



**STATISTIEK & EXCEL**  
**LES 3: REGRESSIE & CORRELATIE**

## HERHALING LES 2

De pH van een oplossing wordt 5 keer gemeten. De gemeten pH is  $7,556 \pm 0,145$ .

De afgeronde waarde is:

- a)  $7,6 \pm 0,1$
- b)  $7,56 \pm 0,14$
- c)  $7,56 \pm 0,15$



# HERHALING LES 2



De pH van een oplossing wordt 5 keer gemeten. De gemeten pH is  $7,556 \pm 0,145$ .

De afgeronde waarde is:

a)  $7,6 \pm 0,1$

b)  $7,56 \pm 0,14$

c)  $7,56 \pm 0,15$

•Afronding van meetwaarden en meetfouten:

1. Schrijf de meetwaarde en meetfout in dezelfde macht van 10.
2. Rond de meetfout af op **1 significant cijfer**.
3. Rond de meetwaarde af op **hetzelfde aantal decimalen** als de meetfout.

$$7,556 \pm 0,145$$

$$1) 75,56 \cdot 10^{-1} \pm 1,45 \cdot 10^{-1}$$

$$2) 75,56 \cdot 10^{-1} \pm 1 \cdot 10^{-1}$$

$$3) 76 \cdot 10^{-1} \pm 1 \cdot 10^{-1}$$

$$\rightarrow 7,6 \pm 0,1$$

## HERHALING LES 2

De lengte van 10 ILC studenten wordt gemeten. De gemiddelde lengte is  $1,70 \cdot 10^2 \pm 0,51 \cdot 10^1$  cm.

De afgeronde waarde is:

- a)  $1,7 \cdot 10^2 \pm 0,05 \cdot 10^2$  cm
- b)  $170 \pm 5,1$  cm
- c)  $1,70 \cdot 10^2 \pm 0,05 \cdot 10^2$  cm



# HERHALING LES 2



De lengte van 10 ILC studenten wordt gemeten. De gemiddelde lengte is  $1,70 \cdot 10^2 \pm 0,51 \cdot 10^1$  cm.

De afgeronde waarde is:

a)  $1,7 \cdot 10^2 \pm 0,05 \cdot 10^2$  cm

b)  $170 \pm 5,1$  cm

c)  $1,70 \cdot 10^2 \pm 0,05 \cdot 10^2$  cm

• Afronding van meetwaarden en meetfouten:

1. Schrijf de meetwaarde en meetfout in dezelfde macht van 10.
2. Rond de meetfout af op **1 significant cijfer**.
3. Rond de meetwaarde af op **hetzelfde aantal decimalen** als de meetfout.

$$1,70 \cdot 10^2 \pm 0,51 \cdot 10^1$$

1)  $1,70 \cdot 10^2 \pm 0,051 \cdot 10^2$

2)  $1,70 \cdot 10^2 \pm 0,05 \cdot 10^2$

3)  $1,70 \cdot 10^2 \pm 0,05 \cdot 10^2$

->  $170 \pm 5$

# HERHALING LES 2

Twee studenten werken samen om de pH van een oplossing te meten. De een meet, de ander schrijft de meting op.

De pH van de oplossing wordt 5 keer gemeten. De gemeten waarden zijn weergegeven in de tabel. De werkelijke pH is 7,6.

Hier is waarschijnlijk sprake van een:

- a) Toevallige fout
- b) Systematische fout
- c) Vermijdbare fout

meting	pH
meting_1	7,45
meting_2	7,55
meting_3	7,77
meting_4	9,01
meting_5	7,65



# HERHALING LES 2

Twee studenten werken samen om de pH van een oplossing te meten. De een meet, de ander schrijft de meting op.

De pH van de oplossing wordt 5 keer gemeten. De gemeten waarden zijn weergegeven in de tabel. De werkelijke pH is 7,6.

Hier is waarschijnlijk sprake van een:

- a) Toevallige fout
- b) Systematische fout
- c) Vermijdbare fout

meting	pH
meting_1	7,45
meting_2	7,55
meting_3	7,77
meting_4	9,01
meting_5	7,65



# HERHALING LES 2

Welk type vraag is de volgende vraag?

Is er relatie tussen de dagelijkse voedselinname en het lichaamsgewicht van hamsters?

- a) Beschrijvende vraag
- b) Verschilvraag
- c) Verbandvraag





# HERHALING LES 2

Welk type vraag is de volgende vraag?

Is er een verband tussen de dagelijkse voedselinname en het lichaamsgewicht van hamsters?

- a) Beschrijvende vraag
- b) Verschilvraag
- c) **Verbandvraag**

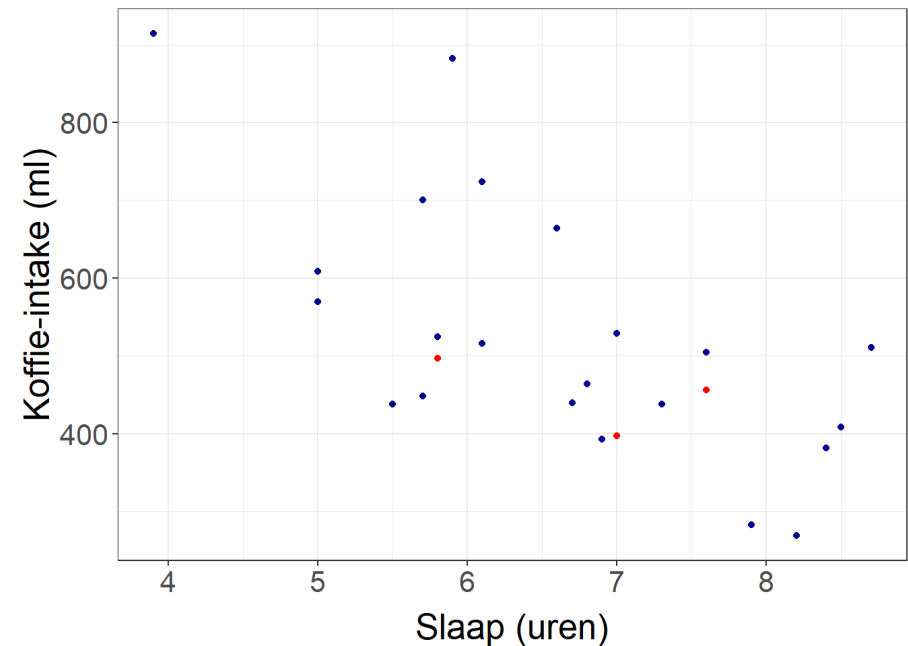


# INHOUD LES 3

- Wat is een correlatieanalyse?
- De correlatiecoëfficiënt
- Wat is een regressieanalyse?
- De determinatiecoëfficiënt
- De ijklijn
- Afronden van correlatie-/determinatiecoëfficiënten

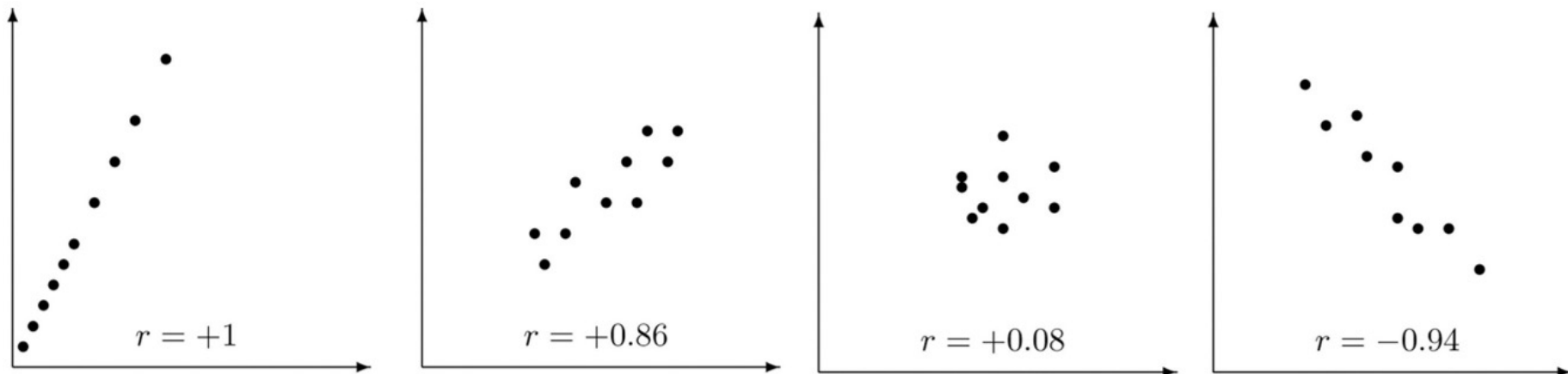
# WAT IS EEN CORRELATIEANALYSE?

- Alleen voor **kwantitatieve data**
- Er is sprake van **twee afhankelijke variabelen** (we gaan niet uit van een causaal verband)
- Vraag: 'Is er een statistisch verband tussen variabele 1 en variabele 2?'
- Voorbeeld: 'Is er een statistisch verband tussen het aantal uren slaap en de koffieconsumptie van docenten?'
- Statistisch verband  $\neq$  causaal verband



# DE CORRELATIECOËFFIËNT

- Correlatiecoëfficiënt is een getal tussen -1 en 1.
- Correlatiecoëfficiënt geeft aan hoe sterk het verband is tussen de twee variabelen:
  - $r = 0 \rightarrow$  er is geen verband
  - $r = -1 \rightarrow$  er is een perfect negatief verband
  - $r = +1 \rightarrow$  er is een perfect positief verband



# DE CORRELATIECOËFFIËNT

$R=1$  of  $-1 \rightarrow$  “hoeveel koffie een docent drinkt hangt volledig samen met hoeveel hij/zij geslapen heeft. Een docent die 6 uur geslapen heeft, drinkt altijd [yyy] ml koffie.”

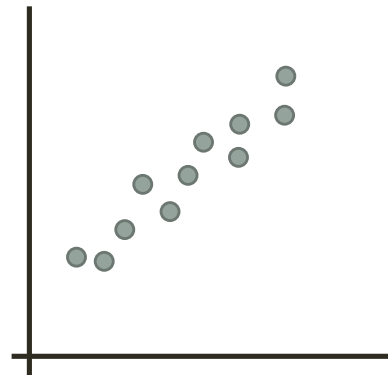
$R=0 \rightarrow$  hoeveel koffie een docent dinkt en hoeveel hij/zij slaap heeft niets met elkaar te maken.

# VRAAG 1



Welke correlatiecoëfficiënt ( $r$ ) hoort waarschijnlijk bij de onderstaande grafiek?

- a) 0,86
- b) -0,86
- c) 0,04

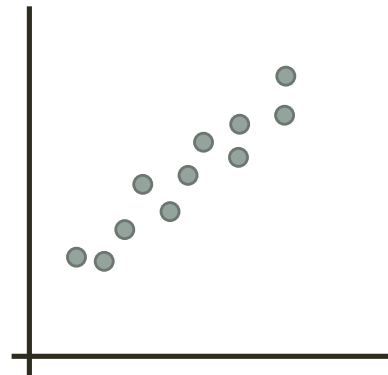


# VRAAG 1



Welke correlatiecoëfficiënt ( $r$ ) hoort waarschijnlijk bij de onderstaande grafiek?

- a) **0,86**
- b) -0,86
- c) 0,04

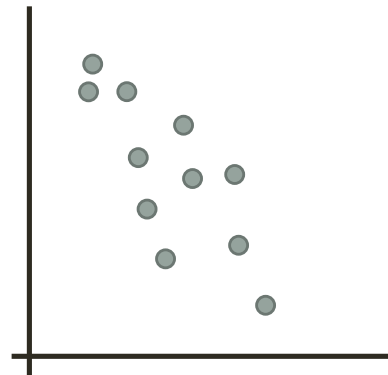


## VRAAG 2



Welke correlatiecoëfficiënt ( $r$ ) hoort waarschijnlijk bij de onderstaande grafiek?

- a) 0,86
- b) 0,64
- c) -0,64



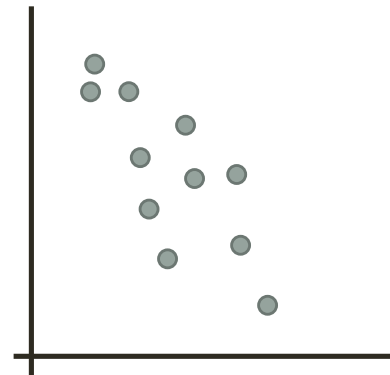


## VRAAG 2



Welke correlatiecoëfficiënt ( $r$ ) hoort waarschijnlijk bij de onderstaande grafiek?

- a) 0,86
- b) 0,64
- c) **-0,64**

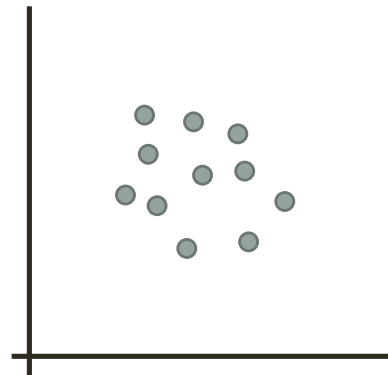


## VRAAG 3



Welke correlatiecoëfficiënt ( $r$ ) hoort waarschijnlijk bij de onderstaande grafiek?

- a) 0,64
- b) -0,86
- c) -0,04

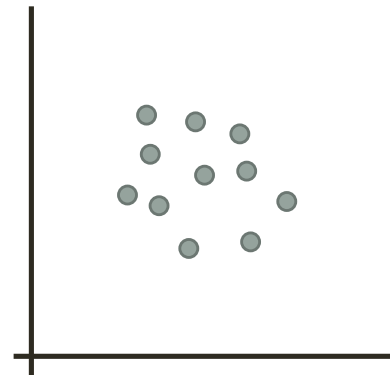


## VRAAG 3



Welke correlatiecoëfficiënt ( $r$ ) hoort waarschijnlijk bij de onderstaande grafiek?

- a) 0,64
- b) -0,86
- c) -0,04



# WAT IS EEN REGRESSIEANALYSE

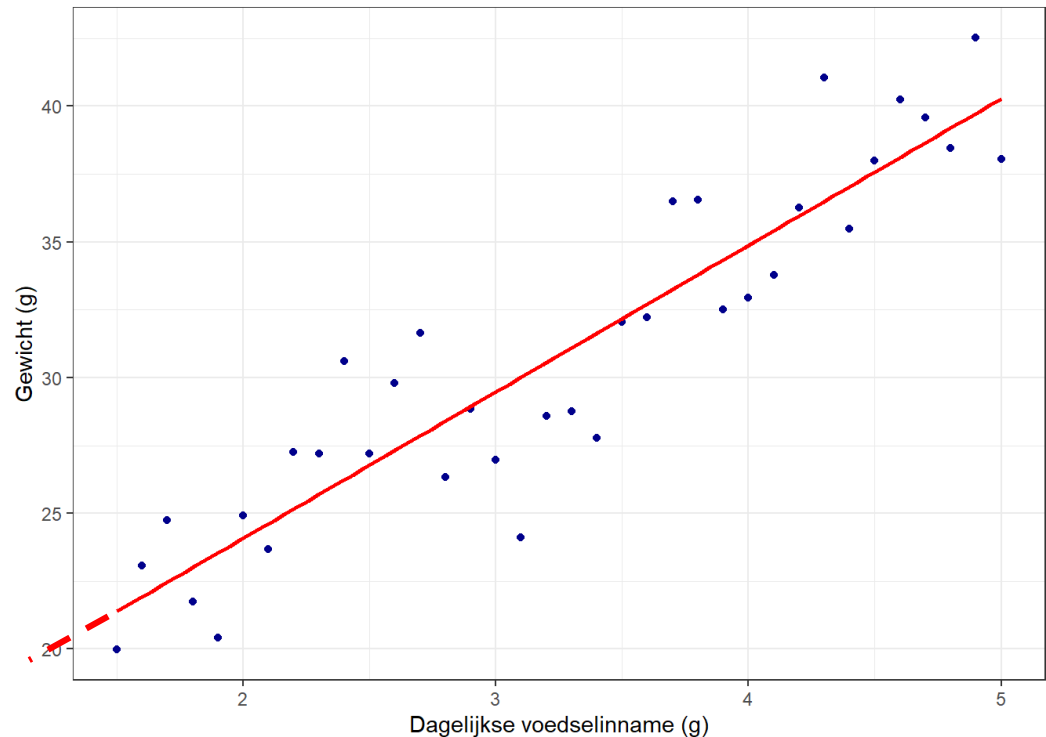
- Alleen voor **kwantitatieve data**
- Er is sprake van **één onafhankelijke variabele (x)** en **één afhankelijke variabele (y)** (we gaan wel uit van een causaal verband)
- **Om uit te gaan van een causal verband is er ondersteunend bewijs, een theorie, of dogma**
- Vraag: ‘Kan ik de waarde van y voorspellen op basis van x?’
- Voorbeeld: ‘Kan ik het gewicht van een hamster voorspellen op basis van zijn dagelijkse voedselinname?’
- Speciaal voorbeeld: **ijklijn** (zie hierna)
- Maak altijd eerst een scattergrafiek om te **controleren of er wel een lineair verband is!**

# WAT IS EEN REGRESSIEANALYSE?

- De rode lijn in de grafiek noemen we een **regressielijn**.
- De formule van de regressielijn is:  
$$y = ax + b$$

Hierin is:

- $a$  de **richtingscoëfficiënt**;
- en  $b$  het snijpunt met de  $y$ -as.



# DE DETERMINATIECOËFFICIËNT

- De **determinatiecoëfficiënt** ( $R^2$ ) is een getal tussen 0 en 1
- De determinatiecoëfficiënt geeft aan hoe goed de regressielijn past bij de data
  - $R^2 = 0 \rightarrow$  regressielijn past helemaal niet bij de data.
  - $R^2 = 1 \rightarrow$  regressielijn past uitstekend bij de data (alle punten liggen op de regressielijn).
- De determinatiecoëfficiënt geeft **geen** richting (positief of negatief) aan
- Bij een lineair verband tussen twee variabelen geldt:

$$r^2 = R^2 \text{ (correlatiecoëfficiënt}^2 \text{ = determinatiecoëfficiënt)}$$

## VRAAG 4



Er wordt een experiment uitgevoerd om te bepalen wat het effect is van de hoeveelheid pauze die studenten krijgen tijdens de lessen op het tentamencijfer. We doen een regressie-analyse. In een scatterplot is te zien dat er sprake is van een lineair verband en we vinden de best passende lineaire functie. Er geldt een determinatiecoëfficiënt van  $R^2 = 0,893$ . Dit alles betekent:

- a) Deze lineaire functie is niet bruikbaar,  $R^2$  is te laag.
- b) Op basis van de lineaire functie die gevonden is, is te zeggen dat mensen die meer pauze krijgen meestal een hoger cijfer halen
- c) Op basis van de lineaire functie die gevonden is, is best goed te voorspellen wat iemands tentamencijfer wordt als je weet hoeveel pauze hij/zij kreeg.

## VRAAG 4



Er wordt een experiment uitgevoerd om te bepalen wat het effect is van de hoeveelheid pauze die studenten krijgen tijdens de lessen op het tentamencijfer. We doen een regressie-analyse. In een scatterplot is te zien dat er sprake is van een lineair verband en we vinden de best passende lineaire functie. Er geldt een determinatiecoëfficiënt van  $R^2 = 0,893$ . Dit alles betekent:

- a) Deze lineaire functie is niet bruikbaar,  $R^2$  is te laag.
- b) Op basis van de lineaire functie die gevonden is, is te zeggen dat mensen die meer pauze krijgen meestal een hoger cijfer halen.
- c) Op basis van de lineaire functie die gevonden is, is best goed te voorspellen wat iemands tentamencijfer wordt als je weet hoeveel pauze hij/zij kreeg.

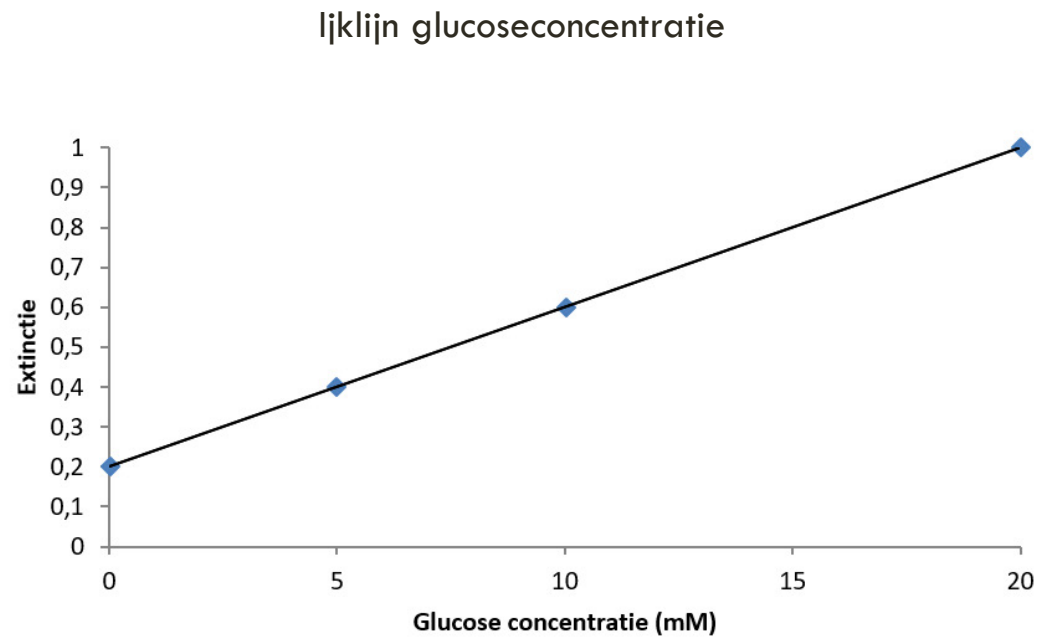


# DE IJKLIJN

- Speciaal voorbeeld van de regressieanalyse is de **ijklijn**.
- Vaak zijn we geïnteresseerd in het berekenen van  $x$ , omdat  $y$  bekend is:

$$y = ax + b \rightarrow x = \frac{y - b}{a}$$

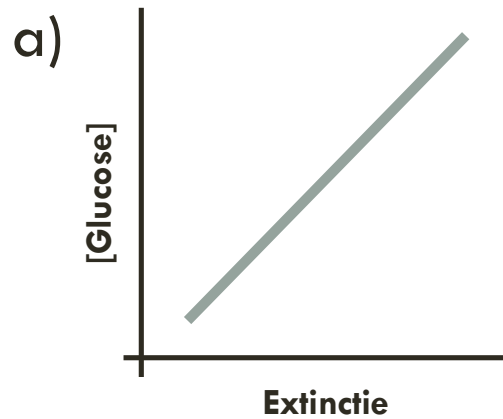
- Voor een ijklijn geldt dat  $R^2 > 0.99$



## VRAAG 5



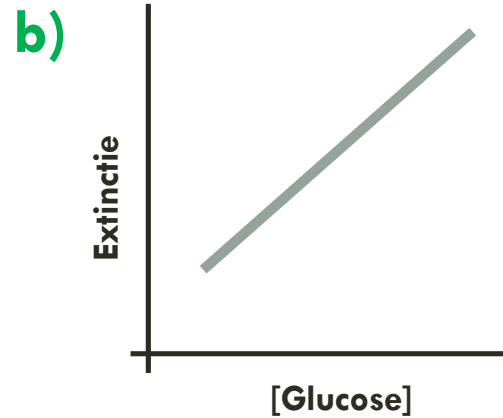
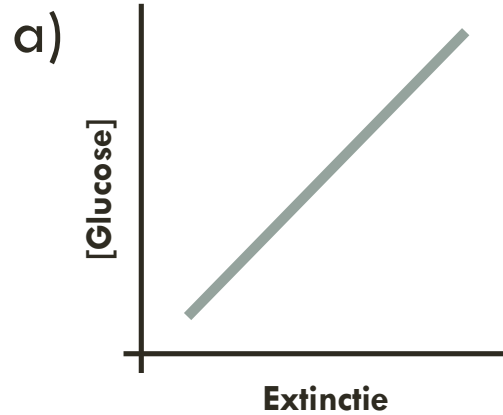
Voor een glucosebepaling wordt een ijklijn opgesteld door de extinctie voor verschillende monsters te meten. Hoe ziet de ijklijn eruit?



## VRAAG 5



Voor een glucosebepaling wordt een ijklijn opgesteld door de extinctie voor verschillende monsters te meten. Hoe ziet de ijklijn eruit?



## VRAAG 6

Gegeven de volgende ijklijn:  $extinctie = 0,543 \cdot [glucose] + 0,034$

Wat is 0,543?

- a) De richtingscoëfficiënt,  $\Delta x / \Delta y$
- b) De richtingscoëfficiënt,  $\Delta y / \Delta x$
- c) Het snijpunt met de x-as



## VRAAG 6

Gegeven de volgende ijklijn:  $extinctie = 0,543 \cdot [glucose] + 0,034$

Wat is 0,543?

- a) De richtingscoëfficiënt,  $\Delta x/\Delta y$
- b) De richtingscoëfficiënt,  $\Delta y/\Delta x$**
- c) Het snijpunt met de x-as



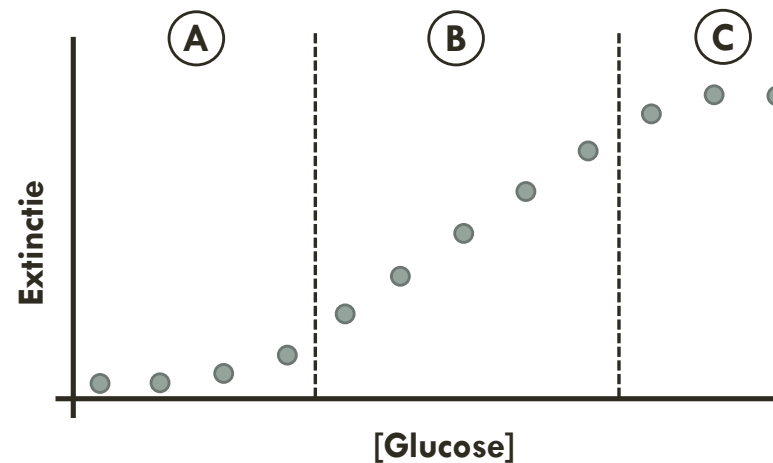
# VRAAG 7



We meten de extinctie voor verschillende bekende glucoseconcentraties. We krijgen de onderstaande grafiek.

Welke punten nemen we mee voor de ijklijn?

- a) De punten in B
- b) De punten in A en B
- c) De punten in A, B en C



# VRAAG 7



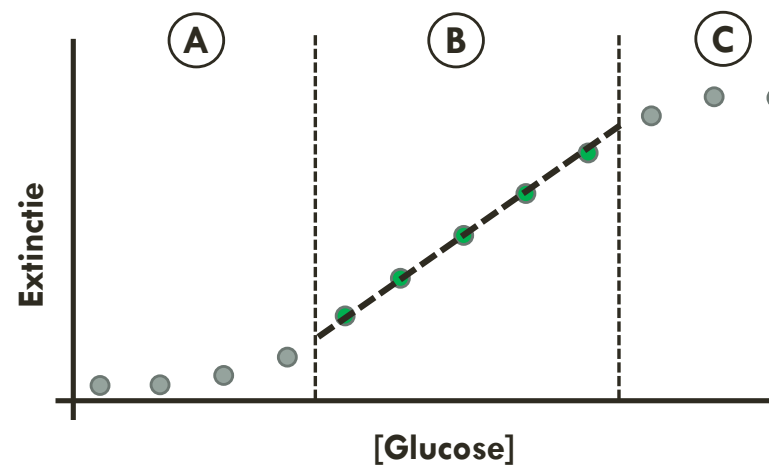
We meten de extinctie voor verschillende bekende glucoseconcentraties. We krijgen de onderstaande grafiek.

Welke punten nemen we mee voor de ijklijn?

a) De punten in B

b) De punten in A en B

c) De punten in A, B en C

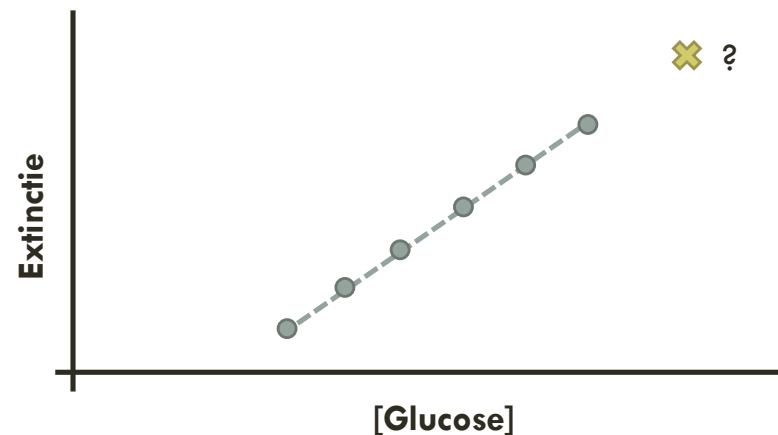


## VRAAG 8



We stellen een ijklijn op voor een glucosemeting. We meten een hoge extinctiewaarde voor een monster. Zie de onderstaande grafiek. Hoe bepalen we de glucoseconcentratie?

- a) We extrapoleren de ijklijn en bepalen met de nieuwe ijklijn de concentratie
- b) We verdunnen het monster en doen de meting opnieuw
- c) We gebruiken de formule van de bestaande ijklijn om het de onbekende waarde te bepalen.



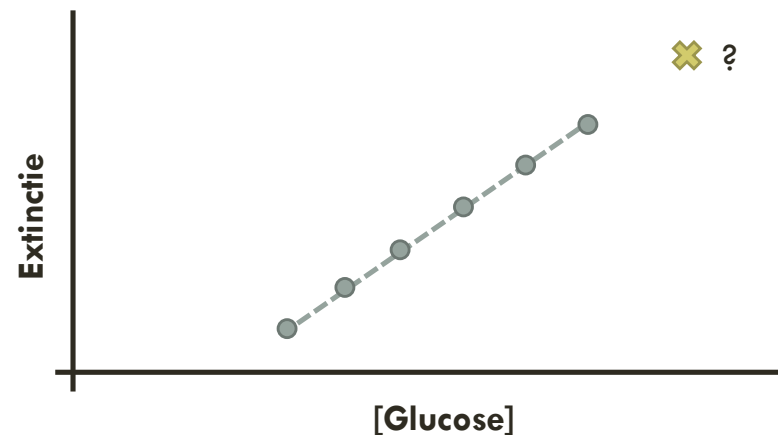


## VRAAG 8



We stellen een ijklijn op voor een glucosemeting. We meten een hoge extinctiewaarde voor een monster. Zie de onderstaande grafiek. Hoe bepalen we de glucoseconcentratie?

- a) We extrapoleren de ijklijn en bepalen met de nieuwe ijklijn de concentratie
- b) We verdunnen het monster en doen de meting opnieuw**
- c) We gebruiken de formule van de bestaande ijklijn om het de onbekende waarde te bepalen.

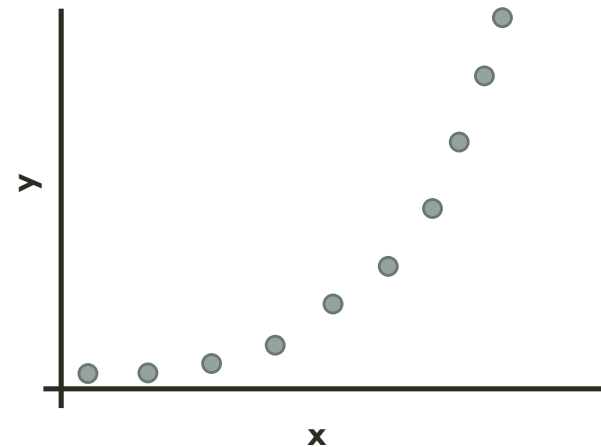


## VRAAG 9



We doen metingen voor een ijklijn en we krijgen de onderstaande grafiek. Hoe kunnen we de formule voor de ijklijn bepalen?

- a) We kunnen de formule voor de ijklijn niet bepalen; er is immers geen lineair verband
- b) We moeten de  $x$  en/of  $y$  waarden eerst transformeren voor we de ijklijnformule bepalen
- c) We trekken een lijn door de punten en bepalen daarvan de formule, zoals we dat altijd doen

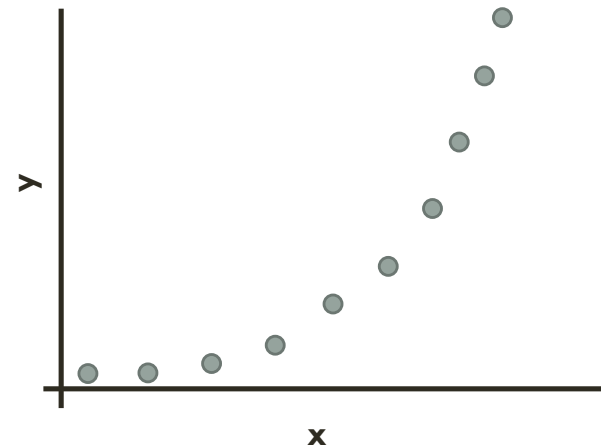


## VRAAG 9



We doen metingen voor een ijklijn en we krijgen de onderstaande grafiek. Hoe kunnen we de formule voor de ijklijn bepalen?

- a) We kunnen de formule voor de ijklijn niet bepalen; er is immers geen lineair verband
- b) We moeten de x en/of y waarden eerst transformeren voor we de ijklijnformule bepalen**
- c) We trekken een lijn door de punten en bepalen daarvan de formule, zoals we dat altijd doen



# AFRONDEN VAN CORRELATIE- EN DETERMINATIECOËFFICIËNTEN

Stappenplan:

- 1) Bereken:  $1 - |r| = x$
- 2) Rond het verkregen getal af op 1 significant cijfer ( $x_{afgerond}$ )
- 3) Bereken:  $1 - x_{afgerond} = r_{afgerond}$   
(en maak de uitkomst zo nodig weer negatief)

Voorbeeld:  $r = -0,992332545$

- 1)  $1 - |-0,992332545| = 0,007667455$
- 2)  $0,007667455 \approx 0,008$
- 3)  $1 - 0,008 = 0,992 \rightarrow$  Negatief maken  $\rightarrow -0,992$

NB: voor  $R^2$  geldt precies hetzelfde.

## VRAAG 10

We vinden  $r = 0,89356$ . Afgerond is dit:

- a) 0,89
- b) 0,90
- c) 0,9



## VRAAG 10

We vinden  $r = 0,89356$ . Afgerond is dit:

- a) 0,89
- b) 0,90
- c) **0,9**



# VRAAG 11

We vinden  $r = -0,999875$ . Afgerond is dit:

- a) 0,9999
- b) -1,0
- c) -0,9999



# VRAAG 11

We vinden  $r = -0,999875$ . Afgerond is dit:

- a) 0,9999
- b) -1,0
- c) **-0,9999**





# AAN DE SLAG MET EXCEL!

In het werkcollege leer je:

- hoe je een regressieanalyse (ijklijn) uitvoert in Excel;
- hoe je de correlatie- en determinatiecoëfficiënt kunt berekenen in Excel;
- en hoe je data kunt transformeren naar een lineair verband in Excel.

# HEEL VEEL PLEZIER!

Ga aan de slag met de opgaves in het werkcollege.

Stel vragen aan elkaar of aan de docent.