$n \geqslant 30$
$np < 5$
Poisson
$\simeq \mathcal{P}(\alpha)$
met $\alpha = np$
$\alpha > 10$
$\Rightarrow \simeq \mathcal{N}(\alpha, \alpha)$
$nq < 5$
Verwissel p en q, en benader met Poisson
[want $X \sim \mathcal{B}(n,p) \Leftrightarrow (n - X) \sim \mathcal{B}(n,q)$]

① Beschrijvende statistiek

Beschrijven van frequenties

3 stappen bij statistische analyse:
1) verzamelen van gegevens → steekproef van n observaties
2) beschrijvende statistiek → waar dit hoofdstuk over gaat
3) statistische interferentie → beslissingen nemen
   ↳ hypothesetoetsen

Wat willen we weten?
- grote verschillen tussen geg. of ± gelijk?
- hoe groot zijn de verschillen?
- uitschieters?
- trends?

⇒ kansmodel bepalen

Soorten variabelen

- kwalitatieve var.: geven categorie aan
  - ↳ nominale var.: geen rangorde (nationaliteit)
  - ↳ ordinale var.: kunnen geordend ω (weinig → veel)
- kwantitatieve var.: numerieke waarden (lengte)
  - ↑ continu of discreet
  - ↑ bewerkingen zinvol (meeteenheid)

Frequentietabellen en grafieken

steekproef: $(x_1, \dots, x_n)$, uitkomstenverz.: $\Omega = \{m_1, \dots, m_u\}$

- absolute frequentie: n_j = # keer dat m_j voorkomt
- relatieve frequentie: $h_j = n_j/n$

  $\sum_j n_j = n$ $\sum_j h_j = 1$

→ in frequentietabel (relatief) staafjesdiagram of taartdiagram zetten

- histogram: voor oneindig universum v. kwantitatieve
  - ↳ geg discretiseren / groeperen in klassen
  - ↳ klassenmiddens c_j, klassebreedte Δ
  - ↳ continue tegenhanger v.h. staafjesdiagram
  - ↳ frequenties ∝ opp.

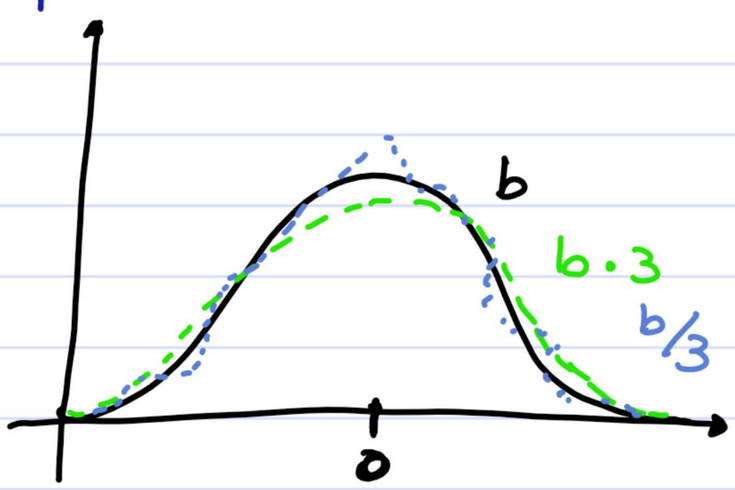
→ klassiek histogram bij vaste Δ
→ klassen met ≠ breedtes Δ_j ⚠ vertekend beeld
  ↳ schalen: hoogte $= \frac{h_j}{\Delta_j}$ bij dichtheidshistogram
  totale opp: $\sum_j \Delta_j \frac{h_j}{\Delta_j} = 1$

Kernel dichtheid schatter

↳ (dichtheids) histogram is ruwe schatter v. dichth. en hangt af v. aantal (en locatie) v. balken
↳ Kernel nauwkeuriger: kernel functie K
  ↑ dichtheidsfunctie symm. rond 0

$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x_i - x}{b}\right)$ met b de bandbreedte

→ optimale bandbreedte kiezen (niet stabiel voor kleine b)
→ weerspiegeling v. verdeling van beschikbare steekproefgeg.
→ correcte vorm v. verd. identificeren



ecdf

Empirische verdelingsfunctie en kwantielfunctie

$\hat{F}_n(x) = \frac{1}{n}(\#\, x_i \leq x\,;\, i = 1,\dots,n) = \frac{\sum_{i=1}^{n} \mathbb{1}_{x_i \leq x}}{n}$ $\quad \mathbb{1}_{x_i \leq x} = \begin{cases} 1 \text{ als } x_i \leq x \\ 0 \text{ als } x_i > x \end{cases}$

→ trapfunctie
→ geen samenvallende waarden => $\hat{F}_n(x_i) = \frac{i}{n}$
→ enkel gegroepeerde geg. => cumulatieve freq. voor klassengrenzen met n_j en h_j
→ empirische kwantielfunctie $\hat{Q}_n$: $\hat{F}_n$ trap, dus $\hat{Q}_n$ (inverse van $\hat{F}_n$) als volgt bepalen:
- geen x met $\hat{F}_n(x) = p$, dan $\hat{Q}_n$ kleinste x met $\hat{F}_n(x) > p$
- interval met $\hat{F}_n(x) = p$, $\hat{Q}_n(p)$ = kleinste x in interval

=> $\hat{Q}_n(p)$ is kleinste x met $\hat{F}_n(x) \geq p$
- $\hat{Q}_n(i/n) = x_i$
- $\hat{Q}_n(p) = x_i$ als $\frac{i-1}{n} < p \leq \frac{i}{n}$
- $\hat{F}(x_i) = \frac{i}{n+1}$ of $\frac{i-0,5}{n}$ (zie slides)

Kenmerken voor Locatie en schaal

gegroepeerde gegevens

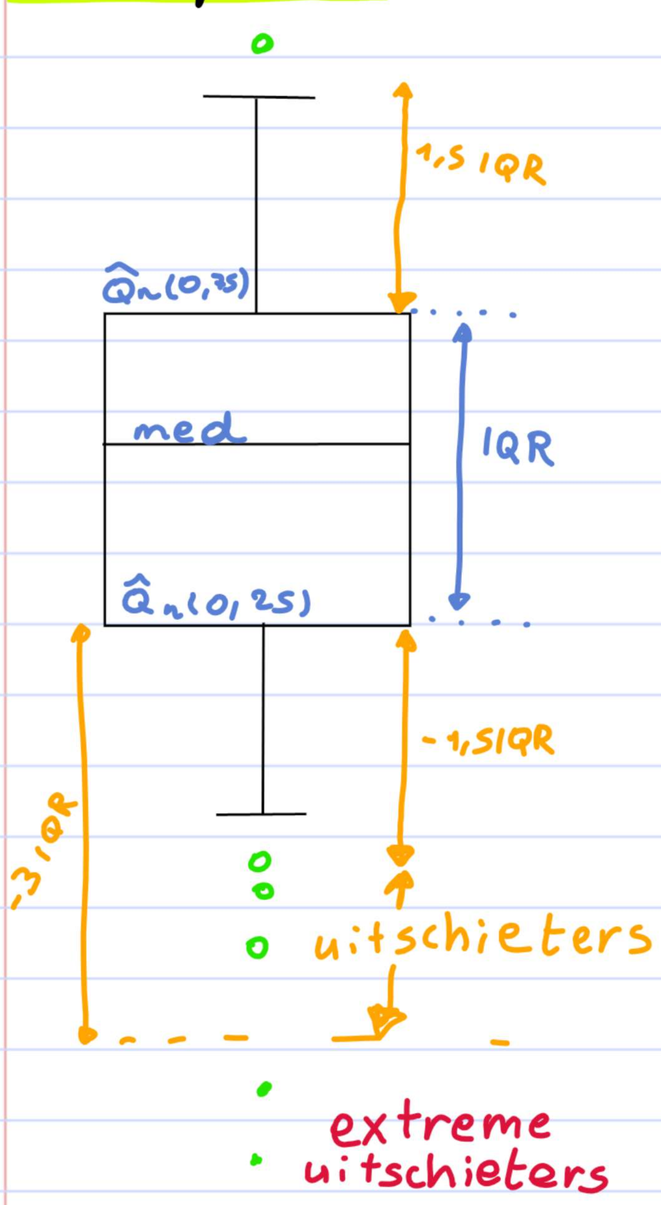
Centrumkenmerken

- steekproefgemiddelde: $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n) = \frac{1}{n}\sum_i x_i = \frac{1}{n}\sum_j n_j c_j$
- steekproefmediaan: $\text{med}(x) = \begin{cases} x_{(n+1)/2} & (n \text{ oneven}) \\ (x_{n/2} + x_{n/2+1})/2 & (n \text{ even}) \end{cases}$
  - benadering: $\frac{y-y_0}{x-x_0} = \frac{y_1-y_0}{x_1-x_0}$ $x = 50\%$ $y = \text{med}$
- modus: meest voorkomende waarde
  - modale klasse: klasse met grootste verh. $\frac{\text{frequentie}}{\text{klassebreedte}}$

Spreidingskenmerken

- empirische var. s^2 en standaardafwijking: $s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$ en $s = \sqrt{s^2}$
- empirische interkwartielafstand: $IQR_n = \hat{Q}_n(3/4) - \hat{Q}_n(1/4)$ → als n groot genoeg: $\frac{IQR_n}{s} \approx 1{,}34$
- median absolute deviation: $Mad(x) = \text{med}(|x - \text{med}(x)|)$
  → als n voldoende groot: $\frac{Mad(x)}{s} \approx 0{,}67$
- bereik (range): $R = x_n - x_1$

Boxplot



zichtbare informatie over verdeling:
- centrum: mediaan (soms ook gem)
- IQR = lengte v.d. doos
- scheefheid: symmetrie & whiskers
- zwaarte v.d. staarten

Covariantie en correlatie

↳ Lineair verband tussen twee (metrische) var. X en Y.?

→ grafisch: via scatterplot

→ analytisch: via steekproef covariantie en -correlatie

- steekproefcovariantie:

$s_{xy} = Cov(x,y) = \frac{1}{n-1} \sum_i (x_i - \bar{x})(y_i - \bar{y})$, $s_{xx} = s_x^2$

→ Lineariteitseigenschap: als $u_i = a_1 x_i + b_1$ en $v_i = a_2 y_i + b_2$, dan

$s_{uv} = a_1 a_2 s_{xy}$

↳ sterk afhankelijk van meeteenheid

- (Pearson) correlatiecoëfficiënt:

$r = r_{xy} = Corr(x,y) = \frac{Cov(x,y)}{s_x s_y}$

als $s_x > 0$ en $s_y > 0$: $r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} = \frac{1}{n-1} \sum_i \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$

→ $-1 \leq r \leq 1$

→ Lineariteitseigenschap: $|r_{uv}| = |r_{xy}|$

→ maat voor lineair verb. $|r|$ ↑ sterker verb. $r>0$ /, $r<0$ \

Kruistabellen

↳ geeft verband tussen twee kwalitatieve variabelen

→ verwachte waarde voor elk onderdeel berekenen en vergelijken. ong. gelijk ⇒ onafhankelijk

verwachte waarde = $\frac{\text{rijtotaal} \cdot \text{kolomtotaal}}{n}$

QQ-plot (kwantielplot)

↳ normaliteit (of andere continue verdeling) nagaan

→ Normale QQ-plot:

- zet empirische kwantielen $\hat{Q}_n(p)$ uit tegenover theoretische kwantielen $Q(p) = \Phi^{-1}(p)$ v.e. standaardnormale verd voor versch. $0 \leq p \leq 1$

↳ Lineair verb. bij normaal verd. geg.

$Q(p) = \mu + \sigma \Phi^{-1}(p)$ $\forall$ $0<p<1$ → $(\Phi^{-1}(p), Q(p))$ vormen rechte

→ bij lin verb.: $(\Phi^{-1}(p), \hat{Q}_n(p)) \approx$ rechte

welke p gebr.? → $\hat{Q}_n(i/n) = x_i$ ⇒ $p = 1/n, \ldots, (n-1)/n, 1$

→ continuïteidscorrectie: $(\Phi^{-1}(\frac{i-0{,}5}{n}), x_i)$

▽ let op welke as waar staat ▽

→ exponentiële verdeling

- $1 - F_\alpha(x) = e^{-\alpha x}$ $(x>0)$
- standaard verd.: $1 - F_1(x) = e^{-x}$
- $Q_\alpha(p) = -\frac{1}{\alpha} \log(1-p)$ $(0<p<1)$
- $Q_\alpha(p) = \frac{1}{\alpha} Q_1(p)$
- plot $(-\log(1-p), \hat{Q}_n(p))$ → $(-\log(1 - \frac{i-0{,}5}{n}), x_i)$
- rechtlijnig patroon met rico $\frac{1}{\alpha}$

algemeen:

- lognormale verd.f.: $F_X(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \Rightarrow \ln(x_i) = \mu + \sigma \Phi^{-1}\left(\frac{i}{n}\right)$

② Schatten v. parameters & bi

Puntschatters

$\bar{X}$: schatter (s.v.), $\bar{x}$: schatting (van 1 steekproef)
Ξ : uitkomstenverzameling / waarnemingsruimte
$\mathcal{B}$: bijbehorende σ-algebra
θ : parameter, Θ : parameterruimte
$\{P_\theta ; \theta \in \Theta\}$ verz. kansverdelingen op meetbare ruimte $(\Xi, \mathcal{B})$
→ 1 kansverdeling per $\theta \in \Theta$

def. $(\Xi, \mathcal{B}, \{P_\theta ; \theta \in \Theta\})$ is een statistisch model a.s.a.
$\forall \theta \in \Theta: (\Xi, \mathcal{B}, P_\theta)$ is een kansruimte

def. Een reële (of univariate) statistiek T is een meetbare afbeelding $T: (\Xi^n, \mathcal{B}^{\otimes n}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$
↪ T is een functie v.d. steekproef en niet v.d. onbekende parameter θ
⇒ T is een S.V. en heeft dus een verdeling

def. Beschouw een statistisch model $(\Xi, \mathcal{B}, \{P_\theta ; \theta \in \Theta\})$ en een willekeurige afbeelding $g: \Theta \to \mathbb{R}$ (estimand).
Een (punt)schatter (estimator) van $g(\theta)$ is een reële statistiek $T: (\Xi^n, \mathcal{B}^{\otimes n}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (onafh. van θ).
Een realisatie $T(x)$ voor een $x \in \Xi^n$ heet een (punt)schatting (estimate) van $g(\theta)$.

zie p. 28-29 voor voorbeeld

Onvertekende schatters

def. Bias (vertekening) van een schatter is
$b_\theta(\hat{\theta}) = E_\theta[\text{error}] = E_\theta[\hat{\theta} - \theta] = E_\theta[\hat{\theta}] - \theta$
Een schatter $\hat{\theta}$ voor θ is onvertekend of zuiver als
$E_\theta[\hat{\theta}] = \theta \quad (b_\theta(\hat{\theta}) = 0)$
→ mean squared error van een schatter $\hat{\theta}$:
$MSE_\theta(\hat{\theta}) = E_\theta[(\hat{\theta} - \theta)^2] = b_\theta(\hat{\theta})^2 + Var_\theta[\hat{\theta}]$ $\quad Var_\theta[\hat{\theta}] = E_\theta[(\hat{\theta} - E_\theta[\hat{\theta}])^2]$

bew. p. 29

↪ zo klein mogelijk voor nauwkeurige $\hat{\theta}$
zie slide 34 voor rekenvoorbeeld (in. comb.

Parameters v. normale populatie schatten

$X_i \sim \mathcal{N}(\mu, \sigma^2)$, $\hat{\mu} = \bar{x}$, want onvertekend ($E[\bar{x}] = \mu$) en $MSE[\bar{x}] = Var[\bar{x}] = \frac{\sigma^2}{n}$
→ onvertekende schatter met kleinste variantie
⇒ $\bar{X}$ is een UMVU (uniform minimum variantie onvertekend) voor μ
! enkel als alle datapunten uit normale verdeling
• σ^2 schatten met $s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$
→ delen door $n-1$ voor onvertekend
↪ 1 v.d. n vrijheidsgr. gebruikt voor $\bar{x}$

st. Zij $X_1, \ldots, X_n$ steekproeven uit $\mathcal{N}(\mu, \sigma^2)$-verdeelde populatie, dan $E[S^2] = \sigma^2$

zie slide 43 - 44 voor bew

Voor $S_*^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$ geldt: $E[S_*^2] = E[\frac{n-1}{n} S^2] = \frac{n-1}{n} \sigma^2$
↪ S_*^2 geen onvertekende schatter voor σ^2

st. Stelling v. Helmert: $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ onafh. met $n \geq 2$, dan
• $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
• $\bar{X}$ en S^2 zijn onafhankelijk
• $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

st. X normaal verdeeld a.s.a. $\bar{X}$ en S^2 onafh.
• $Var[S^2] = \frac{2\sigma^4}{n-1}$

zie oef 1.1 a en c

zie slides 55 - 56

Schatten van proporties of kansen

→ relatieve frequentie / proportie successen in steekproef:

$\hat{P} = \frac{1}{n} \sum_i X_i = \bar{X}$

sl. 6 1 voor rekenvb.

- $E[\hat{P}] = p$ → onvertekend
- $Var[\hat{P}] = \frac{p(1-p)}{n}$
- $\frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{CLS}{\approx} N(0,1)$ indien n voldoende groot

Betrouwbaarheidsintervallen

↳ afhankelijk v.d. steekproef

def. Stel $X_1, \dots, X_n$ steekproeven uit X en dichtheid van X hangt af van onbekende parameter $\theta \in \Theta \subset \mathbb{R}$. Kies $0 < \alpha < 1$.
Een $(1-\alpha)$-betrouwbaarheidsinterval is een interval van de vorm $[L(x_1, \dots, x_n), R(x_1, \dots, x_n)]$ met
$P(L(X_1, \dots, X_n) \leq \theta \leq R(X_1, \dots, X_n)) = 1 - \alpha$
↳ α is het significantieniveau
$|\frac{R-L}{2}|$ is de foutenmarge bij een tweezijdige test

def.
- tweezijdig: $P(\theta < L) = P(\theta > R) = \alpha/2$
- Linkseenzijdig: $]-\infty, R]$
- rechtseenzijdig: $[L, +\infty[$

Strategie om BI te construeren:
1) goede schatter T voor θ
2) verdeling van $h(T, \theta)$ n.o.afh. v. θ
3) a en b zodat $P(a \leq h(T, \theta) \leq b) = 1 - \alpha$
4) herschrijf zodat θ middelste deel is

BI voor μ met gekende σ^2

$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
$\Rightarrow P(-1{,}96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1{,}96) = 0{,}95$
$\Rightarrow P(\bar{X} - 1{,}96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1{,}96 \frac{\sigma}{\sqrt{n}}) = 95\%$
$\Rightarrow$ BI: $[\bar{x}_n - 1{,}96 \frac{\sigma}{\sqrt{n}}, \bar{x}_n + 1{,}96 \frac{\sigma}{\sqrt{n}}]$
→ deze intervallen zullen in 95% v.d. gevallen μ bevatten
$1-\alpha$ = het betrouwbaarheidsniveau
algemeen: $[\bar{x}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}]$
- $\alpha \downarrow \Rightarrow$ BI ← →
- $\sigma \uparrow \Rightarrow$ BI ← →
- $n \uparrow \Rightarrow$ BI → ←

⚠ meestal is σ niet op voorhand gekend

Variantie van normale verdeling

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$, $S^2 = \frac{1}{n-1} \sum_i (x_i - \bar{X}_n)^2$, $E[S^2] = \sigma^2$
→ st. v. Helmert → $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
BI: $[\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}]$
⚠ niet symmetrisch rond s^2
→ wortels nemen voor σ

BI voor μ, σ ongekend

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$, $\bar{X}_n \sim (\mu, \frac{\sigma^2}{n})$

bew p. 41 st. Als $X \sim N(\mu, \sigma^2)$, dan $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
↳ n-1 vrijheidsgraden
BI: $[\bar{x} - t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}}]$
$t_{n-1,1-\frac{\alpha}{2}} = -t_{n-1,\frac{\alpha}{2}}$
→ exact resultaat
$n \to \infty \Rightarrow s \to \sigma \Rightarrow t_{n-1,1-\frac{\alpha}{2}} \to z_{1-\alpha/2}$ (1e resultaat)

Proportie

$\hat{p} = \frac{\#\text{waargenomen successen}}{n} = \frac{1}{n}\sum_i x_i = \bar{X}_n$

$E[\hat{p}] = \frac{1}{n}\sum_i E[x_i] = \frac{np}{n} = p$

$n\hat{p} = \sum_i x_i \sim B(n,p)$

als $n \geq 30$, $np \geq 5$ en $nq \geq 5$, dan:

$B(n,p) \approx N(np, npq)$

$\Rightarrow n\hat{p} \approx N(np, npq)$

$\Rightarrow \hat{p} \approx N(p, \frac{pq}{n})$

n groot genoeg $\Rightarrow \sigma = \sqrt{\frac{p(1-p)}{n}}$ ω benaderd door $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$\Rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$

foutenmarge

$BI: [\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$

& 2 benaderingen

$n\uparrow \Rightarrow$ BI → ←

$\alpha\downarrow \Rightarrow$ BI ← →

③ Hypothesetesten

Algemeen

- hypothese: bewering over verdeling of parameter van een verdeling.
- testen van hypothesen: op basis van steekproef uitspraak doen over geldigheid v. hypothese.

Stappenplan van 6 stappen:

1) Hypothesen opstellen
   - ↳ H_0 en H_1, twee complementaire testen opstellen
     - ↳ ofwel H_0 verwerpen (H_1 waar)
       ofwel H_0 niet verwerpen (geen besluit)
   - → H_0 moet gekozen $\bar{w}$ dat het gedrag v.d. toetsingsgrootheid te bep. is in de veronderstelling dat H_0 waar is (gelijkheid moet mee in H_0 staan).
   - → H_1 is als enige te besluiten of ze waar is.
     => wat we willen bewijzen in H_1

H_0	$\theta = \theta_0$	tweezijdig	$\theta \leq \theta_0$	rechtseenzijdig	$\theta \geq \theta_0$ linkseenzijdig
H_1	$\theta \neq \theta_0$		$\theta > \theta_0$		$\theta < \theta_0$

2) Voorwaarden noteren en nagaan
   - → steekproef willekeurig genomen?
   - → specifieke voorwaarden
3) Bepalen v.d. teststatistiek
   - ↳ statistiek met gekende verdeling
   - → afhankelijk van de situatie
4) Berekenen v.d. testwaarde
   - → significantieniveau α
5) Aanvaardingsgebied / p-waarde
   - → AG tweezijdige test:
     - ↳ we kunnen steeds a_1 en a_2 vinden zodat
       $P(a_1 \leq T \leq a_2) = 1-\alpha \mid H_0$
       => we kunnen steeds een interval opstellen zodat de kans dat, onder H_0, de waarde van de teststatistiek in het interval valt steeds gelijk is aan $1-\alpha$.
     - => Beslissingsregel:
       - $T \in [a_1, a_2]$ → H_0 niet verwerpen: zwak besluit
       - $T \notin [a_1, a_2]$ → H_0 verwerpen: sterk besluit
   - → complement aanvaardingsgebied = verwerpingsgebied
     - ↳ scheidingspunten: kritische punten
   - → testen kunnen ook uitgevoerd $\bar{w}$ a.d.h.v. BI
   - → AG eenzijdige test

	rechtseenzijdig	**linkseenzijdig**
Onder H_0	$P(T \leqslant a_3) = 1-\alpha$	$P(T \geqslant a_4) = 1-\alpha$
Aaanvaardingsgebied	$]-\infty, a_3]$	$[a_4, \infty[$
Verwerpingsgebied	$]a_3, \infty[$	$]-\infty, a_4[$

→ p-waarde
↳ aanvaardingsgeb. niet opn. ber. bij andere α

def. De p-waarde of overschrijdingskans is de kans, onder H_0, dat de teststatistiek T in een even grote steekproef een waarde aanneemt die minstens zo extreem ligt in de richting v.h. alternatief als de waargenomen testwaarde.
$p < \alpha \iff$ verwerp H_0 op significantieniveau α

⁉ p onafh. v. α
↳ kleinste niveau waarop H_0 $\bar{\omega}$ verworpen

6] Besluit formuleren
↳ 3 delen:
- testwaarde in aanvaardingsgeb. / $p > \alpha$?
- H_0 verwerpen of niet?
- inhoudelijk besluit

μ normale verd., σ^2 gekend

Tweezijdige z-test

1] $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

2] willekeurig ?
$\bar{X}$ normaal verdeeld ?

3] $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$ $\quad | H_0 \rightarrow$ "onder nulhypothese"

4] $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

5]

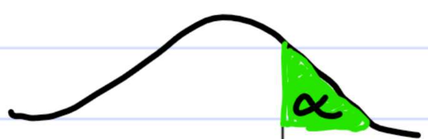
	tweezijdig
Onder H_0	$P(-z_{1-\alpha/2} \leqslant Z \leqslant z_{1-\alpha/2}) = 1-\alpha$
AG	$[-z_{1-\alpha/2}, z_{1-\alpha/2}]$
VG	$]-\infty, -z_{1-\alpha/2}[\cup]z_{1-\alpha/2}, \infty[$

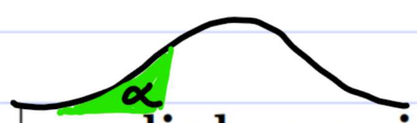
$p = 2P(Z \geq |z|)$

Eenzijdige z-test

1] rechtseenzijdig:
$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

linkseenzijdig:
$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$





5]

	rechtseenzijdig	linkseenzijdig
Onder H_0	$P(Z \leqslant z_{1-\alpha}) = 1-\alpha$	$P(Z \geqslant -z_{1-\alpha}) = 1-\alpha$
AG	$]-\infty, z_{1-\alpha}]$	$[-z_{1-\alpha}, +\infty[$
VG	$]z_{1-\alpha}, +\infty[$	$]-\infty, -z_{1-\alpha}[$

$p = P(Z \geq z)$ $\qquad$ $p = P(Z \leq z)$

μ norm. verd., σ^2 ongekend

t-test

1] en 2] analoog aan vorige

3] $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim_{H_0} t_{n-1}$

4] $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

5]

	tweezijdig
Onder H_0	$\mathrm{P}(-t_{n-1,1-\frac{\alpha}{2}} \leqslant T \leqslant t_{n-1,1-\frac{\alpha}{2}}) = 1-\alpha$
AG	$\left[-t_{n-1,1-\frac{\alpha}{2}}, t_{n-1,1-\frac{\alpha}{2}}\right]$
VG	$]-\infty, -t_{n-1,1-\frac{\alpha}{2}}[\cup]t_{n-1,1-\frac{\alpha}{2}}, \infty[$

	rechtseenzijdig	**linkseenzijdig**
Onder H_0	$\mathrm{P}(T \leqslant t_{n-1,1-\alpha}) = 1-\alpha$	$\mathrm{P}(T \geqslant -t_{n-1,1-\alpha}) = 1-\alpha$
AG	$]-\infty, t_{n-1,1-\alpha}]$	$[-t_{n-1,1-\alpha}, \infty[$
VG	$]t_{n-1,1-\alpha}, \infty[$	$]-\infty, -t_{n-1,1-\alpha}[$

	tweezijdig	**rechtseenzijdig**	**linkseenzijdig**
p-waarde	$2\mathrm{P}(T \geqslant \lvert t \rvert)$	$\mathrm{P}(T \geqslant t)$	$\mathrm{P}(T \leqslant t)$

Fouten

↳ foute conclusie

→ 2 mogelijke fouten: type I en type II fout

	verwerp H_0	verwerp H_0 niet
H_0 waar	Type I fout (α)	geen fout ($1-\alpha$)
H_0 niet waar	geen fout ($1-\beta$)	Type II fout (β)

zie p. 66 bovenaan

H_0: niet zwanger

H_1: zwanger

Type I fout

$P(\text{type-I fout}) = P(H_0 \text{ verwerpen} \mid H_0 \text{ waar})$

$= P(T \text{ in verwerpingsgeb.} \mid \mu = \mu_0)$

$= P(T \in]-\infty, t_{n-1,1-\frac{\alpha}{2}}[\cup]t_{n-1,1-\frac{\alpha}{2}}, \infty[\mid \mu = \mu_0)$

$= \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha \quad (T \sim t_{n-1} \text{ als } \mu = \mu_0)$

Type II-fout

↳ $\beta = P(\text{type II fout})$, $1-\beta$ is de power/onderscheidingsvermogen

↳ kan niet "gekozen" worden

→ enkel bepalen voor $H_1: \mu = \mu_1 \neq \mu_0$

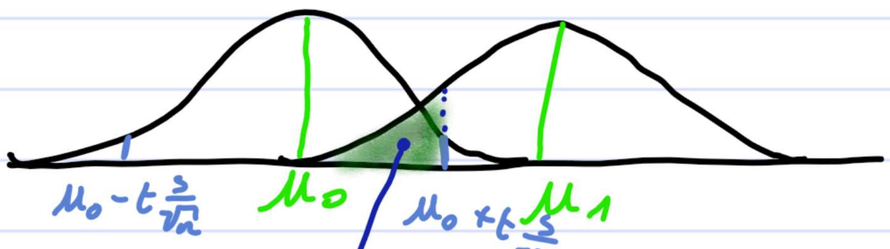
$P(\text{type II-fout} \mid \mu = \mu_1) = P(H_0 \text{ niet verwerpen} \mid H_1: \mu = \mu_1 \text{ waar})$

$= P(-t_{n-1,1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \leq t_{n-1,1-\frac{\alpha}{2}} \mid \mu = \mu_1)$

voor $\mu = \mu_1$: $T = \frac{\bar{x} - \mu_1}{s/\sqrt{n}} \sim t_{n-1}$

$\Rightarrow \beta = P(-t_{n-1,1-\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}} \leq \frac{\bar{x} - \mu_0}{s/\sqrt{n}} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}} \leq t_{n-1,1-\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}} \mid \mu = \mu_1)$

$= P(-t_{n-1,1-\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}} \leq T \leq t_{n-1,1-\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}})$



↳ kans op type II-fout → verkleinen door:

- $s \downarrow$
- $n \uparrow$
- $\alpha \uparrow$ ⁉ type I ↑

→ α beter begr. dan β => H_0 & H_1 zo kiezen zodat fouten v.d. 1e soort "belangrijker zijn.

→ onterecht verwerpen is erger

Normaliteitsassumptie

$\begin{cases} H_0: \text{geg. zijn normaal verdeeld} \\ H_1: \text{geg. zijn niet normaal verdeeld} \end{cases}$

↳ verwerp H_0 als r_Q (correlatiecoëff. v. normale kwantielplot) $<<1$

→ X niet normaal verdeeld?

↳ logaritme wel? → $\begin{cases} H_0: \mu_{\log(x)} = \log(\mu_0) \\ H_1: \mu_{\log(x)} \neq \log(\mu_0) \end{cases}$ $\Rightarrow \text{Med}(X) = e^{\mu_0}$ $\not\Leftrightarrow$ $\begin{cases} H_0: \mu_X = e^{\mu_0} \\ H_1: \mu_X \neq e^{\mu_0} \end{cases}$

Variantie van normale verd.

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ met μ en σ onbehend

1] & 2] zoals voorheen

3] $X^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim_{H_0} X^2_{n-1}$

4] $X^2 = \frac{(n-1)s^2}{\sigma_0^2}$

5]

	tweezijdig
Onder H_0	$P(\chi^2_{n-1,\frac{\alpha}{2}} \leqslant X^2 \leqslant \chi^2_{n-1,1-\frac{\alpha}{2}}) = 1-\alpha$
AG	$\left[\chi^2_{n-1,\frac{\alpha}{2}}, \chi^2_{n-1,1-\frac{\alpha}{2}}\right]$
VG	$[0, \chi^2_{n-1,\frac{\alpha}{2}}[\cup]\chi^2_{n-1,1-\frac{\alpha}{2}}, +\infty[$

$p = 2\,min\{P(X^2 \geq x^2), P(X^2 \leq x^2)\}$

	rechtseenzijdig	linkseenzijdig
Onder H_0	$P(X^2 \leqslant \chi^2_{n-1,1-\alpha}) = 1-\alpha$	$P(X^2 \geqslant \chi^2_{n-1,\alpha}) = 1-\alpha$
AG	$[0, \chi^2_{n-1,1-\alpha}]$	$[\chi^2_{n-1,\alpha}, +\infty[$
VG	$]\chi^2_{n-1,1-\alpha}, +\infty[$	$[0, \chi^2_{n-1,\alpha}[$

$p = P(X^2 \geq x^2)$ $\quad$ $p = P(X^2 \leq x^2)$

Proportie

$X_1, \dots, X_n \sim B(1, p)$

↳ $\hat{P} = \frac{1}{n}\sum X_i$ of als n voldoende groot: $\hat{P} \approx N(p, \frac{p(1-p)}{n})$

=> 2 mogelijke testen

Exacte test

3] $X \sim_{H_0} B(n, p_0)$

4] $x = n\hat{p}$ = #successen in de steekproef

→ enkel p-waarde (AG ingewikkeld door discrete verd.)

$p = P(|X - np_0| \geq |x - np_0|)$

$= P(X - np_0 \geq |x - np_0|) + P(X - np_0 \leq -|x - np_0|)$

2 mogelijkheden:

① $x \geq np_0$

=> $P(X - np_0 \geq x - np_0) + P(X - np_0 \leq -x + p_0) = p$

$= P(X \geq x) + P(X \leq 2np_0 - x$

② $x \leq np_0$

=> $p = P(X \leq x) + P(X \geq 2np_0 - x)$

5]

	tweezijdig
$x \geqslant np_0$	$P(X \geqslant x) + P(X \leqslant 2np_0 - x)$
$x \leqslant np_0$	$P(X \leqslant x) + P(X \geqslant 2np_0 - x)$

rechtseenzijdig	linkseenzijdig
$P(X \geqslant x)$	$P(X \leqslant x)$

Benaderende test

2) ▽ $n \geq 30$, $np \geq 5$ en $n(1-p) \geq 5$ → p benaderen door $\hat{p}$

3) $Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx_{H_0} \mathcal{N}(0,1)$ $\quad E[\hat{P}] = p_0$, $Var[\hat{P}] = \frac{p_0(1-p_0)}{n}$

4) $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

5)

	tweezijdig
Onder H_0	$\mathrm{P}(-z_{1-\alpha/2} \leqslant Z \leqslant z_{1-\alpha/2}) = 1-\alpha$
AG	$]-z_{1-\alpha/2}, z_{1-\alpha/2}]$
VG	$]-\infty, -z_{1-\alpha/2}[\cup]z_{1-\alpha/2}, \infty[$

$p = 2P(Z \geq |z|)$

Onder H_0	rechtseenzijdig $\mathrm{P}(Z \leqslant z_{1-\alpha}) = 1-\alpha$	linkseenzijdig $\mathrm{P}(Z \geqslant -z_{1-\alpha}) = 1-\alpha$
AG	$]-\infty, z_{1-\alpha}]$	$[-z_{1-\alpha}, \infty[$
VG	$]z_{1-\alpha}, \infty[$	$]-\infty, -z_{1-\alpha}[$

$p = P(Z \geq z)$ $\quad p = P(Z \leq z)$

Vergelijken van varianties

$X_1, \ldots, X_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$, μ_1 en σ_1 onbekend

$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu_2, \sigma_2^2)$, μ_2 en σ_2 onbekend

→ σ_1^2 en σ_2^2 schatten:

$S_1^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X}_n)^2$

$S_2^2 = \frac{1}{m-1} \sum_j (Y_j - \bar{Y}_m)^2$

st. als σ_1 en σ_2 onbekend, dan: $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1,m-1}$

$\Rightarrow P(F_{n-1,m-1,\frac{\alpha}{2}} \leq \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq F_{n-1,m-1,1-\frac{\alpha}{2}}) = 1-\alpha$

met $F_{n,m,\frac{\alpha}{2}} = \frac{1}{F_{m,n,1-\frac{\alpha}{2}}}$

$\Rightarrow$ BI voor σ_1^2/σ_2^2:

$\left[\frac{S_1^2}{S_2^2} \frac{1}{F_{n-1,m-1,1-\frac{\alpha}{2}}}, \frac{S_1^2}{S_2^2} F_{n-1,m-1,1-\frac{\alpha}{2}}\right]$

→ hypothesetest

1) $\sigma_1^2 = \sigma_2^2$ komt overeen met $\frac{\sigma_1^2}{\sigma_2^2} = 1$, anderen kunnen op analoge manier herschreven worden

▽ let op nummering groepen

2) • willekeurig ?
• normaal verd. (voor beide groepen)?
• kwantitatief en groepen onafh. ?
• geen uitschieters?

3) omdat $\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n-1}$ en $\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2_{m-1}$ met S_1^2 en S_2^2 onafh.

$F = \frac{S_1^2}{S_2^2} \sim_{H_0} F_{n-1,n-2}$

4) $f = \frac{s_1^2/s_2^2}{}$

5)

	tweezijdig
Onder H_0	$\mathrm{P}(F_{n_1-1,n_2-1,\frac{\alpha}{2}} \leqslant F \leqslant F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}) = 1-\alpha$
AG	$\left[F_{n_1-1,n_2-1,\frac{\alpha}{2}}, F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}\right]$
VG	$[0, F_{n_1-1,n_2-1,\frac{\alpha}{2}}[\cup]F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}, +\infty[$

	rechtseenzijdig	linkseenzijdig
O. H_0	$\mathrm{P}(F \leqslant F_{n_1-1,n_2-1,1-\alpha}) = 1-\alpha$	$\mathrm{P}(F \geqslant F_{n_1-1,n_2-1,1-\alpha}) = 1-\alpha$
AG	$[0, F_{n_1-1,n_2-1,1-\alpha}]$	$[F_{n_1-1,n_2-1,\alpha}, +\infty[$
VG	$]F_{n_1-1,n_2-1,1-\alpha}, +\infty[$	$[0, F_{n_1-1,n_2-1,\alpha}[$

tweezijdig	
$f > 1$	$2P(F \geqslant f)$
$f \leqslant 1$	$2P(F \leqslant f)$

rechtseenzijdig	linkseenzijdig
$P(F \geqslant f)$	$P(F \leqslant f)$

↳ benadering

Vergelijken van gemiddelden

→ onderscheid gepaarde en ongepaarde waarnemingen

- geen verband tussen $X_1,\dots,X_n$ en $Y_1,\dots,Y_m$ => ongepaard
- gepaard als verband tussen elke X_i en een Y_j
  => $n=m$

→ checken via boxplots

Ongepaarde waarnemingen

1 & 2 zie vorige

3) onderscheid gelijke/ongelijke varianties

- gelijke varianties

als $\sigma=\sigma_1=\sigma_2$, dan $\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2, \frac{\sigma^2}{n}+\frac{\sigma^2}{m})$

of $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{1/n+1/m}}\sim N(0,1)$

→ σ^2 schatten met gewogen/gepoolde/gemengde variantie

$S^2=\frac{(n-1)S_1^2+(m-1)S_2^2}{(n-1)+(m-1)}$ → $E[S^2]=\sigma^2$ (onvertekende sch.)

bew. p. 92 St. $T=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t_{(n-1)+(m-1)}=t_{n+m-2}$

=> BI: $[\bar{x}-\bar{y}\pm t_{n+m-2,1-\frac{\alpha}{2}}\,s\sqrt{\frac{1}{n}+\frac{1}{m}}]$ ← voor $\mu_1-\mu_2$

→ hypothesetest

3) $T=\frac{\bar{X}-\bar{Y}}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim_{H_0} t_{n+m-2}$ 4) $t=\frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n}+\frac{1}{m}}}$

5)

	tweezijdig
Onder H_0	$P(-t_{n_1+n_2-2,1-\frac{\alpha}{2}}\leqslant T\leqslant t_{n_1+n_2-2,1-\frac{\alpha}{2}})=1-\alpha$
AG	$\left[-t_{n_1+n_2-2,1-\frac{\alpha}{2}}, t_{n_1+n_2-2,1-\frac{\alpha}{2}}\right]$
VG	$]-\infty,-t_{n_1+n_2-2,1-\frac{\alpha}{2}}[\cup]t_{n_1+n_2-2,1-\frac{\alpha}{2}},\infty[$

$p=2P(T\geq|t|)$

Onder H_0	rechtseenzijdig $P(T\leqslant t_{n_1+n_2-2,1-\alpha})=1-\alpha$	linkseenzijdig $P(T\geqslant -t_{n_1+n_2-2,1-\alpha})=1-\alpha$
AG	$]-\infty,-t_{n_1+n_2-2,1-\alpha}]$	$[-t_{n_1+n_2-2,1-\alpha},\infty[$
VG	$]t_{n_1+n_2-2,1-\alpha},\infty[$	$]-\infty,t_{n_1+n_2-2,1-\alpha}[$

$p=P(T\geq t)$ $p=P(T\leq t)$

→ hypothesetest ongelijke varianties

- n en m groot => $S_1^2\approx\sigma_1^2$, $S_2^2\approx\sigma_2^2$
  => $\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m})$ of $Z=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\sim N(0,1)$
- voor kleinere steekproeven:

3) $T=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}\approx t_r$ met $r=\frac{(\frac{S_1^2}{n}+\frac{S_2^2}{m})^2}{\frac{(S_1^2/n)^2}{n-1}+\frac{(S_2^2/m)^2}{m-1}}$

4) $t=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}$

5)

	tweezijdig
Onder H_0	$P(-t_{r,1-\frac{\alpha}{2}}\leqslant T\leqslant t_{r,1-\frac{\alpha}{2}})=1-\alpha$
AG	$\left[-t_{r,1-\frac{\alpha}{2}}, t_{r,1-\frac{\alpha}{2}}\right]$
VG	$]-\infty,-t_{r,1-\frac{\alpha}{2}}[\cup]t_{r,1-\frac{\alpha}{2}},\infty[$

$p=2P(T\geq|t|)$

Onder H_0	rechtseenzijdig $P(T\leqslant t_{r,1-\alpha})=1-\alpha$	linkseenzijdig $P(T\geqslant -t_{r,1-\alpha})=1-\alpha$
AG	$]-\infty,-t_{r,1-\alpha}]$	$[-t_{r,1-\alpha},\infty[$
VG	$]t_{r,1-\alpha},\infty[$	$]-\infty,t_{r,1-\alpha}[$

$p=P(T\geq t)$ $p=P(T\leq t)$

Gepaarde waarnemingen

→ nieuwe variabele invoeren: $D_i = X_i - Y_i$ $(n = m)$

$\sigma^2 \neq \sigma_1^2 + \sigma_2^2$ want afh.

↳ als (X_i, Y_i) bivariaat normaal verdeeld zijn, dan $D_1, \ldots, D_n \sim N(\mu, \sigma^2)$ met onbekende $\sigma^2 \neq \sigma_1^2 + \sigma_2^2$ en $\mu = \mu_1 - \mu_2$

→ voer t-test voor μ bij norm. verd. uit (voor D, $\mu = 0$)

Vergelijken van twee proporties

→ onafhankelijke experimenten: $p = p_1 - p_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2 = \frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}$

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}} \approx N(0,1)$$

→ BI: $\left[(\hat{p}_1 - \hat{p}_2) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} \right]$

→ hypothesetest

↳ 2 manieren

① gegevens van BI gebr. en met normale verdeling werken

② met gemengde proportie werken

omdat getest wordt of $p_1 = p_2$

→ $\hat{p}_0 = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m}$

3) $$Z = \frac{\hat{P}_1 - \hat{P}_2}{\hat{P}_0(1-\hat{P}_0)\sqrt{\frac{1}{n} + \frac{1}{m}}} \approx N(0,1)$$

Correlatie tussen twee variabelen

Kwantitatieve variabelen

↳ populatiecorrelatiecoëfficiënt ρ, als $\rho \neq 0 \Rightarrow$ X en Y afhankelijk

1) $\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$ → als H_0 verworpen ⇒ lineaire afh.

H_0 niet verw. ⇏ Lin. onafh.

⚠ enkel tweezijdige test

2) · X en Y bivariaat normaal verdeeld?

↑ X en Y normaal verdeeld (grafisch/test)

↑ scatterplot (elliptische wolk)

· willekeurig?

3) R = steekproefcorrelatiecoëfficiënt

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim_{H_0} t_{n-2}$$

4) $$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

5) $\left[-t_{n-2,1-\frac{\alpha}{2}}, t_{n-2,1-\frac{\alpha}{2}} \right]$

$p = 2P(T \geq |t|)$

Nominale variabelen

1) $\begin{cases} H_0: X \text{ en } Y \text{ onafhankelijk} \\ H_1: X \text{ en } Y \text{ afhankelijk} \end{cases}$

⚠ weer enkel tweezijdig

2) · willekeurig?

· X en Y kwalitatief met minstens 2 mog. uithomsten?

· klassen disjunct?

· som v. alle frequenties > 20?

· frequentie per categorie ≥ 5? (samennemen)

· alle waarnemingen onafhankelijk?

↑ met bijdrage tot frequentie

Er moet gelden dat $P(X = m_x, Y = m_y) = P(X = m_x) P(Y = m_y)$
voor onafhankelijkheid.
↳ meestal beschikken we niet over de juiste geg.

zie p. 108

=> benadering: $\frac{n_{xy}}{n} = \frac{n_x}{n} \frac{n_y}{n}$

→ komt overeen met wat we eerder zagen:

verwachte waarde = $n_{xy} = \frac{\text{rijtotaal} \times \text{kolomtotaal}}{n} = \frac{n_x n_y}{n} = \frac{n_x}{n} \frac{n_y}{n} n$

- waargenomen frequenties O_i
- verwachte frequenties E_i

=> $n = \sum_i O_i = \sum_i E_i$

zie redenering p. 109

3) $X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \approx_{H_0} X^2_{(n_1 - 1)(n_2 - 1)}$

#rijen #kolommen in kruistabel

4) $X^2 = \sum_i \frac{(o_i - e_i)^2}{e_i}$

5) $\square$ 0 : $X^2_{(m_1 - 1)(m_2 - 1), 1-}$

Regressie

Introductie

→ onderzoek of/hoe Y op een lineaire manier voorspeld kan worden door X met lineaire regressie
- Y = responsvariabele / afh. var.
- X = verklarende / onafh. var.

→ Hoe ziet dit verband eruit? Welk model gebr?
↳ ideaal: $y_i = \alpha + \beta x_i$
↳ in hoeverre wijkt de realiteit af?

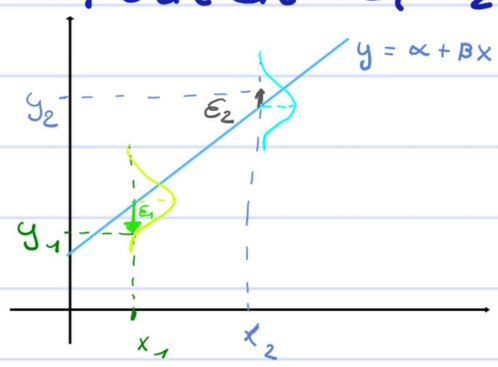
→ Welke vwd^{n} voor goed model?
↳ hoe nagaan?

→ Hoe betrouwbaar?
↳ welke afwijkingen?

Enkelvoudig lineair regressiemodel

$Y_i = \alpha + \beta x_i + \varepsilon_i$

↳ fouten ε_i zijn onafhankelijk en $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$
↳ zelfde voor alle ε_i



→ opmerkingen
- $E(Y_i | X_i) = E(\alpha + \beta x_i + \varepsilon_i) = \alpha + \beta x_i + \overbrace{E[\varepsilon_i]}^{=0} = \alpha + \beta_i$
  ↳ β geeft invloed van x_i op gem. v. Y_i
- $Var[Y_i] = Var[\varepsilon_i] = \sigma_\varepsilon^2$ => homoscedasticiteit
- geen veronderstellingen over x_i en marginale verd. v. Y.

Kleinste kwadratenmethode

→ Hoe $y = \alpha + \beta x$ bepalen?
→ schattingen $\hat{\alpha}$ en $\hat{\beta}$ voor α en β
- definieer $e_i(a,b)$ als vert. afw. van (x_i, y_i) t.o.v. $y = a + bx$
  $e_i(a,b) = y_i - (a + b x_i)$

→ zoek $\hat{\alpha}$ en $\hat{\beta}$ waarvoor $\sum_i e_i^2(a,b) = \sum_i (y_i - (a + b x_i))^2$ minimaal

$$\Rightarrow \hat{\beta} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \underbrace{r}_{\text{correlatie}} \frac{s_y}{s_x}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

⚠ uitschieters weggooien

⇒ geschatte rechte: $y = \hat{\alpha} + \hat{\beta} x$
- schatting voor elke obs: $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$
- residu's: $e_i = e_i(\hat{\alpha}, \hat{\beta}) = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$
- $\sum_i e_i = 0$ altijd!
- $(\bar{x}, \bar{y})$ ligt op $y = \hat{\alpha} + \hat{\beta} x$
- als $r = r(x,y) = 0$, dan horizont. rechte ($\hat{\beta} = 0$ en $\hat{\alpha} = \bar{y}$)
  => geen lin. verb, dan $\bar{y}$ beste voorsp. voor $Y | x_i$ ongeacht waarde van x_i

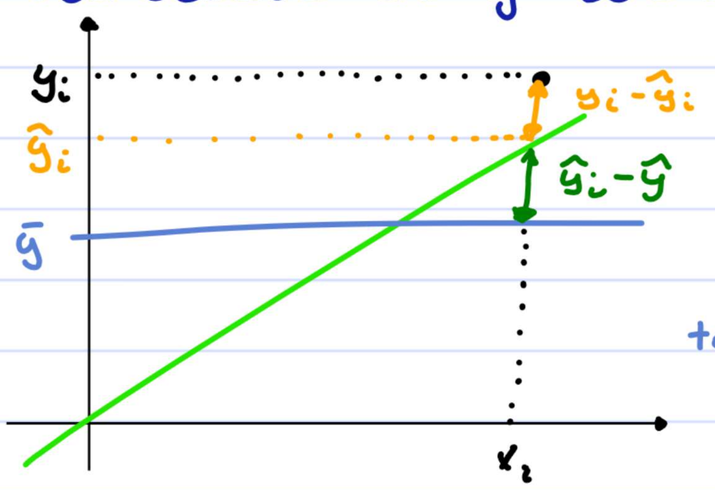
→ Schatting voor σ_ε^2!
↳ $E(\varepsilon_i) = 0 \Rightarrow \sigma_\varepsilon^2 = Var(\varepsilon_i) = E(\varepsilon_i^2)$

$$s_\varepsilon^2 = \hat{\sigma}_\varepsilon^2 = \frac{1}{n-2} \sum_i e_i^2(\hat{\alpha}, \hat{\beta}) = \frac{1}{n-2} \sum_i (y_i - \hat{y}_i)^2$$

↳ (

ANOVA-tabel en determinatiecoëfficiënt

Determinatiecoëfficiënt: geeft weer hoe goed het model de variatie in y-waarden verklaart



$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \text{ met } \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

$$\Rightarrow \sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

$$SST = SSM + SSE$$

total sum of squares = mean sum of squares + error sum of sq.

totale variatie = var verkl. door model + onverkl. var.

→ hoe groter aandeel SSM in SST, hoe beter het model

$\Rightarrow R^2 = \frac{SSM}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$

- enkelvoudige lineaire regressie: $R^2 = r^2$, dus $0 \leq R^2 \leq 1$
- $R^2 = 1 \Rightarrow$ perfecte rechte
- $R^2 \approx 0 \Rightarrow \hat{\beta} = r \frac{s_y}{s_x} \approx 0$

→ andere methode: $f = \frac{MSM}{MSE} = \frac{\text{model mean sq.}}{\text{mean error sq.}} = \frac{\frac{SSM}{1}}{\frac{SSE}{n-2}} = \frac{\sum (y_i - \bar{y})^2}{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$

- $MSE = s_\varepsilon^2$

⇒ samenvatting in **ANOVA-tabel**:

	vrijheidsgraden	SS	MS	F
Model	1	SSM	$MSM = \frac{SSM}{1}$	$f = \frac{MSM}{MSE}$
Error	$n-2$	SSE	$MSE = \frac{SSE}{n-2}$	
Total	$n-1$	SST		

Nagaan v.d. modelveronderstellingen

↳ Lineair regressiemodel soms niet de beste optie

⇒ Wanneer is Lineair goed?

↳ niet goed als R^2 klein

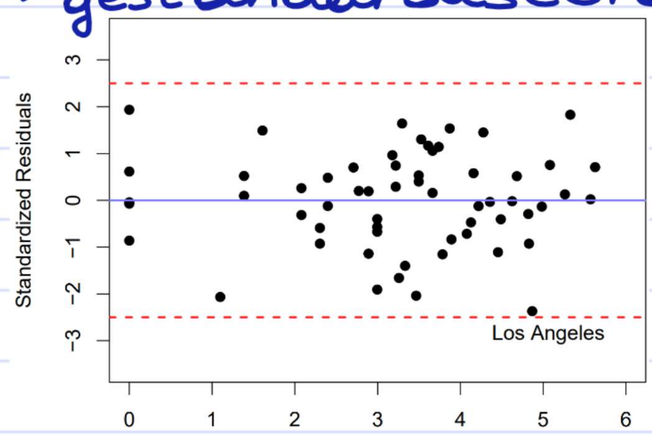
↳ onafhankelijkheid van ε_i en $N(0, \sigma_\varepsilon^2)$-verdeling van ε_i onderzoeken m.b.v. residu's $e_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$

1) ε_i normaal verdeeld met $\mu = 0$: normale kwantielplot v.d. residu's e_i

- intercept ≈ 0
- normaliteit niet formeel testen

2) ε_i onderling onafhankelijk met dezelfde var. σ_ε^2

→ gestandardiseerde residuplot (x_i, e_i)



→ soms ook $(\hat{y}_i, e_i)$

→ ongeordend lukraak persoon

→ gelijke spreiding

3) nagaan of er uitschieters zijn in de gestandaardiseerde residu's: $\frac{e_i}{s_{E_i}} \in [-2,5; 2,5]$?

↳ $E_i = Y_i - (\hat{\alpha} - \hat{\beta} x_i)$ toevalsvariabele

- $s_{E_i} = s_\varepsilon \sqrt{1 - 1/n - \frac{(x_i - \bar{x})^2}{(n-1) s_x^2}}$

Interferentie omtrent regressieparameters

- kleinste kwadratenschatters voor α en β: $\hat{A}$ en $\hat{B}$
  - $\text{Var}[\hat{A}] = \sigma_\varepsilon^2\left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}\right) \Rightarrow \text{s.e.}(\hat{A}) = S_\varepsilon\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}}$
  - $\text{Var}[\hat{B}] = \sigma_\varepsilon^2/(n-1)s_x^2 \Rightarrow \text{s.e.}(\hat{B}) = S_\varepsilon\sqrt{\frac{1}{(n-1)s_x^2}}$
  - $\text{Cov}(\hat{A},\hat{B}) = -\frac{\bar{x}\,\sigma_\varepsilon^2}{(n-1)s_x^2}$
  - $\frac{\hat{A}-\alpha}{\text{s.e.}(\hat{A})} \sim t_{n-2}$
  - $\frac{\hat{B}-\beta}{\text{s.e.}(\hat{B})} \sim t_{n-1}$
  - BI voor α: $\left[\hat{\alpha} \pm t_{n-2,\alpha/2}\, S_\varepsilon\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}}\right]$
  - BI voor β: $\left[\hat{\beta} \pm t_{n-2,\alpha/2}\, \frac{s_e}{\sqrt{(n-1)s_x^2}}\right]$
- Als $\beta = 0$ kan y niet lineair uit x voorspeld worden
  - nagaan of lineaire regressie zinvol is met hypothesetest
    - $H_0: \beta = 0$
    - $H_1: \beta \neq 0$
  - $T = \frac{\hat{B}-\beta}{\text{s.e.}(\hat{B})} \sim t_{n-2}$, $p = 2P(T \geq |t|)$

  of
  - $F = \frac{MSM}{MSE} \sim_0 F_{1,n-2}$, $p = P(F \geq f)$
- $t^2 = f$

	Schatting	Standaardfout	Testwaarde	P-waarde
Intercept	$\hat{\alpha}$	s.e.$(\widehat{A})$	$t = \frac{\hat{\alpha}}{\text{s.e.}(\widehat{A})}$	$2P(T > \lvert t\rvert)$
Slope	$\hat{\beta}$	s.e.$(\widehat{B})$	$t = \frac{\hat{\beta}}{\text{s.e.}(\widehat{B})}$	$2P(T > \lvert t\rvert)$