Bedrijfsstatistiek

TEW 2e fase

Mieke Simons

**H1: Statistiek en kansrekenen**

 **Definitie van Statistiek**

*Statistiek*

*De (hulp)wetenschap die zich bezig houdt met informatie of gegevens
=> analyseren, interpreteren, voorstellen, verzamelen, classificeren,...*

**Enkele voorbeelden**

* Systeem voor vochtigheid P: alle ogenblikken | Var: vochtigheid
* Informatie vliegpassagiers P: alle passagiers | Var: bagage, tijdstip, passagiers,...
* Vulmachine P: dozen | Var: rijst | Pr: vulproces
* Nitraatgehalte groenten P: groenten | Var: nitraatgehalte
* Kwaliteit tennisballen P: tennisballen | Var: diameter | Pr: productiepr
* Risico van bep aandelen P: rendementen | Var: tijd
* Karakteristieken klanten P: klanten | Var: betalingswijze, keuze,...
* Gegevens zijn beschikbaar of moeten verzameld worden
	+ Rechtstreeks beschikbaar
	+ Verkregen via experimenten
	+ Verkregen via enquetes

 **Onderwerp van de Statistiek**

* Over de populatie van elementen of objecten
* Over een proces van elementen of objecten
(=> zie voorbeelden: P (populatie) en Pr (proces) en Var (variabele))
* Gegevens via variabelen: verschilt bij elk element
* Steekproef: slechts een deel wordt bekeken

Beschrijvende Statistiek

= descriptieve statistiek

* Overzichtelijk beeld
* Overzichtelijke beschrijving

Verklarende Statistiek

 = inferentiële statistiek = steekproeftheorie

* Analyseren en interpreteren
* Hypothesen testen en waarde toetsen
* Antwoorden en conclusies veralgemenen naar populatie = inferentie
* Uitspraak niet 100% zeker!
	+ Betrouwbaarheidsgraad => kansuitspraak => kansrekenen

**Opletten met steekproeven**

* Moet voldoende groot zijn
* Moet relevant en representatief zijnµ

**De Kansrekening**

*Kansrekening*

*Bestudeert processen of experimenten waarbij de uitkomst onzeker is.*

* Het gooien van een dobbelsteen, de toekomstige prijs van een product,...
* Statistiek vs Kansrekenen
	+ Statistiek: uitspraken via steekproefgegevens
	+ Kansrekenen: rechtstreekse uitspraken via onderstellingen
* Onderstelling noodzakelijk
	+ Vb dobbelsteen moet homogeen zijn
	+ In de praktijk moeilijk of niet te verifiëren
* Zie voorbeelden HBp 7 voor verschillen bij statistische studie (is) en kansrekening (zal)

*!! Statistiek: beperkte steekproefinformatie: geen absoluut correcte uitspraken!!*

**H2: Data en hun voorstelling**

 Proces of Populatie Census van de populatie (volledige populatie)

 Elementen Steekproef uit de populatie

 Variabelen

 **SOORTEN GEGEVENS**

**Kwalitatieve (categorische) variabelen**

*Nominale variabele
Een nominale variabele laat toe de elementen te classificeren, het element krijgt een “etiket”.*

* Slechts om te categoriseren
* Berekeningen zinloos (behalve percentages)
* Codenummers, codeletters,... (cijfers impliceren dan geen volgorde en/of hoeveelheid!)
* Kleur, merk, verpakking,...

*Ordinale variabele
Een ordinale variabele laat toe de elementen te ordenen.*

* Om te waarderen, te beoordelen
* Geen vaste meeteenheid ( restaurant \* en restaurant \*\* restaurant \*\* en restaurant \*\*\*)
* Berekeningen zinloos
* Zeer goed/ goed/ slecht/ zeer slecht, 1 ster/2 sterren,...

**Kwantitatieve variabele**

*Intervalvariabele*

* Geen natuurlijk nulpunt
* 16h is niet 2x zo laat als 8h
* Tijd op de klok, temperatuur op de Celsiusschaal (nulpunt hebben we zelf gekozen),…

*Ratiovariabele*

* Wel een natuurlijk nulpunt
* 6cm is wel 2x 3cm
* Lengte, volume,Kelvinschaal ,...

*Discrete variabele*

* Eindig: aantal passagiers dat opdaagt, aantal kinderen in een gezin,...
* Oneindig: de aantal worpen nodig om een zes te gooien
* Gehele getallen

*Continue variabele*

* Elk getal tussen de boven en ondergrens: lengte, duurtijd, inhoud,...
* Verdelen in klassen
* Reële getallen

**DE DATAMATRIX**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Waarneming | Gewicht | Vulkop | Uur | Analist |
| O1 | 50,56 | 1 | 8 | A1 |
| O2 | 51,58 | 2 | 8 | A2 |

* Rij: elementen/ waarnemingen = observatievector O
* Kolom: variabelen
* Gewicht: continue variabele (kwantitatief)
* Vulkop, analist: nominale variabele (kwalitatieve)
* Uur: intervariabele/ nominale variabele

**Univariate voorstelling**

* Betrekking op 1 variabele

**Bivariate voorstelling**

* Betrekking op 2 variabelen

**Frequentie**

* Frequentie ve klasse: # elementen vd klasse
* Relatieve frequentie: # elementen vd klasse/ totaal # elementen vd proef

**HET VOORSTELLEN VAN UNIVARIATE KWALITATIEVE VARIABELEN**

**Het staafdiagram**

**Het cirkeldiagram = sectordiagram**

**Het Pareto diagram**

* Klassen volgens dalende frequentie
* Cumulatief voorgesteld
* Kijken naar de klassen met de hoogste frequenties

**Excel** (HBp14-15)

* Frequentie: AANTAL.ALS(celbereik;criterium)

**HET VOORSTELLEN VAN UNIVARIATE KWANTITATIEVE VARIABELEN**

**Discrete kwantitatieve variabelen**

*Gehele getallen*

* Naalddiagram

**Continue kwantitatieve variabelen**

*Reële getallen: verdelen in klassen!*

* Histogram
	+ Klassenbreedte niet altijd gelijk
	+ # klassen staat niet vast (tip: )
	+ Maximaal 15 à 20
	+ Som van de oppervlakten = 1
* Polygoon
	+ Middens van de rechthoeken verbinden
	+ Oppervlakte onder polygoon = 1

**Het stam- en bladdiagram**

* Stam: eerste cijfer van elke waarneming
* Blad: tweede cijfer van elke waarneming

**Excel** (HBp19-21)

* Frequentie: INTERVAL(celbereik,bovengrenzen)

**HET VOORSTELLEN VAN BIVARIATE VARIABELEN**

**Kwalitatieve variabelen**

* Kruistabel
* Meervoudige staafdiagram

**Kwantitatieve variabelen**

* Kruistabel
* Puntenwolk/ puntendiagram

**Gemengd**

* Kruistabel met kwalitatieve en kwantitatieve variabele

**Excel** (HBp19-21)

* Kruistabel: DRAAITABEL

**H3: Beschrijvende Kengetallen van Steeproefgegevens**

**BESCHRIJVENDE KENGETALLEN**

*(steekproef) statistieken = (steekproef) kengetallen
Kenmerkende maten die de gegevens van de steekproef samenvattend beschrijven => functies*

*Parameters
Kengetallen die betrekking hebben op een volledige populatie/ proces*

**UNIVARIATE KWANTITATIEVE VARIABELEN**

**Kengetallen van centrale ligging**

* Zowel voor discrete als continue variabelen

Het gemiddelde

* Streekproefgemiddelde
* Populatiegemiddelde
* : frequentie | : gewichten = relatieve frequentie
* Eigenschappen
	+ *bewijs zie slides*
* Gevoelig aan extreme waarden

De mediaan

* Ongeveer 50% vd waarnemingen zijn groter, 50% kleiner
* Hoe groter de steekproef, hoe nauwkeuriger
* Niet gevoelig aan extreme waarden => robuust

De modus

* Waarneming met de grootste frequentie
* Histogram: modale klasse => unimodaal / bimodaal

**Maatstaven van relatieve ligging**

*Geven de positie van een waarneming relatief tot de andere waarnemingen*

* ie ordekengetal = ie ordestatistiek = ie observatie na rangschikking van klein naar groot
* Percentielen: Cp = waarde zodat p100% (0<p<1) van de gerangschikte waardes kleiner is

Vb C0,76 = X => 76% van de waarnemingen kleiner dan X (en 24% groter)

* Kwartielen
	+ 1e kwartiel: Q1 = c0.25 | 2e kwartiel: Q2 = c0.5 (mediaan) | 3e kwartiel: Q3 = c0.75
	+ Ongeveer ¼ is < Q1 | ongeveer ½ is < Q2 | ongeveer ¾ is < Q3
* Percentiel Cp = pde kwantiel

Vb C0,76 => 0,76ste kwantiel

Vb C0,80 => 0,80ste kwantiel => 8e deciel

**Maatstaven van spreiding**

*Meten de spreiding van de gegevens rond een centrale waarde*

* Want zo kunnen 2 reeksen gegevens hetzelfde gemiddelde hebben, maar toch sterk verschillen!
* Afwijking =

De variatiebreedte = spreidingsbreedte = ‘range’

* Grootste – kleinste waarneming
* Nadeel: hangt alleen af grootste en kleinste
* Nadeel: moeilijk steekproeven met ongelijk # observaties te vgl’en (kan alleen maar stijgen als n stijgt)

De interkwartielbreedte = IKB

* Q3 – Q1 = c0,75 – c0,25

De gemiddelde absolute afwijking

* GAA
* Niet 0 want werken met absolute waarden!!!

De gemiddelde gekwadrateerde afwijking

* GGA
* Niet 0 want werken met kwadraten!!!

De variantie

* Steekproefvariantie
* Populatievariantie
* Altijd positief
0 bij cte waarden, dus geen variantie
* Populatievariantie is dus een GGA,

Steekproefvariantie is geen GGA

De standaarddeviatie

*
*

Variatiecoëfficiënt

* Om de relatieve variabiliteit te bekomen

**Maatstaven van scheefheid**

* Pearson-coëfficiënt
* Scheefheidscoëfficiëntie
	+ Derde moment m3
* Symmetrisch: gemiddelde = mediaan = modus SP = 0 m3 = 0

Links scheef: gemiddelde < mediaan < modus SP < 0 m3 < 0 staart aan de linkerzijde

Rechts scheef: gemiddelde > mediaan > modus SP > 0 m3 > 0 staart aan de rechterzijde

**Transformatie en standaardisatie van gegevens**

* Lineaire transformaties
	+ yi = axi + b met a en b cte
	=> eenvoudige formules: gelden alleen als LINEAIRE transformaties!!!
* Standaardiseren
	+ *De gestandaardiseerde waarneming zi
	Meet de afstand, uitgedrukt in standaarddeviaties, die xi verwijderd liggen van*
	+ Belangrijk voorbeeld van lineaire transformatie
	+ Waarbij en
	+ | |

**Extreme waarden, uitbijters (outliers)**

* Uitzonderlijk grote of kleine steekproefwaarneming
* Valt buiten het patroon
* Oorzaak moeilijk te achterhalen
	+ Toch mogelijk: corrigeren of weglaten
	+ Niet mogelijk: gekwalificeerd oordeel onderzoeker
* Histogram klokvormig symmetrisch
	+ Waarneming extreem als zi < -3 of als zi > +3
* Histogram niet klokvormig symmetrisch: boxplot

**Doos met snorharen = boxplot**

* Centrale deel: box
	+ Begrensd door Q1 en Q3
	+ Mediaan: verticale lijn
	+ Gemiddelde:
* Naast de snorharen: de mogelijke extreme waarden
	+ of
* Nog verder naast de snorharen: de werkelijke extreme waarden
	+ of
* Positie mediaan en gemiddelde (+ snorharen)
* Kan je uit afleiden of ze scheef is of niet en hoe

**UNIVARIATE KWALITATIEVE VARIABELEN**

* Nominale variabel
	+ Modus
* Ordinale variabele
	+ Modus en mediaan
	+ Soms gemiddelde: schaal moet intervalschaal benaderen

**BIVARIATE (KWANTITATIEVE) VARIABELEN**

**Covariantie**

* Steekproefcovariantie sXY
* Populatiecovariantie XY
* Kan – en + zijn: hangt af van de dominantie van de positieve of negatieve termen
* Termen in de covariantie
	+ (+) : kleine waarde voor X en Y / grote waarde voor X en Y (-)(-) of (+)(+)
	+ (-) : kleine waarde voor X en grote waarde voor Y (e.o.) (-)(+) of (+)(-)
* Puntenwolk: assen rond de gemiddelden en
	+ Positieve term: 1e en 3e kwadrant => domineren: positieve samenhang
	+ Negatieve term: 2e en 4e kwadrant => domineren: negatieve samenhang
* Kruistabel van X en Y
	+ Marginale frequenties van X en Y => relatieve frequenties van X en Y
* Nadeel: afhankelijk van meeteenheid
* Variantie van een variabele is dus de covariantie met zichzelf:

**Correlatie**

* Steekproefcorrelatie rXY
* Populatiecorrelatie XY
* Niet afhankelijk van meeteenheid
* Belangrijke eigenschappen
	+ -1 < rXY < +1
	+ -1 < XY < +1
	+ Voor lineaire verband: zie formules
* Puntenwolken op rechte perfecte correlatie
	+ - Rechte met positieve helling r = +1 ‘perfect’ positief
		- Rechte met negatieve helling r = -1 ‘perfect’ negatief
		- Perfect horizontaal/ verticaal s = 0 geen **lineair** verband
* Covarianties zeggen intuïtief niet veel, correlatie wel
* Correlatie meet enkel de graad van **lineair** verband tussen 2 variabelen

**EXCEL FORMULES**

* s² : VAR(bereik) s : STDEV(bereik)
* ² : VARP(bereik) : STDEVP(bereik)
* GAA: GEM.DEVIATIE(bereik) GGA: DEV.KWAD(bereik)
* Vc : STDEV(bereik) / GEMIDDELDE(bereik)
* SCHEEFHEID(celbereik) PEARSON(bereik1;bereik2)
* COVARIANTIE(bereik1;bereik2) CORRELATIE(bereik1;bereik2)

**H4: Kansrekening**

**BEGRIP KANS**

**Benaderend deterministisch**

* Resultaat zal met zekerheid worden voorspeld
* Als aan voorwaarden voldaan
* Slechts in theorie deterministisch want altijd kleine afwijkingen
* Vb. bij druk van 760mm en t° van 100°C (vwen) zal water koken (resultaat)

**Probabilistisch of stochastisch**

* Onzekerheid over de uitkomst
* Kansprocessen = kansexperimenten = stochastische processen
* ‘erg waarschijnlijk dat…’
* Uitkomst kan je interpreteren als steekproefgegevens voor populatie/ proces

**KANSEXPERIMENTEN EN VERZAMELINGENLEER**

**Uitkomstenruimte**

* Uitkomstenruimte = steekproefruimte
* Verzameling van alle uitkomsten waarvoor geldt
* Eindig, oneindig, continu, numeriek, niet-numeriek,…
* Eenzelfde experiment kan meerdere uitkomstenruimten hebben
* Kijken welke vragen gesteld worden over het experiment

**Gebeurtenis** (herbekijk vb HBp58)

* G: een deelverzameling van
* Verzameling van uitkomsten waarbij ook als gebeurtenis worden beschouwd
* Elementaire gebeurtenis: een G die zich naar een uitkomst zelf herleid
* Gebeurtenis G doet zich voor als
	+ Bij uitvoering van het experiment de uitkomst
	+ Bij uitvoering dus 1 van de uitkomsten van G zich voordoet

**Definities, operaties en relaties**

* Unie of vereniging: (A of B of beide)
	+ => dan bevat G alle uitkomsten die behoren tot minsten 1 vd gebeurtenissen *G1,G2,*…
	+ *G* is exhaustief (volledig) als
* Doorsnede: (A en B)
	+ Lege doorsnede: disjunct: sluiten elkaar uit

=> dan bevat G alle uitkomsten die behoren tot al de gebeurtenissen *G1,G2,*…

* Verschil: (A maar niet B)
* Deelverzameling: (B bevat A)
* Complement: (niet in G)
* Partitie van (of…of…of…)
	+ Als ze exhaustief zijn
	+ Als ze elkaar uitsluiten

**DEFINITIE VAN KANS**

*Kans op gebeurtenis G
De kans definiëren dat een gebeurtenis zich voordoet, dus dat de uitkomst een element is van G*

**Frequentiedefinitie**

* Functie P(G)
* Frequentie van G: => relatieve frequentie van G:

Kans dat deze zal stabiliseren voor grote n: P(G)

* Als dan is en dus

Als dan is en dus

* Hoe groot moet n zijn? We zouden n onbeperkt kunnen laten toenemen:

=> is in praktijk moeilijk of zelfs onmogelijk

**Axiomatische definitie**

1. waarbij *G1,G2,*… elkaar uitsluiten (dus alleen of…)
* Nadeel: wordt niets gezegd over de bepaling van P(...) voor een concreet experiment

**Klassieke definitie van Laplace**

* Alle mogelijke uitkomsten zijn even waarschijnlijk
* Slechts mogelijk voor eindig # mogelijke uitkomsten die alle gelijkwaardig zijn

**Klassieke definitie tov frequentiedefinitie**

* Frequentiedefinitie => empirische verdeling
* Klassieke definitie => theoretische verdeling = kansverdeling

**Subjectieve definitie**

* Persoonlijke inschatting
* Verschillende personen zullen dus verschillende inschattingen van de kans hebben

**REKENREGELS VOOR KANSEN**

* Formules gelden voor elk experiment en willekeurige gebeurtenissen *G1,G2,G3,* …

**SIMULATIE IN EXCEL**

**Doel**

* Goed inzicht krijgen
* Kan vaak herhaald worden (op basis van de frequentiedefinitie)
* Systemen optimaliseren

**Programma’s**

* Voor de ingewikkelde processen specifieke simulatietalen ontwikkeld: GPSS, Slam,…
* Excel: ASELECT( )
	+ Lukraak een getal genereren tussen 0 en 1
	+ Pseudolukrake getallen…

**VOORWAARDELIJKE KANS**

**Begrip**

*Voorwaardelijke kans
De kans berekenen op een gebeurtenis, rekening houdend met een andere gebeurtenis, die de kans op de eerste in belangrijke mate zou kunnen veranderen*

* Formule voorwaardelijke kans: P(G2) > 0
* onvoorwaardelijke kans = a priori kans

 voorwaardelijke kans = a posteriori kans = herziene kans

**Vermenigvuldigingsregel**

* Voor 3 gebeurtenissen:
* Dus:

**Onafhankelijke en afhankelijke gebeurtenissen**

* Onafhankelijk: 2e gebeurtenis heeft geen invloed op de eerste: info is neutraal
* Vb. de kans om 1 te gooien met 4 eerlijke dobbelstenen:
* Elkaar uitsluitende gebeurtenissen onafhankelijke gebeurtenissen!!!!
* Informatie in G2 is
	+ Positief :
	+ Negatief:
	+ Neutraal:

**Totale kans en de kansregel van Bayes**

* Partitie: exhaustief en disjunct!
* Bayes: indien totale kans gekend, zoeken waar de kans welke grootte heeft

**H5: Univariate kansvariabelen**

**KANSVARIABELEN EN VERDELINGSFUNCTIE**

**Kansvariabele**

*Kansvariabele
Een functie X() die een reëel getal associeert met elke uitkomst () van een experiment*

* Kansvariabele = stochastische variabele
* Grote letters gebruiken: meestal X,Y of Z
* Kleine letters: verwijzen naar een bepaalde reële waarde voor de kansvariabele
* Betrekking op eigenschap van de elementen of een proces
	+ Univariaat: kansvariabele betrekking op 1 eigenschap
	+ Bivariaat: kansvariabele betrekking op 2 eigenschappen samen
	+ Kansvector: kansvariabele op meerdere eigenschappen samen
* Kansen kansvariabelen kan je afleiden uit kansen gebeurtenissen

 vb. met G3 =

**Verdelingsfunctie**

* Verdelingsfunctie = cumulatieve verdelingsfunctie
	+ Niet dalend want kans is altijd positief
	+ dus ook
* Overlevingsfunctie:

**DISCRETE KANSVARIABELE EN KANSVERDELING**

**Kansverdeling**

* Discrete kansvariabele
	+ Als # mogelijke waarden voor X eindig aftelbaar zijn (som van # ogen)
	+ Als # mogelijke waarden voor X oneindig aftelbaar zijn (opgooien tot uitkomst 6 is)
* Kansverdeling: niet exact in praktijk: onderstellingen maken

Empirische verdeling wel (rechtstreeks uit steekproefgegevens)

* Ook steeds positief
* Kansverdeling verdelingsfunctie! (Zie HB p88)
	+ Kansverdeling:

Waarde dat bij variabele hoort

Gedefinieerd voor reëel getal dat bij variabele hoort

* + Verdelingsfunctie: (cumulatief!)

tussen welke waarden het ligt
gedefinieerd voor elk reëel getal

**CONTINUE KANSVARIABELE EN KANSDICHTHEID**

**Kansdichtheid**

*

* Tegenhanger van de kansverdeling
* Continue kansvariabele
	+ In klassen verdelen!
* Benadering voor de kans (via relatieve frequentie)
	+ Geeft zelf dus geen kans weer!!!
* Oppervlakte onder polygoon tussen a en b benadert de kans van een waarde tussen a en b
* Indien zeer groot => klassenbreedte zeer klein => limiet: continue curve
* Waarde onder polygoon = 1 => waarde onder F ook
* Om na te gaan je een dichtheid hebt
	+ Aanduiden dat
	+ En dat

**Verdelingsfunctie**

* Afleiden en dan heb je de kansdichtheid
* Stijgend tot 1
*

* Uitwerken integraal:

**GRAFISCH OVERZICHTJE**

**Discreet**

* Kansverdeling : naalddiagram Verdelingsfunctie: trapvorm

**Continu**

* Kansdichtheid: klokvormig Verdelingsfunctie: stijgende tot aan 1

**Simuleren van continue kansvariabelen**

* + Met r = ASELECT( )
	+ Functie herleiden tot x

**FUNCTIE VAN EEN KANSVARIABELE**

**Algemeen**

* Kansvariabele
* Functie

**Discreet**

* Waarde van zoeken
* Functie gewoon invullen met de gevonden -waarden
* Of formule aanpassen door te substitueren door
* gewoon zoeken door oorspronkelijke kansverdeling toe te passen met y

**Continu**

* Via de verdelingsfunctie
1. Bepaal het nieuwe domein
2. Bereken de nieuwe verdelingsfunctie
3. Leid deze af tot je krijgt (domein niet vergeten!)
* Via de differentiaalanalyse
1. Bepaal het nieuwe domein
2. Los de nieuwe functie op naar en bepaal de afgeleide
3. Vul deze inverse functie in in de oorspronkelijke en vermenigvuldig met de afgeleide
* Herbekijk voorbeelden HBp98-99

**H6: Kengetallen van populaties en processen**

**VERWACHTE WAARDE VAN EEN KANSVARIABELE**

* Verwachte waarde of
* Wordt ook het gemiddelde van genoemd

*Verwachte waarde voor de populatie of het proces*

 *=*

*Steekproefgemiddelde voor steekproefgegevens*

* Discrete kansvariabele
	+ Met kansverdeling
* Continue kansvariabele
	+ Met kansdichtheid
* Verwachte waarde kan een waarde zijn die niet door de kansvariabele kan aangenomen worden

**VERWACHTE WAARDE VAN EEN FUNCTIE VAN EEN KANSVARIABELE**

**Algemeen**

* Methode 1 (met substititie)
1. Bereken als discreet of als continu van de nieuwe variabele
2. Pas nu de definitie van de verwachte waarde toe
* Methode 2 (direct)
1. Gebruik de kansverdeling of kansdichtheid van zelf en de functie
2. Pas nu de definitie van de verwachte waarde toe
* St Petersburg Paradox
	+ Voor 1x te mogen spelen niet bereid veel in te zetten
	+ Verwachte waarde van de opbrengst is oneindig groot

**Speciale gevallen**

* Lineaire functies => bewijs HBp106
* Constante =>

**VARIANTIE EN STANDAARDDEVIATIE VAN EEN KANSVARIABELE**

**Variantie en standaarddeviatie**

* Nu gedefinieerd voor elke mogelijke kansvariabele (discreet/continu)
* Ook gebruikt als maatstaf van spreiding
* Hoe groter of , hoe groter de spreiding rond de verwachte waarde

**Var(lineaire functie)**

* Bewijs zie HBp109
* Constant b heeft geen invloed op of
* en

**Gestandaardiseerde kansvariabele**

* Heeft altijd verwachte waarde 0 en variantie 1!!!
* Zie ook hoofdstuk 3 (p9)

**ANDERE KENGETALLEN**

**Modus**

* Bestaat niet als er geen maximum bestaat
* Zeer zwakke maat van centrale ligging

**Mediaan**

* Discreet: zie formule
* Continue:
* Ook maat voor centrale ligging want minder gevoelig aan extreme waarden dan

**Kwantielen, parcentielen en kwartielen**

* 100\*pde percentiel = 10\*pde deciel = pde kwantiel

Vb. 0,25e kwantiel = 2,5e deciel = 25e percentiel

* 1e kwartiel | 2e kwartiel | 3e kwartiel
* Voor continue kansvariabelen: zie formules
* Voor discrete kansvariabelen: slechts een benadering!

**Scheefheidscoëfficiënten**

* Pearson populatiecoëfficiënt
* Scheefheidscoëfficiëntie
	+ Vergeljkbaar met derde moment m3
* Symmetrisch: gemiddelde = mediaan = modus SP = 0 m3 = 0

Links scheef: gemiddelde < mediaan < modus SP < 0 m3 < 0 staart aan de linkerzijde

Rechts scheef: gemiddelde > mediaan > modus SP > 0 m3 > 0 staart aan de rechterzijde

**H7: Belangrijke discrete kansvariabelen**

**FAMILIE VAN KANSVERDELINGEN/ KANSDICHTHEDEN**

* Verzameling van verdelingen/ dichtheden
* Geïndexeerd door 1 of meerdere grootheden => parameters
* Elke waarde die de parameter aanneemt (in een bepaald domein)
* resulteert in een kansverdeling/ kansdichtheid

**DE BERNOULLI VERDELING**

* Parameter stelt hier een kans voor =>
* Succes: met kans
* Faling: met kans
* Bernoulli-experiment = kanproces voor 1 element

**DE BINOMIALE KANSVERDELING**

**Voorwaarden**

* Bestaat uit n opeenvolgende Bernoulli experimenten => succes of faling
* Met parameter telkens dezelfde waarde
* Experimenten worden *onafhankelijk* van elkaar uitgevoerd!

**Kans op succes**

* en stelt het aantal successen voor
	+ Kans op succes
	+ Kans op faling

**Parameters en**

* **** en geheeltallig

**Scheefheid**

* Symmetrisch:
* Links scheef:
* Rechts scheef: (figuur =0.2)

**Excel**

* Binomiale kans =BINOMIALE.VERD
* Cumulatieve kans =BINOMIALE.VERD
* Kleinste waarde zodanig dat =CRIT.BINOM
* Herbekijk voorbeelden HBp117-120

**DE POISSON VERDELING**

**Voorwaarden**

* Is een specifiek geval van een binomiale verdeling => succes of faling
* Met per klein interval telkens dezelfde waarde en zeer klein
* In 1 deelinterval kan maximum 1 gebeurtenis voorkomen
* Gebeurtenissen zijn *onafhankelijk* van elkaar!

**Van Binomiaal naar Poisson verdeeld**

* Stel binomiaal verdeeld
	+ Indien en
	+ Dan tendeert de kansverdeling naar de Poisson verdeling met
* Voorbeeld: # fouten in stof als we weten dat gemiddeld 5 fouten per 10m
	+ Zeer veel intervallen =>
	+ Zeer kleine fout per m =>
	+ En dit tendeert dus naar =>

**Excel**

* Poisson kans =POISSON
* Cumulatieve kans =POISSON
* Herbekijk voorbeelden HBp125

**DE GEOMETRISCHE VERDELING**

* Bestaat uit n opeenvolgende Bernoulli experimenten => succes of faling
* Na experimenten een eerste succes

**Excel**

* Geometrische kans =NEG.BINOM.VERD

**DE NEGATIEF BINOMIALE OF PASCAL VERDELING**

* Veralgemening van de geometrische
* Na experimenten een -de succes en dus aantal falingen =

Bij geometrische:

**Excel**

* Pascal kans =NEG.BINOM.VERD

Om grootheden van bepaalde verdeling te bewijzen

* Formules van H6 gewoon invullen met de gepaste waarden
* Gepaste waarden: waarden die de bepaalde verdeling kan aannemen

**KORT OVERZICHT**

**Bernoulli verdeling**

* Binair (0/1)
* Faling/succes, ja/nee, geslaagd/niet geslaagd,…

**Binomiale verdeling**

* Reeks van Bernoulli
* 20 mensen dagen op/dagen niet op, 100 producten falen/zijn succes,…

**Poisson verdeling**

* Gebeurtenis doet zich lukraak voor
* Binomiaal met en
* # klanten per u in apotheker, # brieven in de brievenbus na 2 dagen,…

**Geometrische verdeling**

* Reeks van Bernoulli
* Aantal experimenten nodig voor 1e succes
* Kans om een eerste 6 te gooien, kans om een juiste bloeddonor te vinden,…

**Negatief binomiale of Pascal verdeling**

* Veralgemening van de geometrische
* Aantal experimenten nodig voor e succes
* Kans om een r-de 6 te gooien, kans om r gepaste kleedjes te vinden,…

**H8: Belangrijke continue kansvariabelen**

**INLEIDING**

**Problemen ivm met continue kansvariabele**

* In praktijk niet altijd volwaardig continu!
	+ Niet perfect te meten
	+ Benadering
	+ Eindig in aantal…
* Grote eindige populatie
	+ Eigenschap soms continue verondersteld
	+ Slecht een benadering!
	+ Vb gewicht producten grote voorraad

**DE (CONTINUE) UNIFORME DICHTHEID**

**Excel**

* Genereren van een lukraak getal tussen 2 grenzen
* ASELECT( )

**Wielertoerist**

* Zie slides
* Slechte week referentie bepalen
* Afstand ligt binnen interval
* Formule toepassen

**DE EXPONENTIËLE DICHTHEID**

**Algemeen**

* Stel dat , dan zal de kansvariabele enkel waarden aannemen

**Eigenschap zonder geheugen**

* Zie formule: vb loketten

**Verband exponentiële en Poisson**

* Poisson verdeeld met aankomstenfrequentie

Tijd tussen opeenvolgende aankomsten exponentieel verdeeld met paramater

* Voorbeeld stof
* Bewijs:

**Excel**

* Kansdichtheid =EXPON.VERD
* Verdelingsfunctie =EXPON.VERD





**Duidt aan:**

**H9: De Normale dichtheid**

**DE NORMALE DICHTHEID**

**Voordelen van de normale dichtheid**

* Heel wat processen in praktijk genereren normaal verdeelde gegevens
* Belangrijke functies van normaal verdeelde kansvariabelen zijn normaal verdeeld
* Belangrijke functies van niet normaal verdeelde kansvariabelen zijn benaderend normaal verdeeld

**Belangrijke eigenschappen**

* Voor elke en voor elke , is het een dichtheid
* Klokvormig met asymptoot naar 0 voor
* Symmetrisch rond of
* Modus = mediaan = verwachte waarde =
* Buigpunten:

**DE NORMALE VERDELINGSFUNCTIE**

* Moeilijk te berekenen
* Voor elke waarden anders
* Normale verdeling gebruiken!

**NORMAAL EN STANDAARDNORMAAL**

**Standaardnormaal**

* Als en

**Lineaire transformatie**

* Kansvariabele is normaal verdeeld

Lineaire functie

* is dan ook normaal verdeeld

**Verband normaal en standaardnormaal kansvariabelen: tabel**

* Voor tabel standaardnormaal:
* Alleen de positieve waarden van z
* Altijd tekenen!!
* Meeste elementen bevinden zich dicht bij het gemiddelde
* Als meer dan boven 0 ligt, dan zak de normale verdeling zelden negatieve waarden opleveren

**EXCEL**

* NORM.VERD en NORM.VERD
* STAND.NORM.VERD
* NORM.INV en STAND.NORM.INV

* : grenzen
	+ Eerst kolom
	+ Dan rij aflezen
* : de uiteindelijk waarde





**SOM VAN ONAFHANKELIJKE KANSVARIABELEN**

* Gewoon kansrekenen toepassen
* Zie slides om op te frissen

**Som van 2 onafhankelijke kansvariabelen**

* Met en normaal verdeeld
* Som ook normaal verdeeld met:
	+
	+

**Som van meerdere onafhankelijke kansvariabelen**

* Met , ,…, normaal verdeeld

*Met allemaal dezelfde verwachte waarde en dezelfde variantie:*

* ook normaal verdeeld met:

*Met een lineaire functie:*

* ook normaal verdeeld met:
* Formule hoofdstuk 11

*Met allemaal een verschillende verwachte waarde en verschillende variantie:*

* De centrale limietstelling!

Wat wordt de kansverdeling/ dichtheid indien we een groot # onafhankelijke variabelen optellen?

* Tenderen naar een normale verdeling/dichtheid!
* Ongeacht de kansverdeling/ dichtheid van de oorspronkelijke kansvariabelen!
* benaderend normaal verdeeld
* Voorwaarden:
	+ Geen extreem afwijkende
	+ N moet voldoende groot zijn
		- Kansdichtheid normaal verdeeld =>
		- Kansdichtheid niet normaal, maar geen pieken =>
		- Voor andere gevallen =>
* Aangezien Y dan benaderend normaal => Z(Y) benaderend standaardnormaal

*Met binomiaal verdeeld met*

* Dan ook benaderend normaal verdeeld
* Bewijs doordat de binomiale kans overal hetzelfde is (dus de varianties en verwachte waardes ook)
* n groot => en minimum 5
* continuïteitscorrectie: benaderen via

**H10: Multivariate kansvariabelen**

**BEGRIPPEN**

**Univariate kasvariabele**

* 1 reëel getal wordt geassocieerd met elke uitkomst van het experiment
* Vb dobbelsteen 5x opgooien

**Multivariate kansvariabelen**

* Meerdere reële getallen worden geassocieerd met elke mogelijke uitkomst
* Vb witte en zwarte dobbelsteen *tegelijkertijd* 5x opgooien
* Multivariaat = gezamenlijk = tegelijkertijd!!

**Gezamenlijke kansverdeling**

* Voor discrete bivariate kansvector
* Geld samen als een kans dus voorwaarden:

**MARGINALE KANSVERDELING**

* Kansverdeling van afleiden uit de gezamenlijke kansverdeling
* Kansverdeling = marginale kansverdeling
	+ “marginaal” om aan te tonen dat ze werd afgeleid uit de gezamenlijke
	+ Telkens de totalen optellen
	+ Moet samen 1 geven!
* De 2e kansvariabele wordt buiten beschouwing gelaten
* *Marginale kansverdelingen kunnen afgeleid worden uit de gezamenlijke*

*MAAR de gezamenlijke kan alleen afgeleid worden uit de marginale als onafhankelijk!!!*

**VOORWAARDELIJKE KANSVERDELINGEN**

* De voorwaardelijke kansverdeling van gegeven en

De voorwaardelijke kansverdeling van gegeven

* Wanneer en onafhankelijke kansvariabelen:
	+ Voorwaardelijke kansverdeling van gegeven

= marginale kansverdeling van

* + Voorwaardelijke kansverdeling van gegeven

= marginale kansverdeling van

**H11: Functies van meerdere kansvariabelen**

**FUNCTIE VAN MEERDERE KANSVARIABELEN**

* We willen de reeds gevonden resultaten veralgemenen

**VERWACHTE WAARDE VAN EEN FUNCTIE VAN MEERDERE KANSVARIABELEN**

**Verwachte waarde**

* Zie formule

**Verwachte waarde**

* als onafhankelijk!!!
* als onafhankelijk!!!

**VOORWAARDELIJKE VERWACHTE WAARDEN**

* Hetzelfde als hierboven
* Voorwaardelijke kansverdeling/dichtheid gebruik ipv de gewone afzonderlijke (marginale)
* Eerst voorwaardelijke kansverdeling/dichtheid berekenen, daarna pas verwachte waarde

**KANSVERDELING/KANSDICHTHEID VAN FUNCTIES VAN KANSVARIABELEN**

* Op dezelfde manier als in hoofdstuk 5
	+ Berekeningen gewoon uitvoeren
	+ Met deze resultaten verder werken
* Stellingen en gevolgen reeds gezien in Hoofdstuk 9

**H12: Covariantie - Correlatie - Variantie van lineaire functies**

**COVARIANTIE EN CORRELATIE**

**Begrippen**

* Hetzelfde als in hoofdstuk 3
* Formule voor covariantie hier veralgemeend
* Variantie hier opnieuw een speciaal geval van covariantie
* Covariantie van X met zichzelf is de variantie van X!
* Als en onafhankelijk, dan is
* Omgekeerde niet altijd waar!!!

**Lineair verband**

* Voor lineaire verband:
	+ Rechte met positieve helling ‘perfect’ positief
	+ Rechte met negatieve helling ‘perfect’ negatief
	+ Perfect horizontaal/ verticaal geen **lineair** verband
* Het omgekeerde geldt ook!
* en geven enkel de graad van ***lineair*** verband!!!
	+ Hoe meer de correlatie afwijkt van , hoe minder lineair
* Correlatie zegt iets over de sterkte van het verband, covariantie niet

**Excel**

* COVARIANTIE(celbereik X , celbereik Y)
* CORRELATIE(celbereik X , celbereik Y) of PEARSON(celbereik X , celbereik Y)

**VARIANTIE VAN EEN LINEAIRE FUNCTIE VAN KANSVARIABELEN**

**Formule**

* Herbekijk bewijsje HBp179
* Gevolg
* Je kan de variantie natuurlijk ook gewoon bereken via de methode van substitutie zoals in H11

**Voorbeeld belegging**

* Standaard deviatie als maatstaf voor risico van een portefeuille
	+ Hoe hoger de , hoe groter het risisco
* Gediversifieerde, gedomineerde en efficiënte portefeuilles

**Covariantiematrix**

* Symmetrisch + positief semi-definiet
* Stel de covariantiematrix van het vb van de belegging opnieuw!