

# Samenvatting: Statistiek 1

by

lekkernederlands



The Marketplace to Buy and Sell your Study Material

Buy and sell all your summaries, notes, theses, essays, papers, cases, manuals, researches, and  
many more..

[www.stuvia.com](http://www.stuvia.com)

**College 1**

- **Populatie** = gehele groep elementen
- **Steekproef** = gedeelte van populatie dat onderzocht wordt
- **Eenheden** = elementen waarvan gegevens worden verzameld
- **Variabelen** = element uit steekproef dat wordt bepaald
  - **Kwantitatieve** variabelen: getal met waarde, bijv. telling
    - Continu: alle getallen
    - Discreet: alleen hele getallen
      - Weergeven: frequentie tabel (met klassen), histogram
  - **Kwalitatieve** variabelen: geen getal
    - Nominaal: geen voorkeur
    - Ordinaal: wel gradatie aangegeven
      - Weergeven: frequentie tabel, staafdiagram
- Steekproef heeft kans op vertekening -> dat de gehele populatie niet accuraat wordt weergegeven
  - Vertekening voorkomen:
    - Enkelvoudige Aselecte Steekproef (ESA)
      - Willekeurig aantal eenheden uit populatie genomen
      - Alle mogelijke combinatie even groot om aanwezig te zijn
      - Voorkomt vertekening door toeval gekozen worden
  - Voorbeelden vertekening:
    - Onvolledige dekking: bepaalde groepen niet weergegeven
    - Non-respons: mensen werken niet mee of zijn onbereikbaar
    - Vrijwillige reactie: meestal uitgesproken negatieve reactie
    - Vertekening van reactie: sociale wenselijkheid
- Bij observationeel onderzoek alleen maar waarneming
- Bij experimenteel onderzoek opzettelijke behandeling toekennen om reactie waar te nemen
  - Oorzaak-gevolg reactie vaststellen alleen bij **experimenteel** onderzoek
- **Relatieve frequentie**: fractie van iets gedeeld door het totaal -> komt altijd uit op 1
- Pieken in een histogram: 1 piek= unimodal, 2 pieken = bimodal allemaal gelijk = uniform
- Skewed to the left: heeft links staartje                      skewed to the right: rechts staartje

**College 2**

- Voorselectie? Dan **geen** EAS
- Numerieke maten voor centrum: gemiddelde + mediaan
  - Gemiddelde:  $\frac{1}{n} \sum y_i$                        $y_i = Y_1 + Y_2 + Y_3 \dots Y_n$
  - Mediaan: gemiddelde van alle waarnemingen als ze op grootte geordend zijn
    - **niet** gevoelig voor uitbijters
- Standaard Deviatie =  $\sqrt{\text{variantie}}$ 
  - Variantie =  $\frac{1}{n-1} \sum (y_i - \bar{y})^2$
- Interkwartielafstand = IKA =  $Q_3 - Q_1$                       **Niet** gevoelig voor uitbijters
  - Maximum/minimum waarde kan ook uitbijter zijn
- Empirical rule = 68-95-99,7-regel

## Statistiek 1

- $68 = \bar{y} - s, \bar{y} + s$
- $95 = \bar{y} - 2s, \bar{y} + 2s$
- $99,7 = \bar{y} - 3s, \bar{y} + 3s$
- Wet van grote getallen: relatieve frequentie stabiliseert bij grotere aantallen
- Schatter =  $\hat{p} = k/n$ 
  - $k/n =$  consistentie schatter
    - des te groter steekproef, des te dichter bij onbekende waarde
  - $\hat{p}$  als toevalsvariabele dan **schatter**, specifieke uitkomst dan **schatting**
- $P(\text{gebeurtenis } A) = \frac{\text{aantal uitkomsten } A}{\text{totaal aantal mogelijke uitkomsten}}$
- $\bar{A}$  = uitkomsten die **niet** in A voorkomen
- $A \cup B =$  alleen in A, alleen in B of in A en B
  - **Disjunct** = A en B geen gemeenschappelijke gebeurtenis (no overlap ven diagram)
- $A \cap B =$  uitkomsten in zowel A als B
- Voor **elke** gebeurtenis =  $0 \leq P(A) \leq 1$
- Complementen regel:  $P(\bar{A}) = 1 - P(A)$
- **Range** = verschil tussen grootste en kleinste waarde van een set waarnemingen
- **Percentile** = heeft p% van waarnemingen eronder en 100-p% van waarnemingen erboven

College 3

- Product regel onafhankelijke producten:  
 $P(A \cap B) = P(A) * P(B)$
- Algemene optelregel:  
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Binominaal coëfficiënt  $\sum_k \binom{n}{k} \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k}$   
n = aantal steekproef elementen      k = aantal successen
- Mutually exclusive = disjunct = als disjunct dan kans dat A of B gebeurt =  $P(A) + P(B)$
- Kansverdeling verschilt voor discrete en continue variabelen
  - Bij discrete variabelen kunnen kansverdelingen bij elkaar worden opgeteld

College 4

- Continue toevalsvariabelen
  - Kan alle waardes aannemen
  - Kan niet beschreven worden door losse uitkomsten
    - Exacte verdeling niet bekend
  - Veel waarnemingen >> klassebreedte in histogram kleiner >> histogram wordt een curve
- Bin size = staafjes breedte
- Kans is relatieve frequentie op lange termijn
  - $P(A)$  gegeven door oppervlakte onder de curve en boven A
- **Discreet:**  $P(y \leq 5) \neq P(y < 5)$
- **Continu:**  $P(y \leq 3/50) = P(y < 3/50)$        $P(y=3/50) \Rightarrow 0$
- Normale verdeling: symmetrisch, 1-toppig, klokvormig,      **notatie:**  $N(\mu, \sigma)$
- Standaard normale verdeling:  $Z \sim N(0,1)$

## Statistiek 1

- $P(Z \geq 3,05) = 1 - P(Z \leq 3,05)$       want tabel 1 geeft alleen *kleiner dan* waarden
- Transformatie naar standaard normaal verdeling
  - $Y \sim N(\mu, \sigma) \rightarrow Z \sim N(0,1)$
  - $Y = \mu + \sigma Z$       ook wel       $Z = \frac{y-\mu}{\sigma}$
- Discrete toevalsvariabele is gelijk aan de som van de mogelijke uitkomsten vermenigvuldigd met de bijbehorende kansen
- Een steekproefgemiddelde is consistente schatter voor populatiegemiddelde =  $\bar{y}$

**College 5**

## Rekenregels voor verwachtingen

- **Regel 1:** y toevals variabele, a en b vaste gevallen  
 $\mu_{a+by} = a + b\mu_y$
- **Regel 2:** als x en y toevalsvariabele zijn  
 $\mu_{x+y} = \mu_x + \mu_y$
- Als je alles hetzelfde doet, blijft spreiding hetzelfde

## Rekenregels voor variantie

- **Regel 1:**  $\sigma_{a+by} = |b| \sigma_y$       en dus  $\sigma_{a+by}^2 = b^2 \sigma_y^2$
- **Regel 2:** x en y onafhankelijke toevalsvariabelen  
 $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$        $\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2$   
**varianties altijd bij elkaar optellen**
- Regels 1 en 2 gecombineerd  
 $\mu_{ax+by+c} = a\mu_x + b\mu_y + c$
- Onafhankelijk  
 $\sigma_{ax+by+c}^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2$
- Verwachting en standaardafwijking van som van **onafhankelijke** trekkingen uit een zelfde verdeling:  
 $\mu_{\Sigma y} = n\mu_y$        $\sigma_{\Sigma y} = \sqrt{n} * \sigma_y$
- $\bar{y}$  = steekproef gemiddelde
- Standaard afwijking steekproef  
 $\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}$
- Verdeling gemiddelde  
 $\bar{y} \sim N\left(\mu_y * \frac{\sigma_y}{\sqrt{n}}\right)$
- Tel je bij toevalsvariabele constante op, komt ie er wel bij, bij verwachting, **niet** bij variantie
- Vermenigvuldig je met een toevalsvariabele een constante, dan komt die in het kwadraat in de variantie terug

**Werkcollege 6**

- Binominale verdeling  
 $y \sim \text{Bin}(n, \pi)$
- Alternatieve hypothese: bewering waarvan men hoopt dat die juist is
  - In plaats dat nulhypothese juist is
- $H_0 : \pi = 0.5$                        $H_a : > 0.5$
- Toetsingsgrootheid: aantal in de proef die voor de getoetste situatie kiezen
- P-waarde → overschrijdingskans
  - Kans berekend onder aanname dat  $H_0$  waar is
    - Hoe kleiner P-waarde, hoe sterker door data tegen  $H_0$  geleverd bewijs
- Significantie niveau =  $\alpha$  = onbetrouwbaarheidsdrempel
  - False positive rate: in 5%  $H_0$  verwerpen terwijl dit niet moet
- P-waarde  $\leq \alpha$                       :  $H_0$  verwerpen,  $H_a$  aangetoond  
P-waarde  $> \alpha$                       :  $H_0$  niet verwerpen,  $H_a$  is niet aangetoond
- **P-waarde gaat hand in hand met significantie niveau**

**Werkcollege 7**

- Als de alternatieve tweezijdig is kan die groter of kleiner zijn  
 $H_0 = x$                       maar  $H_a \neq x$
- Verwachtingswaarde =  $N \cdot \pi$
- Tweezijdige P-waarde =  $2 \cdot$  eenzijdige P-waarde  
**Alleen als het symmetrisch is**
- Als significantie niveau niet aangegeven is, dan altijd 0,05
- Verwachting en variantie Binominale verdeling
  - **Aantal successen**  $y \sim \text{Bin}(n, \pi)$   
 $\mu_y = n \cdot \pi$   
 $\sigma_y^2 = n\pi(1-\pi)$                        $\rightarrow \sigma_y = \sqrt{n\pi(1-\pi)}$
  - **Steekproeffractie**  
 $\hat{\pi} = \frac{y}{n}$   
 $\mu_{\hat{\pi}} = \mu_{(1/n)y} = 1/n \mu_y = 1/n \cdot n \cdot \pi = \pi$   
oftewel: schatting steekproeffractie is gewoon  $\pi$   
  
 $\sigma_{(1/n)y}^2 = (1/n)^2 \sigma_y^2 = (1/n)^2 \cdot n \cdot \pi \cdot (1-\pi) = \pi(1-\pi)/n$   
  
dus:  $\sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}$
  - Vuistregel benadering:  $n\pi \geq 5$                       en  $n(1-\pi) \geq 5$

Statistiek 1

- **Continuïteitscorrectie**

- Verbetert benadering van een discrete verdeling met continue verdeling
- Bin alleen hele getallen maar dan tel je niet alles van de rechthoeken mee dus:  
benedenwaarde – 0.5 (van bin size)  
bovenwaarde + 0.5 (van bin size)

**College 7**

- Z-toets

$$\frac{\hat{\pi} - \pi_0}{\sigma_{\hat{\pi}}}$$

-