

Academie Toegepaste Biowetenschappen en Chemie

Statistiek BI2a-Stat 2023-2024

OWE2a: Opsporen van genetische mutaties & Python II



Pleun Zomer

pleun.zomer@han.nl

Week 4 Herhaling week 3 + opdracht 2.6.11 en 12
Binomiale kansverdeling

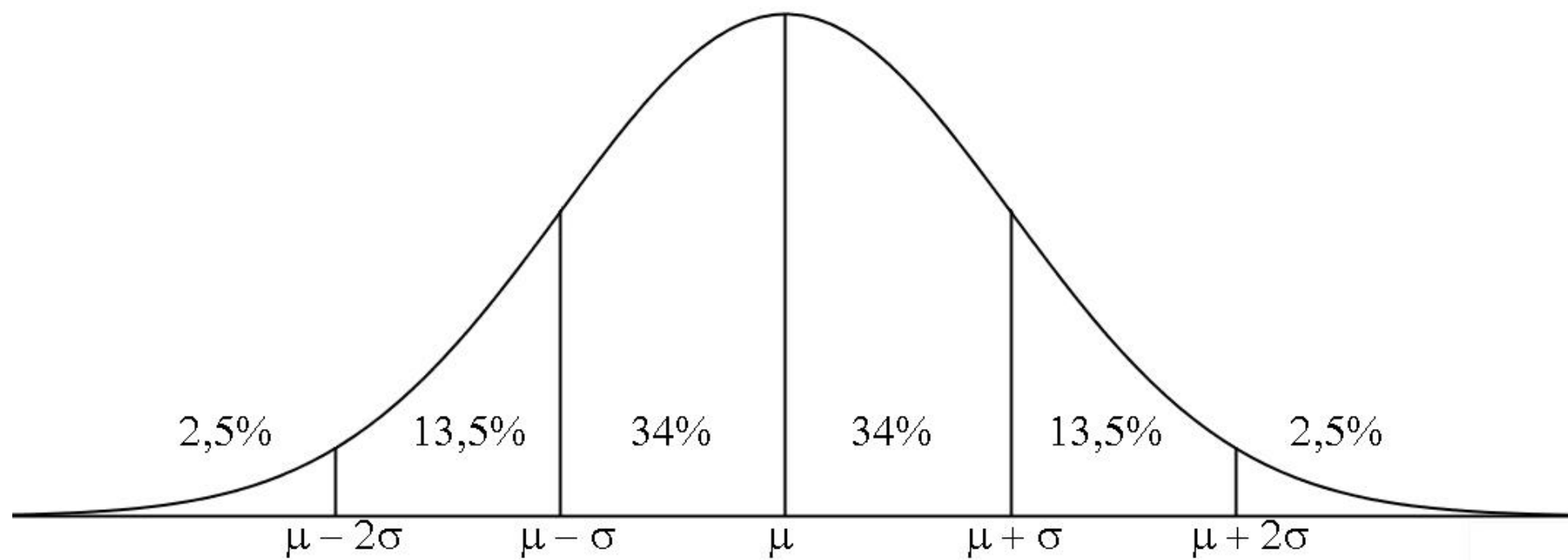
Herhaling week 3

Boxplot

- Eerste en derde kwartiel
- Uitbijters

Statistische uitkomsten en parameters

- Standaarddeviatie *van 2, 4, 6, 7, 8*



Herhaling week 3

$$\bar{x} = 98,3 \text{ U/l}$$

$$s = 40,4 \text{ U/l}$$

$$\bar{x} + s = 138,7 \text{ U/l}$$

$$\bar{x} - s = 57,9 \text{ U/l}$$

$$\bar{x} + 2s = 179,1 \text{ U/l}$$

$$\bar{x} - 2s = 17,5 \text{ U/l}$$

$$\bar{x} + 3s = 219,5 \text{ U/l}$$

$$\bar{x} - 3s = \text{onder } 0$$

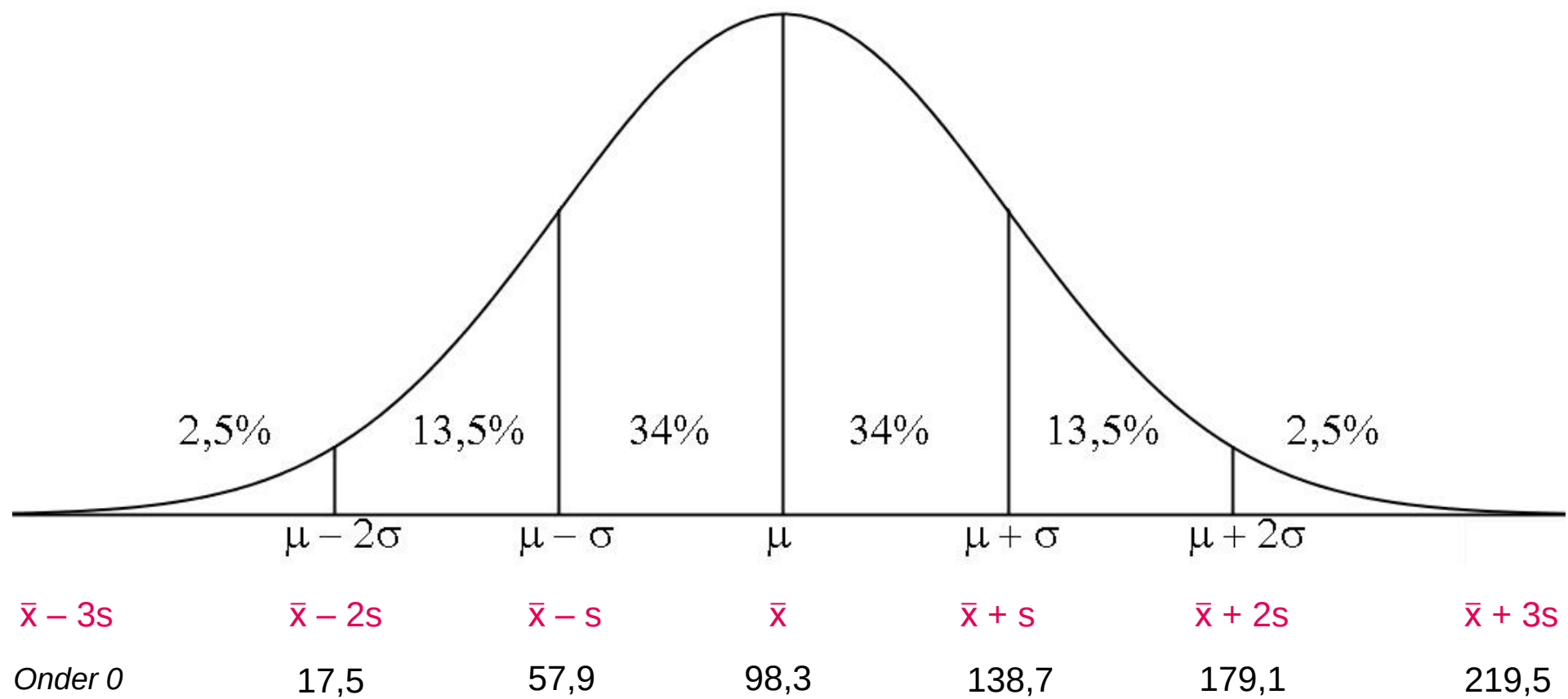
2.6.11 Listed in increasing order are the serum creatine phosphokinase (CK) levels (U/l) of 36 healthy men (these are the data of Example 2.2.6):

25	62	82	95	110	139
42	64	83	95	113	145
48	67	84	100	118	151
57	68	92	101	119	163
58	70	93	104	121	201
60	78	94	110	123	203

The sample mean CK level is 98.3 U/l and the SD is 40.4 U/l. What percentage of the observations are within

- (a) 1 SD of the mean?
- (b) 2 SDs of the mean?
- (c) 3 SDs of the mean?

2.6.12 Compare the results of Exercise 2.6.11 with the predictions of the empirical rule.



Kansen

Boek: Een kans is een numerieke grootte die de waarschijnlijkheid van een gebeurtenis uitdrukt. De kans op een gebeurtenis E wordt geschreven als:

$$P(E) \text{ or } \Pr(E)$$

De kans P is een getal tussen de 0 en 1 $0 \leq P(E) \leq 1$

De kans op een gebeurtenis (Event) $\Leftrightarrow \frac{\text{Aantal keren dat } E \text{ optreedt}}{\text{Totaal aantal keren dat experiment wordt herhaald}}$

Kansen

Boek: Een kans is een numerieke grootte die de waarschijnlijkheid van een gebeurtenis uitdrukt. De kans op een gebeurtenis E wordt geschreven als:

$$P(E) \text{ or } \Pr(E)$$

De kans P is een getal tussen de 0 en 1 $0 \leq P(E) \leq 1$

De kans op een gebeurtenis (Event) $\Leftrightarrow \frac{\text{Aantal keren dat } E \text{ optreedt}}{\text{Totaal aantal keren dat experiment wordt herhaald}}$

Voorbeeld: Werpen met een zuivere dobbelsteen

$E = \text{je werpt een } 6, \text{ dan } P(E) = \frac{1}{6}$

Kansen

Voorbeeld:

Een dobbelsteen heeft 3 rode vlakken, 2 blauwe vlakken en 1 wit vlak.

E = je werpt rood

$$P(E) \text{ of } P(\text{rood}) = \frac{3}{6} = \frac{1}{2}$$

En:

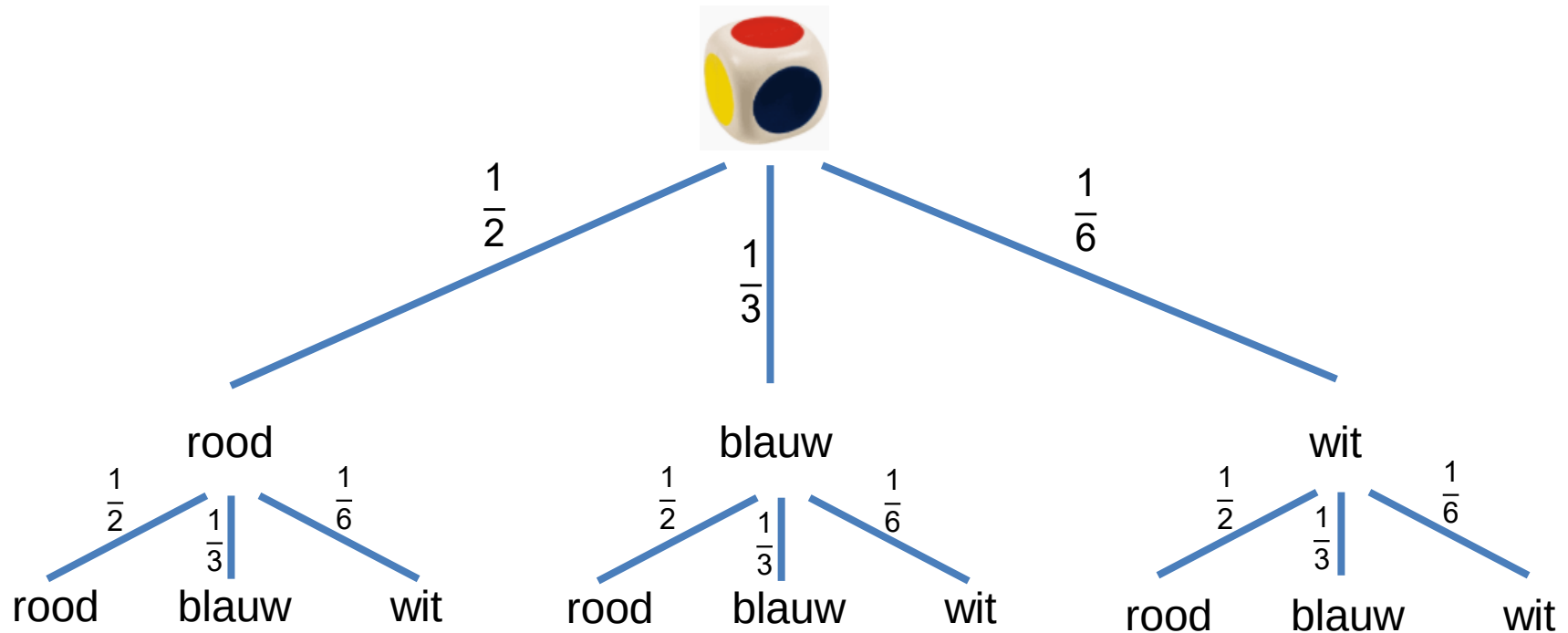
$$P(\text{blauw}) = \frac{2}{6} = \frac{1}{3}$$

$$P(\text{wit}) = \frac{1}{6}$$

Wat als we twee dobbelstenen werpen?



Kansboom!



P(rood en blauw)

Twee mogelijkheden:

eerst rood dan blauw

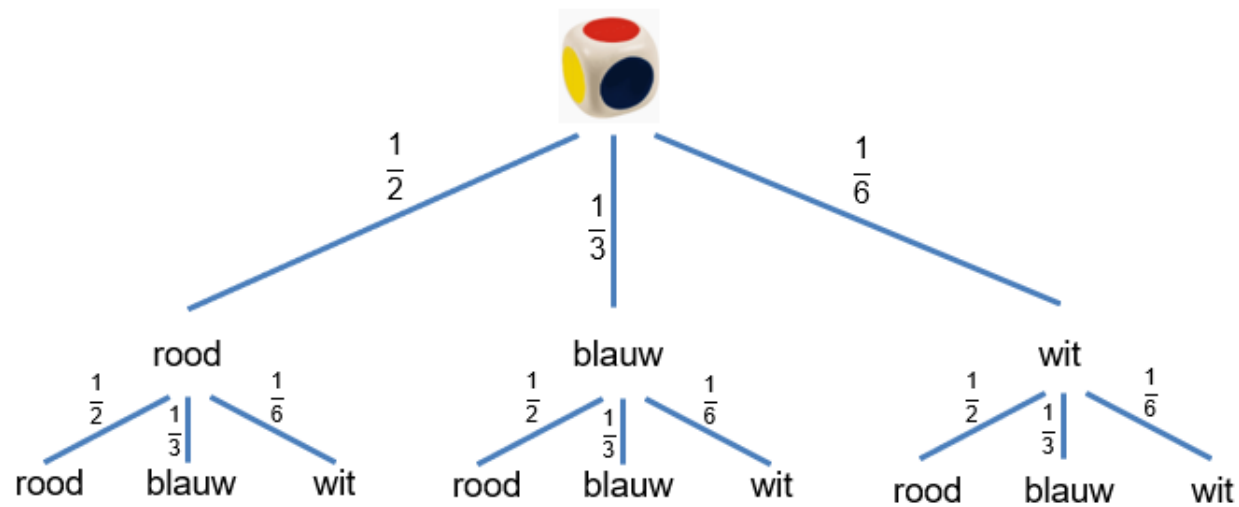
kans: $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$

eerst blauw dan rood

kans: $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6} +$

P(rood en blauw) =

$$\frac{2}{6} = \frac{1}{3}$$



$P(\text{minstens wit}) =$

$$P(w + b) + P(b + w) + P(w + r) + P(r + w) + P(w + w)$$

$$P(\text{géén wit}) = 1 - P(\text{minstens wit})$$

Je hebt een vaas met 5 rode knikkers, 3 blauwe knikkers en 6 groene knikkers.

1. Wat is de kans dat je een rode knikker trekt? En blauw? En groen?
2. Wat is de kans dat je twee rode knikkers trekt, wanneer je de knikkers **niet** teruglegt?
3. Wat is die kans als je de knikkers **wel** teruglegt?
4. Wat is de kans op een rode en een groene knikker, wanneer je de knikkers **niet** teruglegt?

Aan de slag

- Opgave 3.2: 1, 3, 4

Toepassing: medisch testen

Patiënt wordt getest op ziekte

Ziekte aanwezig → positieve test

Ziekte niet aanwezig → negatieve test

Mogelijkheid: test werkt niet 100% nauwkeurig

Positieve test bij afwezigheid ziekte → false positive

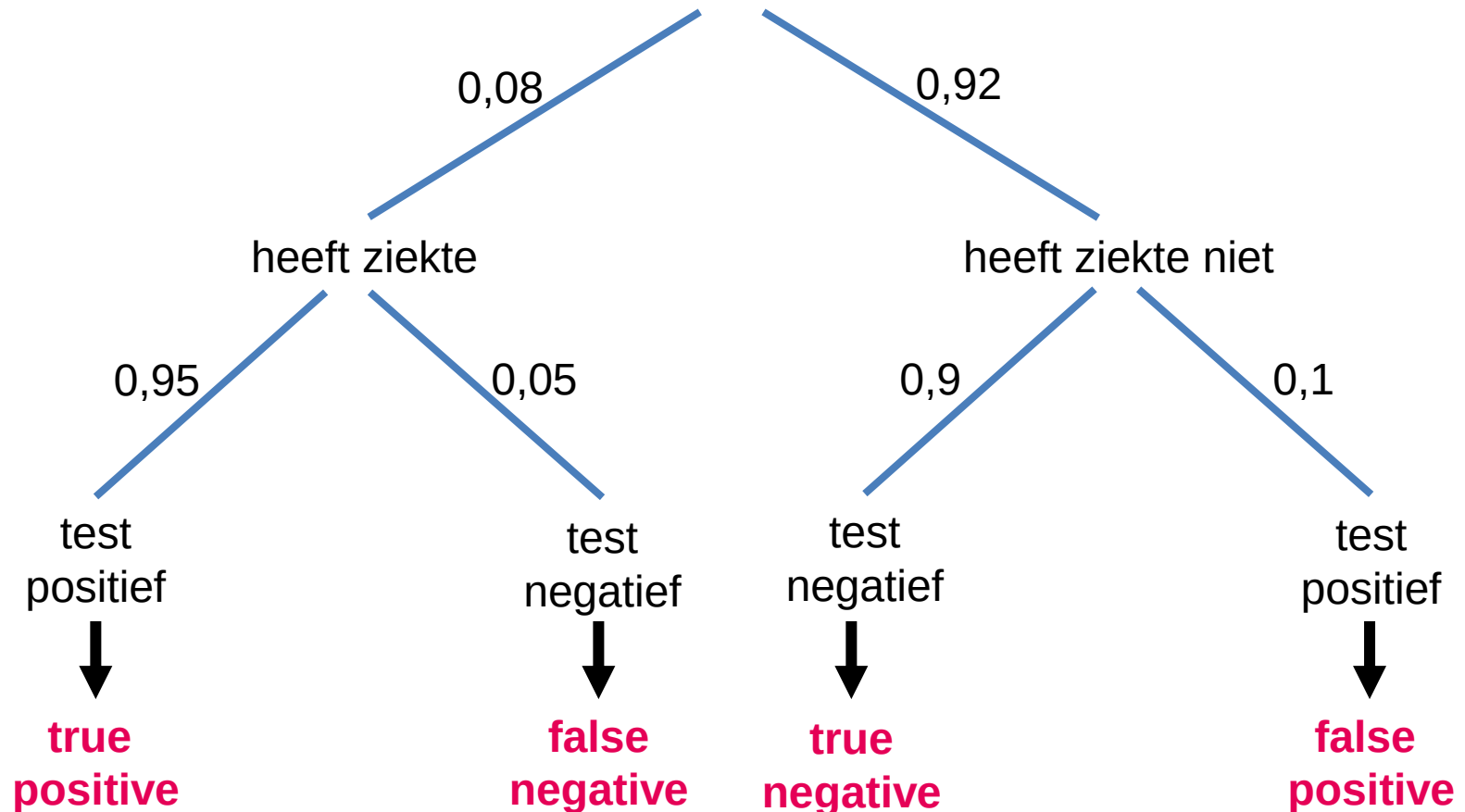
Negatieve test bij aanwezigheid ziekte → false negative

Sensitiviteit is de kans dat de test een ziekte detecteert als de patiënt ook echt ziek is

Specificiteit is de kans dat de test negatief is als de patiënt de ziekte ook niet heeft

Toepassing: medisch testen

Stel: **Sensitiviteit** van de test is 95%
Specificiteit van de test is 90%
8% van de onderzoekspopulatie heeft de ziekte

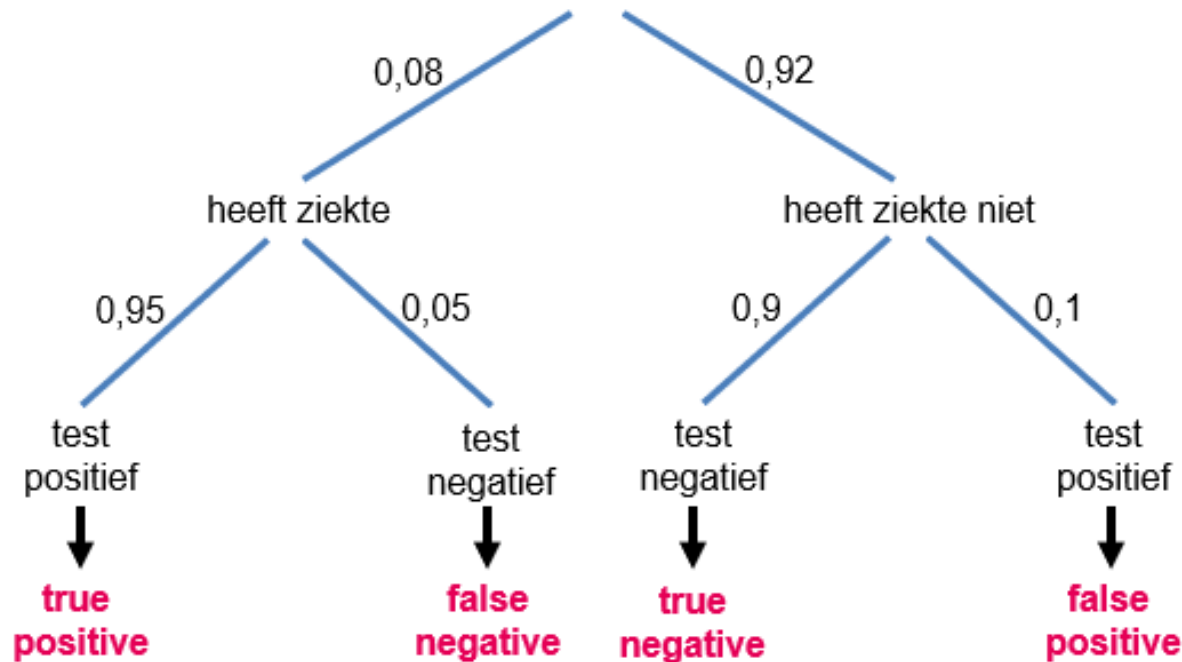


Toepassing: medisch testen

Stel: **Sensitiviteit** van de test is 95%

Specificiteit van de test is 90%

8% van de onderzoekspopulatie heeft de ziekte



$$P(\text{true positive}) = 0,08 \cdot 0,95 = \mathbf{0,076}$$
$$P(\text{false negative}) = 0,08 \cdot 0,05 = \mathbf{0,004}$$

$$P(\text{true negative}) = 0,92 \cdot 0,9 = \mathbf{0,828}$$
$$P(\text{false positive}) = 0,92 \cdot 0,1 = \mathbf{0,092}$$

$$\begin{aligned} P(\text{true positive}) &= 0,08 \cdot 0,95 = \mathbf{0,076} \\ P(\text{false negative}) &= 0,08 \cdot 0,05 = \mathbf{0,004} \end{aligned}$$

$$\begin{aligned} P(\text{true negative}) &= 0,92 \cdot 0,9 = \mathbf{0,828} \\ P(\text{false positive}) &= 0,92 \cdot 0,1 = \mathbf{0,092} \end{aligned}$$

Stel: 1000 personen worden getest

Positieve tests: $76 + 92 = 168$

Echter

76 van de 168 mensen zijn echt ziek, 92 niet

$$\text{Dus } \frac{92}{168} \cdot 100\% \approx 54,8\%$$

van de positief geteste mensen is helemaal niet ziek.

Hoeveel procent van de negatief geteste mensen is wél ziek?

$$\begin{aligned} P(\text{true positive}) &= 0,08 \cdot 0,95 = \mathbf{0,076} \\ P(\text{false negative}) &= 0,08 \cdot 0,05 = \mathbf{0,004} \end{aligned}$$

$$\begin{aligned} P(\text{true negative}) &= 0,92 \cdot 0,9 = \mathbf{0,828} \\ P(\text{false positive}) &= 0,92 \cdot 0,1 = \mathbf{0,092} \end{aligned}$$

Stel: 1000 personen worden getest

Positieve tests: $76 + 92 = 168$

Echter

76 van de 168 mensen zijn echt ziek, 92 niet

$$\text{Dus } \frac{92}{168} \cdot 100\% \approx 54,8\%$$

van de positief geteste mensen is helemaal niet ziek.

Hoeveel procent van de negatief geteste mensen is wél ziek?

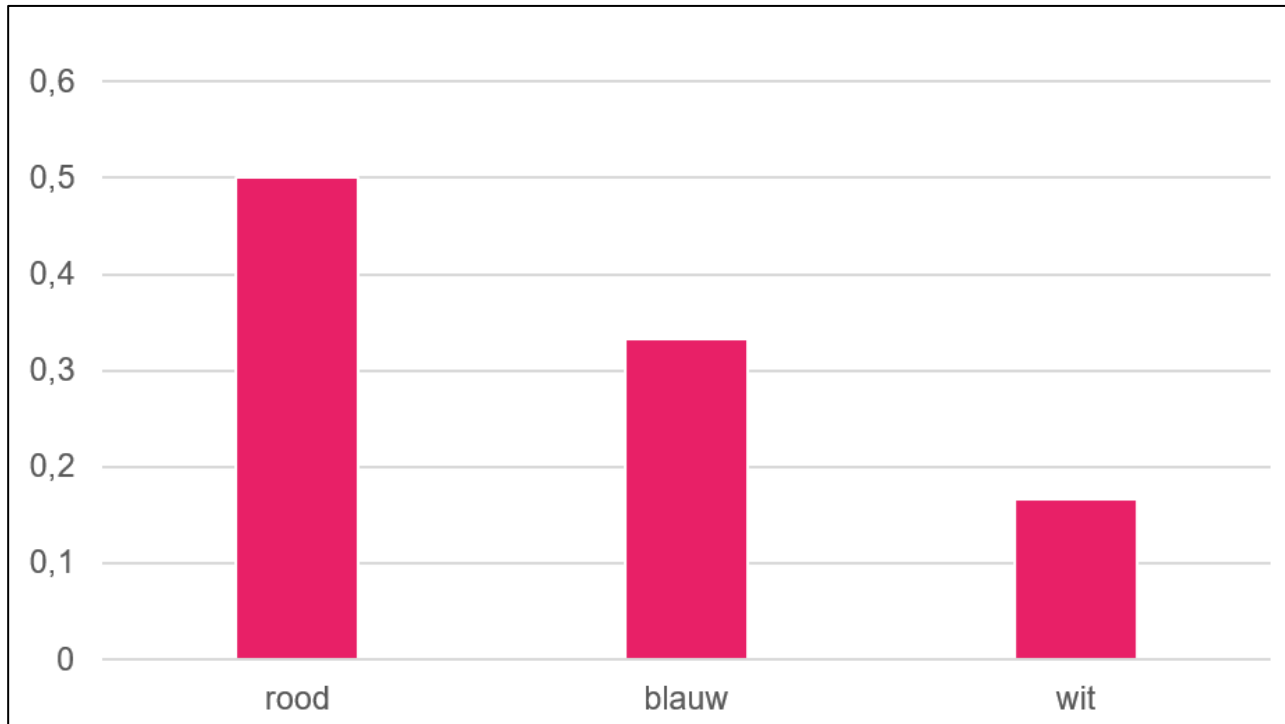
Negatieve tests: $4 + 828 = 832$

$$\frac{4}{832} \cdot 100\% \approx 0,5\%$$

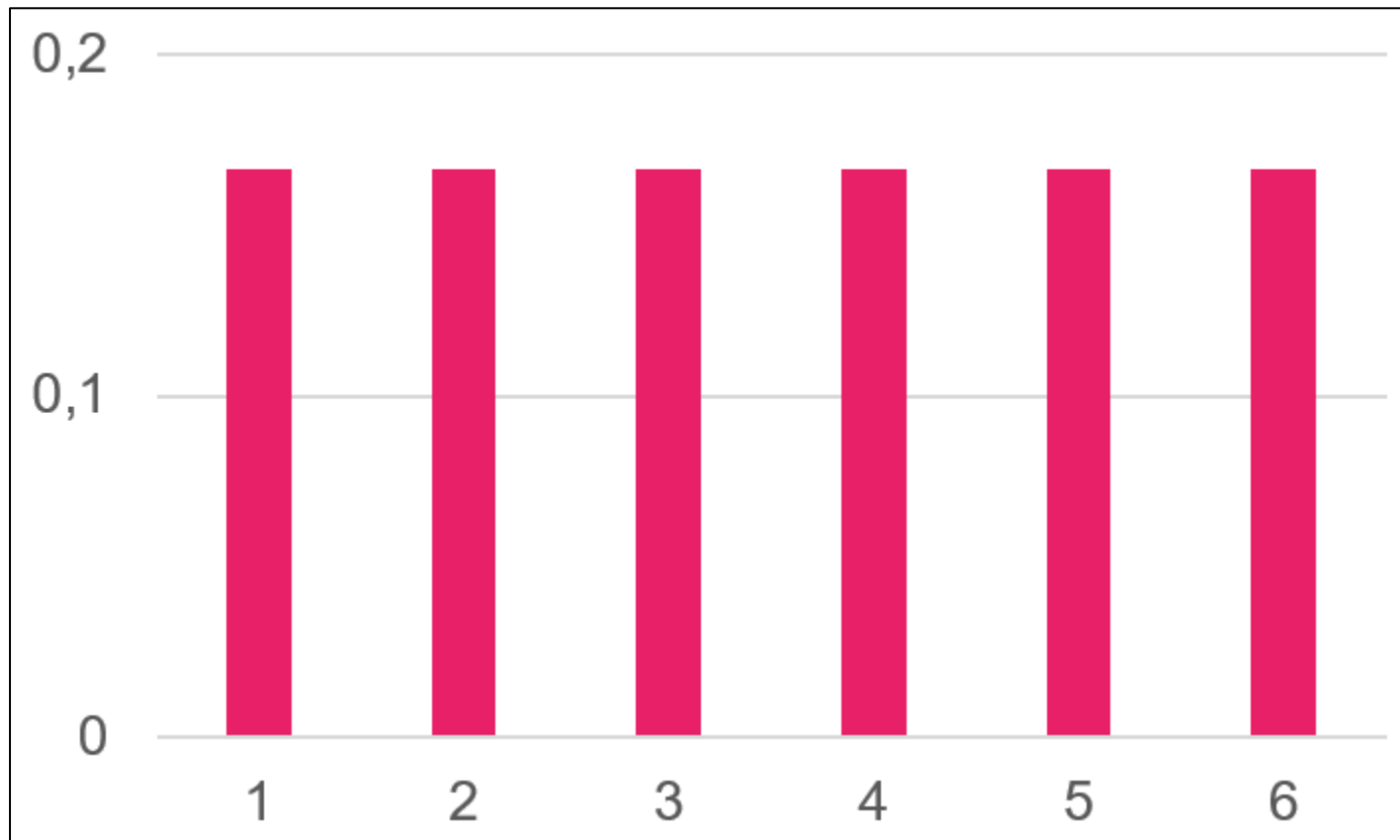
Kansverdeling

Kansen kunnen weergegeven worden in een relatieve frequentie diagram.

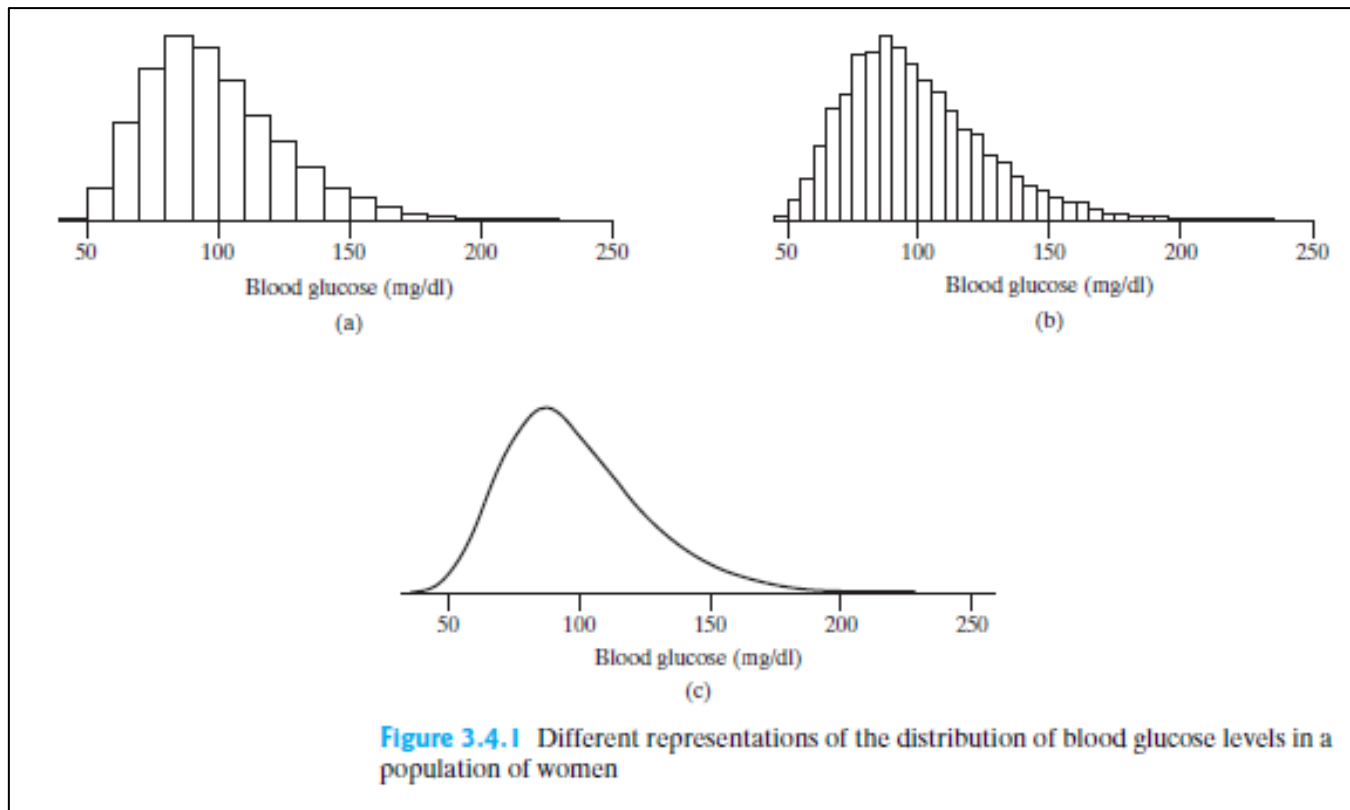
Dobbelsteen met kleuren:



Zuivere (ogen)dobbelsteen:



Zie voorbeeld 3.4.1:



Noot. Herdrukt van "Statistics for the Life Sciences", Samuels, M. L., Witmer, J. A., Schaffner, A. A., 2016, p. 109, Edinburgh Gate, Verenigd Koninkrijk: Pearson Education Ltd.

De kans op alle mogelijkheden is altijd gelijk aan 1

De oppervlakte onder de curve = 1

Zie voorbeeld 3.4.3:

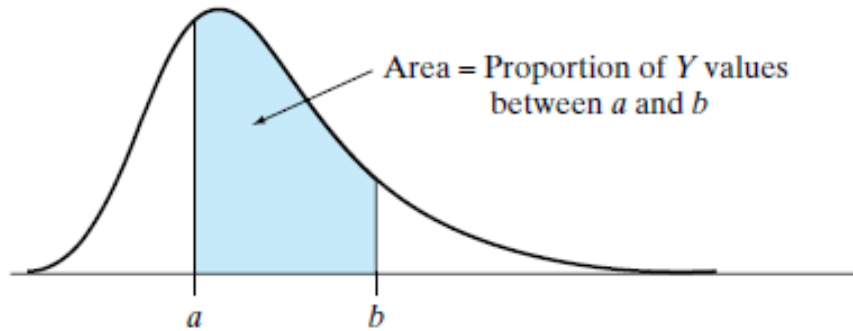


Figure 3.4.2 Interpretation of area under a density curve

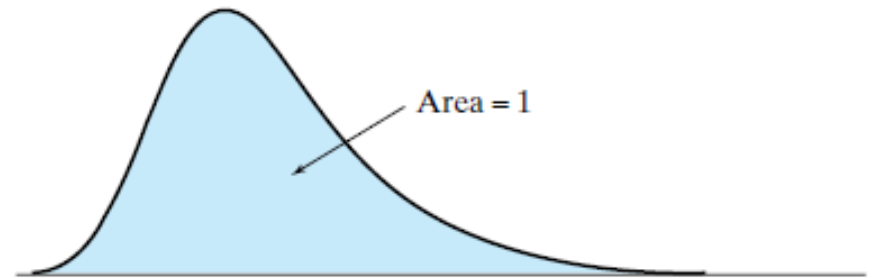


Figure 3.4.3 The area under an entire density curve must be 1

Noot. Herdrukt van "Statistics for the Life Sciences", Samuels, M. L., Witmer, J. A., Schaffner, A. A., 2016, p. 110, Edinburgh Gate, Verenigd Koninkrijk: Pearson Education Ltd.

De kans op alle mogelijkheden is altijd gelijk aan 1

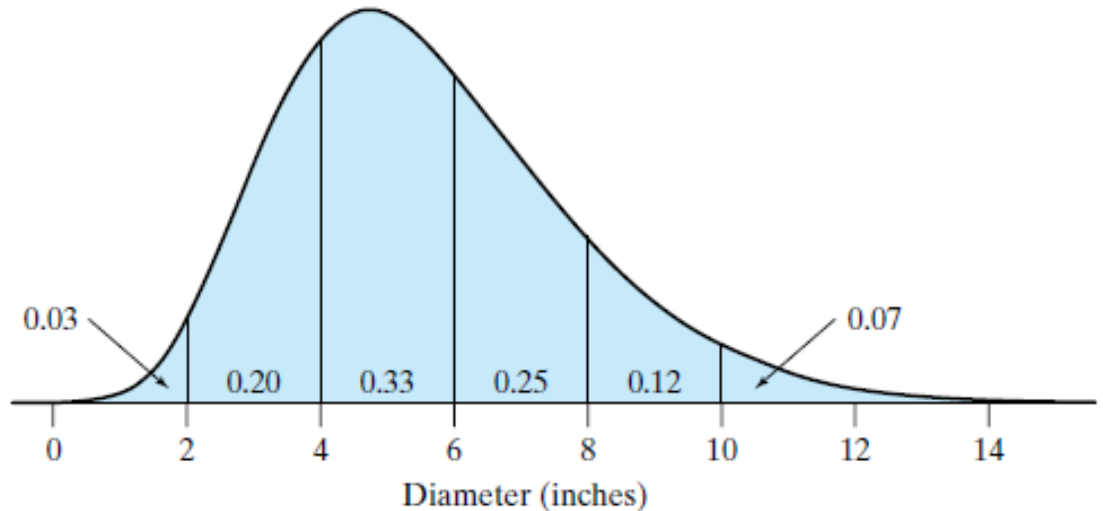
Dat betekent: de oppervlakte onder de curve is 1

Met de curve kan je kansen berekenen

Zie voorbeeld 3.4.5:

Met een curve en gegeven oppervlaktes kan je kansen berekenen.

Figure 3.4.5 Diameters of 30-year-old Douglas fir trees



Noot. Herdrukt van “Statistics for the Life Sciences”, Samuels, M. L., Witmer, J. A., Schaffner, A. A., 2016, p. 111, Edinburgh Gate, Verenigd Koninkrijk: Pearson Education Ltd.

Wat is de kans op een diameter kleiner dan 6?

Notatie:

Bereken: $P(X < 6)$

$X = \text{diameter}$

Bereken: $P(X > 8)$

Aan de slag

- Opgave 3.2: 6, 8
- Opgave 3.4: 1, 2, 3, 5

Huiswerk

- Maak:
 - 3.2: 1, 3, 4, 6, 8
 - 3.4: 1, 2, 3, 5
- Opdrachten bespreken