Stochastiek

www.karelappeltans.be

March 28, 2024

1 Inleidende vragen

Probeer eerst een antwoord te geven op onderstaande vragen. Na het voltooien van de cursus doe je dit opnieuw.

1. Stel je een vrouw voor die Linda heet, eenendertig jaar oud is, alleenstaand, extravert en zeer intelligent. Ze heeft filosofie gestudeerd. Als student was ze zeer begaan met sociale rechtvaardigheid, streed ze tegen discriminatie en nam ze deel aan antinucleaire demonstraties. Wat is het minst waarschijnlijk:

   (a) Linda is actief in de feministische beweging

   (b) Linda is baliemedewerkster bij een bank en actief in de feministische vrouwenbeweging

   (c) Linda is balie medewerkster bij een bank

2. Er zitten 26 lln in onze klas. Hoe groot is de kans dat er minstens twee lln op dezelfde dag verjaren?

3. Thomas heeft reeds een intuïtief begrip van kansen, en hoe die berekend kunnen worden. We nodigen Thomas uit om mee te spelen in een dobbelspelletje dat als volgt gaat. Nadat hij tien euro heeft ingezet mag hij driemaal met een dobbelsteen gooien. Gooit hij minstens eenmaal een zes, dan krijgt hij zijn inzet plus twaalf euro terug. In het andere geval is hij zijn inzet kwijt. Thomas redeneert: ‘Als ik driemaal mag gooien, verdrievoudigt mijn kans op succes ook, dus heb ik $3/6$ kans om te winnen. Aangezien ik in de helft van de gevallen twaalf euro win (en in de andere helft van de gevallen tien euro verlies) zal het spelletje dus wel in mijn voordeel zijn. Akkoord?

4. In een zekere stad zijn er twee hospitalen, een klein, waarin men dagelijks gemiddeld vijftien geboorten telt en een groot met gemiddeld 45 geboorten per dag. De kans op de geboorte van een jongen bedraagt ongeveer 50% (Uiteraard zijn er dagen waarop er meer dan 50% jongens worden geboren en dagen waarop er minder dan 50% jongens worden geboren.) In het kleine hospitaal noteert iemand gedurende een gans jaar de dagen waarop er meer dan negen jongens worden geboren, hetgeen dus meer is dan 60% van het totaal aantal geboorten in dat hospitaal. In het grote hospitaal noteert iemand gedurende een gans jaar de dagen waarop er meer dan 27 jongens worden geboren, hetgeen dus meer is dan 60% van het totaal aantal geboorten in dat hospitaal. In welk van de twee hospitalen komen zulke dagen het vaakst voor?

5. Stel dat de deelnemers aan een spelshow uit drie deuren moeten kiezen: achter de ene deur staat een auto, achter de andere twee een geit. Nadat een deelnemer een deur heeft gekozen, opent de presentator, die weet wat er zich achter alle deuren bevindt, een van de niet-gekozen deuren, waarachter een geit te zien is? Dan zegt hij tegen de kandidaat: ‘wil je toch niet liever de andere deur openen?’ Wordt de kandidaat er beter van als hij voor de andere deur kiest?

6. Sally Clark bevalt in 1996 van een zoon, die hetzelfde jaar nog overlijdt (diagnose wiegendood). In 1998 overlijdt echter ook haar tweede zoon na acht weken, opnieuw naar verluidt aan wiegendood. Toen dat gebeurde werd Sally gearresteerd en beschuldigd van het verstikken van beide kinderen. Een ‘expert’ schatte de kans op wiegendood op $1/8513$, dus de kans dat beide baby’s eraan zouden zijn overleden op 1 op 73 miljoen $\left(\frac{1}{8513} \cdot \frac{1}{8513}\right)$. Deze kans

is zo klein dat het wel moord moest zijn. De jury vond dit ook en in november 1999 werd Sally Clark naar de gevangenis gestuurd. Een terechte beslissing van de rechtbank?

7. Tijdens de wereldbeker voetbal van 2010 in Zuid-Afrika wist Paul de octopus alle zeven de uitslagen van de Duitse ploeg correct op voorhand te voorspellen. En de uitslag van de finale. Een voetbalkennende octopus? (zie bijv. http://www.youtube.com/watch?v=Ya85knuDzp8)

8. In 2007 kreeg een groep van 160 gynaecologen in Engeland onderstaande informatie voorgelegd over de nauwkeurigheid van mammogrammen en borstkankerprevalentie onder de populatie:

   (a) De waarschijnlijkheid dat een vrouw borstkanker heeft is 1% (prevalentie)

   (b) Als een vrouw borstkanker heeft, is de waarschijnlijkheid van een positieve testuitslag 90%

   (c) Als een vrouw geen borstkanker heeft, is de waarschijnlijkheid van een positieve testuitslag desondanks 9%

   Hierna kregen de artsen een multplechoicevraag voorgelegd met het verzoek aan te geven welke constatering het beste de kans weergeeft dat een patiënt met een positief mammogram inderdaad borstkanker heeft:

   (a) De waarschijnlijkheid dat ze borstkanker heeft is ongeveer 81%.

   (b) van de 10 vrouwen met een positief mammogram hebben er ongeveer 9 borstkanker.

   (c) Van de 10 vrouwen met een positief mammogram heeft er ongeveer 1 borstkanker.

   (d) De waarschijnlijkheid dat ze borstkanker heeft is ongeveer 1%.

   Het populairste antwoord bij de gynaecologen was (a). Zien ze het goed?

9. Onderstaand artikel komt uit de Knack en geeft toelichting bij een opiniepeiling. Waarop slaat die maximale fout?

> De online enquête werd uitgevoerd door Kantar bij 1.057 Belgen van 18 jaar of ouder, van 7 t.e.m. 11 januari 2020. De maximale statistische fout bedraagt 3,1% boven en onder het bekomen resultaat voor uitspraken op de gehele steekproef. De steekproef werd herwogen op geslacht, leeftijd, beroep, opleiding en regio op basis van meest recente CIM cijfers voor deze socio-demografische variabelen.

2 Begripsvorming

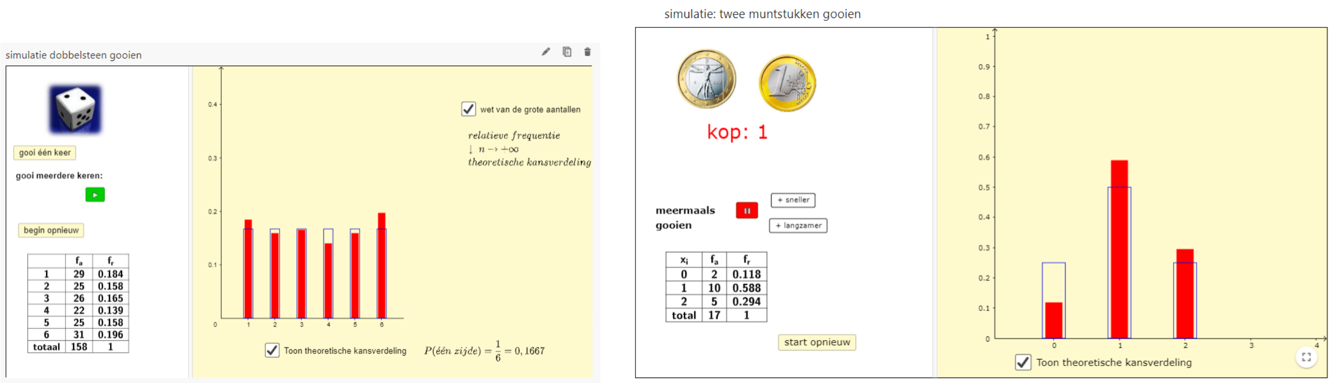


Figure 1: https://www.geogebra.org/m/b5a93xkf

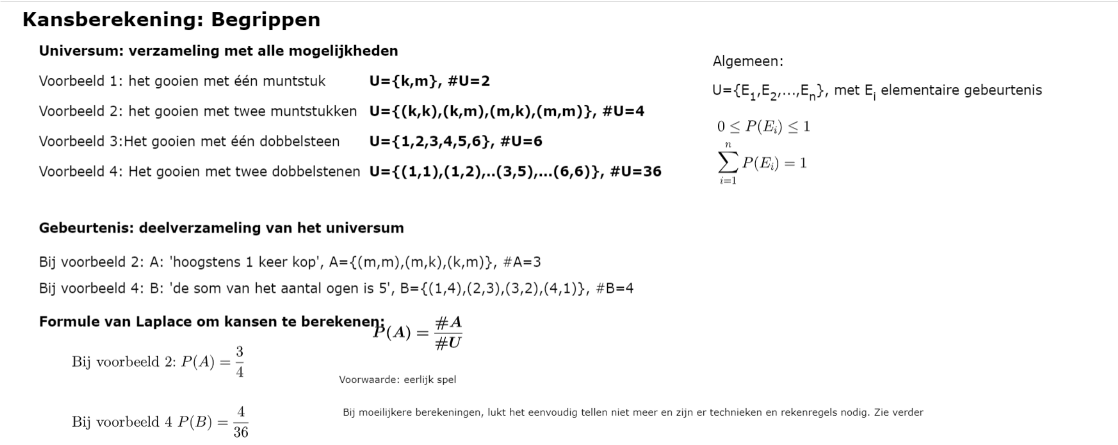


Figure 2: https://www.geogebra.org/m/b5a93xkf

3 Technieken om kansen te berekenen

3.1 raster

Figure 3: https://www.geogebra.org/m/WqZ3Cj2d

3.2 Venndiagrammen

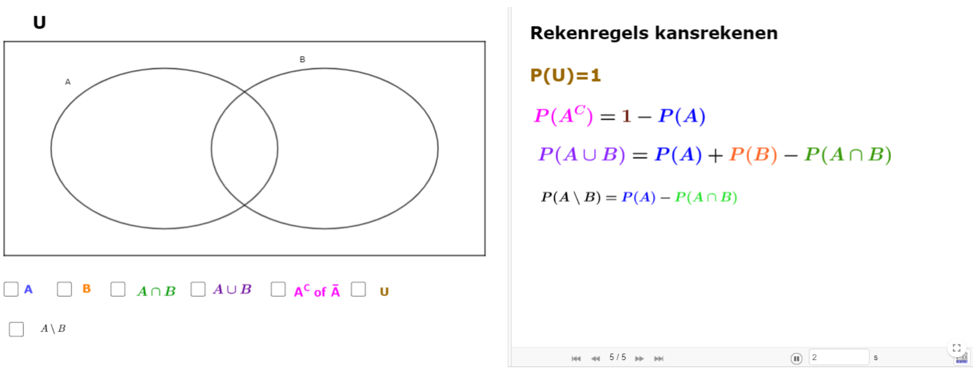


Figure 4: https://www.geogebra.org/m/WqZ3Cj2d

3.3 kansboom

3.3.1 basis

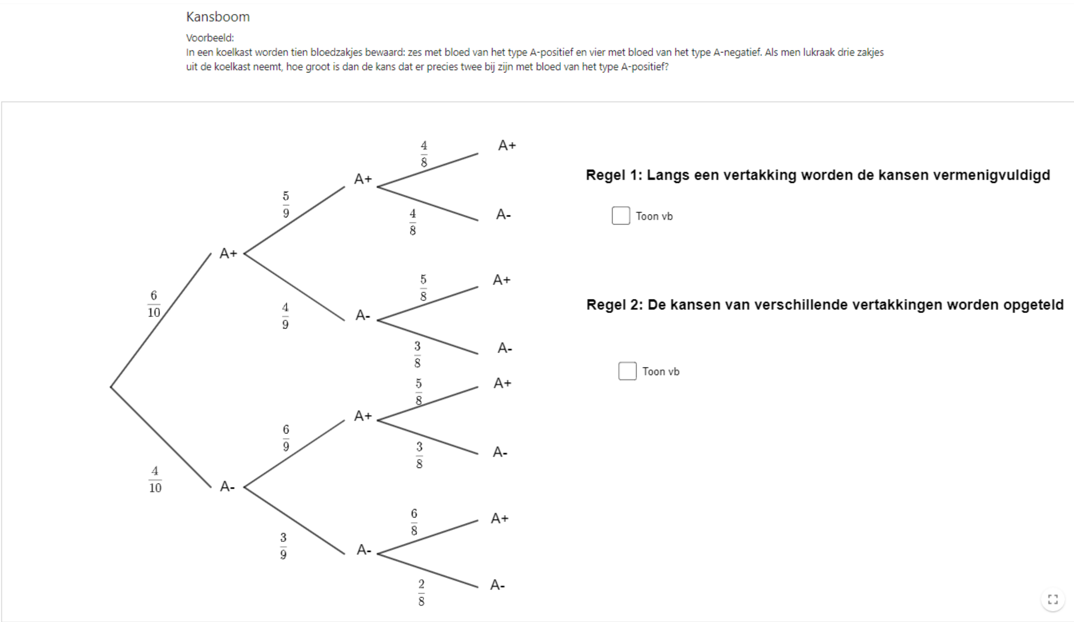


Figure 5: https://www.geogebra.org/m/WqZ3Cj2d

3.3.2 vertakkingen tellen

stap 1: bepaal de kans langs één vertakking stap 2: bepaal via de techniek van anagram het aantal vertakkingen

3.4 technieken uit telproblemen

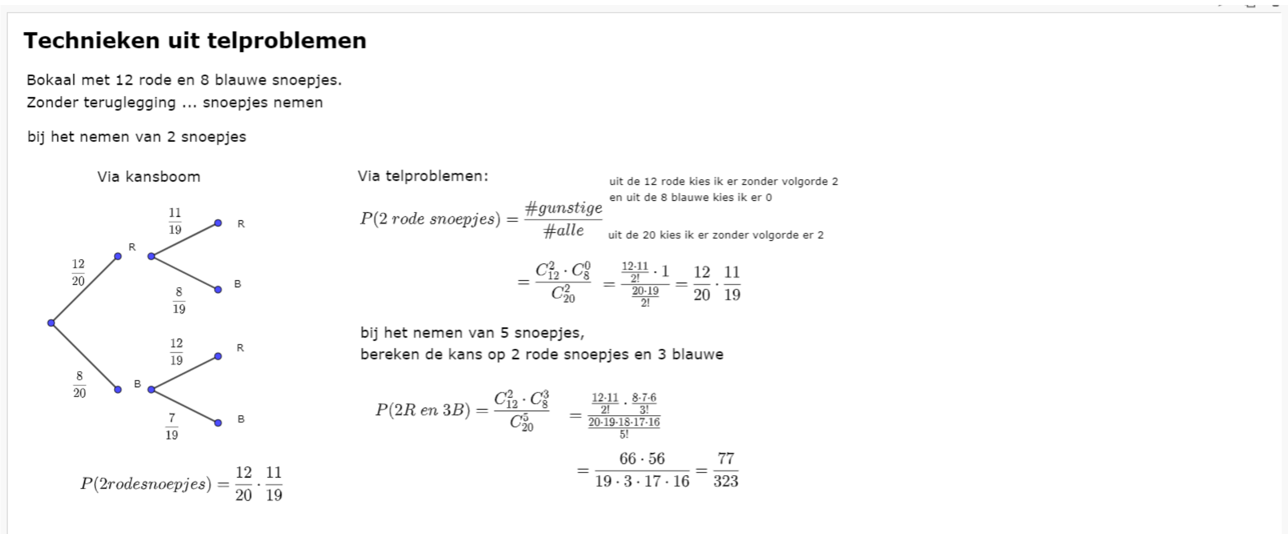
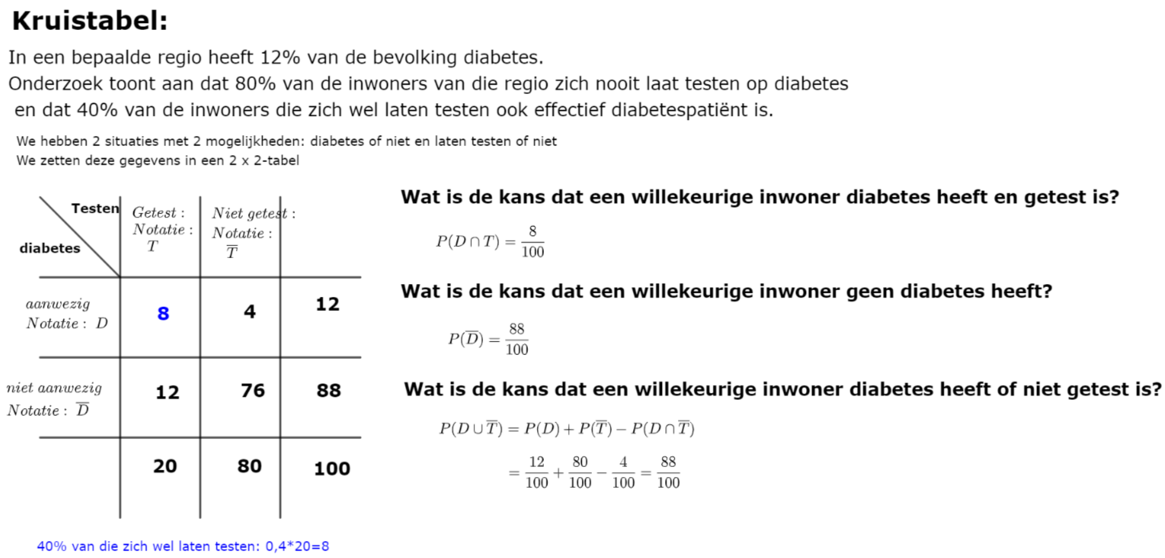


Figure 6: https://www.geogebra.org/m/WqZ3Cj2d

3.5 kruistabel



Kruistabel:

In een bepaalde regio heeft 12% van de bevolking diabetes.
Onderzoek toont aan dat 80% van de inwoners van die regio zich nooit laat testen op diabetes
en dat 40% van de inwoners die zich wel laten testen ook effectief diabetespatiënt is.

We hebben 2 situaties met 2 mogelijkheden: diabetes of niet en laten testen of niet
We zetten deze gegevens in een 2 x 2-tabel

Testen / diabetes	Getest : Notatie : T	Niet getest : Notatie : $\overline{T}$	
aanwezig Notatie : D	8	4	12
niet aanwezig Notatie : $\overline{D}$	12	76	88
	20	80	100

40% van die zich wel laten testen: 0,4*20=8

Wat is de kans dat een willekeurige inwoner diabetes heeft en getest is?

$P(D \cap T) = \frac{8}{100}$

Wat is de kans dat een willekeurige inwoner geen diabetes heeft?

$P(\overline{D}) = \frac{88}{100}$

Wat is de kans dat een willekeurige inwoner diabetes heeft of niet getest is?

$$P(D \cup \overline{T}) = P(D) + P(\overline{T}) - P(D \cap \overline{T})$$
$$= \frac{12}{100} + \frac{80}{100} - \frac{4}{100} = \frac{88}{100}$$

Figure 7: https://www.geogebra.org/m/QtdpYKT3#material/WqZ3Cj2d

3.6 oefeningen

1. We gooien met twee eerlijke dobbelstenen. Hoe groot is de kans de som van de gegooide ogen deelbaar is door 5?

2. We gooien met twee dobbelstenen. Zij A de gebeurtenis 'de som van de ogen is deelbaar door 3' en B de gebeurtenis 'de som van de ogen is deelbaar door 4'. Bereken volgende kansen: $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \cup B)$, $P(\bar{B})$, $P(B \setminus A)$

3. Gegeven: $P(A \cap B) = \frac{1}{4}$, $P(A \cup B) = \frac{5}{8}$, $P(A \setminus B) = \frac{1}{8}$
   Gevraagd: $P(A)$, $P(B)$, $P(B \setminus A)$ en $P(A^C \cap B)$

4. Gegeven: $P(C) = 0,25$, $P(D) = 0,45$ en $P(C \cap D) = 0,1$
   Gevraagd: $P(\bar{C} \cap D)$ (antw: 0,35)

5. Toon aan dat voor de kans op de unie van drie gebeurtenissen geldt:

   $$P(A \cup B \cup C) = P(A) + P(B) + P(C) + P(A \cap B \cap C) - P(A \cap B) - P(B \cap C) - P(A \cap C)$$

6. Een geldstuk is vervalst zodat kop dubbel zoveel kan voorkomen als munt. Als het muntstuk drie keer geworpen wordt, wat is de kans om juist 2 keer munt te hebben?

7. Bereken de kans op 4 juist nummers bij een trekking van de Belgische lotto (6 nummers uit 45).

8. Bereken de kans op rang 5 bij Euromillions (rang 5: 4 winnende nummers en één ster). Euromillions zijn twee trekkingen tegelijkertijd 5 uit 50 (de nummers) en 2 uit 12 (de sterren).

9. Bereken de kans op drie jongens in een gezin met acht kinderen? (A: $\frac{56}{256}$)

10. Een doos bestaat 12 stukken fruit: 7 appels en 5 peren. Wat is de kans dat het laatste stuk fruit dat genomen wordt een appel is (A. 7/12)

11. Twee vrouwen en drie mannen gaan willekeurig naast elkaar zitten aan een ronde tafel. Wat is de kans dat de twee vrouwen naast elkaar zitten? (A. $\frac{1}{2}$)

12. Vijf vrienden zitten op een terras. Twee hebben bier besteld en de andere drie een frisdrank. De serveerster is vergeten wie wat besteld heeft en zet de drankjes blindelings neer. Wat is de kans dat het goed gaat? (A. $\frac{1}{10}$)

13. In een wachtzaal van de dokter zitter er acht personen, waarvan 4 vrouwen en 4 mannen. Alle personen komen willekeurig aan de beurt voor een consultatie. Wat is de kans dat de eerste persoon een vrouw is en dat ook de laatste persoon die aan de beurt is een vrouw is.

14. In een woonzorgcentrum lijdt 8% van de mannen en 4% van de vrouwen aan de ziekte van Parkinson. Onder de bewoners kiest men lukraak één man en één vrouw. Hoe groot is de kans dat precies één van beiden aan deze ziekte lijdt?

15. Twee jongens en 6 meisjes nemen in willekeurige volgorde plaats op één van de acht stoelen die naast elkaar op een rij staan. Hoe groot is de kans dat er precies twee meisjes tussen de twee jongens zitten (A. $\frac{3}{14}$)

16. In een koelkast worden tien bloedzakjes bewaard: zes met bloed van het type A-positief en vier met bloed van het type A-negatief. Als men lukraak zonder te kijken drie bloedzakjes uit de koelkast neemt, hoe groot is dan de kans dat er precies twee bij zijn met het bloed van type A-positief. (A: zie website)

17. Een hotel telt tien verdiepingen (gelijkvloers niet meegerekend). Vijf personen stappen in de lift op het gelijkvloers. Hoe groot is de kans dat zij elk op een andere verdieping uitstappen? (A: zie website)

18. In een woonzorgcentrum zijn 40% van de bewoners mannelijk en 60% vrouwelijk. Er breekt een besmetting door een virus uit. Voor de vrouwelijke bewoners is de kans op besmetting $\frac{1}{4}$ en voor de mannelijke bewoners $\frac{1}{2}$. Als een vrouw besmet is, is de kans op overlijden $\frac{1}{3}$. Als een man besmet is, is de kans op overlijden $\frac{1}{4}$. Hoe groot is de kans dat een lukraak gekozen bewoner van het centrum overlijdt ten gevolge van dit virus? (A. 10%).

19. De bovenbloeddruk van mannen, gemeten in mmHg, is normaal verdeeld met een gemiddelde van 128.5 en een standaardafwijking van 11.5. Hoe groot is de kans dat bij een willekeurig gekozen koppel mannen de bovenbloeddruk bij beiden hoger is dan 140 mmHG? (Antw. 2.56%)

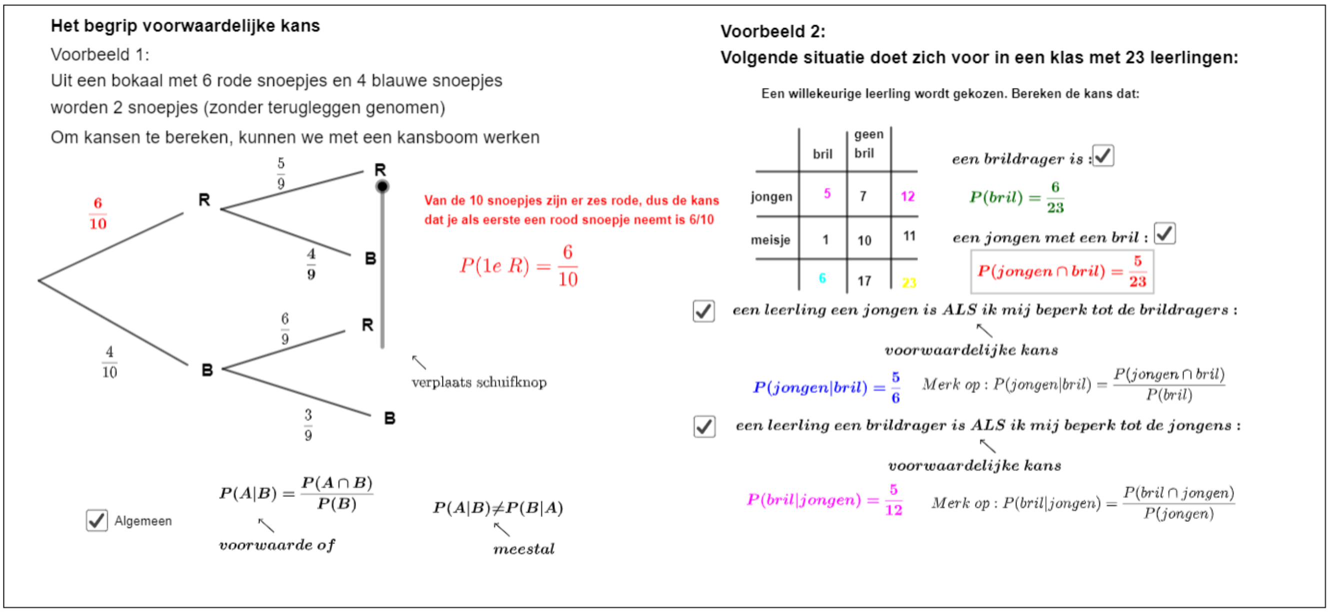
20. Wat is de kans dat in een bus met 50 passagiers minstens 2 mensen op dezelfde dag verjaren?

Figure 8: https://www.geogebra.org/m/WqZ3Cj2d

21. Op 7 maart 2019 won een man uit North Carolina een miljoen dollar bij een loterij. Niks bijzonders, want iemand moet dat miljoen winnen. Deze man had echter in april 2017 ook al eens een miljoen gewonnen. Hoe uitzonderlijk is dit in volgende loterij: Van de getallen 1 tot en met 39 moet je er vijf aankruisen (ga na dat dit op 575757 manieren kan). Dagelijks is er een trekking. Hoe groot is de kans dat er in een jaar tijd (365 trekkingen) minstens twee keer dezelfde vijf getallen worden getrokken (de man speelt namelijk altijd met dezelfde 5 getallen)? (A. bijna 11%)

4 Voorwaardelijke kans

4.1 Begripsvorming



	bril	geen bril	
jongen	5	7	12
meisje	1	10	11
	6	17	23

Figure 9: https://www.geogebra.org/m/m2unubtm

4.2 Statistische (on)afhankelijkheid

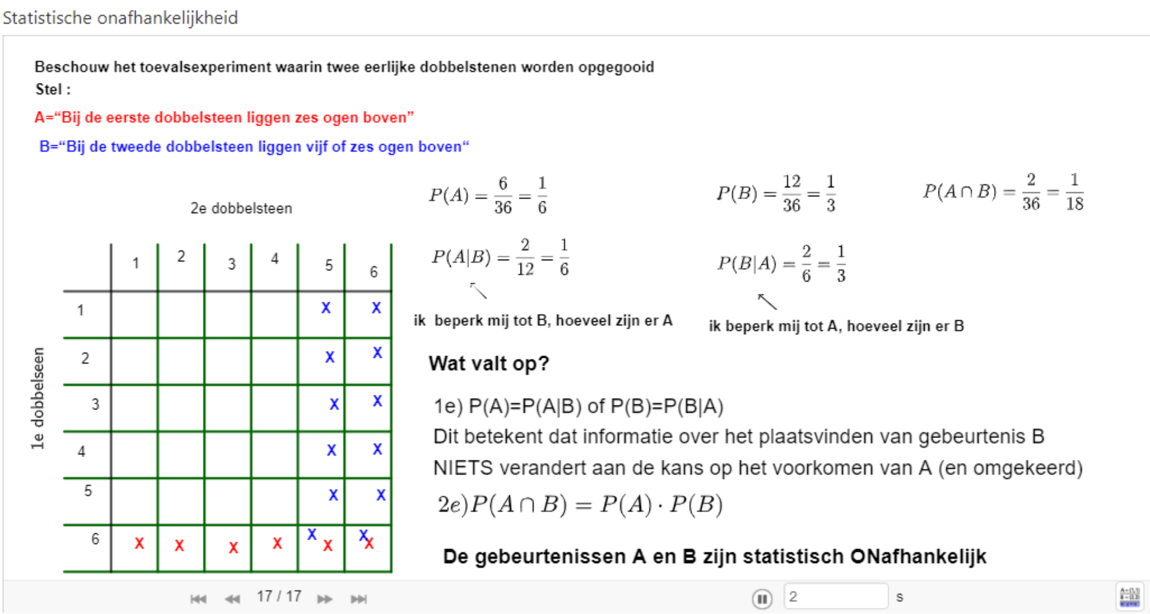


Figure 10: https://www.geogebra.org/m/m2unubtm

4.3 regel van Bayes

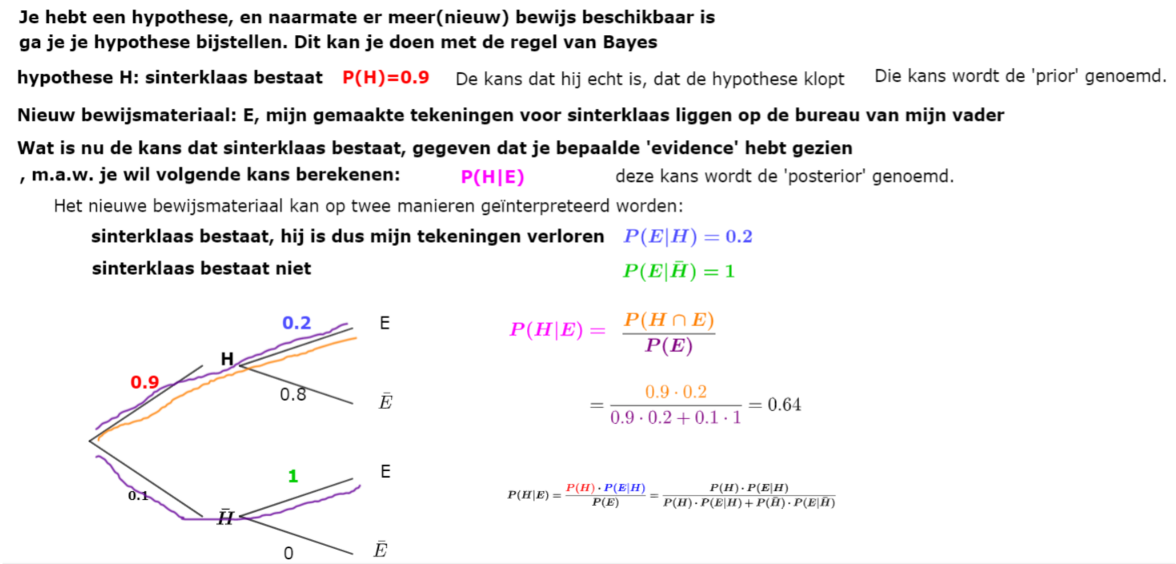


Figure 11: https://www.geogebra.org/m/hF46f3Js

Tijdens de corona-epidemie had ik de hypothese dat ik besmet was met het coronavirus, $P(H) = P(C) = 0.2$. Ik liet mij testen en de test was positief, op welke manier zal dit de priorkans doen veranderen, m.a.w. wat is $P(H|E) = P(C|T+)$?

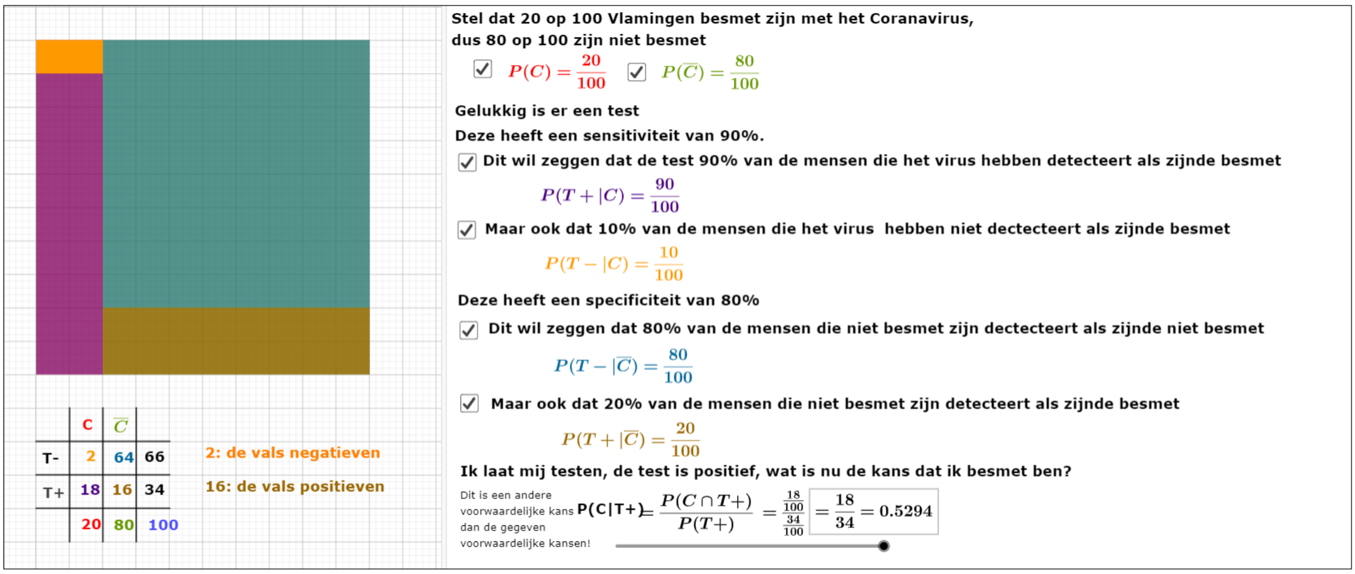


Figure 12: https://www.geogebra.org/m/hF46f3Js

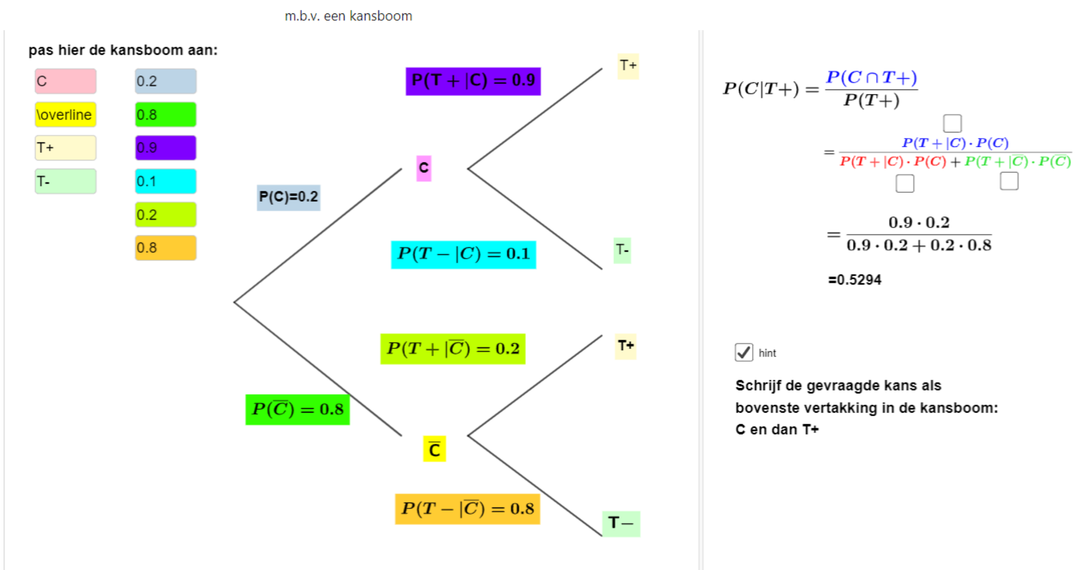


Figure 13: https://www.geogebra.org/m/hF46f3Js

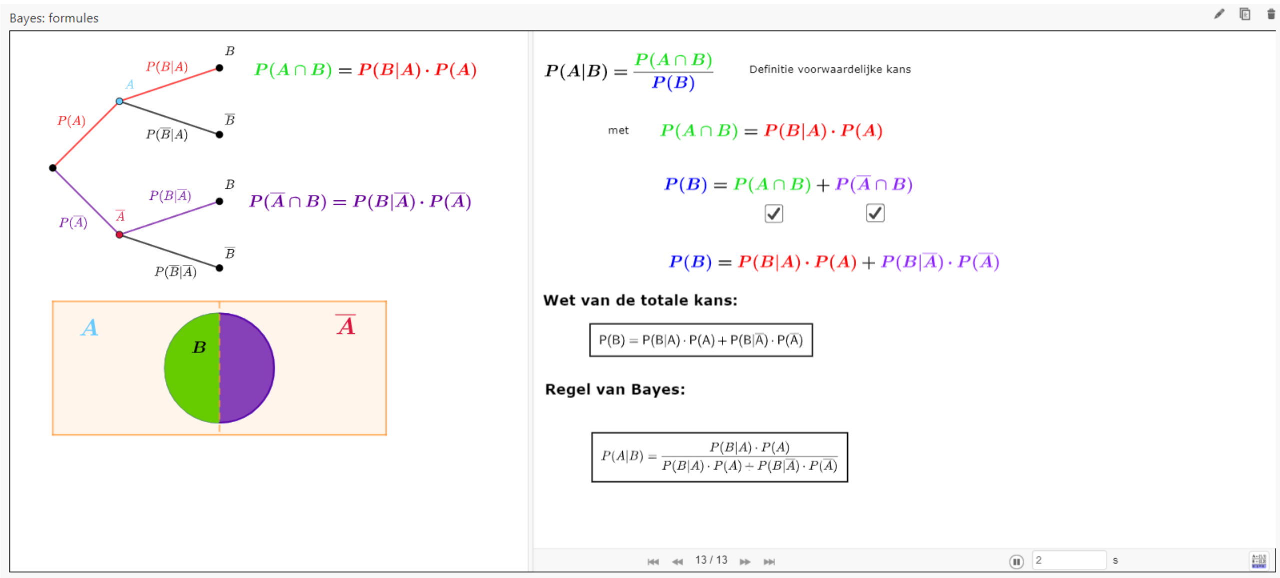


Figure 14: https://www.geogebra.org/m/hF46f3Js

4.4 oefeningen

1. kaartspel

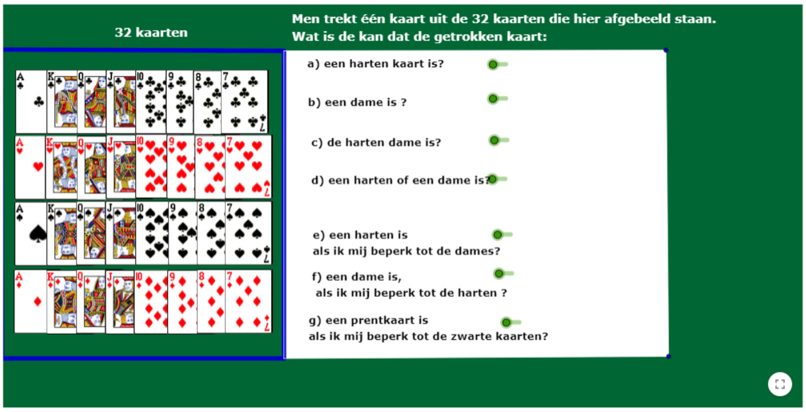


Figure 15: https://www.geogebra.org/m/m2unubtm

2. Het operatief verwijderen van een niersteen. Hangt de kans op complicaties af van de keuze van de dokter?

	complicatie	geen complicatie	
dr. A.	130	360	490
dr. B.	200	310	510
	330	670	1000

   Kan je dat eigenlijk wel concluderen op basis van deze cijfers?

3. Er wordt met 2 dobbelstenen gegooid. Bekijk volgende gebeurtenissen:
   A='som van de ogen bedraagt 3'
   B='som van de ogen is 7'
   C='tenminste 1 van de dobbelstenen toont 1'
   (a) Bereken $P(A|C)$
   (b) Bereken $P(B|C)$
   (c) Zijn A en B statistisch (on)afhankelijk? Zijn B en C statistische (on)afhankelijk?

4. Veronderstel een gezin met twee kinderen. Beschouw de gebeurtenissen A: het gezin bestaat uit 1 jongen en 1 meisje en B: hoogstens 1 meisje. Zijn deze gebeurtenissen statistisch (on)afhankelijk?

5. Zelfde vraag voor een gezin met drie kinderen.

6. In een woonzorgcentrum lijdt 8% van de mannen en 4% van de vrouwen aan de ziekte van Parkinson. Onder de bewoners kiest men lukraak één man en één vrouw. Hoe groot is de kans dat precies één van beiden aan deze ziekte lijdt?

7. De vaders van de huidige koningen van België en Spanje gingen vroeger samen op olifantenjacht. De kans dat Carlos een olifant doodschoot was 0,2. Voor Albert bedroeg die kans 0,3. Deze kansen zijn onafhankelijk van elkaar. Wat was de kans dat een olifant die hun pad kruiste werd neergeschoten?

8. A en B zijn onafhankelijke gebeurtenissen. De kans dat beiden samen voorkomen is $\frac{1}{6}$ en de kans dat geen beide voorkomt is $\frac{1}{3}$. Bereken $P(A)$ en $P(B)$

9. Aan Amerikaanse universiteiten moet je een aanvraag indienen om toegelaten te worden. Bij Jhon Hopkins medical school (JH) heb je 50% kans om toegelaten te worden, aan de Harvard University medical school (UH) is dit 40%.
   (a) Als deze kansen onafhankelijk zijn van elkaar, bereken dan de kans om toegelaten te worden bij JH of HU. (A. 70%

(b) Als deze kansen afhankelijk zijn (zij zijn immers op zoek naar hetzelfde type student), is de kans om toegelaten te worden, kleiner of groter dan de zojuist berekende kans? Verklaar jullie antwoord. (A. kleiner)

10. In de correctionele rechtbank hier in Hasselt wordt in 45% van de zaken een gevangenisstraf uitgesproken. Van deze 45% blijkt slechts 40% schuldig gepleit te hebben. 20% van degenen die geen gevangenisstraf gekregen, hadden ook schuldig gepleit. Bij een willekeurige zaak, als ik weet dat er schuldig gepleit werd, wat is de kans dat er geen gevangenisstraf was uitgesproken?

11. Een bepaalde ziekte heeft een prevalentie van 2% Het vals negatief percentage bedraagt 10% en het vals positief percentage bedraagt 1%. Bereken de waarschijnlijkheid dat een persoon, die positief test, effectief de ziekte heeft. (A. 0,647)

12. Een bedrijf licht zijn aanwervingsbeleid door. Het blijkt achteraf dat 80% van het aangeworven personeel voldoet aan de verwachtingen. Van deze 80% had 75% verkoopservaring. Van het nieuw personeel dat niet voldeed had 55% verkoopservaring. Als ik een willekeurig nieuw personeelslid selecteer en deze blijkt verkoopservaring te hebben, wat is de kans dat hij/zij voldoet aan de verwachtingen? (A. 0.845)

13. In een bepaalde regio heeft 12% van de bevolking diabetes . Onderzoek toont aan dat 80% van de bevolking in die regio zich nooit laat testen op diabetes en dat 40% van diegenen die zich wel laten testen ook effectief diabetespatiënt zijn. Wat is de kans dat iemand die zich niet laat testen op diabetes toch een diabetespatiënt is?

14. Frank Deboosere en Sabine Haegedoren voorspellen samen het weer. Als de weersvoorspelling fout is, dan blijkt dat in 65% van de gevallen de voorspelling kwam van Frank. Wie het weer voorspelt, wordt bepaald door het opgooien van een muntstuk. Van één vierde van alle voorspellingen kan men zeggen dat ze én juist én door Frank worden gedaan.

    (a) Wat is de kans dat als Frank het weer voorspelt, dat dit een foute voorspelling is?

    (b) Wat is de kans dat als Sabine het weer voorspelt, dit een foute voorspelling is?

    (c) Wat is de kans dat als een voorspelling juist is, deze werd gemaakt door Frank?

15. In een bepaalde stad zijn er twee taxibedrijven: het ene heeft groene taxi's, het andere blauwe: 85% van de taxi's zijn blauw, de overige 15% zijn groen. Op een nacht raakt een taxi betrokken in een auto-ongeluk en pleegt vluchtmisdrijf. Er was echter een getuige op de plaats van het gebeuren: deze beweert dat de taxi groen was. Het gerecht onderzoekt de kleuren-onderscheidingsvermogen van de getuige, gezien de duisternis op het ogenblik van de feiten en rekening houdend met de plaats van het gebeuren. Ze stellen tijdens die experimenten vast dat de getuige in 80% van de gevallen de juiste kleur ziet, maar zich in 20% van de gevallen vergist. Wat is de kans dat de taxi die in het auto-ongeluk betrokken was inderdaad groen was, gegeven het feit dat die getuige 'groen' zegt?

16. PTA is een proteïne die geproduceerd wordt door cellen in de prostaatklier. Door het opmeten van de PSA-waarde in het bloed kan men bij mannen het risico op prostaatkanker bepalen. In een medisch labo gebruikt men drie toestellen om de PSA-waarden te bepalen.

    (a) met toestel T1 is er een kans van 1% op een foute analyse en dit toestel wordt bij 60% van de analyses gebruikt;

    (b) met toestel T2 is er een kans van 2% op een foute analyse en dit toestel wordt bij 30% van de analyses gebruikt;

    (c) met toestel T3 is er een kans van 4% op een foute analyse en dit toestel wordt bij 10% van de analyses gebruikt.

    Als men vaststelt dat de PSA-analyse van een bepaald bloedstaal onjuist is, hoe groot is dan de kans dat men hierbij toestel T1 of T2 heeft gebruikt?

17. Er werd een onderzoek opgezet naar de werkzaamheid van een nieuwe test voor het opsporen van een parasiet op beukenbomen. Elke beuk reageert positief of negatief op de test. De bevindingen van het onderzoek zijn de volgende: 14% van de beuken test positief; 2% van de

beuken test positief, maar heeft de parasiet niet; 4% van de beuken test negatief, maar heeft toch de parasiet. Wat is de kans dat, als een beuk de parasiet heeft, de beuk ook positief zal testen? (A. 75%)

18. Voor een bepaald examen geldt dat 80% van de goede studenten en 25% van de zwakke studenten slaagt. De groep bestaat uit 70% goede studenten. Wat is de kans dat een student die slaagt in werkelijkheid een goede student is

19. In een bepaald land zijn er 10% zware rokers, 20% matige rokers en 70% rookt niet. De kans dat een willekeurig geselecteerde inwoner longkanker krijgt is 0.03. Verder is het een algemeen gegeven dat matige rokers 5 keer meer kans en zware rokers 10 keer meer kans op longkanker hebben dan niet rokers.

    (a) Bereken de kans dat een niet roker longkanker krijgt. (A. $\frac{1}{90}$)

    (b) Bereken de kans dat iemand die longkanker krijgt een zware roker is. (A. $\frac{10}{27}$)

20. De Niet Invasieve prenatale test (NIPT) is een test of een foetus oa al dan niet het syndroom van Down heeft. Bij sreening en diagnostiek zijn twee waarden van belang: specificiteit van een test zegt iets over de kans dan een zwangere van een gezonde foetus inderdaad een niet-afwijkend testresultaat krijgt. Een hoge specificiteit betekent dus: weinig fout-positieve uitslagen. Gemeten in een groot aantal studies onder vrouwen met een verhoogd risico op een ongeboren kind met het downsyndroom was de kans op een fout-positieve uitslag bij de NIPT slechts 0,09%. De gevoeligheid van een test geeft aan hoe groot de kans is dat een kind met een bepaalde aandoening ook echt wordt gevonden. Die 'sensitiviteit van de NIPT was, gemeten in dezelfde studies, 99,2% Trisomie 21 komt voor bij ongeveer 1 op 700 levend geboren baby's (bron: NIPT en de screening op aangeboren afwijkingen, Gezondheidsraad 2016). Wat is de kans op Trisomie 21, gegeven dat de NIPT positief is.

21. The prosecutor's fallacy (dwaling van de aanklager) in de zaal Sally Clark:
    Definieer volgende begrippen: H voor de hypothese: Sally Clark is schuldig aan dubbele kindermoord en E: het bewijsmateriaal, de kinderen zijn dood. De rechtbank moet de volgende kans beoordelen $P(\bar{H}|E)$, m.a.w. wat is de kans dat zij onschuldig is aan kindermoord, gegeven het feit dat haar twee kinderen gestorven zijn. Deze kans wordt echter verward met $P(E|\bar{H})$ (de kans dat haar twee kinderen dood zijn, gegeven dat zij onschuldig is, hier door 'specialisten' geschat op $\frac{1}{8513} \cdot \frac{1}{8513}$. Deze kans is al fout berekend, omdat men ten onrechte uitgaat van onafhankelijkheid. In een artikel in het British Medical Journal werd plausibel gemaakt dat een factor 5 van toepassing is op de verhoogde kans, m.a.w. $P(E|\bar{H}) = \frac{1}{8513} \cdot \frac{5}{8513}$. Wil je nu de juiste kans berekenen dat moet je ook nog P(H) weten, wat is de kans dat iemand één of meer van haar eigen kinderen vermoord. Op basis van Amerikaanse gegevens bekwam men (als overschatting): $P(H) = \frac{1}{2.4X10^6}$

22. "Why does the NSA engage in mass surveillance of Americans when it's statistically impossible for such spying to dectect terrorists?"-casus

23. Netflix-casus

5 Discrete kansverdeling

5.1 Begripsvorming

5.1.1 Verwachtingswaarde

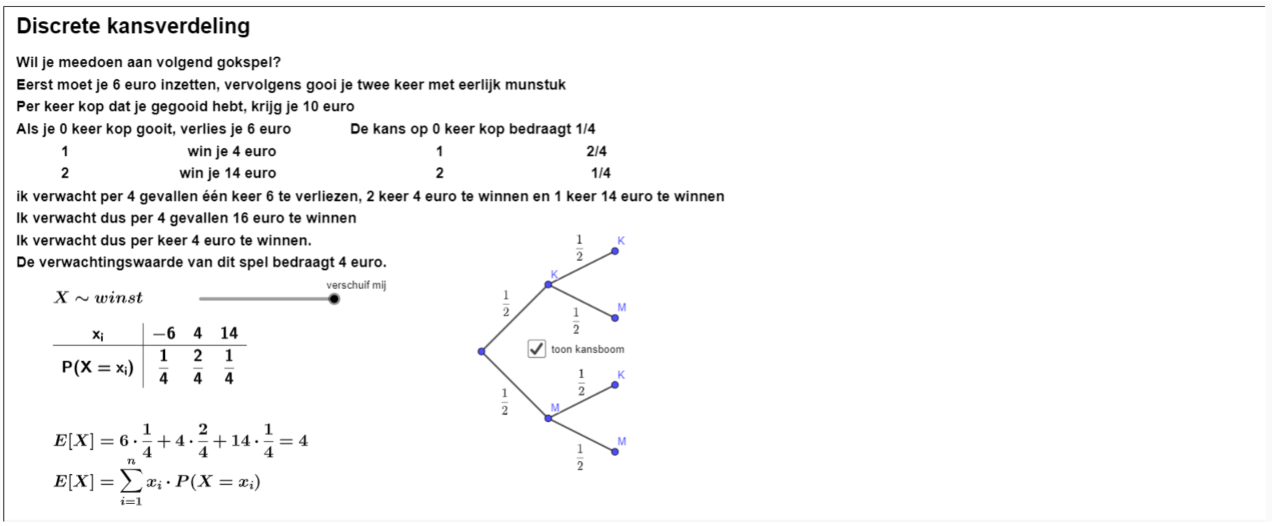


Figure 16: https://www.geogebra.org/m/wfavdjcz

5.1.2 Variantie

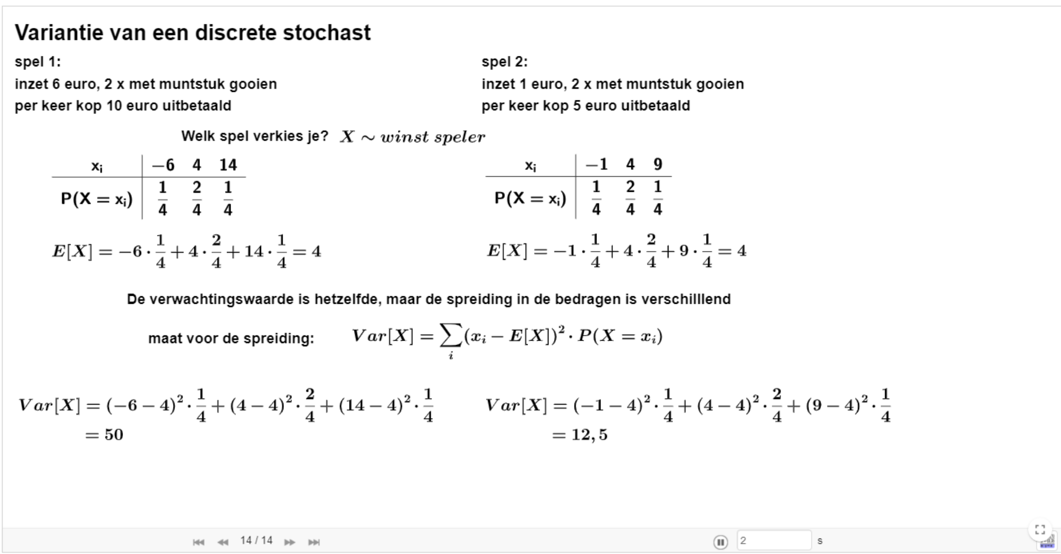


Figure 17: https://www.geogebra.org/m/wfavdjcz

5.1.3 Rekenen met E[X] en Var[X]

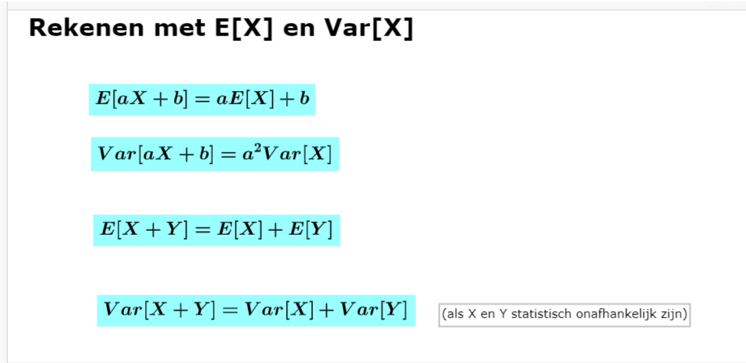


Figure 18: https://www.geogebra.org/m/wfavdjcz

5.2 oefeningen

1. Jouw vriend(in) vraagt je mee te doen aan volgend gokspel: er wordt er met één dobbelsteen gegooid. Ligt een even getal boven dan moet jij dit even getal aan hem/haar betalen. Ligt een oneven getal boven dan betaalt hij/zij dit bedrag aan jou uit. Ga je meedoen?

2. Jouw vriend(in) vraagt je mee te doen aan een gokspel met drie dobbelstenen. De inzet bedraagt 5 €. Als je 1 keer een 6 gooit, krijg je jouw inzet terug; bij twee keer een zes gooien, krijg je twee keer je inzet terug; en bij drie keer een zes gooien, krijg je driemaal jouw inzet terugbetaalt. In de andere gevallen verlies je jouw inzet Ga je meedoen?

3. Ik wil een toets met meerkeuzevragen maken. Per vraag zijn er 4 mogelijke antwoorden. Welke giscorrectie moet ik toepassen om het gokken te ontmoedigen?

4. Bereken de verwachtingswaarde van een inzet naar keuze bij de Franse roulette. Vergelijk jouw resultaat met die van je buurman (die een andere inzet heeft gekozen) (https://www.vustat.eu/apps/roulette/index.html

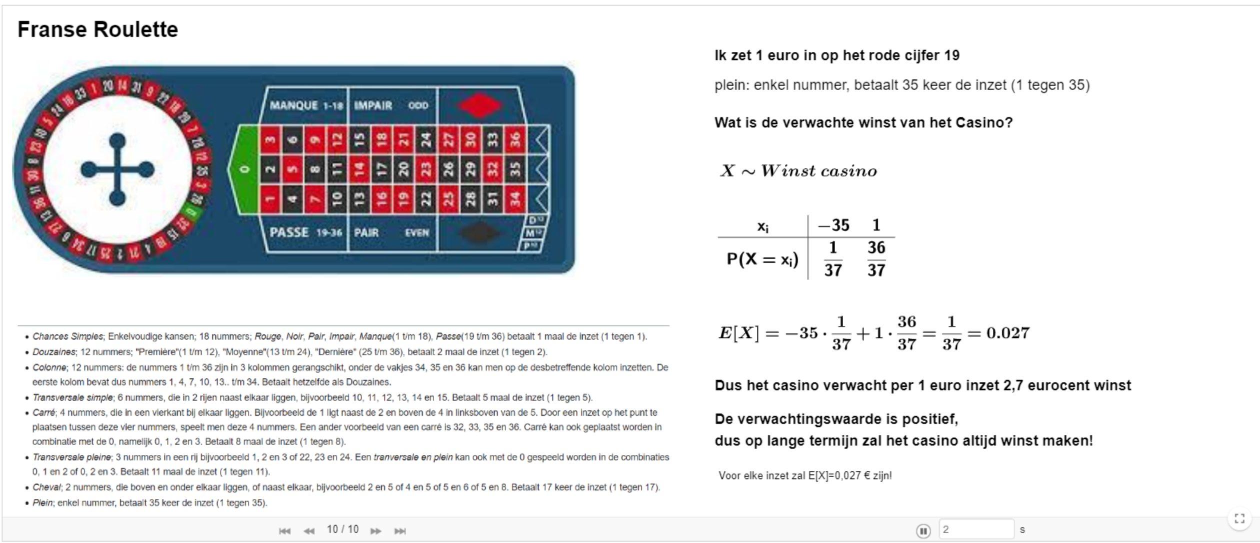


Figure 19: https://www.geogebra.org/m/wfavdjcz

5. "Unders and Overs" was vroeger in Amerika een populair spel dat op opendeur dagen van scholen gespeeld werd om de schoolkas te versterken. Het spel werd gespeeld met twee dobbelstenen en een spelbord met de drie secties "Under 7", "7" en "Over 7". Twee dobbelstenen worde geworpen en je kunt fiches plaatsen op één of meer van de secties. Voor elke geplaatste

fiche betaal je 1 dollar. Voor elke fiche op de sectie "Under 7" is de uitbetaling 2 dollars als onder de 7 geworpen wordt, dezelfde uitbetaling voor elke fiche op "Over 7" als boven 7 geworpen wordt, en een uitbetaling van 5 dollars voor elke fiche op "7" als precies 7 geworpen wordt. Je strategie is om 1 fiche in te zetten op "Under 7" en 1 fiche op "7". Wat zijn de verwachtingswaarde en de standaardafwijking van het netto bedrag dat de school overhoudt aan jouw weddenschap? (A. $\frac{1}{3}$ dollar en $\sqrt{\frac{103}{18}}$ dollar)

6. Er worden 6000 loterijlotjes verkocht aan €5 elk. Er wordt een auto van €10000 verloot. Wat is de verwachte winst van een persoon die een lot koopt?

7. Een vaas bevat witte en zwarte ballen, die niet van elkaar te onderscheiden zijn. Een spel bestaat er uit om zonder te kijken en zonder teruglegging, twee ballen uit deze urne te nemen. Als de twee genomen ballen wit zijn, dan verliest de speler 9 euro. Als de twee genomen ballen zwart zijn, dan verliest de speler 1 euro. Bij verschillende kleur zal de speler 5 euro winnen.
   (a) In een eerste spel zitten er 2 zwarte en 3 witte ballen in de urne.
      i. Bepaal de kansboom voor dit spel
      ii. Bereken de verwachtingswaarde van dit spel. Welke conclusie kan je trekken?
   (b) In een tweede spel zitten er 3 witte ballen en een onbekend aantal zwarte ballen (wel minstens twee).
      i. Bepaal de kansboom voor dit spel
      ii. Bereken de verwachtingswaarde van dit spel (in functie van het onbekend aantal zwarte ballen ($E[X] = \frac{-x^2+30x-81}{(x+3)^2}$)
      iii. Bepaal het aantal zwarte ballen zodat het spel in het voordeel van de speler is. ($x \in [4;26]$)
      iv. Bepaal het aantal zwarte ballen om een maximale gemiddelde winst te hebben.($x = 7$)

8. Jouw vriend(in) vraagt jou mee te doen aan volgend spelletje: op voorhand moet je één euro betalen; er liggen drie kaarten van dezelfde soort op tafel. Nu worden drie dezelfde kaarten van een andere soort volledig willekeurig op deze kaarten gelegd, per kaart 'die juist ligt' krijg je één euro. Bereken de verwachtingswaarde van dit spel om uit te maken of je mee zal doen. En wat zou je doen als er gespeeld wordt met vier kaarten?

9. Oefening op technisch berekenen:

Berekening $\mu = \mathrm{E[X]}$ en $\sigma = \sqrt{\mathrm{Var[X]}}$ van een stochast

x_i	8	16	25
$P(X = x_i)$	0.173	0.253	0.574

$$E[X] = \sum_{i=1}^{3} x_i \cdot P(X = x_i) = (8)(0.173) + (16)(0.253) + (25)(0.574) = 19.777$$

$$\begin{aligned}\sqrt{\mathrm{Var}[X]} &= \sqrt{\sum_{i=1}^{3} (x_i - E(X))^2 \cdot P(X = x_i)} \\ &= \sqrt{(8-19.777)^2(0.173) + (16-19.777)^2(0.253) + (25-19.777)^2(0.574)} \\ &= 6.582\end{aligned}$$

$$\begin{aligned}Var[X] &= E[X^2] - (E[X])^2 \\ &= \sum_{i=1} x_i^2 \cdot P(X = x_i) - (E[X])^2 \\ &= (64)(0.173) + (256)(0.253) + (625)(0.574) - (19.777)^2 \\ &= 434.459 - 391.14 = 43.319 \\ \Rightarrow \sqrt{Var[X]} &= \sqrt{43.319} = 6.582\end{aligned}$$

Figure 20: https://www.geogebra.org/m/wfavdjcz

5.3 Binomiale verdeling

5.3.1 Begripsvorming

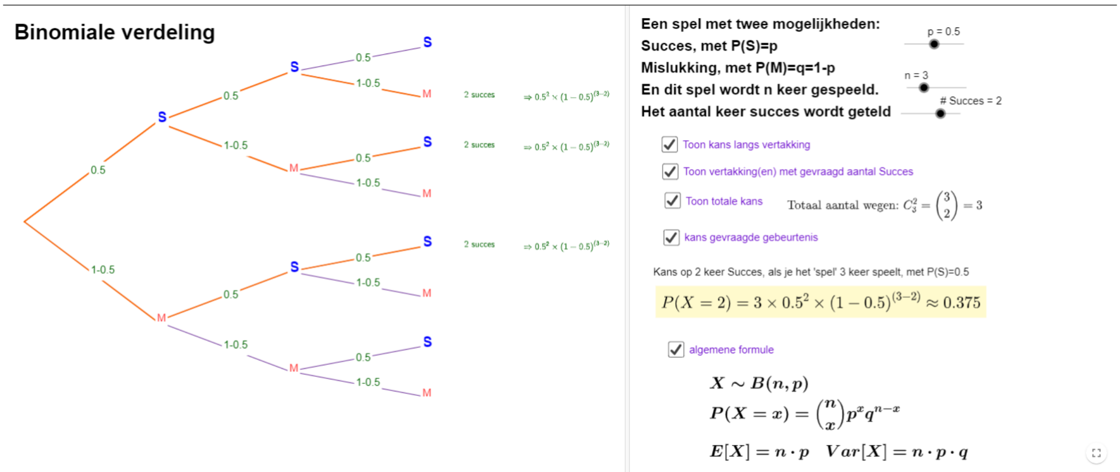


Figure 21: https://www.geogebra.org/m/Taz3435b

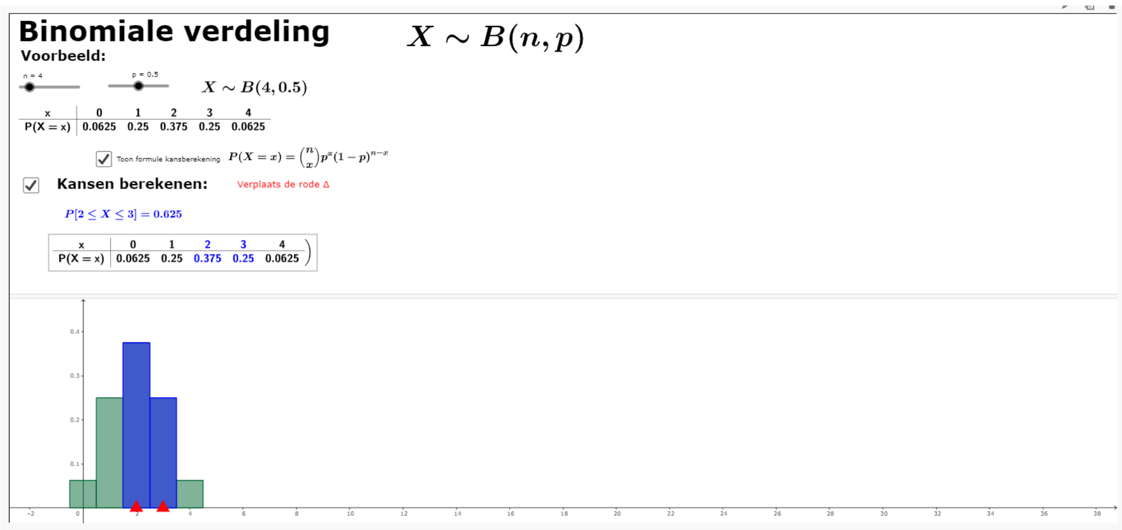


Figure 22: https://www.geogebra.org/m/Taz3435b

5.3.2 Oefeningen

1. In drie van de vier gevallen slaagt een basketbalspeler in een 'driepunter'. Veronderstel dat hij in een wedstrijd 10 maal voor een driepunter kan gaan. Wat is de kans dat hij minstens negen maal scoort. En wat is de kans dat hij bij zijn derde poging voor de eerste keer scoort?

2. Thomas is graag gezien bij de meisjes: gemiddeld driekwart die hij ten dans vraagt op een fuif zeggen ja. Op de laatste fuif is hij niet in vorm. Hij vraagt 20 meisjes ten dans maar niet meer dan 10 willen meedansen. Wat is de kans op deze prestatie?

3. Ongeveer 10% van alle mensen wereldwijd zijn linkshandig. Beschouw een groep van 15 mensen. Bereken de kans dat

   (a) niemand linkshandig is

   (b) 7 personen linkshandig zijn

   (c) hoogstens 3 linkshandig zijn

(d) minstens 2 linkshandig zijn

(e) hoeveel linkshandige kan je verwachten?

4. De vliegtuigmaatschappij Bernoulli-lines vliegt dagelijks tussen Brussel en Basel. Het vliegtuig heeft een capaciteit van 150 plaatsen. De kans dat een passagier niet komt opdagen is 10%. Daarom verkoopt men 160 tickets per vlucht. Wat is de kans dat iedereen mee kan?

5. Om wat zakgeld bij te verdienen, solliciteert Pieter op 6 gelijke, maar onafhankelijke jobs. Pieter heeft 40% kans om een job te krijgen en verdient €200 bij elke job die hij krijgt. De kosten die hij maakt bij het solliciteren voor de 6 jobs samen, raamt hij op €300.

   (a) Wat is het verwachte aantal jobs dat hij zal krijgen?

   (b) Wat is de verwachte winst?

   (c) Wat is de kans dat hij winst maakt?

   (d) Wat is de kans dat hij eraan verliest?

   (A. 1. 2.4 2. 180 3. 0.767 4. 0.233)

6. Een favoriet snoepje van Frank zijn muisjes. En van alle muisjes vindt hij de rode het lekkerste. de muisjes worden verkocht in pakjes van 50 stuks. In een pak zitten muisjes van 6 verschillende kleuren, waarvan we aannemen dat ze allemaal evenveel geproduceerd worden. Maar in de fabriek worden ze eerst gemengd voor ze verpakt worden. Hoeveel muisjes er van elke kleur in een pak terecht komen, is dus onderheving aan toeval.

   (a) Bereken de kans dat er minstens 10 rode muisjes in een willekeurig pak zitten

   (b) Bereken de kans dat in een pak minder dan 5 rode neusjes zitten.

   (c) Bereken de kans dat een pak precies 8 rode neusjes bevat

   (d) Bereken de kans dat als Frank 8 pakjes muisjes koopt, er minstens 6 pakjes zijn met minstens 10 rode muisjes.

   (e) Bereken de kans dat als Frank 8 pakjes muisjes koopt, er maximaal 4 zijn met minder dan 5 rode muisjes.

7. Het genotype van een bepaald kenmerk bestaat uit 2 allelen en kan AA, Aa en aa zijn. Als we weten dat de prevalentie van het A-allel werd geschat op 29%, wat is dan de kans dat er in een steekproef van 10 personen minstens 3 personen voorkomen die minstens 1 A-allel hebben. Stel hiervoor eerst de kansverdeling voor de verschillende genotypes op.(A. $X \sim B(10; 0.4959); P(X \geq 3) = 0,9424$)

5.4 Poissonverdeling

5.4.1 begripsvorming

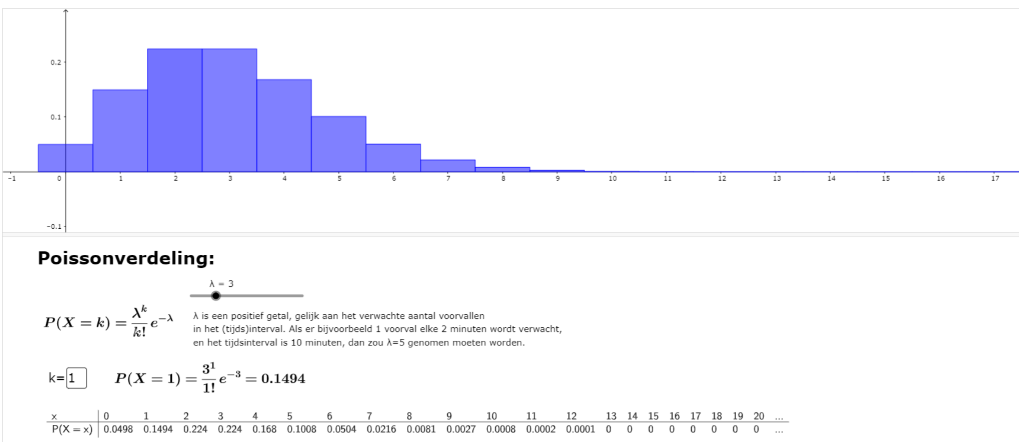


Figure 23: https://www.geogebra.org/m/dvm9rheq

5.4.2 oefeningen

1. In een textielfabriek worden rollen stof vervaardigd met een lengte van 50 meter per rol. Het aantal weeffouten per rol is Poisson-verdeeld met een bijbehorende verwachtingswaarde van 1 weeffout per rol. Bij de kwalitatieve keuring van de rollen stof worden deze gescheiden in rollen van A-kwalitiet (met 0 of 1 weeffout per rol) en rollen van B-kwaliteit (met twee of meer weeffouten per rol)

   (a) Bereken de kans dat een willekeurige rol de aanduiding B-kwaliteit krijgt.

   (b) De productieomvang per dag is gelijk aan 2000 meter stof. Hoe groot is de kans dat er op een willekeurige dag tenminste 30 rollen met A-kwaliteit worden gemaakt?

2. Bij een callcenter komen gemiddeld 10 telefoontjes per uur binnen.

   (a) Bereken de kans dat er in een uur meer dan 15 telefoontjes binnen komen

   (b) Bereken de kans dat er in een uur minder dan 5 telefoontjes binnen komen
   Het komt wel eens voor dat er gedurende een kwartier geen telefoontjes binnen komen

   (c) Bereken de kans dat er een kwartier lang geen telefoontjes binnen komen.

3. De bezoekers van een vogelreservaat kunnen aan de ingang van het park een verrekijker huren. Een bezoek duurt anderhalf uur. Er zijn gemiddeld 7,5 bezoekers per uur die een verrekijker huren. Dit aantal is Poisson verdeeld. het park beschikt echter slechts over 5 verrekijkers.

   (a) Het vogelreservaat opent om 10 uur 's ochtends. Wat is de kans dat een bezoeker die om 10h40 toekomt geen verrekijker meer kan huren. (A. 0.3840, $\lambda = 5$)

   (b) Over hoeveel verrekijkers moet het park minstens beschikken om ten minste 90% zekerheid te hebben dat alle bezoekers die voor 10h20 toekomen nog een verrekijker kunnen huren? (A. minstens 5, $\lambda = 2.5$)

4. Het verjaardagenprobleem

6 Continue Kansverdelingen

6.1 willekeurige verdeling

6.1.1 Begripsvorming

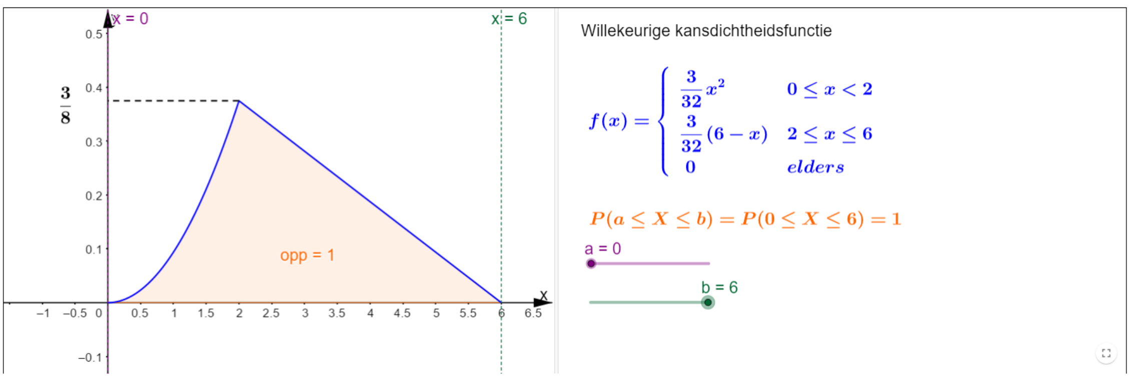


Figure 24: https://www.geogebra.org/m/dmsaqrpq

6.1.2 Oefeningen

1. Gegeven de toevalsvariabele x met kansdichtheidsfunctie $f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & elders \end{cases}$.

   Bepaal de waarde van de constante c

2. Gegeven de toevalsvariabele x met kansdichtheidsfunctie $f(x) = \begin{cases} kx^2 & -1 \le x \le 1 \\ 0 & elders \end{cases}$.

   (a) Bepaal de waarde van k
   (b) Bereken $P\left(-\frac{1}{2} \le x \le \frac{1}{2}\right)$
   (c) Bereken $E[X]$

3. Een benzinestation wordt eenmaal per week bevoorraad. De wekelijkse vraag (met als eenheid 5000 liter) is verdeeld als $f(x) = \begin{cases} \frac{5}{4}(1 - x^4) & 0 < x < 1 \\ 0 & elders \end{cases}$

   (a) Maak een schets van deze kansdichtheidsfunctie
   (b) Wat moet de capaciteit van de tank zijn dat de kans dat deze voor het einde van de week leeg zou zijn, minder dan 1% is?

6.2 Uniforme verdeling

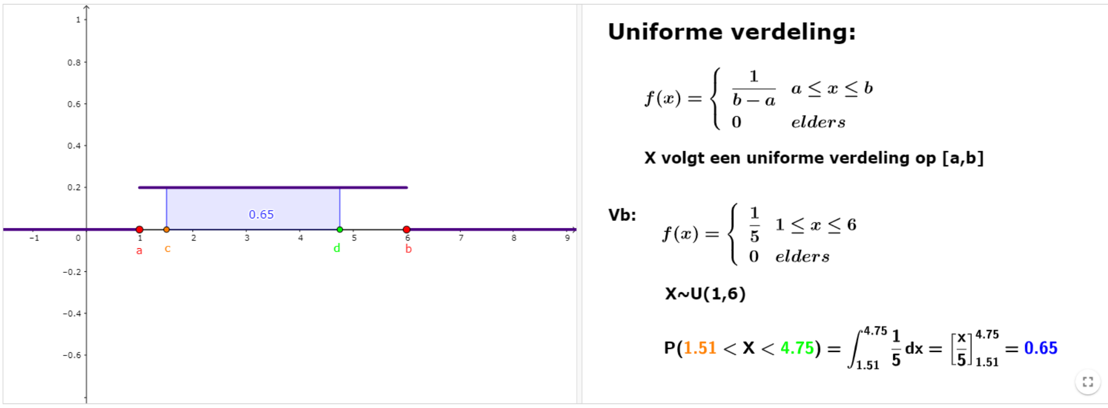


Figure 25: https://www.geogebra.org/m/qjmadwxu

6.3 Normale verdeling

6.3.1 Begripsvorming

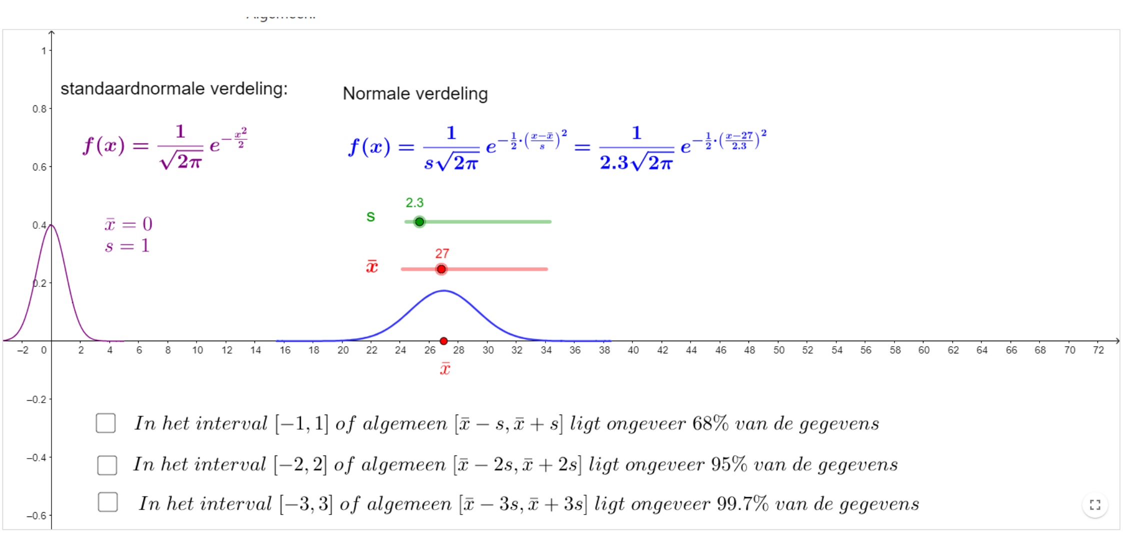


Figure 26: https://www.geogebra.org/m/muzjMBZq

6.3.2 De standaardnormale verdeling

de z-transformatie

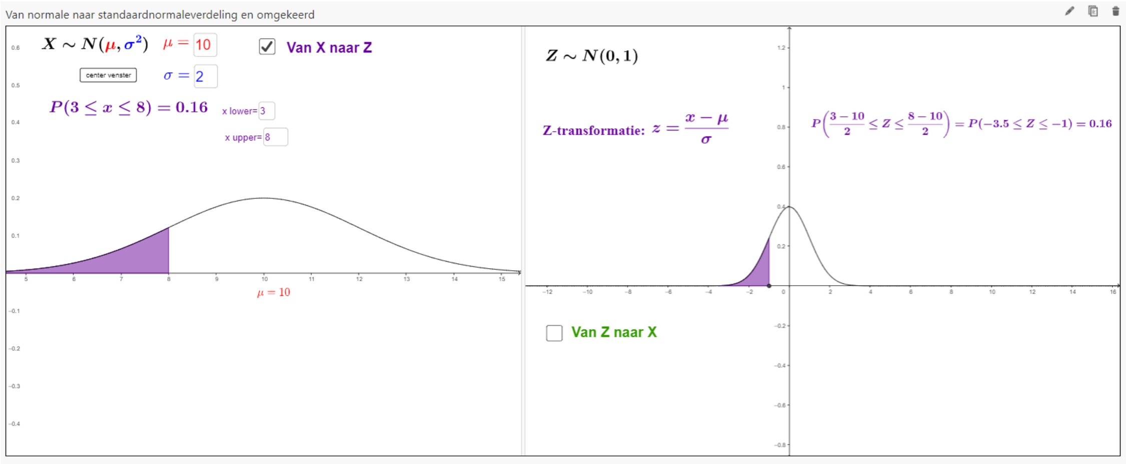


Figure 27: https://www.geogebra.org/m/muzjMBZq

z-waarden

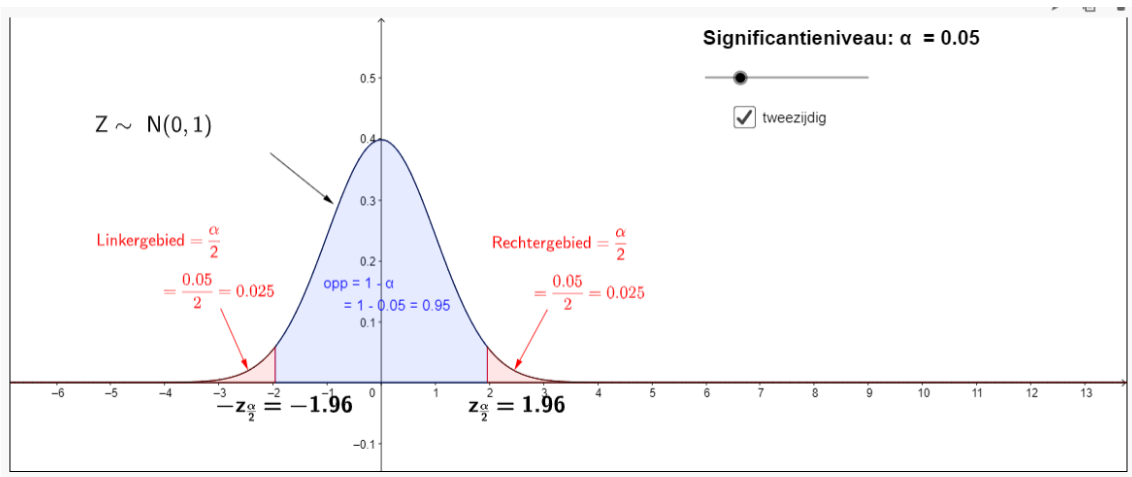


Figure 28: https://www.geogebra.org/m/gWCxY2Kt

6.3.3 centrale limietstelling

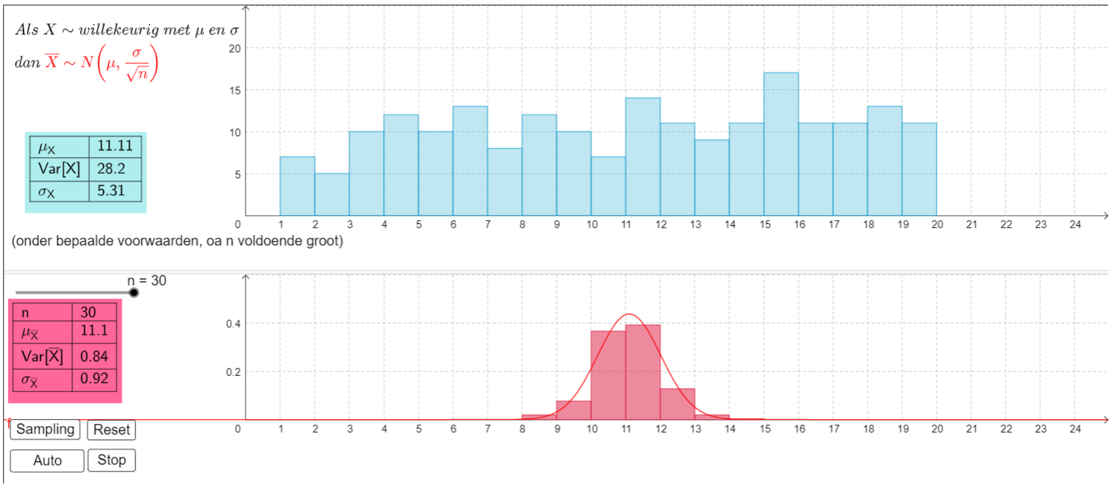


Figure 29: https://www.geogebra.org/m/gWCxY2Kt

6.3.4 Normale verdeling als benadering van de binomiale verdeling

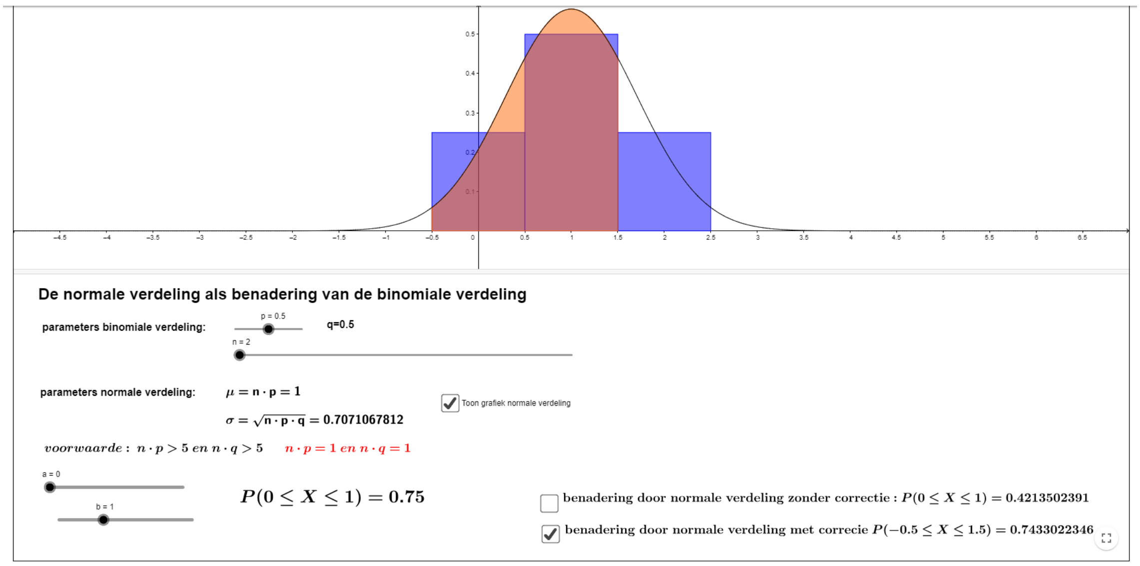


Figure 30: https://www.geogebra.org/m/Taz3435b

6.3.5 Oefeningen

1. Los op m.b.v. tabellen

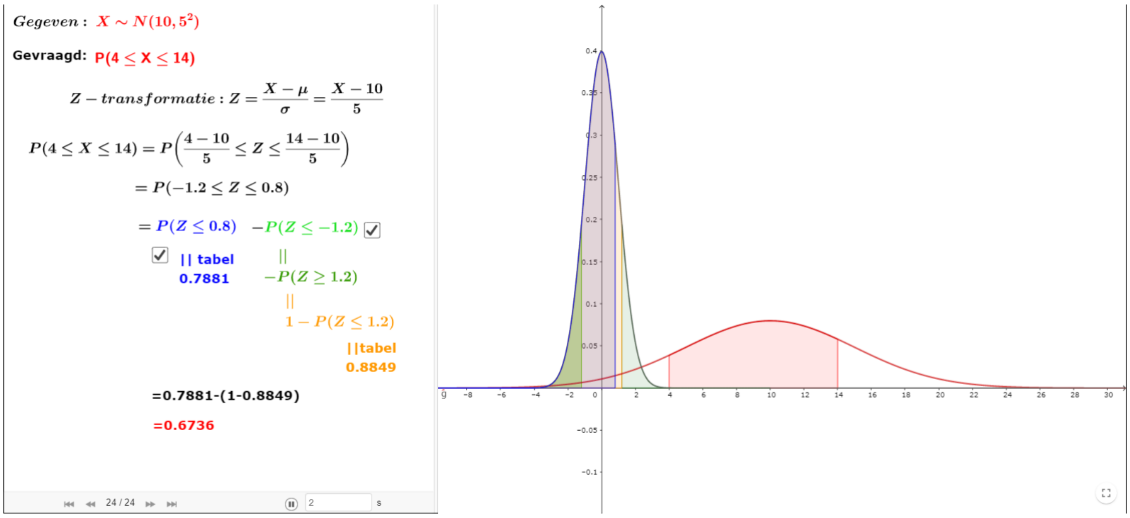


Figure 31: https://www.geogebra.org/m/muzjMBZq

2. Bereken de volgende kansen als $Z \sim N(0,1)$
   (a) $P(0 \leq Z \leq 1,2)$
   (b) $P(Z < -0.6)$
   (c) Bepaal z zodat $P(Z \leq z) = 0.8621$

3. Een makelaar stelt vast dat de koers/winst verhouding van aandelen op de New York Stock Exchange kan beschreven worden met een stochastische variabele die normaal verdeeld is met een gemiddelde van 10.2 en een standaardafwijking van 3.3.
   (a) Bereken de kans dat een willekeurig gekozen aandeel een koers/winst verhouding heeft tussen 2.5 en 8.7.
   (b) Als de makelaar de 5 procent beste aandelen wil kennen door de koers/winst verhouding als criterium te beschouwen, wat moet hij dan als minimale koers/winst verhouding nemen ?

4. Onderstel dat de lichaamslengte van een 20-jarige normaal verdeeld is met een gemiddelde $\mu = 170cm$ en een variantie van $\sigma^2 = 100cm^2$. Wat is bijgevolg de kans dat de lichaamslengte kleiner is dan 150 cm (A. 0.0228)

5. Onderstel dat de IQ-scores op een middelbare school normaal verdeeld zijn met gemiddelde 105 en variantie 81
   (a) Wat is het percentage studenten met een IQ score tussen 90 en 115?
   (b) Wat is het percentage studenten met een IQ score dat hoger ligt dan 85?
   (c) Hoe hoog moet de IQ score van een leerling minimaal zijn zodat deze zou behoren tot de 5% best scorende leerlingen?

6. De dosis van een bepaald prduct nodig voor een algemene anesthesie is normaal verdeeld met gemiddelde 50 mg en standaardafwijking 10 mg. De letale dosis (dit is de dosis die de dood kan veroorzaken) van dat product is eveneens normaal verdeeld met gemiddelde 110 mg en standaardafwijking 20 mg. Veronderstel dat men -zonder verdere controle- bij een anesthesie een dosis zou gebruiken die in 90% van de gevallen voldoende is om de patiënt te verdoven. Hoeveel sterfgevallen zou men dan hebben? (A. $\approx 0.92\%$)

7. De gemiddelde levensduur van een polshorloge is 25 maanden met een standaardafwijking van 5 maanden. Je mag veronderstellen dat de levensduur van een polshorloge normaal verdeeld is. Voor hoeveel maanden moet de garantie gelden als de fabrikant van de polshorloge maximum 10% van de horloges wil omruilen? (A. 18.6 maanden)

8. De score van een examen in eerste zittijd is normaal verdeeld met gemiddelde μ_1 en standaardafwijking σ_1. De score van het examen in tweede zittijd is ook normaal verdeeld met gemiddelde $\mu_2$en standaardafwijking σ_2. Stel dat $\mu_2 = \mu_1 + \sigma_1$, en $\sigma_2 = 2\sigma_1$, en beschouw de score $x = \mu_2 + \sigma_2$. Welke van de volgende vier uitspraken is waar?

   (a) De kans om in de tweede zittijd de score x te behalen is ongeveer 10 keer kleiner dan de kans om in de eerste zittijd minstens de score x te behalen.

   (b) De kans om in de tweede zittijd minstens de score x te behalen is ongeveer gelijk aan de kans om in de eerste zittijd minstens de score x te behalen.

   (c) De kans om in de tweede zittijd minstens de score x te behalen is ongeveer 10 keer groter dan de kans om in de eerste zittijd minstens de score x te behalen.

   (d) De kans om in de tweede zittijd minstens de score x te behalen in ongeveer 100 keer groter dan de kans om in de eerste zittijd minstens de score x te behalen.

9. Bij bevestiging van de autogordel in de wagen moet de fabrikant ervoor zorgen dat het vastzetdraaimoment minimaal 30Nm is. Het blijkt dat draaimomenten normaal verdeeld zijn met gemiddelde $\mu = 32Nm$ en standaardafwijking $\sigma = 1.6Nm$ Welk percentage autogordels zal dan niet voldoen aan de vooropgestelde kwaliteitseisen?. Om dit percentage te beperken tot maximaal 2% moeten de ingenieurs aan de slag. De machines moeten zo verfijnd worden dat μ hetzelfde blijft maar σ verkleind. Welke waarde voor σ moet gehaald worden?

10. Een vulmachine van 1-liter flessen is zo afgesteld dat dit proces een normale verdeling volgt met een gemiddelde van 1002 ml en een standaardafwijking van 1.5 ml.

    (a) Hoeveel procent van de gevulde flessen bevat minder dan 1 l?

    (b) Het percentage uit de vorige vraag is eigenlijk veel te groot. Met welk gemiddelde (en zelfde standaardafwijking) moeten de flessen gevuld worden zodat maar 1% van de flessen minder dan 1 liter bevat

11. Een frisdrankautomaat gebruikt bekers die maximaal 200 ml kunnen bevatten. De hoeveelheid frisdrank die de automaat per keer produceert is normaal verdeeld met een standaardafwijking van 6 ml. Bepaal het gemiddelde waarop men de automaat moet instellen opdat de kans dat een beker overloopt hoogstens 2% is.

12. Veronderstel dat de populatie van de gewichten van de festivalgangers van Rock Werchter een gemiddelde van 68 kg met een standaarddeviatie van 11.5 kg heeft. Wat kan je dan zeggen over het gemiddelde en de standaardafwijking van de steekproevenverdeling van het gemiddeld gewicht van 150 bezoekers?

13. Het gewicht van kippen is normaal verdeeld met een gemiddelde μ en een standaardafwijking σ. Bepaal deze parameters als je weet dat het percentage kippen dat lichter is dan
    1.83 kg gelijk is aan het percentage kippen zwaarder dan 3.43 kg. Ook is gegeven dat 89.04% een gewicht heeft tussen deze grenzen.
    Als 9 kippen willekeurig geselecteerd worden, wat is dan de kans dat hun gemiddeld gewicht ligt tussen de 2.5 kg en de 3.1 kg.? (A. $\mu = 2.63$, $\sigma = 0.5$, 0.7799)

14. Van abrikozen is bekend dat 6% van de abrikozen wegens een vervelende schimmel niet te eten zijn. Helaas is dit aan de buitenkant niet waar te nemen. Van een partij van 100 abrikozen wordt onderzocht hoeveel abrikozen aangetast zijn.

    (a) Bereken met de binomiale verdeling de kans dat er meer dan 3 abrikozen aangetast zijn.

    (b) Benader deze kans met de normale verdeling

7 Betrouwbaarheidsintervallen

7.1 Inleidend voorbeeld

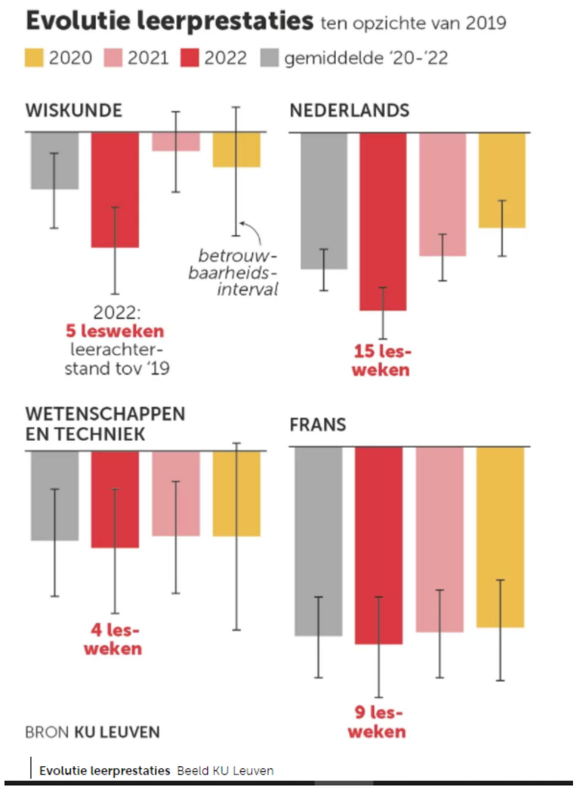


Figure 32: https://www.demorgen.be/nieuws/en-nu-donderen-zelfs-de-topleerlingen-achteruit-de-vier-b

7.2 Betrouwbaarheidsinterval voor een proportie

7.2.1 formule

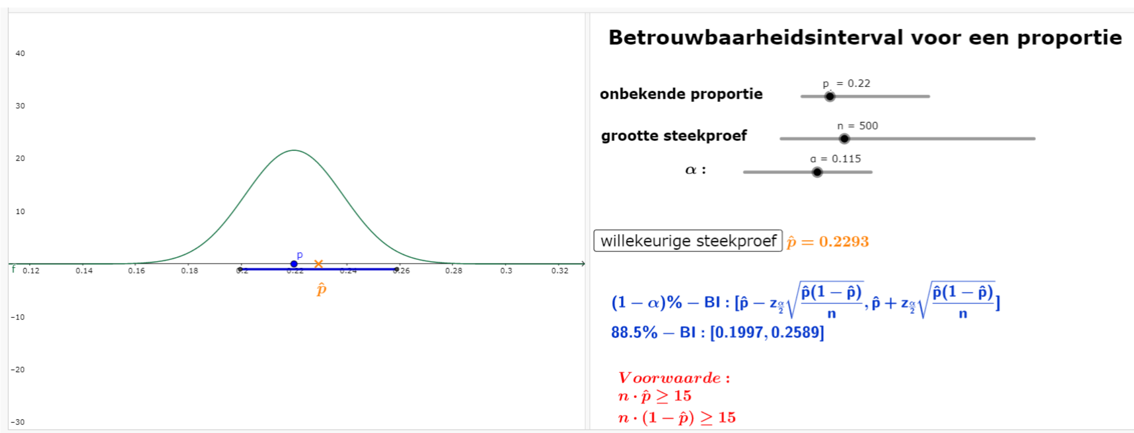


Figure 33: https://www.geogebra.org/m/gWCxY2Kt

7.2.2 oefeningen

1. Uit een onderzoeksrapport. Herbereken de (maximale) foutenmarge.

De resultaten van dit onderzoek zijn gebaseerd op een online survey met 1.020 respondenten van het Brussels Hoofdstedelijk Gewest (zone Nielsen III) ouder dan 18 jaar, gebruik makend van een gestructureerde vragenlijst van gemiddeld 4 minuten. De foutenmarge bedraagt +/- 3.07%.

2. Neem een actuele opiniepeiling en bepaal en/of herbereken het 95%- BI.
3. Een kleine politieke partij wil na een reclamecampagne de belangstelling voor deze partij onderzoeken. Zij wenst een tweezijdig 95%-betrouwbaarheidsinterval met een breedte van hoogstens 0.1 voor de fractie personen die op deze partij zouden stemmen. Op voorhand is bekend dat de fractie zeker niet groter is dan 0.15. Bepaal het minimale aantal personen dat ondervraagd dient te worden.
4. Volgens een onderzoek aangevraagd door het nationaal expertisecentrum tabaksontmoediging in Nederland (2015) rookte 211 vrouwen van de 1740 vrouwelijke deelnemers aan een enquête rond roken tijdens de zwangerschap van hun laatste kind. Construeer een $90\% - BI$ voor het werkelijk percentage rokende Nederlandse vrouwen tijdens de zwangerschap en leg hiervan de betekenis uit.
5. Uit een enquête van begin 2021 blijkt dat 27% van de Vlaamse scholieren liever online les krijgt dan naar school te gaan. Het 95%-betrouwbaarheidsinterval is $[0.242; 0.298]$. Bepaal de grootte van de steekproef.

7.3 BI voor populatiegemiddelde

7.3.1 Begripsvorming

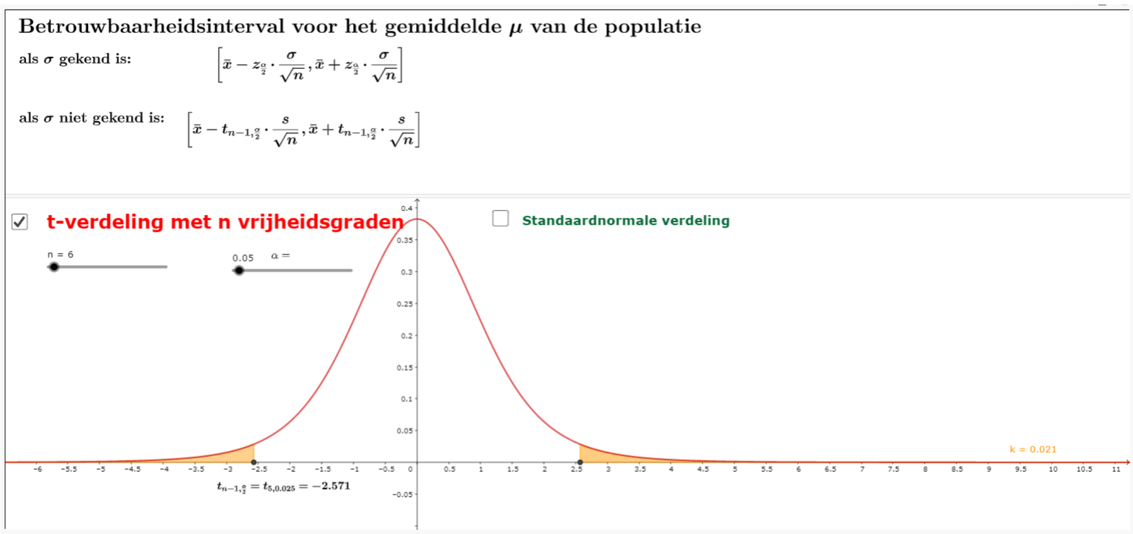


Figure 34: https://www.geogebra.org/m/gWCxY2Kt

7.3.2 Oefeningen

1. In een advertentie-campagne wordt gesteld dat spouwmuurisolatie gemiddeld 30% besparing op de verwarmingskosten oplevert. Een geïnteresseerde econoom die erover denkt zijn huis te laten isoleren, informeert bij een aantal kennissen die deze isolatie hebben aangebracht. Hij vindt de volgende besparingen (in % per jaar): 18 20 37 24 23 27 21 30
   (a) Beschouw de procentuele besparing als normaal verdeeld en geef een tweezijdig 95% betrouwbaarheidsinterval voor de gemiddelde besparing. Geef aan welke veronderstelling u moet maken.
   (b) Is het in de advertentie genoemde percentage redelijk (gebruik (a)) ?
   (A. a) (10.37,39.63) b)Ja)

2. Een consumentenorganisatie wil een betrouwbaarheidsinterval opstellen voor het gemiddeld aantal punaises in doosjes van een zeker merk die volgens het opschrift 100 punaises zouden moeten bevatten. In negen aselect gekozen doosjes vindt zij achtereenvolgens de volgende aantallen punaises:
90 94 88 92 90 86 94 90 86
Veronderstel dat het aantal punaises in een doosje normaal verdeeld is
   (a) Geef een 95% betrouwbaarheidsinterval voor μ (A. $[87.69, 92.31]$)
   (b) Indien men via inside information erachter is gekomen dat de door de vulmachine veroorzaakte standaardafwijking van het aantal punaises per doosje zes bedraagt, wat is dan het 95% BI voor μ (A. 86.08,93.92)
   (c) Is het aannemelijk dat er gemiddeld 100 punaises in een doosje zitten?
3. In een hospitaal wou men nagaan hoeveel dagen moeders gemiddeld na hun bevalling nog in het hospitaal blijven. Hiervoor werd een steekproef van grootte 30 beschouwd uit de populatie van alle moeders die in het afgelopen jaar bevallen zijn. De volgende resultaten werden gevonden:
3,3,4,3,2,5,3,1,4,3,4,2,3,5,3,2,4,3,2,4,1,6,3,4,3,3,5,2,3,2
Bepaal een 95% betrouwbaarheidsinterval voor het gemiddeld aantal dagen dat moeders na hun bevalling in het hospitaal verblijven. Leg kort uit wat het betrouwbaarheidsinterval betekent. (A. $\bar{x} = 3.1667$, $s^2 = 1,3848$, $t(29) \to t_{\frac{0.05}{2}} = 2.045$, $[2.7273; 3.6061]$
4. Uit een steekproef waarin aan 55 Vlamingen gevraagd werd hoeveel uur televisie ze per week kijken, berekent men een 95% betrouwbaarheidsinterval voor het populatiegemiddelde. Dit levert het interval $[17.66, 22.34]$ op.
   (a) Leg de betekenis uit van dit betrouwbaarheidsinterval
   (b) Waaraan is de schatting van dit populatiegemiddelde gelijk? (A. 20)
   (c) Hoe groot is de foutenmarge? (A. 2.34)
5. NASA wilde de verwachte levensduur van ratten in de ruimte bepalen. Daarom werden 24 ratten in een experiment opgenomen. De levensduur(gemeten in dagen) voor elk van de verschillende ratten werd genoteerd,en NASA berekende het gemiddelde $\bar{x} = 595$ en met een standaardafwijkinge $s = 84$ van deze steekproef. Wat zou NASA moeten opgeven als schatting van de gemiddelde levensduur, indien zij een 95%-betrouwbaarheidsinterval wil geven? (A. $[559.5; 630.5]$)
6. De gestyleerde letter e, die je op heel veel verpakkingen tegenkomt, wil zeggen dat de inhoud/hoeveelheid meer of minder dan de aangegeven hoeveelheid mag zijn, en dit volgens Europese afspraken. Als controle zegt Europa:

2.3. Controle van de gemiddelde werkelijke inhoud van de afzonderlijke eenheden van een partij voorverpakkingen

2.3.1. Een partij voorverpakkingen wordt voor deze controle als aanvaardbaar beschouwd indien het gemiddelde $\mathbf{x}^- = \frac{\sum \mathbf{x_i}}{\mathbf{n}}$ van de werkelijke inhouden x_i van n voorverpakkingen van een steekproef meer bedraagt dan:

$$\mathbf{Q_n} - \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} \cdot \mathbf{t_{(1-\alpha)}}$$

In deze formule is:

Q_n : de nominale hoeveelheid van de voorverpakking,

n : het aantal voorverpakkingen waaruit de steekproef voor deze controle bestaat,

s : de schatting van de standaardafwijking van de werkelijke inhouden van de partij,

$t_{(1-\alpha)}$: de stochastische variabele van de Studentverdeling, functie van het aantal vrijheidsgraden $v = n — 1$ van de betrouwbaarheidsgrens $(1 — \alpha) = 0{,}995$.

Voor een 1 liter fles melk, bekom ik volgende resultaten bij een steekproef bij een verpakking van 6 flessen: 1001, 998, 995, 996, 993, 1000
   (a) voldoet deze verpakking?
   (b) wat is het verband van deze formule met BI?

7.4 Schema

Confidence Interval for	Sample Statistic	Margin of Error	Use When
Population mean (μ)	$\overline{x}$	$\pm z^* \frac{\sigma}{\sqrt{n}}$	X is normal, or $n \geq 30$; σ known
Population mean (μ)	$\overline{x}$	$\pm t^*_{n-1} \frac{s}{\sqrt{n}}$	σ unknown
Population proportion (p)	$\hat{p}$	$\pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$n\hat{p} \geq 10$; $n(1-\hat{p}) \geq 10$
Difference of two population means ($\mu_1 - \mu_2$)	$\overline{x}_1 - \overline{x}_2$	$\pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	Both normal distributions or $n_1, n_2 > 30; \sigma_1, \sigma_2$ known
Difference of two population means ($\mu_1 - \mu_2$)	$\overline{x}_1 - \overline{x}_2$	$\pm t^*_{n_1+n_2-2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$	$n_1, n_2 > 30; \sigma_1, \sigma_2$ unknow
Difference of two proportions ($p_1 - p_2$)	$\hat{p}_1 - \hat{p}_2$	$\pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$n\hat{p} \geq 10$; $n(1-\hat{p}) \geq 10$ for each group

8 Toetsen van hypothesen voor 1 populatieparameter

8.1 begripsvorming

Bij een betrouwbaarheidsinterval ga je kijken welke waarden van de gezochte parameter mogelijk zijn bij jouw dataset. Bij het toetsen van hypothesen ga je na in welke mate jouw data samengaan met de (hypothetische) waarde van de parameter. Bij toetsen vertrekken we van de veronderstelling dat de nulhypothese juist is. Hierdoor kun je een kansverdeling voor de toetsingsgrootheid opstellen. Binnen dat model kun je berekeningen maken. Je kunt berekenen hoe extreem het resultaat van de steekproef is. Hiervoor bereken je met de gegevens van de steekproef de overschrijdingskans van de uitkomst. Die kans wordt ook de p-waarde genoemd. Indien de p-waarde groter is dan een vooraf vastgelegd niveau α, wordt de nulhypothese verworpen.

8.2 hypothesetoetsing voor een proportie

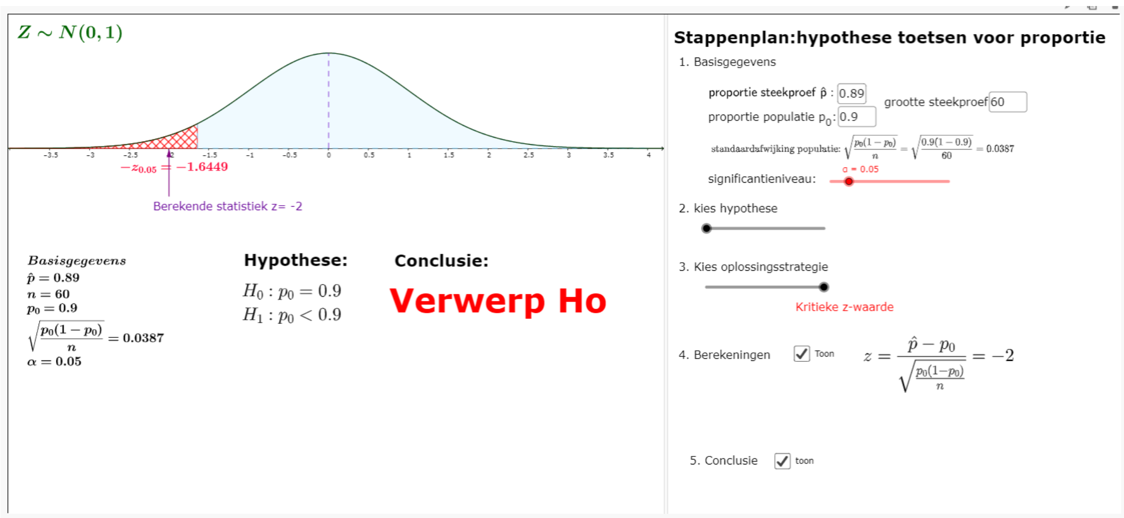


Figure 35: https://www.geogebra.org/m/rx7zragy

8.3 hypothesetoetsing voor een gemiddelde

8.3.1 als σ gekend is

Figure 36: https://www.geogebra.org/m/ds7rrxk5

8.3.2 als σ NIET gekend is

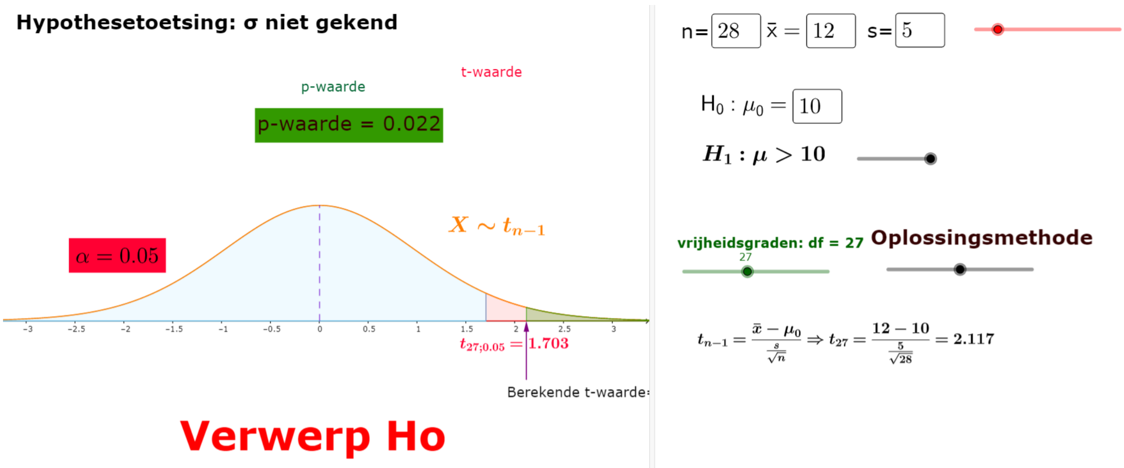
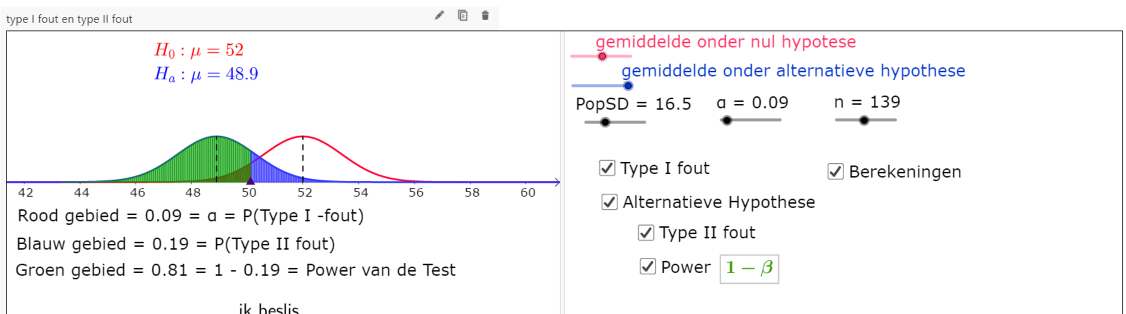


Figure 37: https://www.geogebra.org/m/ds7rrxk5

8.4 type I en type II fout



	H_0 te aanvaarden	H_0 te verwerpen
H_0 is juist	juiste beslissing	type I fout $= \alpha$
H_0 is fout	type II fout $= \beta$	juiste beslissing

Figure 38: https://www.geogebra.org/m/u7xvhpc4

8.5 Schema

Test for	Null Hypothesis (H_0)	Test Statistic	Distribution	Use When
Population mean (μ)	$\mu = \mu_0$	$\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	Z	Normal distribution or $n \geq 30$; σ known
Population mean (μ)	$\mu = \mu_0$	$\frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	t_{n-1}	$n < 30$ and/or σ unknown
Population proportion (p)	$p = p_0$	$\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	Z	$n\hat{p} \geq 10$; $n(1-\hat{p}) \geq 10$

8.6 Oefeningen

1. Een fabrikant heeft een medicijn ontwikkeld waarvan de fabrikant van mening is dat het medicijn in 90% van de gevallen de gewenste genezing brengt. Een onderzoeker betwijfelt dit en wil controleren of deze hoge werkzaamheid niet een beetje hoog is ingeschat en doet een onderzoek bij 50 patiënten.

   (a) Leg uit dat je hier $H_0 : p = 0.9$ tegen $H_1 : p < 0.9$ zal worden getoetst.

   (b) Wat zal de conclusie zijn (bij een significantie van 5%) als de onderzoeker constateert dat bij 46 patiënten het medicijn werkt?

   (c) Wat zal de conclusie zijn (bij een significantie van 5% als de onderzoeker constateert dat bij 41 patiënten het medicijn werkt?

2. Een machine snijdt viltstiften. De voorgestelde lengte bedraagt 10 cm. met $\sigma = 0,12$ cm. Bij een steekproef van 16 stuks blijkt $\bar{x} = 9.95$ cm. Moet de machine bijgesteld worden bij een significantieniveau van 5% (A. Neen)

3. De scores op een welbepaald examen worden verondersteld normaal verdeeld te zijn met een gemiddelde van 70 op 100 en een standaardafwijking van 18. Een willekeurige steekproef van 50 studenten genereert een gemiddelde score van 67. Kan men op basis hiervan besluiten dat het werkelijke gemiddelde lager ligt dan 70? (A. neen)

4. Bij de productie van houten stoelen mag de vochtigheidsgraad van het hout niet hoger zijn dan 10%. Bij een nieuwe lading hout wordt van een lukraak getrokken staal de vochtingheidsgraad gemeten.

   (a) Formuleer een nul- en alternatieve hypothese omtrent de bruikbaarheid van het hout

   (b) Interpreteer de type I en de type II fout. Welke van deze fouten ziet de leverancier (resp. de producent liever niet gemaakt worden?)

5. Een fabrikant van frisdranken houdt een actie om zijn marktaandeel te vergroten. Bij deze actie kan de consument doppen verzamelen. De onderkant van een dop heeft één van de kleuren rood, groen, blauw of oranje. Lever je vier verschillende gekleurde doppen in, dan krijg je een fles gratis.
   De fabrikant beweert dat 50% van de doppen rood is, 30% groen, 15%blauw en 5% oranje.

   (a) Bereken de kans dat iemand na het kopen van vier flessen frisdrank een gratis fles krijgt (A. 0.027)

   (b) Bereken de kans dat iemand na 8 flessen twee gratis flessen krijgt (A. 0.0033)

   (c) Bereken de kans dat iemand die 30 flessen heeft gekocht minstens 1 oranje dop heeft. (A. 0.7854)

   Volgens een klant is het percentage flessen met een rode dop groter dan 50%. Van de 100 flessen die hij heeft opengemaakt hadden er 57 een rode dop.

   (d) Moet je hem bij een significatieniveau van 10% gelijk geven? (A. 0.0808<0.1, dus H_0 verwerpen)

Een groothandel in frisdranken krijgt klachten binnen over het percentage oranje doppen; dat zou minder zijn dan 5%. Er wordt besloten een steekproef te nemen van 200 flessen.

(e) Hoeveel oranje doppen moet men in deze steekproef aantreffen opdat de bewering van de fabrikant kan worden verworpen? Neen een significantieniveau van 10% (A. minder dan 6 doppen)

6. In het algemeen wordt er van uit gegaan dat het IQ van de 'gemiddelde Belg' normaal verdeeld is met een gemiddelde van 100 en een standaardafwijking van 15. Zijn mijn wiskundeleerlingen slimmer? Ik neem een willekeurige steekproef van 25 wiskunde studenten en onderwerp ze aan een IQ-test. Het gemiddelde IQ uit deze steekproef noemen we $\bar{x}$.

   (a) Bij welke grenswaarde ($\alpha = 0,05$) kan ik H_0 verwerpen? (A. 104.9)

   Uit mijn onderzoek blijkt het gemiddelde IQ in de steekproef gelijk te zijn aan 107.

   (b) Bereken de kans dat ik ten onrechte H_0 verwerp. (A. 0.0098)

7. Er bestaat ook een EQ-test (emotionele intelligentie. Het gemiddelde van alle Vlamingen zou 90 zijn. Ik heb deze test hier op school door 9 lln laten afleggen met als resultaat $\bar{x} = 87.2$ en $s = 5$. Kan ik aan de hand van deze gegevens de nulhypotese verwerpen (A. neen)

9 Toetsen voor 2 populatieparameters

9.1 Vergelijken van twee gemiddelden bij gepaarde steekproeven

9.2 Vergelijken van twee populatiegemiddelden bij ongepaarde steekproeven

9.2.1 schema

Test for	Null Hypothesis (H_0)	Test Statistic	Distribution	Use When
Difference of two population means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Z	Both normal distributions or $n_1, n_2 \geq 30$; σ_1, σ_2 known
Difference of two population means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	t-distribution with $df = min\{n_1 - 1, n_2 - 1\}$	$n_1, n_2 < 30$; σ_1, $\sigma_2$unknown
Difference of two proportions ($p_1 - p_2$)	$p_1 - p_2 = 0$	$\frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ width $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$	Z	$n\hat{p} \geq 10$; $n(1 - \hat{p}) \geq 10$ for each group

9.3 oefeningen

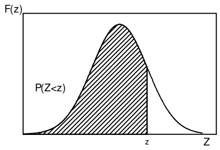
1. Je kunt afvallen door te sporten of door te diëten. Onderzoekers willen weten of er een verschil in gewichtsverlies is tussen de beide methoden. Ze volgen daartoe een groep proefpersonen die willen afvallen. 42 kandidaten doen dat door enkel een dieet te volgen en 47 door alleen te sporten. Na afloop van de testperiode heeft de dieet groep een gemiddeld gewichtsverlies van 5,9 kg met standaardafwijking 4.1 kg. De sportgroep verloor gemiddeld 4,1 kg met standaardafwijking 3,7 kg. Men wil het het verschil in het gemiddelde gewichtsverlies tussen beide groepen onderzoeken.

2. Ooit was gastrische bevriezing een aanbevolen behandeling van zweren in het darmkanaal. Een gerandomiseerd vergelijkend experiment stelde vast dat de toestand bij 28 patiënten die gastrische bevriezing ondergingen verbeterde, terwijl dat bij 30 van de 78 patiënten in de controlegroep, die een andere behandeling kregen, het geval was. Is er een verschil in behandeling?

10 Geogebra

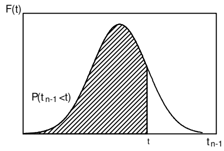
11 tabellen

Tabel 9.1: Standaard normale verdeling



	Tweede decimaal van z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Tabel 9.2: t-verdeling



	P						
vg	.750	.900	.950	.975	.990	.995	.999
1	1.000	3.078	6.314	12.706	31.821	63.656	318.289
2	0.816	1.886	2.920	4.303	6.965	9.925	22.328
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.894
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Tabel 9.2 voortgezet

	P						
vg	0.983	0.988	0.992	0.993	0.994	0.995	0.996
1	19.081	25.452	38.189	44.557	50.922	63.656	76.392
2	5.339	6.205	7.649	8.277	8.860	9.925	10.886
3	3.740	4.177	4.857	5.138	5.392	5.841	6.232
4	3.186	3.495	3.961	4.148	4.315	4.604	4.851
5	2.912	3.163	3.534	3.681	3.810	4.032	4.219
6	2.749	2.969	3.287	3.412	3.521	3.707	3.863
7	2.642	2.841	3.128	3.238	3.335	3.499	3.636
8	2.566	2.752	3.016	3.117	3.206	3.355	3.479
9	2.510	2.685	2.933	3.028	3.111	3.250	3.364
10	2.466	2.634	2.870	2.960	3.038	3.169	3.277
11	2.431	2.593	2.820	2.906	2.981	3.106	3.208
12	2.403	2.560	2.779	2.863	2.934	3.055	3.153
13	2.380	2.533	2.746	2.827	2.896	3.012	3.107
14	2.360	2.510	2.718	2.796	2.864	2.977	3.069
15	2.343	2.490	2.694	2.770	2.837	2.947	3.036
16	2.328	2.473	2.673	2.748	2.813	2.921	3.008
17	2.316	2.458	2.655	2.729	2.793	2.898	2.984
18	2.304	2.445	2.639	2.712	2.775	2.878	2.963
19	2.294	2.433	2.625	2.697	2.759	2.861	2.944
20	2.285	2.423	2.613	2.683	2.744	2.845	2.927
21	2.278	2.414	2.601	2.671	2.732	2.831	2.912
22	2.270	2.405	2.591	2.661	2.720	2.819	2.899
23	2.264	2.398	2.582	2.651	2.710	2.807	2.886
24	2.258	2.391	2.574	2.642	2.700	2.797	2.875
25	2.252	2.385	2.566	2.634	2.692	2.787	2.865
26	2.247	2.379	2.559	2.626	2.684	2.779	2.856
27	2.243	2.373	2.552	2.619	2.676	2.771	2.847
28	2.238	2.368	2.546	2.613	2.669	2.763	2.839
29	2.234	2.364	2.541	2.607	2.663	2.756	2.832
30	2.231	2.360	2.536	2.601	2.657	2.750	2.825
40	2.204	2.329	2.499	2.562	2.616	2.704	2.776
60	2.178	2.299	2.463	2.524	2.575	2.660	2.729
120	2.153	2.270	2.428	2.486	2.536	2.617	2.683
∞	2.128	2.241	2.394	2.450	2.498	2.576	2.638