**Samenvatting inductieve statistiek**: formules, definities,werkwijzen en eigenschappen

De **populatie** is de verzameling van alle objecten of experimentele eenheden waarover men uiteindelijk uitspraken wil doen.

Een **steekproef** is een geordende verzameling van elementen uit een populatie.

Steekproef is als geordende verzameling aangeduid met ronde haken: (,).

De omvang van een steekproef wordt aangeduid met het symbool n.

**Gegevensverzameling**:
Stap 1: Uitvoering van een toevalsexperiment
*Volgens vaste regels wordt een procedure uitgevoerd, welke inhoudt dat uit een populatie een steekproef wordt getrokken, waarop vervolgens een aantal operaties worden toegepast die na afloop gegevens opleveren.*
Stap 2: Bepaling van de uitkomst
*Van alle informatie die het toevalsexperiment oplevert, wordt een beperkt deel, de uitkomst, geregistreerd.*
Stap 3: Bepaling van de score op een toevalsvariabele
*Voor elke uitkomst kan men de numerieke waarde bepalen van de variabele(n) X waarin men uiteindelijk geïnteresseerd is.*

Een **toevalsexperiment** is een procedure die:
(1) volgens vaste regels wordt uitgevoerd
(2) in principe herhaalbaar is
(3) gegevens oplevert die niet met zekerheid kunnen voorspeld worden vanuit de condities
 waaronder het experiment plaatsvindt

**Steekproeftrekking met/zonder teruglegging**:
na trekking van elk element van de steekproef wordt dit wel/niet terug toegevoegd aan de populatie alvorens een volgend element wordt getrokken.

De trekking van een steekproef met omvang n gebeurt op **zuiver toevallige wijze** (ZTW) als dit gebeurt met teruglegging en op een dusdanige manier dat elke geordende deelverzameling van n elementen uit de populatie evenveel kans heeft om als steekproef getrokken te worden.

**Gestratificeerde steekproeftrekking**: de populatie wordt op basis van beschikbare informatie opgesplitst en homogene deelpopulaties of strata.

**Resultaat**: elke uitvoering van een toevalsexperiment dat leidt tot een, in principe unieke, reeks van gegevens.

**Uitkomst**: het gedeelte van deze gegevens dat informatief is in verband met de vragen die men wil beantwoorden.

De verzameling van alle mogelijke uitkomsten van een TE noemt men de **uitkomstenverzameling of steekproefruimte** van dat TE.

Een **toevalsvariabele X** is een functie van de uitkomstverzameling van een TE naar ℝ.

Een toevalsvariabele die slechts een eindig of aftelbaar oneindig aantal waarden kan aannemen heet **discreet**, een toevalsvariabele die een niet-aftelbaar oneindig aantal waarden kan aannemen noemen we **continu**.

Een **gebeurtenis** A is een deelverzameling van de uitkomstenverzameling.

**Bijzondere gebeurtenissen**:
∅: de lege gebeurtenis
Ω: de zekere gebeurtenis (welke hier fungeert als universele verzameling
{ω}: **elementaire gebeurtenis** (alle singletons of deelverzamelingen van Ω die slechts 1 uitkomst bevatten)

A⋂B = ∅ (gebeurtenissen A en B zijn disjunct of mutueel exclusief)

Gegeven een uitkomstenverzameling Ω en de verzameling van alle gebeurtenissen 𝒢. **Een kans P is een functie**:
 P: 𝒢🡪 [0,1]
 G 🡪 P(G)
met (1) P(Ω) = 1
 (2) als G1, …, Gm gebeurtenissen zijn met Gi⋂Gi’ = ∅ als i ≠ i’
 dan P(G1⋃…⋃Gm) = P(G1)+…+P(Gm).

P(G) = met ωj ∈ G

**Regel van Laplace**:
Als alle uitkomsten in een eindige uitkomstenverzameling Ω even waarschijnlijk zijn dan geldt voor elke gebeurtenis G ⊂ Ω:

**De conditionele probabiliteit** op A gegeven B, P(A|B), is gedefinieerd als:
 als P(B)≠0
Als P(B)=0 is P(A|B) onbepaald

**Regel van Bayes**:

Als P(A) ≠ 0 ≠ P(B) dan zijn de volgende eigenschappen equivalent:
(1) P(A|B) = P(A)
(2) P(B|A) = P(B)
(3) P(A⋂B) = P(A) x P(B)
Als deze eigenschappen gelden, dan zegt men dat de gebeurtenissen A en B **statistisch onafhankelijk** zijn.

Als één van deze eigenschappen geldt, gelden alle anderen ook.
Als één niet geldt, idem.

A1,…,An heten **mutueel statistisch onafhankelijk** als voor elke deelverzameling Ai1,…Aim geldt:
P(Ai1⋂…⋂Aim) = P(Ai1)x…xP(Aim).

Paarsgewijze statistische onafhankelijkheid is **géén voldoende voorwaarde** voor mutuele statistische onafhankelijkheid.

**Nagaan of iets paarsgewijs statistisch onafhankelijk is**:
P(A) = P(B) = P(C)
**Nagaan of iets mutueel statistisch onafhankelijk is**:
P(A⋂B) = P(A) x P(B)
P(B⋂C) = P(B) x P(C)
P(A⋂C) = P(A) x P(C)
Daarna trio’s, etc etc

**Regel (Vermenigvuldigheidsstelling)**:
Als de uitkomst van een TE bestaat uit het gezamenlijk resultaat van het na mekaar uitvoeren van k operaties, waarbij de eerste operatie N1 verschillende deelresultaten kan hebben, waarbij de tweede operatie na het uitvoeren van de eerste N2 verschillende deelresultaten kan hebben, enz. dan geldt:
#Ω=N1 x N2 x … x Nk

**Regel (aantal permutaties)**:
Het aantal mogelijkheden om n objecten te ordenen is gelijk aan n x (n – 1) x … x 1;
Dit noteren we kortweg als n!

**Regel (aantal combinaties)**:
Het aantal mogelijkheden om zonder teruglegging uit een verzameling met N verschillende elementen een deelverzameling van n elementen te trekken (waarvan de volgorde niet belangrijk is) is gelijk aan:

Dit noteren we kortweg als

P(Gc) = 1 – P(G)

In het algemeen geldt dat:
P(A⋂B) = P(A) x P(B|A)
Alleen als A en B statistisch onafhankelijk zijn geldt dat:
P(A⋂B) = P(A) x P(B)

In het algemeen geldt dat:
P(A⋃B) = P(A) + P(B) – P(A⋂B)
Als A en B disjunct zijn geldt dat:
P(A⋃B) = P(A) + P(B)

**POPULATIEKARAKTERISTIEKEN VAN 1 TOEVALSVARIABELE**

|  |  |
| --- | --- |
| Discrete toevalsvariabele | Continue toevalsvariabelen |
| Kansmassa  | Dichtheid φ |
| Φ trap | Φ continu/vloeiend |
| : maakt uit | : maakt niet uit |
| Kwantielen 🡪 tabel of grafiek | Kwantielen 🡪 grafiek |

DISCRETE TOEVALSVARIABELE

Als X een discrete toevalsvariabele is met waardengebied {x1, x2, …} dan is de **kansmassafunctie** (of x) gedefinieerd als volgt:
 : ℝ 🡪 ℝ
 x 🡪(x)
met (x) = P({ω | X(ω) = x})

Hieruit volgt dat:
-∀ x: 0 (x) 1
-Als x ∉ waardengebied van X dan (x) = 0
-

Als X een toevalsvariabele is dan is de **cumulatieve verdelingsfunctie Ф** (of Фx) gedefinieerd als volgt:
 Ф: ℝ🡪ℝ
 x 🡪Ф(x)
met Ф(x) = P({ω | X(ω) x})

**r-de populatiekwantiel, xr\* zoeken:**
GEVAL 1: er is een x met Ф(x) = r
🡪 xr\* het gemiddelde van de kleinstex-waarde met Ф(x)=r en de kleinste x-waarde met Ф(x)r
GEVAL 2: er is géén x met Ф(x) = r
🡪 xr\* de kleinste x met Ф(x) r

CONTINUE TOEVALSVARIABELE

Als X een continue toevalsvariabele is dan is een **dichtheidsfunctie φ** (of φx) gedefinieerd als volgt:
 φ: ℝ🡪ℝ
 x 🡪φ(x)
met φ 0 en als a < b ∈ ℝ: P(a X b) =

CENTRALE TENDENSMATEN

**Populatiemodus**: elke waarde x waarvoor (x) resp. φ(x) maximaal is

**Populatiemediaan** ():Pc50\* = D5\* = Q2\*

**Populatiegemiddelde/verwachte waarde** (x of E[X]):
 X discreet: x = als
 X continu: x = als

Eigenschappen van het populatiegemiddelde:
E[X -x] = 0
als c ≠ x dan E[(X - x)2] < E[(X – c)2]
Regel van Steiner: E[(X – c)2] = E[(X - x)2] + (x – c)2

SPREIDINGSMATEN

**Bereik**: max X – min X

**Interkwartielbereik**:

**Populatievariantie** (): E[(X - x)2]

**Populatie standaardeviatie** ():

Eigenschappen van :
 = E[X2] -

**Ongelijkheid van Tchebychev**:
Als k > 1 dan P(|X - x| k)
Als k’ > 1 dan P ((X - x)2 k’)

TRANSFORMATIES VAN TOEVALSVARIABELEN

Als Y = aX + b dan E[Y] = a E[X] + b

Stel Y = aX + b dan

Ϛ-transformatie: Ϛx =

 = 0 en = 1

**POPULATIEKARAKTERISTIEKEN VAN 2 TOEVALSVARIABELEN**

TWEE DISCRETE TOEVALSVARIABELEN

Als (X,Y) een discrete toevalsvariabele is met waardengebied {(x1,y1), (x2,y2),…} dan is de **bivariate kansmassafunctie** gedefinieerd als volgt:
 : ℝ2 🡪 ℝ
 (x,y) 🡪 (x,y)
met = P({ω | (X,Y)(ω) = (x,y)})

Als (X,Y) een discrete toevalsvariabele is met waarden (xj, yj’) en met kansmassafunctie (xj,yj’), dan is voor elke xj in het waardengebied van Y met (xj)≠0
 : ℝ 🡪 ℝ
 y 🡪 P(Y = y|X = xj) =
een **conditionele kansmassafunctie**,
Analoog is ook voor elke Yj’ in het waardegebied van Y met (Yj’) ≠ 0
 : ℝ 🡪 ℝ
 y 🡪 P(X = x|Y = yj’) =
een **conditionele kansmassafunctie**

Als (X,Y) een discrete toevalsvariabele is, met X-waarden xj en Y-waarden yj’, dan zijn de volgende eigenschappen equivalent:
 (1) ∀j, j’: (xj,yj’) = (xj) (yj’)
 (2) ∀j:
 (3) ∀j’:
Als deze eigenschappen gelden, dan zegt men dat X en Y **statistisch onafhankelijk** zijn.

TWEE CONTINUE TOEVALSVARIABELEN

Als (X,Y) een continue toevalsvariabele is, dan is een **bivariate dichtheidsfunctie** φX,Y gedefinieerd als:
 φX,Y : ℝ2 🡪 ℝ
 (x,y) 🡪 φX,Y (x,y)
met φX,Y  0 en met, als a < b en c < d:
P(aXb en c Y d) =

Als (X,Y) een continue toevalsvariabele is met bivariate dichtheidsfunctie φX,Y, dan impliceert φX,Y dichtheidsfuncties φX  en φY voor X en Y, welke we **marginale dichtheidsfuncties** noemen. Er geldt dat:

Als (X,Y) een continue toevalsvariabele is met bivariate dichtheidsfunctie φX,Y, dan is voor elke xj met (xj)≠0
 : ℝ 🡪 ℝ
 y 🡪 met
een **conditionele dichtheidsfunctie.**
Analoog is ook voor elke Yj’ met (yj’)≠0
 : ℝ 🡪 ℝ
 y 🡪 met
een **conditionele dichtheidsfunctie.**

Als (X,Y) een continue toevalsvariabele is, met dichtheidsfunctie φX,,Y, dan zijn de volgende eigenschappen equivalent:
 (1) ∀ x,y : φX,,Y(x,y) = φX (x) x φY(y)
 (2) ∀ x met φX,(x) ≠ 0 : φY|X=x = φY
 (3) ∀y met φY (y) ≠ 0 : φX,|Y=y = φX
Als deze eigenschappen gelden, dan zegt men dat X en Y **statistisch onafhankelijk** zijn.

CENTRALE TENDENSMATEN

**Conditionele verwachte waarden of conditionele populatiegemiddelden**

(X,Y) discreet:

(X,Y) continu: analoog opgebouwd

SPREIDINGSMATEN

**Conditionele populatievarianties** :

(X,Y) discreet:

(X,Y) continu: analoog opgebouwd

SAMENHANGS- OF ASSOCIATIEMATEN

**Populatiecovariantie** ():
Met als X,Y discreet:

en met als X,Y continu:

**Eigenschappen van populatiecovarianties**:

**Populatiecorrelatie**

**Eigenschappen van populatiecorrelaties**:

 Gevolg:
(a)
(b)
Als X en Y statistisch onafhankelijk zijn dan

SOMVARIABELEN

**Eigenschappen**:

**RELATIE TUSSEN STEEKPROEF- EN POPULATIEKARAKTERISTIEKEN VAN TOEVALSVARIABELEN**

**Probleem van parameterschatting** : de grootte van een (onbekende) parameter willen kennen

**Probleem van hypothesetoetsing** : willen weten of een veronderstelling of hypothese m.b.t. parameters wel degelijk klopt

**iid** = independent and identically distributed (onafhankelijk en identiek verdeeld)

**Zuivere schatter** : een schatter die als verwachte waarde de te schatten parameter heeft

**Verwachte kwadratische fout van** :
Gezien een zuivere schatter is 🡪