Stochastiek

Kansrekenen + statistiek

6 TW

3^{e} trimester

Naam leerling

………………………………………………

Deel 1: Kansrekenen

1 Kansrekenen

1.1 Leren tellen

“EN”	=>
“OF”	=>

vb Op de drankkaart staan er 5 soorten bier en 4 soorten frisdrank.

a Hoeveel keuzes heb je voor het kiezen van een drankje als je een biertje of een frisdrank mag kiezen?

b Hoeveel keuzes heb je voor het kiezen van een drankje als je een biertje en een frisdrank mag kiezen?

Algemeen: keuze van p elementen uit een verzameling van n elementen:

- Herhaling of niet?
- Volgorde telt of niet?

	volgorde speel een rol	volgorde speelt geen rol
geen herhaling		
herhaling		

vb 1 Je hebt 6 cijfers 1, 2, 3, 4, 5, 6 en 6 letters A, B, C, D, E, F. Op hoeveel manieren kan je een cijfer aan een letter koppelen?

vb 2 Hoeveel verschillende antwoorden kan je indienen met een lotto-formulier (waar je 6 cijfers uit de 42 moet kiezen)?

vb 3 Hoeveel “woorden” (moeten niet bestaan) zijn er met 6 letters?

vb 4 Je hebt 10 paaseieren en 3 smekende kinderen A, B en C. Op hoeveel manieren kan je al de eieren verdelen onder de kinderen?

Oefeningen

1 Een zaal heeft 5 deuren. Op hoeveel verschillende manieren kan iemand binnenkomen langs een deur en langs een andere deur weer naar buiten gaan?

2 Hoeveel anagrammen heeft het woord

a "ZOMER"?

b "PARALLELLOGRAM"?

3 Een klas van 20 leerlingen moet verdeeld worden in een groep van 8 en een groep van 12. Op hoeveel verschillende manieren kan dit?

4 Voor een cursus van het Rode Kruis worden acht lesuren voorzien: drie uur EHBO, één uur anatomie, twee uur gewondentransport en twee uur verbandenleer.

Op hoeveel manieren kan men het lessenrooster opstellen?
Een lessenrooster is bijvoorbeeld: E – E – A – G – E – G – V – V

5 Een volksdansgroep bestaat uit 24 personen waarvan één derde jongens zijn.

a Op hoeveel manieren kan een bestuur van 6 personen worden gevormd?

b Hoeveel mogelijkheden zijn er voor het bestuur als er 2 jongens en 4 meisjes in moeten zitten?

c Eens het bestuur gekozen zijn er zijn 10 opdrachten te verdelen onder de 6 bestuursleden. Hoeveel mogelijkheden zijn er in principe?

6 Op hoeveel manieren kan een sportleraar twee sportzakken verloten onder 7 leerlingen (Andy, Bert, Cindy, Dany, Eefje, Freddy en Gert)?

a De sportzakken zijn duidelijk verschillend en er wordt afgesproken dat iedere leerling slechts één sportzak kan winnen.

b De sportzakken zijn duidelijk verschillend en er wordt afgesproken dat één leerling de twee sportzakken kan winnen.

c De sportzakken zijn identiek en er wordt afgesproken dat één leerling slechts één sportzak kan winnen.

d De sportzakken zijn identiek en er wordt afgesproken dat één leerling de twee sportzakken kan winnen.

7 Op hoeveel manieren kan men 8 kaarten trekken uit een spel van 52 kaarten als er bij de acht kaarten …

a … precies drie azen moeten zitten?

b … precies twee azen en twee heren moeten zitten?

8 In een tenniswedstrijd is de setstand voorlopig 3-3.

Op hoeveel manieren kan deze tussenstand tot stand zijn gekomen?

1.2 Kansrekenen

Een **stochastisch experiment** bestaat uit:

- een **universum** = de uitkomstenverzameling U
- een **gebeurtenis** $G \subset U$

vb Je gooit met een dobbelsteen en je wil nagaan of je een even getal gooit.

$U =$

$G =$

1.2.1 Kans

- proefondervindelijk: relatieve frequentie voor $n \to +\infty$.
- theoretisch: kansregels:
  - $P(\emptyset) =$
  - $P(U) =$
  - $\quad \leq P(G) \leq$

<u>Formule van LAPLACE</u> (enkel als alle elementaire gebeurtenissen even waarschijnlijk zijn: geen vervalsingen):

$$\boxed{P(G) = \qquad\qquad\qquad\qquad}$$

vb Een dobbelsteen is zodanig vervalst dat $P(6) = 30\%$. De kans op de rest van de getallen is even groot.
Wat is de kans op een even getal?

1.2.2 Somregel (“of”)

$P(G_1 U G_2) =$

vb Je gooit met een niet-vervalste dobbelsteen. Wat is de kans dat je een even getal of een getal groter dan 3 gooit?

Opmerking

- ‘OF’ = niet exclusief: 1 of 2 of allebei
- Als $G_1 \cap G_2 = \emptyset$ dan noemen we dit een **disjuncte** gebeurtenis en alleen dan geldt $P(G_1 U G_2) =$

2 Kansverdelingen

2.1 Stochast en kansverdeling

Stochast = kansvariabele

- Discreet: eindig aantal antwoorden, vaak geteld (vb. aantal broers)
- Continu: oneindig aantal antwoorden, vaak gemeten (vb. lengte)

Kansverdeling: overzicht van alle waarden van je stochast met daarbij de overeenkomstige kans (tabel of grafische voorstelling).

vb kansexperiment = met 2 dobbelstenen gooien

Stochast X = som van de ogen (discreet)

Noteer de kansverdeling (tabel en grafische voorstelling)

2.2 Verwachtingswaarde en standaardafwijking van een stochast

2.2.1 Formules

Vergelijking beschrijvende statistiek (beschrijft verleden) ten opzichte van kansrekenen (beschrijft de toekomst): de kans bij kansrekenen is te vergelijken met de relatieve frequentie bij n keer ($n \to +\infty$) herhalen van het experiment.

vb

x_i	n_i
2	7
3	10
4	5

Bereken het gemiddelde.

In kansverdeling: **verwachtingswaarde:**

Variantie: $\sigma^2 = \frac{\sum_{i=1}^{p}(x_i - \bar{x})^2 \cdot n_i}{n}$

In kansverdeling:

Standaardafwijking: $\sigma = \sqrt{\sigma^2}$

2.2.2 Eigenschappen

Als X_1 en X_2 stochasten zijn, dan geldt:

$$\boxed{E(X_1 + X_2) = E(X_1) + E(X_2)}$$

$$\boxed{Var(X_1 + X_2) = Var(X_1) + Var(X_2)}$$

vb Je checkt in op het vliegveld. Je valies weegt 14,5 kg (met een σ van 0,5 kg) en van je partner 14,6 kg (met een σ van 0,2 kg). Bepaal de verwachtingswaarde en de standaardafwijking van de beide valiezen samen (stel X = totaal gewicht).

2.3 Discrete kansverdeling

2.3.1 Uniforme kansverdeling

= alle waarden x_i hebben dezelfde kans

vb 1 keer met een dobbelsteen gooien

2.3.2 Bernoulliverdeling

= er zijn slechts 2 mogelijke uitkomsten (stel: succes en mislukking)

$P(succes) = p$

$P(mislukking) = q =$

X = aantal successen

Kansverdeling:

$E(X) =$

$\sigma(X) =$

2.3.3 Binomiale verdeling

= Bernoulli-experiment n keer onafhankelijk uitvoeren

vb Je trekt (met teruglegging) 10 keer een kaart uit een boek van 52.

Geef de kansverdeling van de stochast X = aantal prentjes

$$\boxed{P(X = x_i) = }$$

GRM: 2nd VARS

Binompdf(n, p, k)

$E(X) =$

$\sigma(X) =$

2.3.4 Hypergeometrische verdeling

= herhaald Bernoulli-experiment, maar <u>NIET onafhankelijk</u>

vb In een doos van 30 lampen zijn er 10 defect. Wat is de kans dat er 3 stuk zijn als er 10 lampen uitgenomen worden …

a … met teruglegging?

b … zonder teruglegging?

2.4 Continue kansverdeling

<u>Discreet</u>	vs	<u>Continu</u>
(telbaar)		(oneindig)

2 onmiddellijke gevolgen:

$$\int_{-\infty}^{+\infty} pdf(x)dx =$$

$$\int_{a}^{a} pdf(x)dx =$$

2.4.1 Voorbeeld

Gegeven is de lengte van 8500 18-jarige jongens (normaal verdeeld):

Klasse	m_i	n_i
[150,5;155,5[	153	5
[155,5;160,5[	158	30
[160,5;165,5[	163	162
[165,5;170,5[	168	582
[170,5;175,5[	173	1371
[175,5;180,5[	178	2100
[180,5;185,5[	183	2108
[118,5;190,5[	188	1372
[190,5;195,5[	193	583
[195,5;200,5[	198	157
[200,5;205,5[	203	28
[205,5;210,5[	208	2

<u>Beschrijvende statistiek</u>

GRM: Lijsten ingeven

Stat – Calc

1-Var Stats L_1, L_2

$\bar{x} =$

$\sigma_x =$

Tekenen histogram:

2nd STAT Plot

Normale verdeling (kansmodel): klokcurve van Gauss

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

GRM: Y_1 = normalpdf$(X, \bar{X}, \sigma_X) * n * Xscl$

2.4.2 Toepassingen

1 Kansen berekenen

a Wat is de kans dat een 18-jarige jongen een lengte heeft tussen 170,5 cm en 180,5 cm op basis van het voorbeeld uit 2.4.1?

b Wat is de kans dat een jongen groter is dan 190 cm (vb uit 2.4.1)?

c De inhoud van flessen frisdrank is normaal verdeeld met $\mu = 50{,}5$ cl en $\sigma = 0{,}25$ cl (notatie: $X \sim N(50{,}5; 0{,}25)$).
Bepaal de kans dat de inhoud meer dan $49{,}9$ cl maar minder dan $51{,}1$ cl is.

Wat is de kans dat de inhoud van de fles meer dan $50{,}8$ cl is?

Wat is de kans dat de inhoud van de fles minder dan de geafficheerde 50 cl bevat?

2 Grenzen berekenen

Het IQ van mensen is normaal verdeeld met een gemiddelde van 100 en een afwijking van 12.

Tussen welke grenzen ligt het IQ van de middelste helft van de mensen?

3 *z-score*

Leerling A behaalde 84 % op een test met $\mu = 76\ \%$ en $\sigma = 6\ \%$.

Leerling B behaalde 90 % op een test met $\mu = 82\ \%$ en $\sigma = 10\ \%$.

Wie scoorde relatief gezien het best?

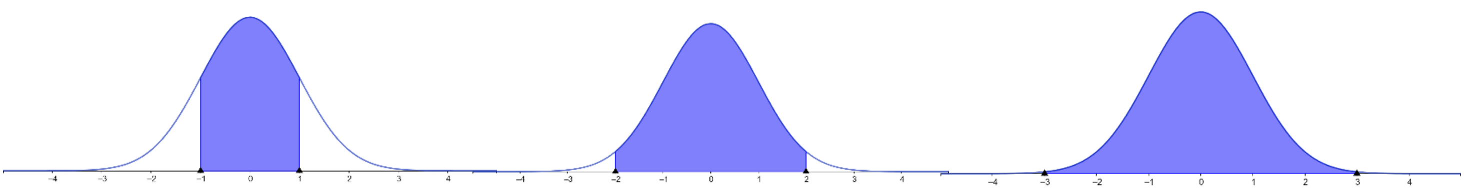
4 *Betekenis van de standaardafwijking*

Standaardnormale verdeling: normaalverdeling met $\mu = 0$ en $\sigma = 1$: $X \sim N(0,1)$

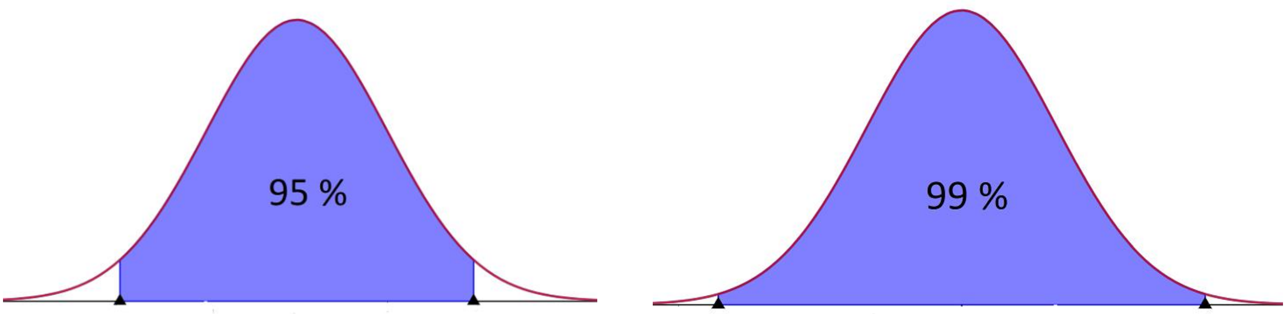
$$P(-1 < X < 1) =$$

$$P(-2 < X < 2) =$$

$$P(-3 < X < 3) =$$



Omgekeerd: grenzen zoeken:



<u>Samenvatting</u> normale verdeling:

- $normalpdf(x, \mu, \sigma)$: tekenen Gauss-curve
- $normalcdf(x_1, x_2, \mu, \sigma)$: berekenen kansen
- $invNorm(opp\ links,\ \mu, \sigma)$: berekenen grenzen
- Specifieke gevallen:
    - $[\mu - \sigma, \mu + \sigma] \rightarrow 68\%$
    - $[\mu - 2\sigma, \mu + 2\sigma] \rightarrow 95\%$
    - $[\mu - 3\sigma, \mu + 3\sigma] \rightarrow 99{,}7\%$

vb Onderscheid discrete en continue benadering.

Je gooit 100 keer met een muntstuk. Wat is de kans dat je tussen de 50 en 60 keer kop hebt (incl. 50 en 60).

a Discreet:

b Continu:

Oefeningen

1 Voor de jaarlijkse neerslag geldt dat deze normaal verdeeld is met een gemiddelde van 830 mm en een standaardafwijking van 60 mm. Wat is de kans dat er meer dan 915 mm neerslag valt?

2 Het gemiddelde gewicht in kg van een doos waspoeder is normaal verdeeld met parameters 7,92 kg en 0,20 kg. Wat is de kans dat een doos waspoeder minder dan 8 kg weegt?

3 De levensduur, gemeten in uren, van een bepaald type lampen is normaal verdeeld met $\mu = 1000$ en $\sigma = 120$. Wat is de kans dat een lukraak gekozen lamp meer dan 900 uren brandt?

4 De pH van het bloed bij mensen is normaal verdeeld met parameters 7,4 en 0,2.

a Wat is de kans dat, bij een willekeurige geteste persoon, de pH-waarde tussen 7,25 en 7,55 ligt?

b Indien de pH-waarde kleiner is dan 7,15 of groter dan 7,7 moet er een extra test worden gedaan. Bereken de kans op een extra test.

5 Een stochast is normaal verdeeld met verwachtingswaarde 100 en standaardafwijking 10. Hoe groot is de kans dat een resultaat meer dan 4 afwijkt van 100?

6 In een fabriek worden flessen automatisch gevuld. De inhoud van de gevulde flessen heeft een normale verdeling rond de inhoud waarop de machine wordt ingesteld. De standaardafwijking is 5 cc. De machine moet nu zo worden ingesteld dat 90% van de flessen een inhoud heeft van minstens 500 cc. Op welke inhoud moet de machine worden ingesteld?

7 Op pakjes margarine vind je meestal als gewicht 250 g E. Dit betekent dat volgens een Europese norm, niet meer dan 5 % van de pakjes margarine minder dan 250 g mag bevatten.

a De gewichten van margarine A zijn normaal verdeeld met standaardafwijking 7 g. Bereken het gemiddelde gewicht van een vlootje margarine opdat aan de Europese norm zou worden voldaan.

b De gewichten van margarine B zijn eveneens normaal verdeeld met een gemiddeld gewicht van 255 g. Bereken de maximale standaardafwijking opdat ook deze vlootjes aan de Europese norm zouden voldoen.

8 Het gewicht van een reiskoffer bij mannen is normaal verdeeld met en gemiddelde van 16 kg en een standaardafwijking van 2,5 kg. Bij vrouwen is dit ook normaal verdeeld, maar met een gemiddelde van 18 kg en een standaardafwijking van 1,5 kg. Michiel en Amber gaan op reis en elkeen pakt zijn/haar eigen koffer in. In totaal mogen ze 40 kg meenemen in het vliegtuig. Wat is de kans dat ze deze grens overschrijven (uitgaande van de veronderstelling dat het gewicht van beide koffers onafhankelijk is)?

Deel 2: Statistiek

- In ICT: beschrijvende statistiek: gegevens
  - verzamelen;
  - ordenen;
  - voorstellen;
  - berekenen.
- In 5TW en deel 1: Kansen en kansverdelingen
- Dit deel: verklarende inferentiële statistiek
  - Welke conclusies kan je halen uit metingen?
  - Welke bevindingen uit je steekproef kan je doortrekken naar je populatie?

3 Uitspraken op basis van gegevens

3.1 Het belang van een goeie steekproef

Onbetrouwbaar (= vertekend)

- Manipulatie van resultaten
- Vrijwillige respons
- Opportunistische steekproef
- Suggestieve vragen
- ...

Betrouwbaar (= onvertekend)

EAS: enkelvoudige aselecte steekproef

Wij gaan steeds uit van een EAS als die door een peilingbureau is uitgevoerd (ook al staat het niet expliciet vermeld).

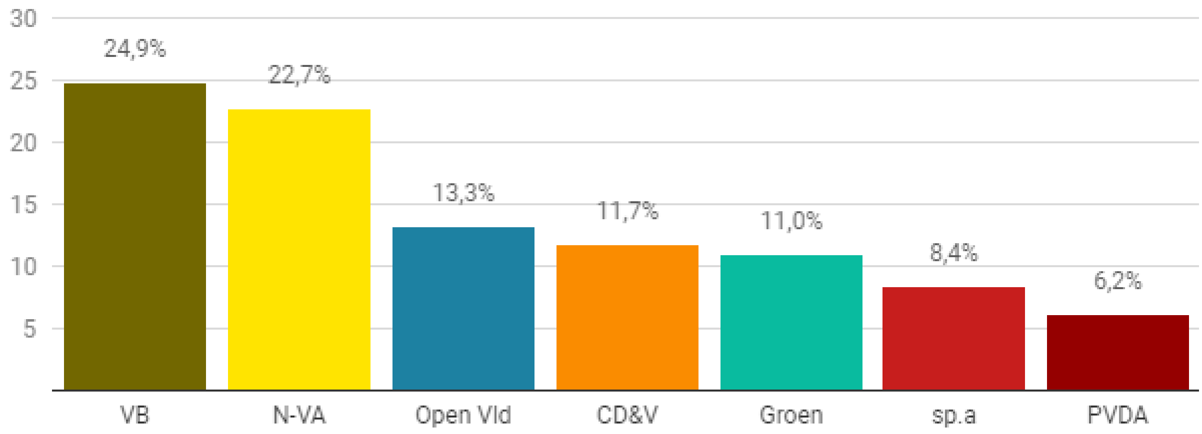
vb Oef 6 blz. 15

AIPO (democraten)	LD (republikeinen)
Populatie:	Populatie:
Steekproef:	Steekproef:

3.2 Betrouwbaarheid van een steekproef

3.2.1 Inleidend voorbeeld

Peiling verkiezingen[1]:



Onderaan staat:

Deze peiling bij 2.540 respondenten, die representatieve steekproeven vormen van de Belgen van 18 jaar en ouder, a rato van 992 in Wallonië, 1000 in Vlaanderen en 548 in de 19 gemeentes van het Brussels Hoofdstedelijk Gewest werd uitgevoerd van 2 tot en met 10 september 2019. De interviews werden online afgenomen. De maximale foutenmarge, voor een percentage van 50% en een vertrouwensgraad van 95% bedraagt +-3,1 in Wallonië en in Vlaanderen en +-4,2 in Brussel. Aangesloten bij: ESOMAR, Consumer Understanding Belgium.

Bij een uitspraak over statistische gegevens hoort steeds een **foutenmarge** en een **betrouwbaarheidsniveau**.

Er kan ook een **betrouwbaarheidsinterval** opgesteld worden.

$$verwachte\ waarde\ \pm foutenmarge$$

Wanneer het betrouwbaarheidsniveau bij een opiniepeiling niet wordt vermeld, dan is het gelijk aan 95 %.

[1] Bron: HLN, Le Soir, VTM, RTL TVI

3.2.2 Betrouwbaarheidsniveau en foutenmarge

We willen een populatieparameter (populatieproportie p) schatten aan de hand van deze parameter uit de steekproef $\hat{p}$. Om inzicht te krijgen, doen we niet 1, maar veel steekproeven en bekijken we het (wispelturig?) gedrag van $\hat{p}$ en kunnen we verder conclusies trekken bij het uitvoeren van 1 steekproef.

vb Oef 9 p 21-22

Populatie: ....

Steekproef: ....

p: relatieve frequentie van bloedgroep O (in de populatie, dus onbekend)

Proportie (in %)	n_i	f_i
$[0,5[$		
$[5,10[$		
$[10,15[$		
$[15,20[$		
$[20,25[$		
$[25,30[$		
$[30,35[$		
$[35,40[$		
$[40,45[$		
$[45,50[$		
$[50,55[$		
$[55,60[$		
$[60,65[$		
$[65,70[$		
$[70,75[$		
$[75,80[$		
$[80,85[$		
$[85,90[$		
$[90,95[$		
$[95,100[$		

Besluiten:

....

....

....

....

....

....

Conclusies:

<u>Stap 1</u>: Globale uitspraak

....

....

....

....

<u>Stap 2</u>: Uitspraak over 1 steekproef

....

....

<u>Stap 3</u>: Uitspraak over p

....

....

4 Steekproefverdelingen

4.1 Uitschieters

Uitschieters moet je altijd met zorg behandelen: zie ook vb Newcomb blz. 51 versus vb ozonlaag blz. 53.

We stellen dit ook grafisch (m.b.v. GRM) voor a.d.h.v. het vb van Newcomb op blz. 51 en tekenen een boxplot.

Definitie uitschieters (Tukey): alle gegevens die op $1{,}5 \cdot IKA$ buiten de box liggen, worden als uitschieters aanzien.

We vergelijken het histogram en de Gausscurve eens met en zonder de uitschieters als we aannemen dat de gegevens normaal verdeeld zijn m.b.v. de GRM.

4.2 Centrale limietstelling

CENTRALE LIMIETSTELLING

$\hat{p}$: steekproefverdeling van een proportie

normaal verdeeld met:

- gemiddelde $\mu_{\hat{p}} = p$
- standaardafwijking $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Wanneer mag dit?

- $n_{populatie} \geq 10 \cdot n_{steekproef}$
- $n \cdot p \geq 10$
- $n \cdot (1-p) \geq 10$

Volgens de limietstelling ligt p dan met een betrouwbaarheid van 95 % in het interval:

Probleem: heel vaak is p niet gekend en zullen we dit benaderen door gebruik te maken van $\hat{p}$.

vb Een steekproef onder 120 studenten leverde 37 rokers op. Doe een uitspraak met een betrouwbaarheid van 95 % over de proportie rokers.

5 Uitspraken op basis van steekproefverdelingen

5.1 Betrouwbaarheidsintervallen

Dankzij de Centrale limietstelling weten we dat $\hat{p}$ normaal verdeeld is met $\mu =$ en $\sigma =$

Een correcte statistische uitspraak luidt:

“met een betrouwbaarheid van ligt de populatieproportie in het interval

Merk op: p komt voor in de foutenmarge

=> we zullen $\hat{p}$ gebruiken als onvertekende schatter van p.

5.2 Significantietesten

vb Je vermoedt dat een dobbelsteen vervalst is (6 komt meer voor). Daarom gooide je de dobbelsteen 600 keer op en noteerde je dat 6 115 keer voorkwam.

Heb je argumenten genoeg op je hypothese sterk te maken of kan het aan toeval liggen?

Als P klein is ($< 0{,}05, < 0{,}01$) =>

Als P groot is ($\geq 0{,}05, \geq 0{,}01$) =>