Fysik C/B/A

Elevhæfte

HVERDAGSFYSIK

Naturvidenskabelige undersøgelser

Databehandling

Forfatter: Jens Christian Hansen
Redaktør: Søren Storm
Korrekturlæst og faktatjekket af:
Vibeke Axelsen (Egaa gymnasium)
Kim Vedel Pedersen (Nørre Gymnasium)

Støttet af: novo nordisk fonden

Hej elev!

Så er det tid til databehandling af dataene fra det forsøg I selv har udført. Vi regner med at I allerede har arbejdet med de første to hæfter, der behandler hypoteser og eksperimenter.

Der er hjælp til dig her, hvis du bruger regneark eller Logger Pro.

1. Hjælp til regneark på side 2. Uddybende hjælp til regneark her.
2. Hjælp til Logger Pro i videoen her.

Fremgangsmåde

Start med at åbne dit datasæt i enten regneark eller Logger Pro. Find dit niveau herunder, og følg punkterne.

C niveau

1. Bestem middelværdierne af jeres afhængige variabel
2. Lav en tabel med middelværdierne og jeres uafhængige variabel
3. Lav en graf med (uafhængig variabel, <afhængig variabel>)
4. Undersøg om en regressionskurve passer med modellen fra jeres hypotese. Er forklaringsgraden over 0,95 (R^2 > 0,95), kan du bekræfte din hypotese

B/A niveau

Følg først punkt 1-3 i C niveau.

4. Undersøg om en regressionskurve kan passe med forskriften fra jeres hypotese
5. Lav et residualplot, for at vurdere om data er tilfældigt eller systematisk fordelt omkring din regressionskurve
6. Er residualerne tilfældigt fordelt omkring 0, og er R^2 > 0,95 kan du bekræfte din hypotese

Ekstra for B/A niveau: Beregn residualspredningen. Kommentér fordelingen af residualerne under antagelse af at de er normalfordelte

Hjælp til punkt 1 og 2

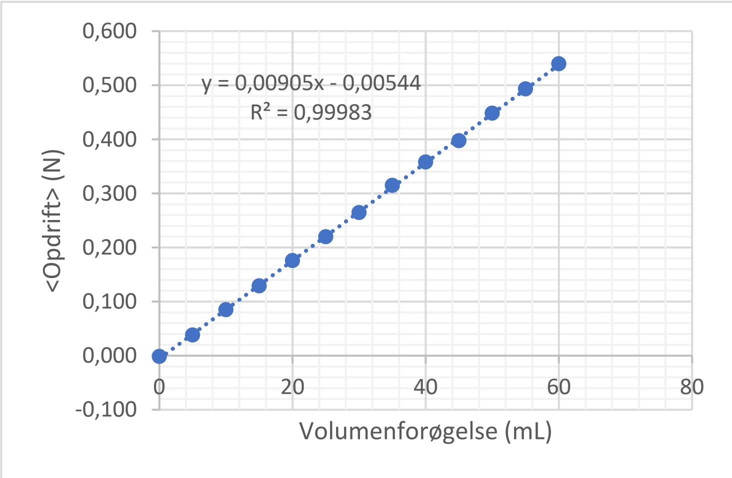
Her har vi bestemt middelværdierne for den nulstillede opdrift og 6 volumenændringer. For hver <Opdrift> er der midlet over 50 målinger. Det fulde datasæt kan hentes her.

Volumenforøgelse (mL)	0	10	20	30	40	50	60
<Opdrift> (N)	-0,002	0,085	0,176	0,264	0,358	0,448	0,539

I regnearket beregnede vi middelværdien med =MIDDEL().

Hjælp til punkt 3 og 4

Som eksempel undersøger vi hypotesen: ”*Opdriften er ligefrem proportional med volumenet, ved konstant masse*”. Det gør vi ved at lave en lineær regression på volumenforøgelserne og middelværdierne af opdriften.

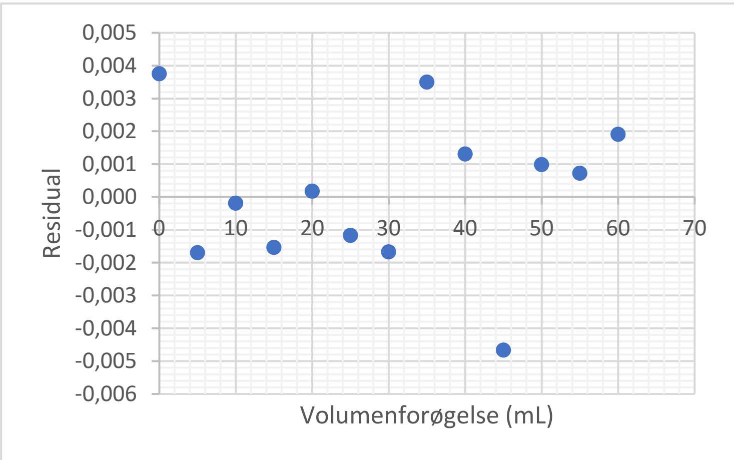


Regressionslinjen er her bestemt i et regneark. I videoen om databehandling, forklarer vi hvordan den bestemmes i Logger Pro. Modellen har en flot forklaringsgrad (R^2) på 0,999.

Bemærk at alle datapunkter fra 0-60 mL indgår. I forskriften for regressionslinjen er konstanten -0,00544 N, hvilket svarer til 0,6 g vand eller ~0,6 mL vand, hvilket ligger indenfor den usikkerhed der er på volumenændringerne.

Hjælp til punkt 5+6

Residualplottet her er lavet efter vejledningen på side 5 i dette hæfte.



Residualplottet afslører ikke nogen systematik, der gør at vi skal revurdere modellen i punkt 4. Så da $R^2 > 0{,}95$ kan vi bekræfte vores hypotese.

Databehandling

Nu skal du bruge data fra videoen, som du kan hente her. Dataene du skal arbejde med, stammer fra hypotesen:

"Opdriften er omvendt proportional med legemets masse, for samme volumen"

Der er hjælp til dig her, hvis du bruger regneark eller Logger Pro.

1. Hjælp til regneark på side 2. Uddybende hjælp til regneark her.
2. Hjælp til Logger Pro i videoen her.

Opgave 1 (C/B/A niveau)

A. Bestem middelværdierne af F_{op} (herfra skriver vi $< F_{\mathrm{op}} >$) for hver af de 6 forsøgsrækker
B. Hvorfor er det en fordel at bruge middelværdien fra en serie målinger som målepunkt, i stedet for en enkelt måling?

Opgave 2 (B/A niveau)

A. Bestem standardafvigelsen for hver af de 6 de forsøgsrækker. Bruger du Logger Pro har du allerede fået resultatet i opgave 1A
B. Hvad er standardafvigelsen et mål for?

Opgave 3 (C/B/A niveau)

A. Lav to nye kolonner "*m* (g)" og "<*F*op> (N)" i regnearket og indtast værdierne for massen og de tilhørende værdier af $< F_{op} >$
B. Lav et $(m, < F_{op} >)$-plot

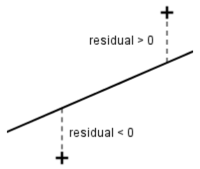
Opgave 4 (C/B/A niveau)

Se plottet fra opgave 3.

A. Hvordan burde målepunkterne fordele sig, hvis $< F_{op} >$ ikke afhænger af m?
B. Er du enig i at hypotesen fra videoen skal afkræftes?

Repetition af statistik

Herunder en oversigt over de statistikbegreber der bruges i de tre videoer.

Begreber	Beskrivelse
Middelværdi	Middelværdien eller gennemsnittet m af datasættet $x_1, x_2, \ldots . x_n$ $m = \frac{1}{n} \cdot \sum_{i=1}^{i=n} x_i$
Standardafvigelse	Standardafvigelsen s for datasættet $x_1, x_2, \ldots . x_n$ er et mål for usikkerheden på enkeltmålingerne. $s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{i=n} (x_i - m)^2} \quad ; \quad n \geq 2$ Gentages én måling, vil det forventes at den med en sandsynlighed på 68% hhv. 95% ligger i intervallerne $[m - s; m + s]$ hhv. $[m - 2s; m + 2s]$.
Standardfejl (SE)	Standardfejlen s_m er et mål for usikkerheden på middelværdien $s_m = \frac{s}{\sqrt{n}}$ Gentages en måleserie, vil middelværdien m for den nye serie med med en sandsynlighed på 68% hhv. 95% ligge i intervallerne $[m - s_m; m + s_m]$ hhv. $[m - 2s_m; m + 2s_m]$. Et måleresultat angives derfor som $m \pm s_m$ eller $m \pm 2s_m$
Lineær regression	Regression hvor den forventede sammenhæng mellem variablerne er lineær; dvs af formen $y = ax + b$.
Forklaringsgraden/R^2	Et relativt mål for, hvor tæt datasættets punkter i gennemsnit ligger omkring regressionslinjen. Hvis alle punkter ligger på regressionslinjen, så er $R^2 = 1$, hvis der ikke er nogen sammenhæng er $R^2 = 0$
Residual	Den lodrette afstand – regnet med fortegn - mellem målepunktet og regressionslinjen kaldes et residual 
Residualspredning (RMSE)	Et mål for hvor tæt målepunkterne ligger på regressionslinjen $s_r = \sqrt{\frac{1}{n-2} \cdot \sum_{i=1}^{i=n} residual_i^2}\,; n \geq 3$ hvor n er antallet af målepunkter og $residual_i$ er residualet svarende til den i'te måling. Ca. 68% af residualerne bør ligge i intervallet $[-s_r; s_r]$ og 95% i intervallet $[-2s_r; 2s_r]$