

# Data Analyse en Statistiek

HELLA SNOEK & MARTHE SCHUT



# Inhoud

<b>1</b>	<b>Introductie module 2</b>	<b>5</b>
<b>2</b>	<b>Foutenpropagatie</b>	<b>7</b>
2.1	Basisregel . . . . .	8
2.2	Som en verschil . . . . .	13
2.3	Vermenigvuldigen met constante . . . . .	13
2.4	Vermenigvuldigen met variabelen . . . . .	14



# Hoofdstuk 1

## Introductie module 2

Deze week verdiepen we ons in het concept meetonzekerheid. We leren hoe we onzekerheden kunnen doorrekenen in vergelijkingen. We zien hoe onzekerheden veranderen als we meer meetpunten aan onze dataset toevoegen. We kijken ook naar relaties tussen onzekerheden in meerdimensionale datasets, en we introduceren de laatste kans rekenregels die vooral voor multidimensionale data belangrijk is.

We werken in de laptopcolleges aan de opdrachten van deze module M2. Je vindt in het schema wanneer je aan welke opdrachten werkt en wanneer je deze moet inleveren. Vergeet ook niet te kijken naar het oefenmateriaal voor de tweede tussentoets. De tweede tussentoets volgt aan het einde van het derde hoorcollege.



## Hoofdstuk 2

# Foutenpropagatie

Vaak kunnen we de grootheid die we willen weten niet direct meten, maar meten we een observabele die zich via een bepaalde functie verhoudt tot de gezochte grootheid. Of meten we zelfs twee of meer variabelen die we nodig hebben om de gewilde grootheid te bepalen.

Dit is bijvoorbeeld het geval als we de gemiddelde snelheid van een auto willen bepalen. Dit zouden we kunnen doen door de tijd te meten die de auto nodig heeft om een bepaald traject af te leggen. We meten dan de door de auto gebruikte tijd,  $T$  en de lengte van het traject,  $L$ , en die zetten we dan om in snelheid via de bekende formule  $v = L/T$ . Of we bepalen bijvoorbeeld de massa van een elementair deeltje (in rust) en willen dit omzetten naar de energie van het deeltje via de formule  $E = mc^2$ .

Als we de onzekerheid weten op de gemeten grootheden dan kunnen we deze omzetten naar de grootheid die we eigenlijk willen bepalen. Dit noemen we het propageren van fouten. In dit hoofdstuk leren we je de basisregels voor het propageren van **ongecorreleerde** fouten. Dat wil zeggen dat als er meerdere onzekerheden worden gepropageerd deze onzekerheden onafhankelijk zijn; De meting van de ene observabele heeft geen invloed op de meting van de andere observabele; de fout die we maken in het meten van de ene grootheid hangt niet af van de fout die we maken op de andere gemeten grootheid.

Het is goed om alvast te beseffen dat er ook gecorreleerde fouten bestaan. Er zijn twee oorzaken voor het ontstaan van gecorreleerde fouten:

- Doordat er in de meting een correlatie is. Een voorbeeld van een gecorreleerde fout is als we een oppervlakte van een tafel willen weten en we meten de lengte en de breedte met hetzelfde meetlint op. Als het meetlint een afwijking heeft waardoor we de lengte te groot opmeten, dan zullen we waarschijnlijk ook de breedte te groot opmeten.
- Doordat er een onderliggende parameter is waar beide gemeten grootheden vanaf hangen.

Hier behandelen we dus alleen ongecorreleerde fouten.

## 2.1 Basisregel

We beginnen met de **algemene regel voor het propageren van ongecorreleerde fouten**. Daarna zullen we laten zien hoe deze regel eruitziet voor eenvoudige relaties. Deze zou je apart kunnen leren, maar je kunt ook altijd de basisregel gebruiken. Het resultaat behoort hetzelfde te zijn. We noteren de onzekerheid op variabele  $x$  in dit hoofdstuk met  $\Delta x$  waar we eerder ook wel  $\sigma_x$  hebben gebruikt.

Als  $q = q(x, y, z, \dots)$  een functie is met meerdere ongecorreleerde variabelen, dan wordt de onzekerheid op  $q$  gegeven door:

$$\Delta q = \sqrt{\left(\frac{\delta q}{\delta x} \Delta x\right)^2 + \left(\frac{\delta q}{\delta y} \Delta y\right)^2 + \left(\frac{\delta q}{\delta z} \Delta z\right)^2 + \dots} \quad (2.1)$$

Hierbij zijn  $\frac{\delta q}{\delta x}$ ,  $\frac{\delta q}{\delta y}$  etc. de partiële afgeleiden van  $q$  naar de betreffende variabele.

We zullen laten zien hoe deze formule werkt aan de hand van een paar voorbeelden.

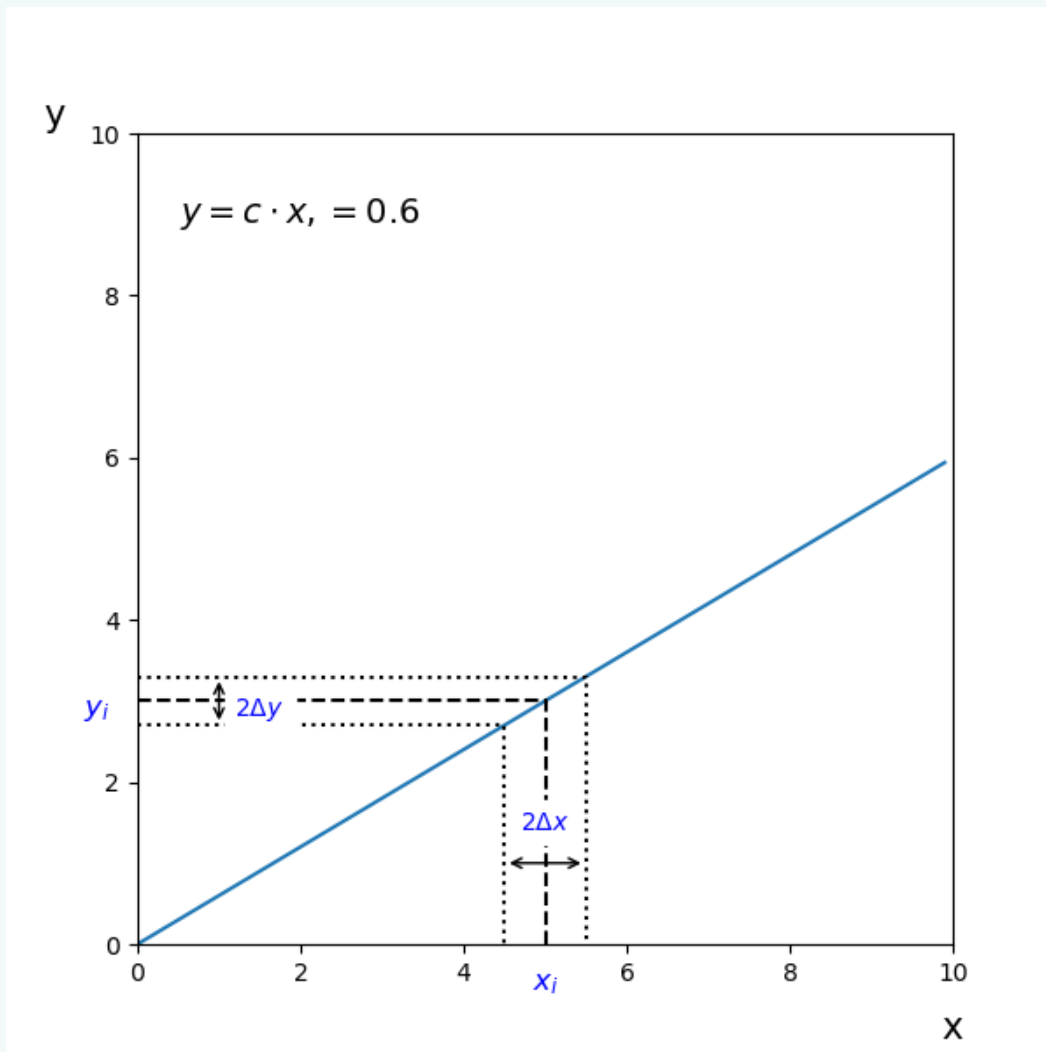


**Voorbeeld 1: Factor**

Stel we hebben een vergelijking  $y = c \cdot x$  met een standaarddeviatie op  $x$  van  $\Delta x$ . Dan is de standaarddeviatie op  $y$ , ( $\Delta y$ ), gelijk aan:

$$\Delta y = \sqrt{\left(\frac{\delta y}{\delta x} \Delta x\right)^2} = c \cdot \Delta x.$$

In dit geval schaalt de onzekerheid op  $x$  ( $\Delta x$ ) dus met dezelfde factor  $c$  tot de onzekerheid op  $y$  ( $\Delta y$ ). In het plaatje hieronder wordt voor een willekeurige waarde  $x_i$  het effect van de propagatie van  $\Delta x$  rond de waarde  $x_i$  naar de fout  $\Delta y$  rond  $y_i$  visueel weergegeven. Je kunt duidelijk zien dat de grootte van  $\Delta y$  veranderd is met de factor  $c$ .

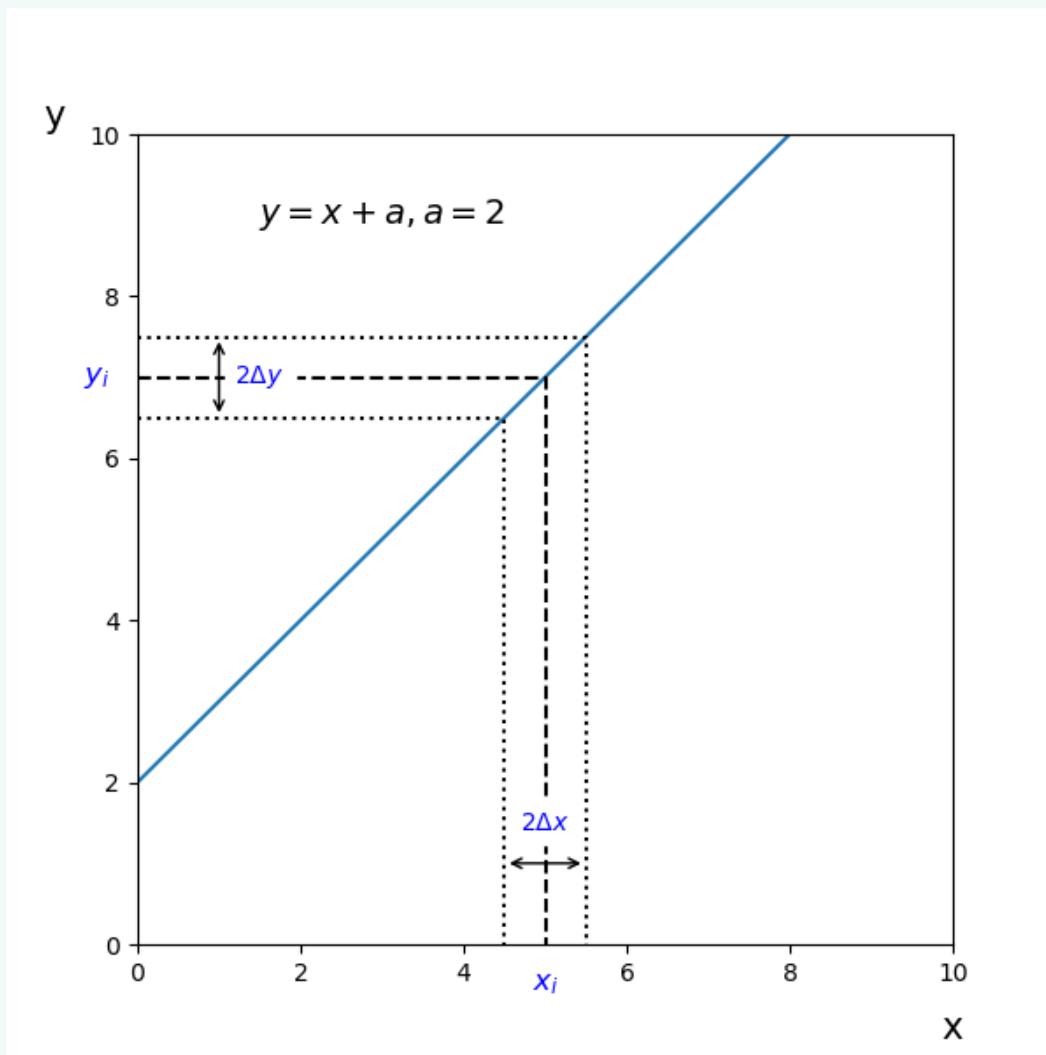


**Voorbeeld 2: Translatie**

Stel we hebben een vergelijking  $y = x + a$  met een standaarddeviatie op  $x$  van  $\Delta x$ . Dan is de standaarddeviatie op  $y$ ,  $(\Delta y)$ , gelijk aan:

$$\Delta y = \sqrt{\left(\frac{\delta y}{\delta x} \Delta x\right)^2} = \Delta x.$$

Wederom geven we het effect van de foutenpropagatie van  $\Delta x$  rond  $x_i$  naar  $\Delta y$  rond  $y_i$  grafisch weer in het plaatje hieronder. Je ziet dat de translatie geen effect heeft op de grootte van de onzekerheid.

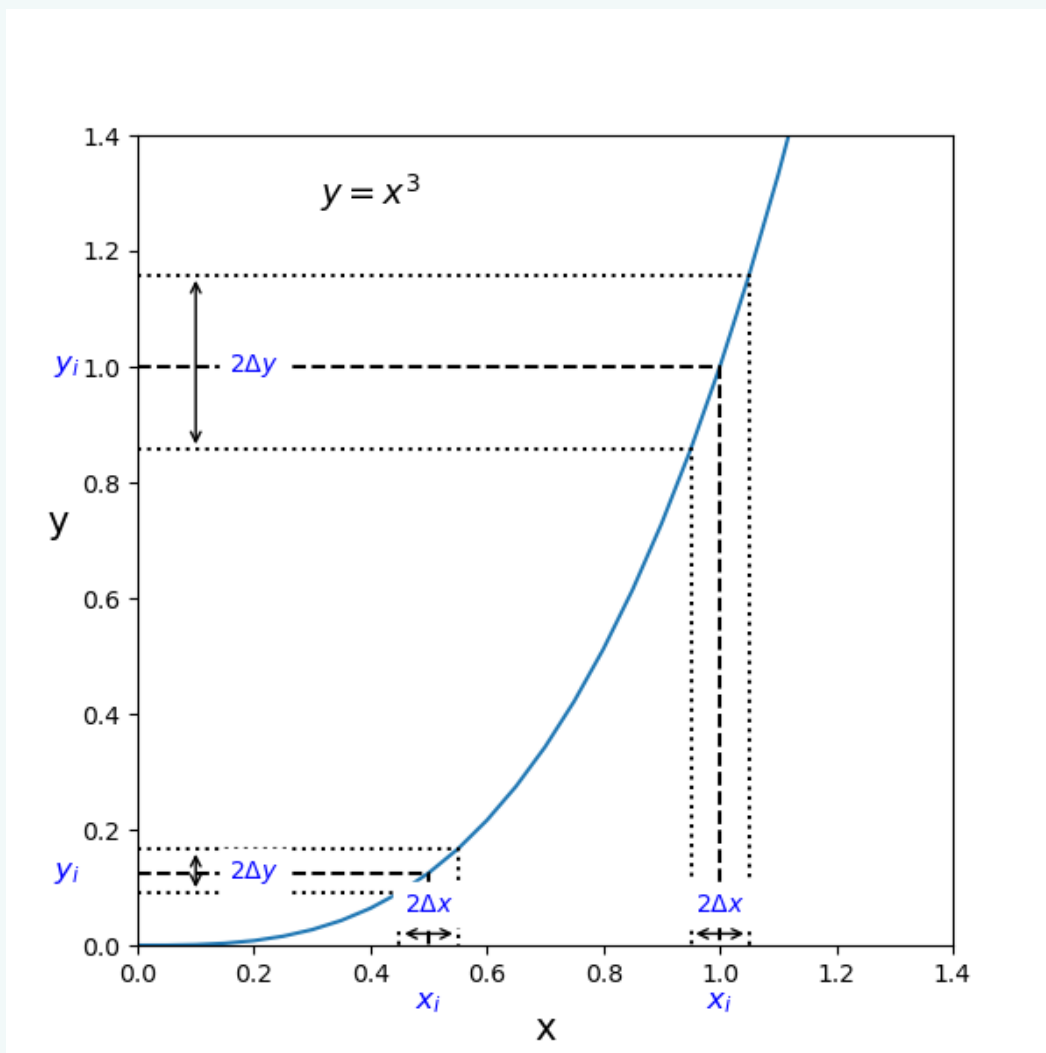


**Voorbeeld 3: Macht**

Stel we hebben een vergelijking  $y = x^3$  met een standaarddeviatie op  $x$  van  $\Delta x$ . Dan is de standaarddeviatie op  $y$ , ( $\Delta y$ ), gelijk aan:

$$\Delta y = \sqrt{\left(\frac{\delta y}{\delta x} \Delta x\right)^2} = 3x^2 \cdot \Delta x.$$

Het effect van de foutenpropagatie volgens deze formule van  $\Delta x$  rond  $x_i$  naar  $\Delta y$  rond  $y_i$  wordt weer grafisch weergegeven in het plaatje hieronder. Je kunt zien dat de mate waarin de grootte van  $\Delta x$  verandert afhangt van de gekozen waarde van  $x_i$ , op sommige plekken is hij kleiner geworden, op andere plekke groter.

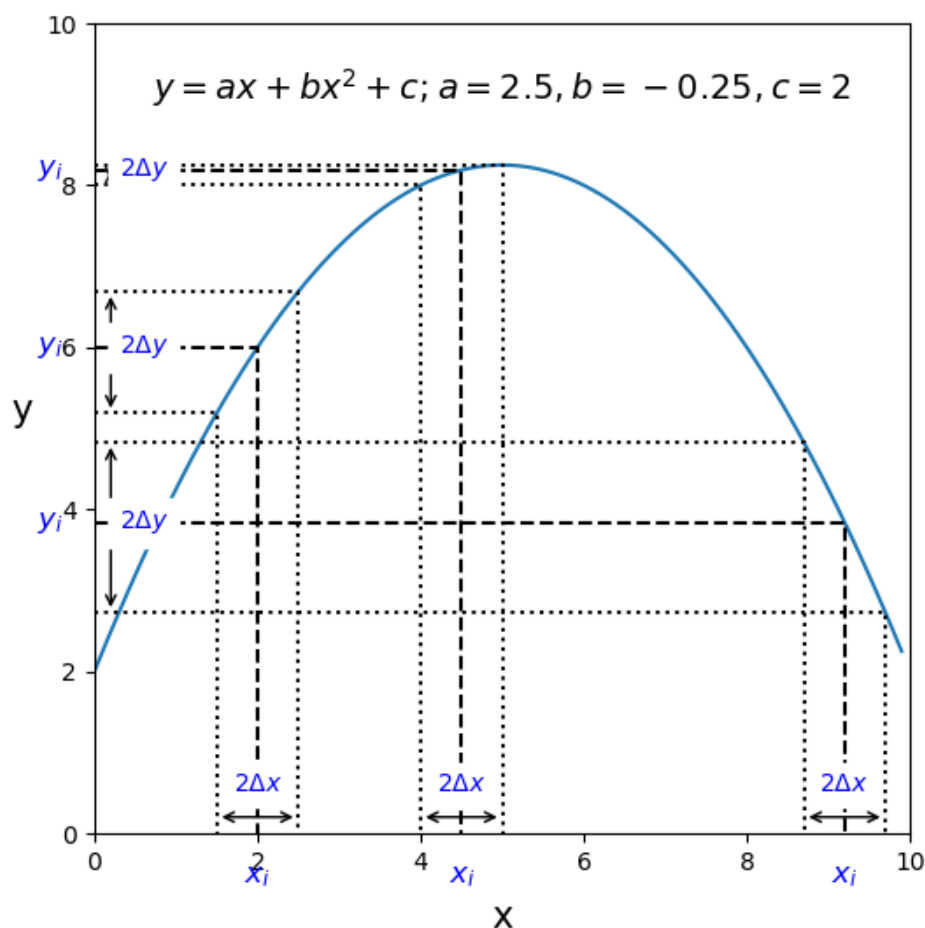


**Voorbeeld 4**

Stel we hebben een vergelijking  $y = ax + bx^2 + c$  met een standaarddeviatie op  $x$  van  $\Delta x$ . Dan is de standaarddeviatie op  $y$ , ( $\Delta y$ ), gelijk aan:

$$\Delta y = \sqrt{\left(\frac{\delta y}{\delta x} \Delta x\right)^2} = (a + 2bx)\Delta x.$$

In het plaatje hieronder geven we nu voor verschillende waarden  $x_i$  de foutenpropagatie van  $\Delta x$  naar  $\Delta y$  de grafische interpretatie. We zien dat het niet alleen de relatieve grootte van  $\Delta y$  afhangt van de gekozen waarde van  $x_i$  maar dat op sommige plaatsen de boven en ondergrens van de onzekerheid zijn geïnverteerd.



**Voorbeeld 5**

Stel we hebben een vergelijking  $z = ax + y^2$  met standaarddeviaties  $\Delta x$  en  $\Delta y$ . Dan is de standaarddeviatie op  $z$ ,  $(\Delta z)$ , gelijk aan:

$$\Delta z = \sqrt{\left(\frac{\delta z}{\delta x} \Delta x\right)^2 + \left(\frac{\delta z}{\delta y} \Delta y\right)^2} = \sqrt{(a\Delta x)^2 + (2y\Delta y)^2}.$$

**Voorbeeld 6**

Stel we hebben een vergelijking  $z = ax + y^2 + 2xy$  met standaarddeviaties  $\Delta x$  en  $\Delta y$ . Dan is de standaarddeviatie op  $z$ ,  $(\Delta z)$ , gelijk aan:

$$\Delta z = \sqrt{\left(\frac{\delta z}{\delta x} \Delta x\right)^2 + \left(\frac{\delta z}{\delta y} \Delta y\right)^2} = \sqrt{((a + 2y) \cdot \Delta x)^2 + ((2y + 2x) \cdot \Delta y)^2}.$$

## 2.2 Som en verschil

De algemene regel kan eenvoudig worden uitgeschreven naar de regel voor som en verschil. Als  $q = x + y$  of  $q = x - y$  dan wordt de onzekerheid op  $q$  gegeven door:

$$\Delta q = \sqrt{\left(\frac{\delta q}{\delta x} \Delta x\right)^2 + \left(\frac{\delta q}{\delta y} \Delta y\right)^2} = \sqrt{(\Delta x)^2 + (\Delta y)^2}. \quad (2.2)$$

We mogen de varianties  $(\Delta x)^2$  en  $(\Delta y)^2$  in het geval van een vergelijking met enkel sommen en/of verschillen dus optellen.

## 2.3 Vermenigvuldigen met constante

Als  $q$  een exacte veelvoud  $c$  is van de gemeten waarde  $x$ , dus  $q = c \cdot x$ , dan geldt:

$$\Delta q = \sqrt{\left(\frac{\delta q}{\delta x} \Delta x\right)^2} = |c| \Delta x. \quad (2.3)$$

De onzekerheid op  $q$  is dus gelijk aan de onzekerheid op  $x$  geschaald met dezelfde factor  $c$ .

## 2.4 Vermenigvuldigen met variabelen

Als  $q$  een vermenigvuldiging is van meerdere variabelen, dus bijvoorbeeld  $q = x \cdot y \cdot z$  dan geldt:

$$\Delta q = \sqrt{\left(\frac{\delta q}{\delta x} \Delta x\right)^2 + \left(\frac{\delta q}{\delta y} \Delta y\right)^2 + \left(\frac{\delta q}{\delta z} \Delta z\right)^2} = \sqrt{\left(\frac{q}{x} \Delta x\right)^2 + \left(\frac{q}{y} \Delta y\right)^2 + \left(\frac{q}{z} \Delta z\right)^2}. \quad (2.4)$$

Dit kan je eenvoudiger schrijven als:

$$\frac{\Delta q}{q} = \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2 + \left(\frac{\Delta z}{z}\right)^2}. \quad (2.5)$$

Ofwel de relatieve fout  $\frac{\Delta q}{q}$  is gelijk aan de kwadratische som van de variabelen.

**Voorbeeld - foutenpropagatie en afronding van de getallen**

Stel dat we de lengte van het blokje hebben gemeten en we lezen de volgende waarde af:

- De lengte ( $l$ ) =  $7.60 \pm 0.10$  cm
- De breedte ( $b$ ) =  $4.10 \pm 0.20$  cm
- De hoogte ( $h$ ) =  $2.00 \pm 0.20$  cm

Het volume van het blokje wordt gegeven door:

$$V = l \cdot b \cdot h = 7.60 \cdot 4.10 \cdot 2.00 = 62.32 \text{ cm}^3$$

We gebruiken de regel dat als  $q = x \cdot y \cdot \dots$  dan:

$$\frac{\Delta q}{|q|} = \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2 + \left(\frac{\Delta z}{z}\right)^2}$$

Dus:

$$\begin{aligned} \frac{\Delta V}{|V|} &= \sqrt{\left(\frac{\Delta l}{l}\right)^2 + \left(\frac{\Delta b}{b}\right)^2 + \left(\frac{\Delta h}{h}\right)^2} \\ &= \sqrt{\left(\frac{0.1}{7.6}\right)^2 + \left(\frac{0.2}{4.1}\right)^2 + \left(\frac{0.2}{2.0}\right)^2} \\ &= 0.01255 \dots \end{aligned}$$

We ronden dit nog niet af, dat doen we pas als we de absolute fout hebben:

$$\begin{aligned} \Delta V &= \frac{\Delta V}{|V|} \cdot |V| \\ &= 0.01255 \dots \cdot 62.32 \\ &= 0.78228 \dots \\ &\approx 0.78 \end{aligned}$$

Het gemeten volume van het blokje is dus  $V = 62.32 \pm 0.78 \text{ cm}^3$