

Introduktion til

DataMeter

Bjørn Felsager

Variabelsammenhænge
med DataMeter



Forlag Malling Beck
Læhegnet 71
2620 Albertslund



Titel: *Introduktion til DataMeter*

Samhørende titel: *DataMeter – den danske udgave af "Fathom"*

Forfatter: *Bjørn Felsager*

Forlagsredaktion: *Frants Toft*

© Forlag Malling Beck A/S, Bjørn Felsager 2006

1. udgave, 1. oplag

Nærværende version af manuskriptet kan frit gøres til genstand for kopiering til undervisningsformål. Al anden mekanisk, fotografisk, elektronisk eller andre former for gengivelser af denne bog eller dele heraf er her ud over kun tilladt efter Copy-Dans regler.

Nærværende introduktion er tænkt som elevmateriale.

Den bør læses med programmet tændt, så man kan arbejde eksemplerne igennem og få en god førstehåndsforfølelse for hvordan programmet virker.

Den bør læses i flere omgange efterhånden som klassen når frem til de forskellige emner.

Emnerne er af forskellig sværhedsgrad, så det er ikke alle eksempler, der egner sig til alle studieretninger.

Den nærværende version er læst igennem af Morten Birk Christensen (Oure Idrætsgymnasium) og Hans Vestergaard (Frederikshavn Gymnasium), der er kommet med mange nyttige kommentarer og forslag til ændringer. Hvad der resterer af u hensigtsmæssigheder og fejl er alene den uopmærksomme forfatters ansvar!

Hæftet kan downloades som en pdf-fil fra programmets hjemmeside:

www.datameter.dk

Illustrationen side 36 er gengivet efter en eksamensopgave i matematik: Opg. 6b, hn99aab, Undervisningsministeriet

Illustrationen side 40 er gengivet efter en eksamensopgave i fysik: Opg. 3, 2006-7-1, Undervisningsministeriet

Illustrationen side 65 er gengivet efter en eksamensopgave i matematik: Opg. 3, hf9985, Undervisningsministeriet

Illustrationen side 67 er gengivet efter en eksamensopgave i matematik: Opg. 3, 2006-8-2ST.MED, Undervisningsministeriet

Indhold

Forord.....	4
1 Indføring i DataMeter.....	5
1.1 Grundlæggende DataMeter-teknikker.....	5
Teknisk bemærkning: Kategorilister.....	13
Teknisk bemærkning: Datasættet som udstillingsvindue.....	20
1.2 Fremskaffelse af data.....	22
Øvelse 1.2.1 Jupiters måner.....	25
Øvelse 1.2.2 Tilfældige rektangler.....	26
2 Variabelsammenhænge.....	27
2.1 Lineære sammenhænge.....	27
Teknisk bemærkning: Funktioner for lineær regression.....	31
2.2 Simple proportionaliteter.....	33
Ligefrem proportionalitet.....	33
Teknisk bemærkning: Nulpunktsfejl.....	35
Omvendt proportionalitet.....	36
Omvendt kvadratrodsproportionalitet.....	41
2.3 Fibonaccital og eksponentiel vækst.....	44
Teknisk bemærkning: Den rekursive formel for fibonaccital.....	45
Øvelse 2.4.1 Kvadratet på fibonaccital.....	47
Øvelse 2.4.2 Den direkte formel for fibonaccital.....	47
2.4 Variabelkontrol.....	48
Øvelse 2.4.1 Isbjørnespillet.....	51
Øvelse 2.4.2 Flyde-synke-spillet.....	52
Øvelse 2.4.3 Variabelkontrol med andengradspolynomier.....	55
3 Regressionsmodeller.....	61
3.1 Lineær regression.....	61
3.2 Potensregression.....	65
Teknisk bemærkning: Formlen for a.....	67
Teknisk bemærkning: Om enheder og potensproportionalitet.....	70
3.3 Eksponentiel regression.....	71
Teknisk bemærkning: Formlen for a.....	72
Teknisk bemærkning: Om enheder og eksponentiel vækst.....	75
3.4 Dynamisk parametertilpasning.....	76
Øvelse 3.4.1 Lysintensitetens afstandsafhængighed.....	79
Øvelse 3.4.2 Lydens fart i luft.....	79
Indeks.....	80

Forord

DataMeter er et dynamisk databehandlingsprogram, der især egner sig til at arbejde med databehandling i gymnasiets undervisning. Denne introduktion til **DataMeter** er skrevet til brug for undervisningen i matematik på det indledende niveau i gymnasiet, dvs. Mat C. Men der lægges samtidigt op til samarbejde med især de naturvidenskabelige fag, herunder det naturvidenskabelige grundforløb på STX. Dels er mange af eksemplerne hentet fra de naturvidenskabelige fag, dels er der konsekvent brugt enheder. Brugen af enheder er et af matematiks naturlige bidrag til et sådant samarbejde, ligesom de naturvidenskabelige fag fx kan være leveringsdygtige i eksperimentelle data. Et samarbejde med samfundsfag er også ønskværdigt. Her vil statistik være et oplagt bud, men det er *ikke* emnet for denne introduktion, der koncentrerer sig om variabelsammenhænge.

I det første kapitel 'Indføring i **Datameter**' introduceres de grundlæggende værktøjer til at arbejde med tabeller og grafer gennem et større eksempel omhandlende planetsystemet, der munder ud i en undersøgelse af Keplers tredje lov. Dernæst følger et afsnit om hvordan man kan fremskaffe data ved at kopiere dem, ved at trække websider ind i **DataMeter** og ved at generere dem med formler.

I andet kapitel 'Variabelsammenhænge' arbejdes der først med lineære sammenhænge, herunder hvordan man dels kan arbejde med frie linjer, dels med mindste kvadraters linje. Dernæst vises det først, hvordan man kan håndtere simple proportionalitetsmodeller, og derefter, hvordan man kan håndtere eksponentiel vækstmodeller. Der afsluttes med en længere diskussion af variabelkontrol, der også omfatter konfundering og en kort omtale af Simpsons paradoks.

I tredje kapitel 'Regressionsmodeller' arbejdes der mere dybtgående med de simple regressionsmodeller – lineær, potens og eksponentiel. Det sker med henblik på dels at opnå en bedre forståelse af mindste kvadraters metode – inklusive værdien af at inddrage residualplot –, dels at opnå fortrolighed med en generel metode til regression, der kan bruges på alle modeller med to parametre. Specielt bygger potens- og eksponentiel regression altså *ikke* på logaritmiske transformationer. Det er en klar fordel på Mat C niveauet, hvor eleverne sjældent opnår den fornødne fortrolighed med logaritmer. Der afsluttes med nogle mere varierede anvendelser af mindste kvadraters metode til bl.a. dynamisk parametertilpasning af systematiske fejl.

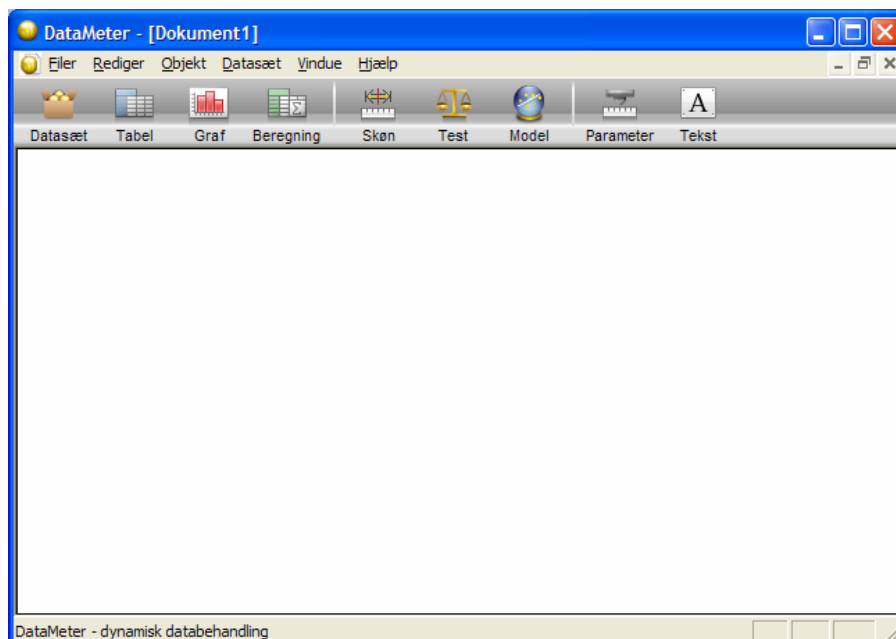
Undervejs i teksten er der nogle bokse kaldet 'Tekniske bemærkninger', der nok mest er af interesse for de naturvidenskabelige studieretninger. De er lidt mere krævende og kan overspringes uden tab af sammenhæng. Endelig kan man finde mange gode øvelser til at arbejde videre med stoffet blandt undervisningsministeriets eksamensopgaver i især matematik, men også fysik har gode tilbud.

Bjørn Felsager

1. Indføring i DataMeter

1.1 Grundlæggende DataMeter-teknikker

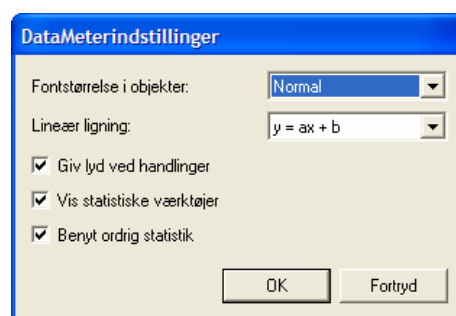
Når du åbner for **DataMeter** viser der sig følgende skærbillede



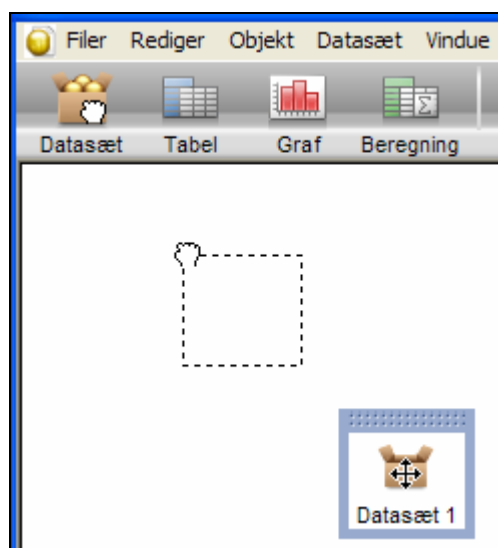
Det er især værktøjslinjen, der er interessant



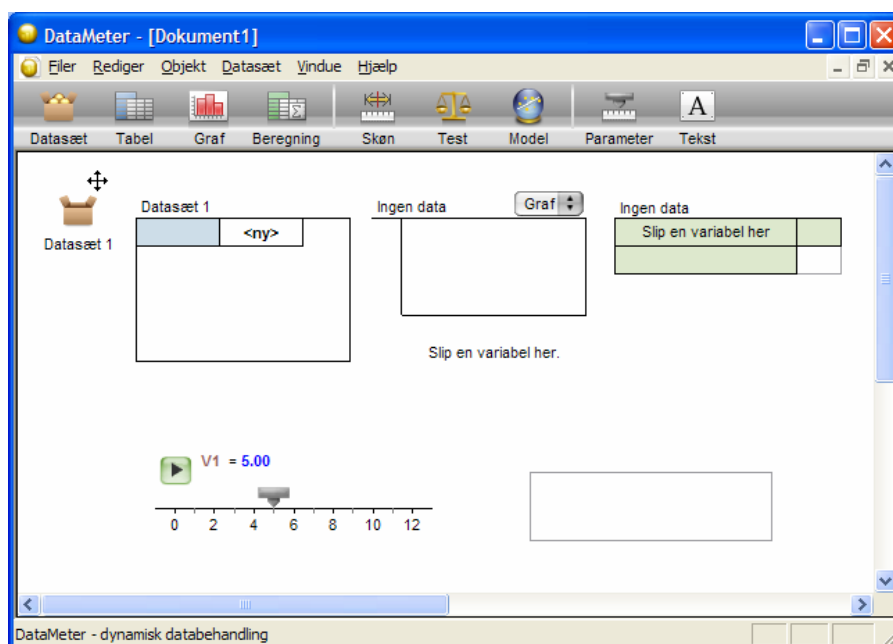
På den finder du tre rum: I det første er der simple værktøjer til at arbejde med tabeller, grafer og beregninger. I det sidste er der tilsvarende simple værktøjer til at arbejde med parametre og tekstbokse. Men i det midterste rum findes der avancerede værktøjer til statistik. Dem vil vi slet ikke komme ind på i dette hæfte. Faktisk kan man fjerne dem helt under menupunktet **Indstillinger** i **Rediger**-menuen



Men tilbage til de simple værktøjer: Dem kan vi hente ind i dokumentet ved at trække dem ned med markøren – der forvandles til en gripende hånd, når vi klikker på et værktøj – og slippe dem præcis, hvor det passer os. Her har vi fx – vist i en kombination af flere billeder – trukket et datasæt ind i dokumentet

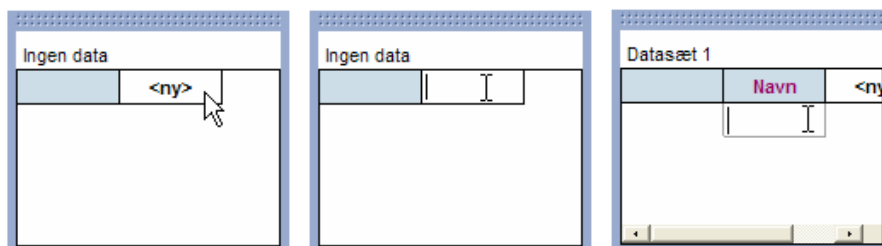


Men præcis den samme teknik virker med alle de andre værktøjer



De to første – **Datasæt** og **Tabel** – er uløseligt forbundne og bærer den samme titel, mens de resterende – **Graf**, **Beregning**, **Parameter** og **Tekst** – kan bruges for sig selv uafhængigt af et datasæt.

Men lad os komme i gang med et eksempel: Kernen i **DataMeter** er dets unikke evne til at håndtere variable, så lad os gøre nogle observationer og knytte variable til dem. Det kunne være om *klassen*, hvor vi kunne se på datasættet bestående af de enkelte elever karakteriseret ved forskellige egenskaber, såsom navn, køn, alder, højde osv. Det kan I imidlertid selv lege med. Her tager vi som et andet eksempel *planetsystemet* i vores solsystem. Det består af 9 planeter¹, som vi passende kan skrive ind i en tabel. Vi trækker derfor et tabelværktøj ned i dokumentet og starter med at give et navn til den første variabel – planetens navn – ved at klikke på titlen <ny> og skrive løs

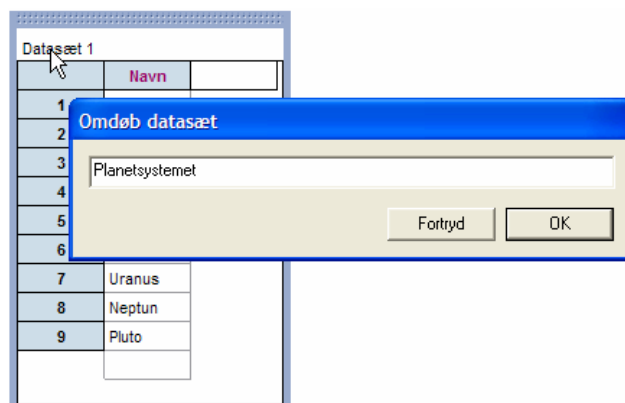


Straks begynder der at ske noget: Dels dukker der en celle op lige under titlen på variabelen, dels får tabellen selv en titel – Datasæt 1 – og der dukker et datasæt op i form af en tom skattekiste



Derefter skriver du bare løs i de tomme celler der dukker op hver gang en celle er udfyldt og trykker **ENTER**. Så snart de første data i form af navne på planeterne dukker op i tabellen, kommer der guld-kugler i skattekisten.

Til sidst kan du også gå ind og dobbeltklikke på tabeltitlen og ændre navnet på datasættet til fx Planetsystemet



¹ Pluto mistede i august 2006 efter en dramatisk afstemning i den internationale astronomiske union sin planetstatus og bør egentlig klades en dværgplanet.

Men planeter er jo karakteriseret ved andet end deres navne, så vi fortsætter med at indtaste andre typiske egenskaber for planeterne fx deres type, dvs. om det er sten-, is- eller gasplaneter, om de har måner, deres størrelse i form af deres radius, deres masse, deres afstand til Solen og deres omløbstid. De første variable er nemme nok at indtaste, da de bare består af tekst

Planetsystemet

	Navn	Type	Måner	<ny>
enhed				
1	Merkur	Sten	Nej	
2	Venus	Sten	Nej	
3	Jorden	Sten	Ja	
4	Mars	Sten	Ja	
5	Jupiter	Gas	Ja	
6	Saturn	Gas	Ja	
7	Uranus	Gas	Ja	
8	Neptun	Gas	Ja	
9	Pluto	Is	Ja	

Sådanne **tekstvariable** kaldes også for **kategoriserede variable**, fordi de opdeler planeterne i forskellige kategorier, fx sten-, gas- og isplaneter. Tekstvariable er venstrestillede, ligesom i et almindeligt regneark. I det hele taget minder tabellen i sin struktur meget om et almindeligt regneark med nummererede rækker, idet hver planet har fået tildelt sit eget **indeks**, der angiver dets plads i tabellen. Læg mærke til at søjlerne, dvs. de enkelte variable, alle er navngivne. De er altså ikke født med bogstavnavne – **A**, **B**, **C** osv. – som i et almindeligt regneark. Og det er ikke nemt at referere til en enkelt celle. I **DataMeter** – som er et dynamisk regneark – arbejder man med hele søjler/lister ad gangen, sådan som du måske også kender det fra din grafregner.

Vi er så kommet til **talvariablene**: planeternes størrelse, masse osv. De kaldes også for **numeriske variable**. Men her skal vi nu være opmærksom på at der ikke blot er tale om talværdier, men også om **enheder**, idet de fleste størrelser er forsynet med en naturlig enhed. Da **DataMeter** er konstrueret til at arbejde med enheder er der ingen grund til at se bort fra dette. **DataMeter** kender de fleste grundlæggende enheder såsom meter, sekund, kilogram osv., men når der kommer præfix på, kilometer, millisekund osv. så er det ikke altid de er indbyggede på forhånd. Der kan man prøve sig lidt frem og om nødvendigt erstatte præfikset med en passende tierpotens. Det er også vigtigt at gøre sig klart at tallene skal skrives ind som tal, men gerne med brug af videnskabelig notation. Jordens masse er fx givet ved $5.976 \cdot 10^{24}$ kg, men det skal indskrives som 5.976e24 kg, da det ellers vil blive opfattet som et regnestykke (dvs. som en tekst) og ikke som et tal.

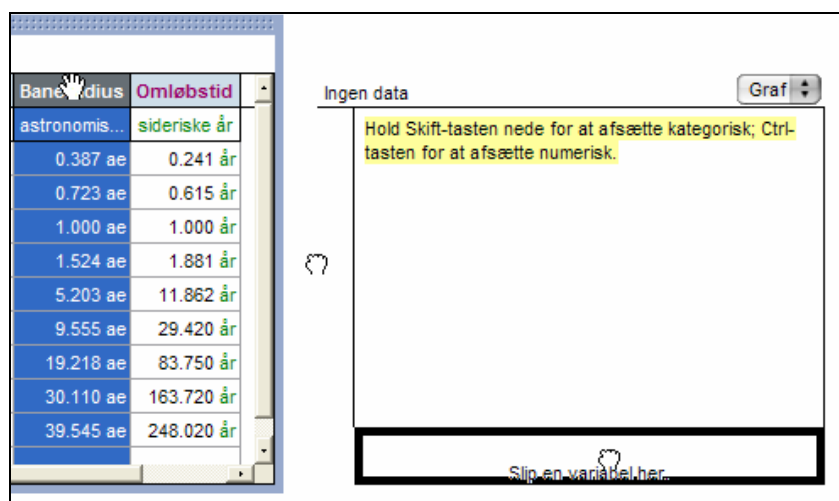
1.1 Indføring i DataMeter: Grundlæggende DataMeter-teknikker

Planetsystemet

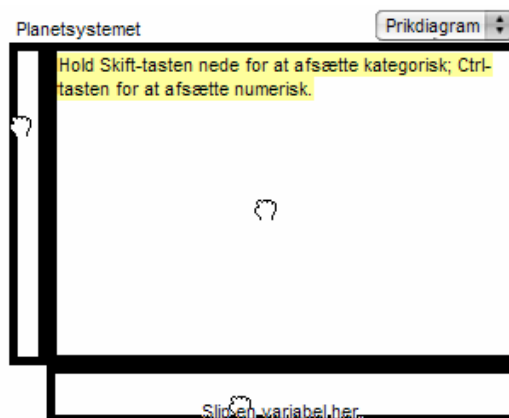
	Navn	Type	Måner	Radius	Masse	Baneradius	Omløbstid
enhed				kilometer	kilogram	astronomis...	sideriske år
1	Merkur	Sten	Nej	2439 km	3.287e+23 kg	0.387 ae	0.241 år
2	Venus	Sten	Nej	6052 km	4.870e+24 kg	0.723 ae	0.615 år
3	Jorden	Sten	Ja	6378 km	5.976e+24 kg	1.000 ae	1.000 år
4	Mars	Sten	Ja	3397 km	6.394e+23 kg	1.524 ae	1.881 år
5	Jupiter	Gas	Ja	71492 km	1.899e+27 kg	5.203 ae	11.862 år
6	Saturn	Gas	Ja	60268 km	5.687e+26 kg	9.555 ae	29.420 år
7	Uranus	Gas	Ja	25559 km	8.665e+25 kg	19.218 ae	83.750 år
8	Neptun	Gas	Ja	24764 km	1.028e+26 kg	30.110 ae	163.720 år
9	Pluto	Is	Ja	1151 km	1.494e+22 kg	39.545 ae	248.020 år

Læg mærke til rækken med enheder, der dukker op helt af sig selv, når man skriver størrelser op med enheder. Læg også mærke til, at man kun behøver forsyne det første tal med en enhed. Derefter indsættes enheden af sig selv ned gennem søjlen. Læg endeligt mærke til de to sidste enheder, idet programmet også kan håndtere astronomiske enheder. Inden for Solsystemet kan man med fordel bruge Jordens baneradius som afstandsenhed. Den kaldes den astronomiske enhed *ae*. Tilsvarende kan man med fordel bruge Jordens omløbstid som tidsenhed. Den kaldes det *sideriske år*.

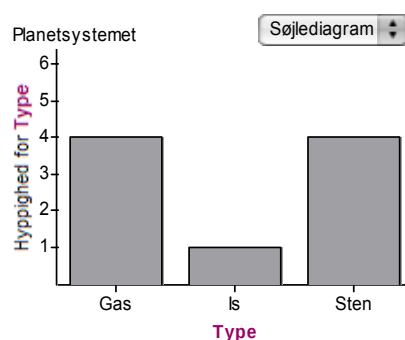
Dermed har vi fundet en masse variable vi kan lege med. Det vil være praktisk at kunne afbilde disse variable grafisk. Vi trækker derfor et grafværktøj ned fra værktøjshylden



Vi trækker nu de variable ind fra tabellen – ved som vist at gribe deres navn – og slipper dem i grafrummet. Der er tre steder vi kan slippe dem: Langs førsteaksen (som vist) som en uafhængig variabel, langs andenaksen som en afhængig variabel og inde i selv grafrummet, som en variabelsignatur (dvs. en kodning af datapunkterne i grafrummet som afspejler værdien af variabelen)

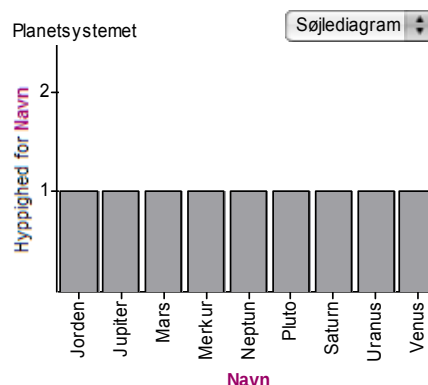


Lad os forsøge med nogle grafer: Først kan vi trække *tekstvariablen* **Type** ind på førsteaksen. Vi får da et *søjlediagram* med optælling af hyppigheden for gas- henholdsvis is- og stenplaneter. Der er altså 4 gasplaneter, 1 isplanet og 4 stenplaneter. Et sådant søjlediagram over hyppighederne er standardgrafen for en kategoriseret variabel.



tæl ()

Lad os i stedet prøve at trække variabelen **Navn** ind på førsteaksen. I første omgang bliver man nok lidt skuffet, for vi får som standard et søjlediagram over hyppighederne, dvs. et diagram der viser at der netop er én planet svarende til hvert navn. Men læg mærke til funktionen **tæl()** i nederste venstre hjørne. Det er den der afgør hvad søjlerne står for, i dette tilfælde altså en simpel sammentællning. Men denne funktion kan vi skifte ud ved at

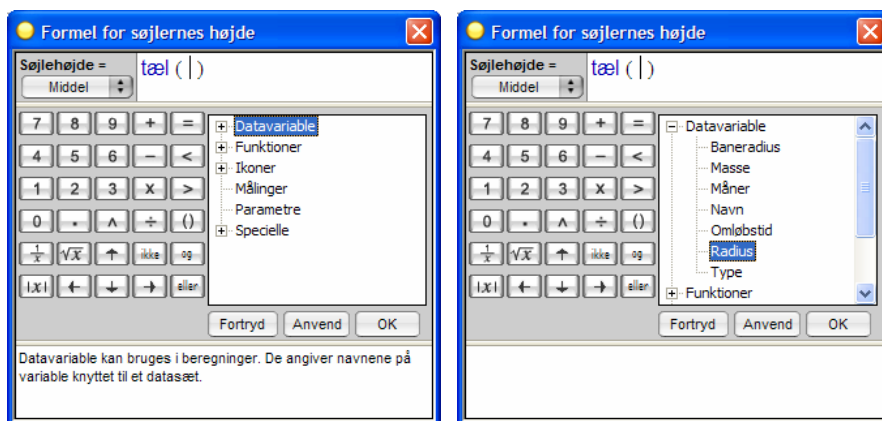


tæl ()

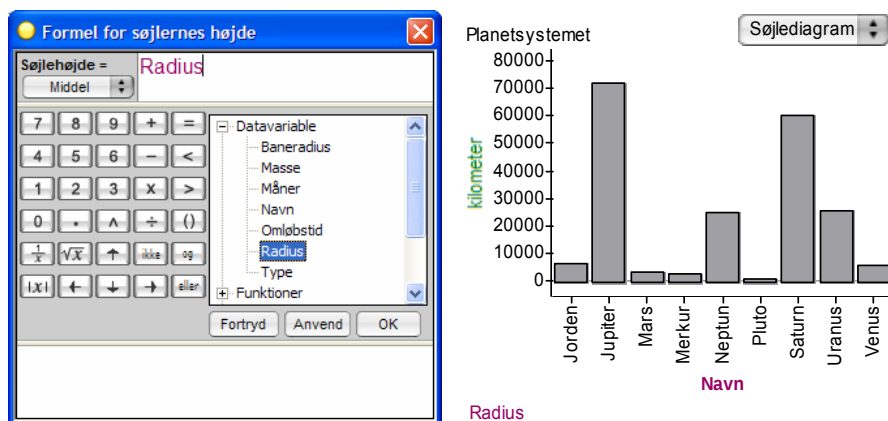
klikke på den hvorved vi får adgang til **DataMeters** lommeregner (se næste side). Med undtagelse af tekstværktøjet, der kun kan bruges til indskrivning af tekster og formler, men ikke til at foretage beregninger, er alle de øvrige værktøjer tilknyttet denne lommeregner, der kraftigt forøger vores muligheder for at arbejde med dataene. Den er som vist forsynet med en stifinder til indbyggede funktioner mm.

Datavariabel – Funktioner – Ikoner – Målinger – Parametre – Specielle

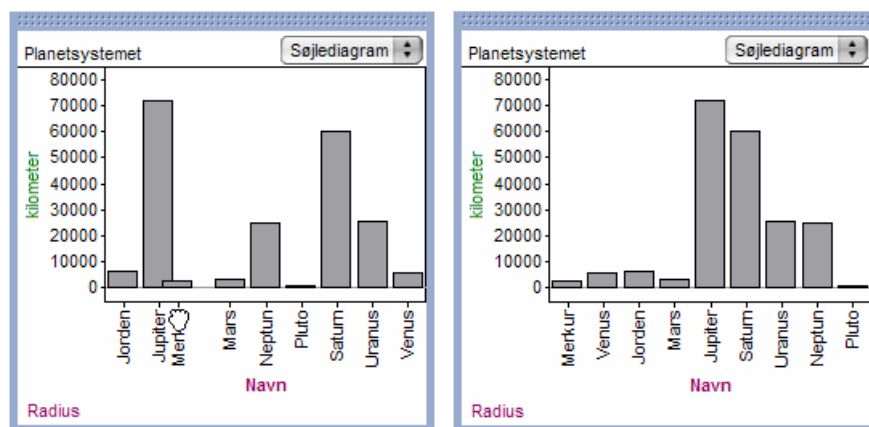
Her vil vi se nærmere på datavariabel, der giver adgang til alle de datavariabel, der er tilknyttet datasættet



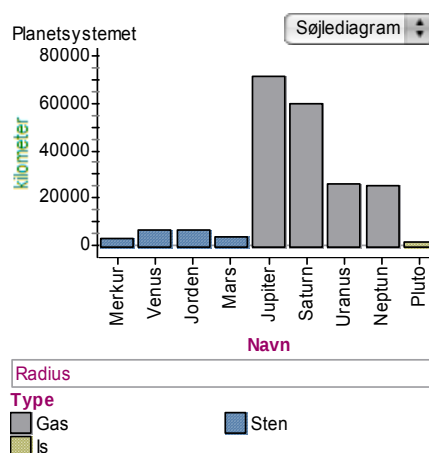
Hvis vi dobbeltklikker på datavariablen **Radius** får vi i stedet afsat en søjlehøjde svarende til planetens radius



Læg mærke til at rækkefølgen af planeterne er alfabetisk i stedet for den rækkefølge de blev indskrevet med i tabellen, dvs. efter voksende afstand fra Solen. Men det kan vi rette op på ved at gribe planeterne navne med musen og flytte rundt på dem som det passer os

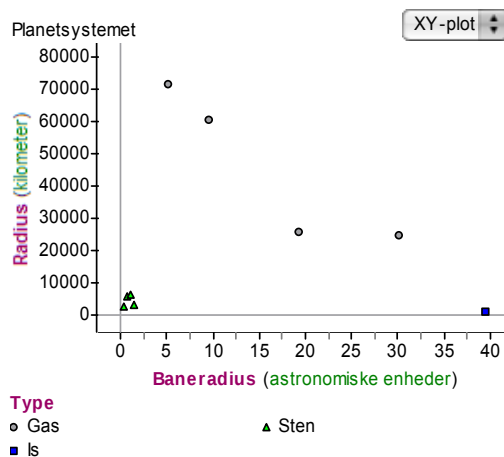


Vi kan endda trække variabelen **Type** ind i grafrummet for at få afsat den som signatur på søjlerne



Vi ser da at de fire inderste planeter er små stenplaneter, derefter kommer der fire store gasplaneter, og endelig har der til sidst sneget sig en enkelt isplanet ind i systemet. Men Plutos rolle som en planet er faktisk omstridt efter man har fundet flere 'isplaneter' længere ude, hvoraf nogen endda er større end Pluto. Pluto har dog nok vundet historisk hævd på at være en planet, så vi lader den stå!

Vi går så over til at se på grafer for *numeriske variable*. Her kan vi trække **Baneradius** ind som den grundlæggende variabel på førsteaksen. Den fortæller hvor langt væk fra Solen planeten befinder sig og afspejler den naturlige rækkefølge for planeterne. Vi kan så trække **Radius** ind på andenaksen, som afspejler planetens størrelse og endelig trække **Type** ind i grafrummet som en *variabelsignatur*



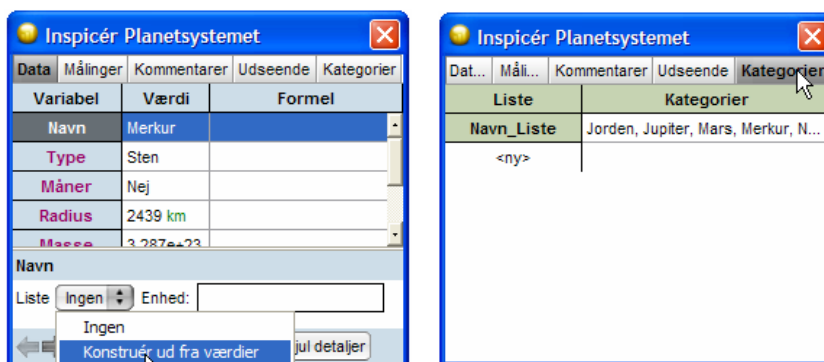
Den kan passende sammenlignes med søjlediagrammet foroven, idet den rummer den samme information. Men denne gang er datapunkterne afsat i et rigtigt koordinatsystem som et *XY-plot*.

Teknisk bemærkning: Kategorilister

Når man arbejder med kategoriserede variable kan det godt virke forvirrende at de afsættes alfabetisk i stedet for efter en naturlig fast rækkefølge. Man kan imidlertid gennemtvinge en fast rækkefølge ved at indføre en *kategoriliste* i datasættet. Vi viser teknikken her i detaljer. Ved at dobbeltklikke på datasættet kan man som vist vække inspektøren for datasættet



Vi får da adgang til de såkaldte kartotekskort for de enkelte planeter og kan som vist udpege variabelen **Navn** og klikke på **Vis detaljer**

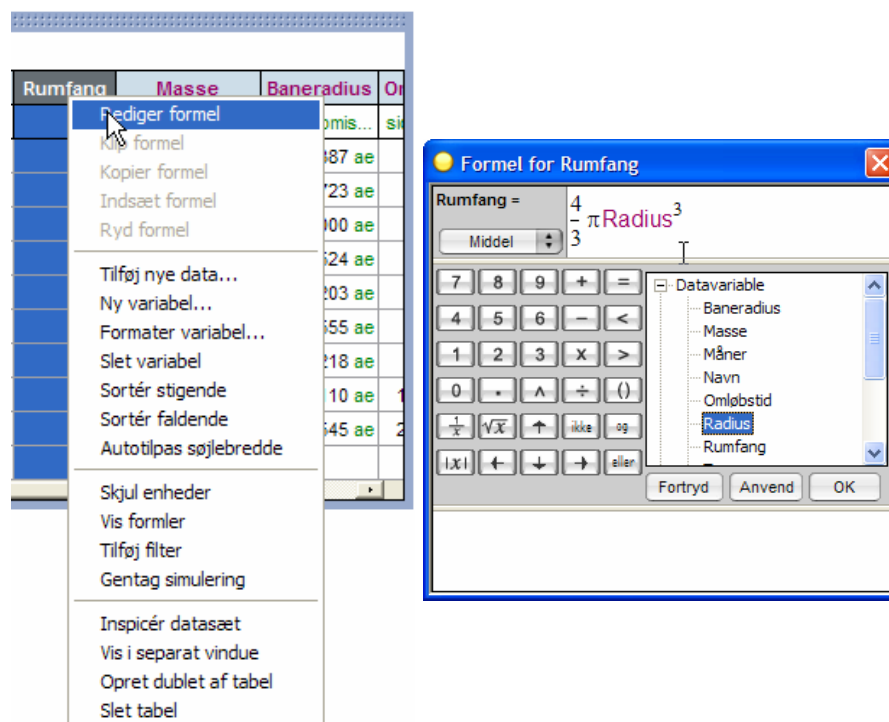


Vi kan da vælge **Konstruér ud fra værdier**, hvorefter **DataMeter** som vist under fanebladet **Kategorier** opretter en alfabetisk liste over kategorier. Men pointen er nu at denne kategoriliste **Navn_liste** kan man gå ind og rette i efter forgodtbefindende, herunder tilføje flere kategorier, hvis man senere regner med at udvide tabellen og ønsker at tage højde for fremtidige kategorier. Har man først oprettet en sådan kategoriliste vil programmet respektere den i alle sammenhænge, hvor man grafisk eller i tabeller får vist oversigter hvor den pågældende variabel indgår. Sådanne kategorilister er et meget stærkt værktøj i større undersøgelser.

Med adgang til diverse data for planeterne kan vi nu begynde at undersøge diverse sammenhænge. Her vil vi først se lidt nærmere på diverse simple *sammensatte variable*. De fremkommer af de grundlæggende variable ved at man ganger og dividerer. Fx kan man konstruere en variabel for rumfanget ud fra planetens radius, idet vi antager at planeten er kugleformet og benytter formlen for kuglens rumfang

$$V = \frac{4}{3} \pi \cdot r^3.$$

Vi opretter derfor en ny variabel med navnet **Rumfang** og højreklikker på den for at få adgang til menupunktet **Rediger formel**, der åbner lommeregneren for programmet



Men med både **Masse** og **Rumfang** til rådighed kan vi så på tilsvarende måde beregne planetens (gennemsnitlige) densitet, dvs. dens massefylde som den følgende simple sammensatte variabel

$$\text{Densitet} = \frac{\text{Masse}}{\text{Rumfang}}.$$

Læg mærke til at formlerne ikke automatisk bliver vist i tabellen (se figur næste side). Her skal man altså aktivt vælge menupunktet **Vis formler** fra **Tabel**-menuen. Tilsvarende skal højden af formelfeltet tilpasses manuelt for at få det hele vist. Det sker ved som vist at trække i formelfeltets nederste kant. I modsætning hertil kan man automatisk få justeret bredden af variabelsøjlen ved at dobbeltklikke på navnefeltets højre kant (hvilket er standard i regneark)

Planetsystemet				
	Radius	Rumfang	Masse	Densitet
enhed	kilometer	km ³	kilogram	kg/km ³
=		$\frac{4}{3}\pi \text{Radius}^3$		$\frac{\text{Masse}}{\text{Rumfang}}$
1	2439 km	6.07749e+10 km ³	3.287e+23 kg	5.40816e+12 kg/km ³
2	6052 km	9.28507e+11 km ³	4.870e+24 kg	5.24545e+12 kg/km ³
3	6378 km	1.08678e+12 km ³	5.976e+24 kg	5.49881e+12 kg/km ³
4	3397 km	1.64201e+11 km ³	6.394e+23 kg	3.89421e+12 kg/km ³
5	71492 km	1.5306e+15 km ³	1.899e+27 kg	1.24093e+12 kg/km ³
6	60268 km	9.16957e+14 km ³	5.687e+26 kg	6.20171e+11 kg/km ³
7	25559 km	6.99391e+13 km ³	8.665e+25 kg	1.23896e+12 kg/km ³
8	24764 km	6.36137e+13 km ³	1.028e+26 kg	1.61618e+12 kg/km ³
9	1151 km	6.38726e+09 km ³	1.494e+22 kg	2.33903e+12 kg/km ³

Her kan man indvende at enheden kg/km³ ikke er den mest gennemskuelige enhed for densitet, men det kan man ændre på efter for-godtbefindende ved at indtaste en anden enhed i enhedsfeltet. Det eneste man skal passe på er betegnelsen for gram, da g også står for tyngdeaccelerationen, og gram derfor som en af de meget få undtagelser afkortes til *gm* i **DataMeter**. Når man vil bruge gram i en enhed er det derfor sikrest at skrive den helt ud fx som gram/cm³

Planetsystemet				
	Radius	Rumfang	Masse	Densitet
enhed	kilometer	km ³	kilogram	gm/cm ³
=		$\frac{4}{3}\pi \text{Radius}^3$		$\frac{\text{Masse}}{\text{Rumfang}}$
1	2439 km	6.07749e+10 km ³	3.287e+23 kg	5.41 gm/cm ³
2	6052 km	9.28507e+11 km ³	4.870e+24 kg	5.25 gm/cm ³
3	6378 km	1.08678e+12 km ³	5.976e+24 kg	5.50 gm/cm ³

Formater variabel

☐ Dynamisk
 ☒ Fast antal decimaler
 ☐ Videnskabeligt format
 ☐ Ingeniørformat

2 Antal decimaler

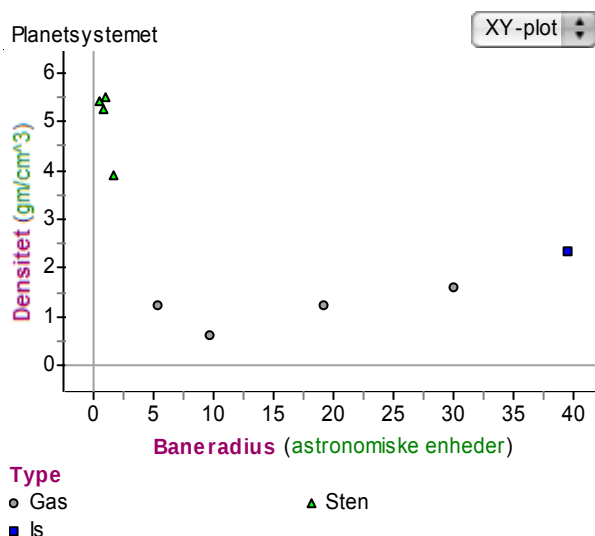
Eksempel
12345.68

Bemærkning: Formateringen har kun indflydelse på hvordan tallene vises. Der afrundes ikke i beregninger eller hvor dataene afsættes i et grafrum.

OK Fortryd

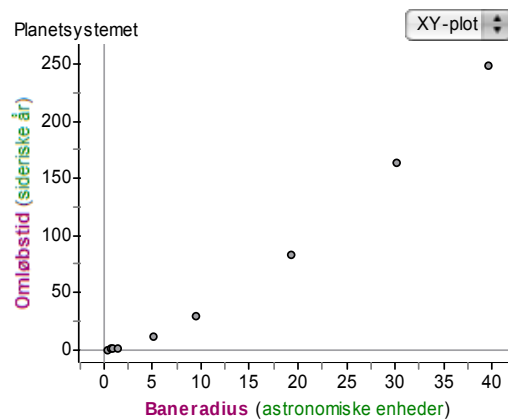
Som vist har vi også formateret den variable til kun at vise 2 decimaler. Som altid får man adgang til menupunktet ved at højreklikke på variablen.

Densitet kan nu trækkes ind som den afhængige variabel på andenaksen med planetens baneradius på førsteaksen og planetens type i grafrummet.



De fire indre stenplaneter har altså en relativt høj massefylde som er 4-5 gange så stor som vands massefylde, mens de fire gasplaneter har en relativ lav massefylde (sammenlignelig med vands) og endelig har Pluto en massefylde, der er ca. dobbelt så stor som vands. Alt dette antyder noget om hvordan planetsystemet er dannet i sin tid ud af den roterende sky af gas og sten og støv, der omgav Solen i tiden lige efter solsystemets dannelse. Men faktisk forstår vi ikke særligt godt planetsystemets dannelse, så det er et meget aktivt forskningsområde, hvor man især prøver at trække informationer ud af de ny-opdagede planetsystemer omkring andre stjerner i mælkevejen.

Som et andet eksempel på en variabelsammenhæng i planetsystemet vil vi se på sammenhængen mellem **Baneradius** og **Omløbstid**. Det ses tydeligt at omløbstiden vokser jo længere væk fra Solen vi kommer, samt at omløbstiden vokser hurtigere end baneradien idet grafen krummer opad. Men hvor meget hurtigere vokser den?



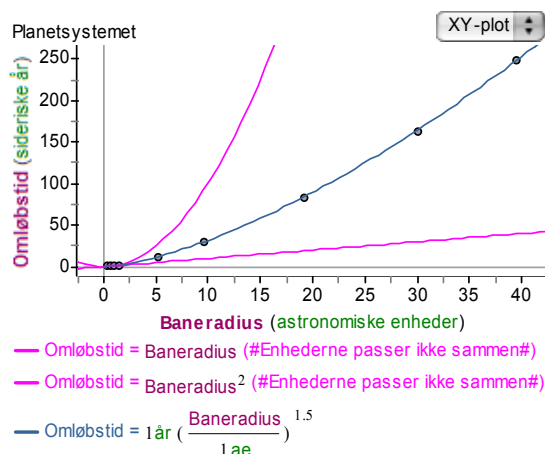
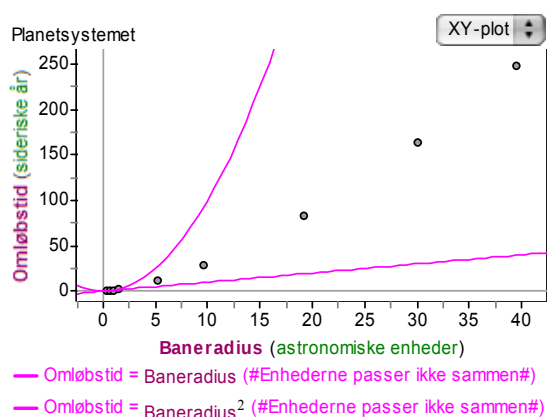
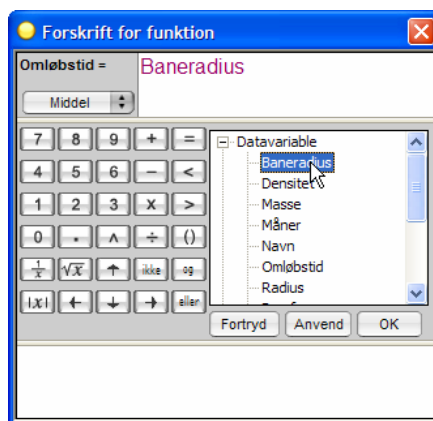
Vi kan fx prøve at tilføje de følgende grafer

- 1) Omløbstiden vokser lige så hurtigt som baneradien
- 2) Omløbstiden vokser med kvadratet på baneradien

Det sker ved at vælge menupunktet **Plot funktion** i grafrummet, hvorved vi endnu engang får adgang til den indbyggede lommeregner og kan indskrive ligningerne for de givne variablers sammenhænge. De tilhørende grafer får så en bemærkning med på vejen om at de ikke passer sammen med enhederne. Det ville vi godt kunne reparere på, men lad os ignorere det i denne omgang, da det er nogle andre pointer vi vil trække frem.

Vi bemærker da, at omløbstiden ikke blot tydeligt vokser hurtigere end en ligefrem proportionalitet, men den vokser også tydeligt langsommere end en kvadratisk proportionalitet.

Det gør det nærliggende at forsøge sig med en ligning midt imellem, dvs. vi forsøger os i stedet med potensen 1.5. Og lad os for fuldstændighedens skyld tilføje passende enheder i ligningen, så vi undgår de 'sure opstød' fra **DataMeter**



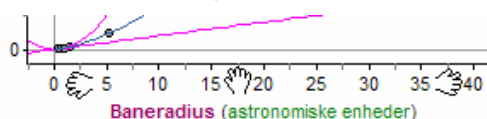
Den er ramt lige på kornet: Vi har genopdaget Keplers tredje lov!

Ganske vist kunne vi måske være lidt i tvivl om hvorvidt sammenhængen også gælder for de indre planeter, der ligger meget tæt sammen nede ved koordinatsystemets begyndelsespunkt. Men det kan vi efterprøve ved at zoome ind på dette område. Der er mange måder hvormed vi kan regulere vinduesgrænserne for et koordinat-udsnit

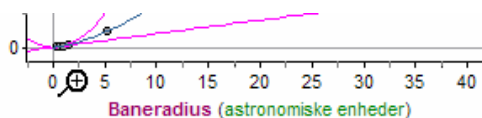
- 1) Vi kan dobbeltklikke i grafrummet, hvorved vi vækker *graf-
rummets inspektør* og får adgang til at skrive vinduesgrænserne ind direkte

Inspicér graf hørende til Planetsystemet	
Data	Egenskaber
Parameter	Værdi
punktstørrelse	5
xNedre	-2.5
xØvre	42.5
yNedre	-10
yØvre	270
xModsatskala	falsk
yModsatskala	falsk
xAutotilpas	sand
yAutotilpas	sand

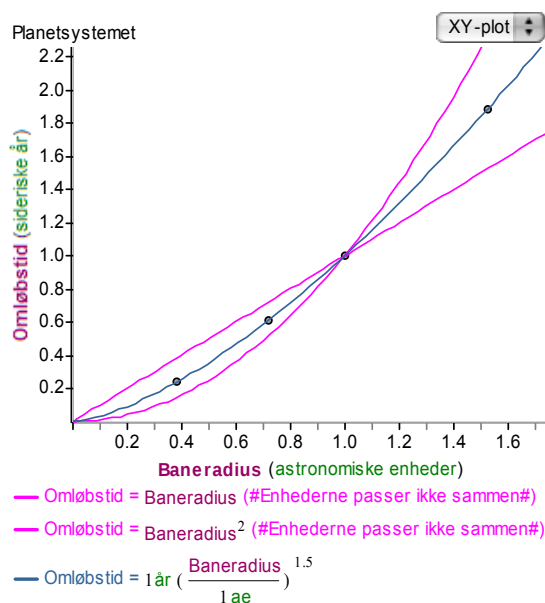
- 2) Vi kan gribe fat i tallene på koordinataksene med musen og trække dem til siderne, når vi griber fat i en af siderne, eller flytte hele akse, når vi griber fat i akse midtpå



- 3) Vi kan zoome ind (eller ud) ud fra et punkt på akse ved at holde **CTRL**-tasten nede (eller **SKIFT-CTRL**-tasten nede) samtidigt med at vi klikker på akse, hvorved markøren skifter til et forstørrelsesglas med et plustegn (eller et minustegn)



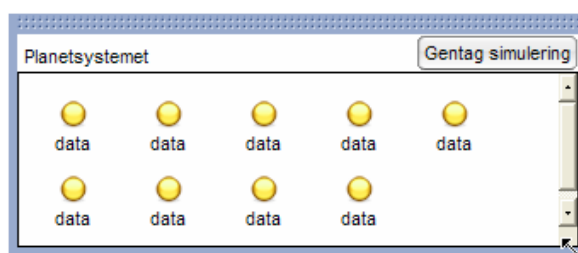
Alle metoderne er meget fleksible så man kan roligt eksperimentere med dem, men hvis man gerne vil kunne vende tilbage til det originale grafrum kan det betale sig at arbejde på en *dublet* i stedet for. Alle værktøjerne kan dubleres, så man kan fortsætte ud fra en skabelon, der allerede har mange af de egenskaber man ønsker sig. Det kan ske ved at taste **SKIFT CTRL D** eller ved at højreklikke på værktøjet og vælge menupunktet **Opret dublet af ...**. I vores tilfælde zoomer vi ind på både førsteaksen og andenaksen og trækker derved de indre planeter frem, hvilket kun bekræfter at vi har fundet en meget overbevisende variablsammenhæng



Læg mærke til at alle graferne netop går gennem enhedspunktet (1,1) i første kvadrant. Der ligger Jorden netop i kraft af vores snedige valg af enheder.

Den fundne sammenhæng mellem baneradius og omløbstid gælder faktisk for alle planetligningerne systemer, fx også for Jupiters måner, der kredser omkring Jupiter, eller de menneskeskabte satellitter, der kredser omkring Jorden.

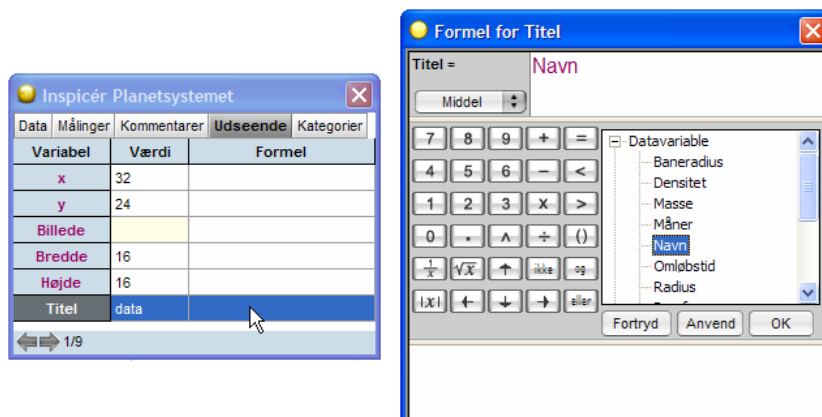
Vi har nu set på de to vigtigste værktøjer i **DataMeter**: **Tabel** og **Graf**. Samtidigt er der dukket et **Datasæt** op helt af sig selv i form af en skattekasse med guldkugler og tilhørende kartotekskort, der rummer alle de indtastede oplysninger. Det er i datasættet, at dataene ligger beskyttet. Sletter man datasættet forsvinder alt. Sletter man en tabel eller en graf sker der derimod ingenting med dataene. Men kan blot trække en ny tabel ned og knytte forbindelsen til datasættet ved enten at have markeret datasættet på forhånd eller ved at trække titlen for datasættet ind i tabellen. Straks dukker alle dataene op igen. Man kan også åbne et datasæt ved at trække i et hjørne



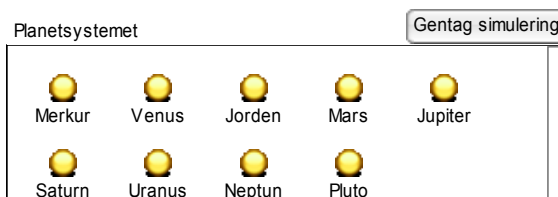
Men man kan godt blive lidt skuffet for man finder bare ni guldkugler: Én for hver af de ni planeter.

Teknisk bemærkning: Datasættet som udstillingsvindue

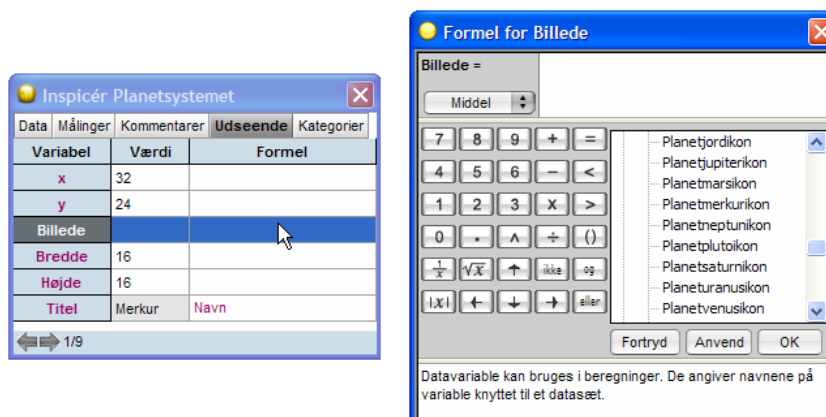
Har man mod på det kan man også bruge datasættet til nogle raffinerede grafiske fremstillinger. Hertil dobbeltklikker vi i datasættet og vækker inspektøren. Men denne gang åbner vi fanebladet **Udseende**



Der dukker da som vist mulighed for at vælge placering, billedikon, størrelse og titel for 'gulduglen'. Det kan alt sammen styres af formelfeltet. Ved at dobbeltklikke i formelfeltet åbnes lommeregneren og her kan vi som det første skifte parameteren **Titel** ud med en passende variabel, fx **Navn**, der netop angiver planetens navn

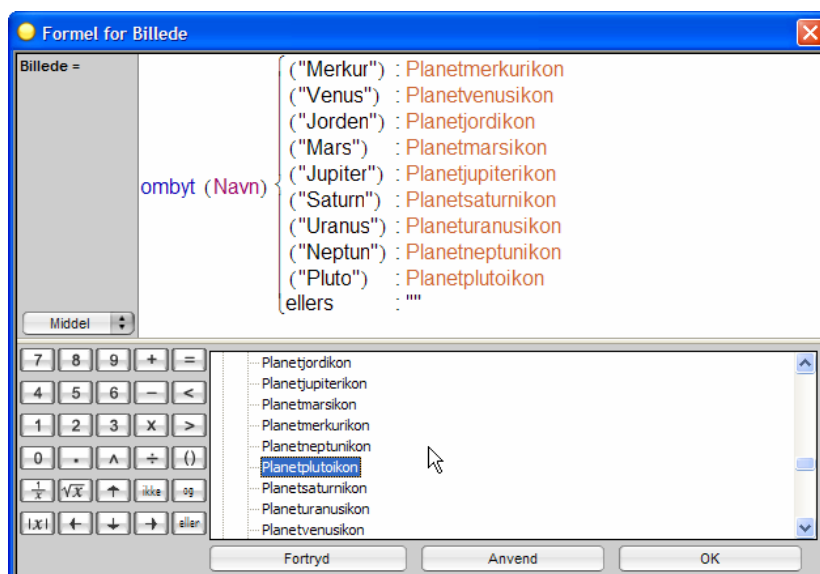


Herefter klikker vi på formelfeltet for **Billede**. Det åbner igen lommeregneren, hvor vi denne gang kigger på afdelingen for ikoner

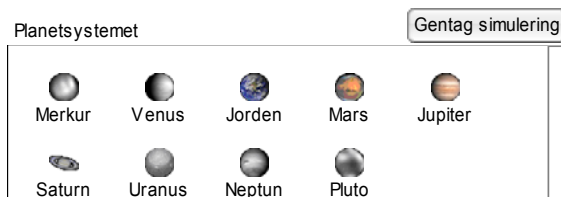


Minsandten om **DataMeter** ikke har planetikonerne indbygget. Men for at kunne bruge dem skal de indbygges i en formel, hvor pla-

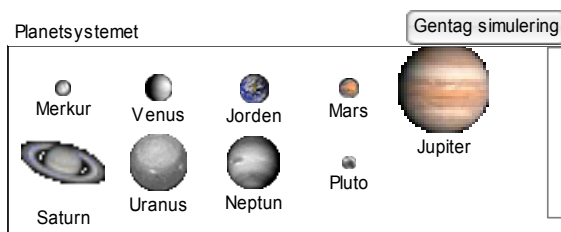
neterne én for én tildeles deres planetikon. Det sker nemmest ved hjælp af den fleksible **ombyt**-funktion, der ombytter planetens navn med det tilhørende ikon. Først skal man dog frembringe det rigtige antal tilfælde ved gentagne tryk på **INS**-tasten, mens markøren pe- ger på en af forgreningerne for denne stykvise funktion



Har man ikke prøvet det før kan man godt få lyst til at løbe skrigende væk, men man vænner sig hurtigt til **ombyt()**-funktionen, der er en afgørende styrke ved **DataMeter** i mange forskellige sammenhænge.



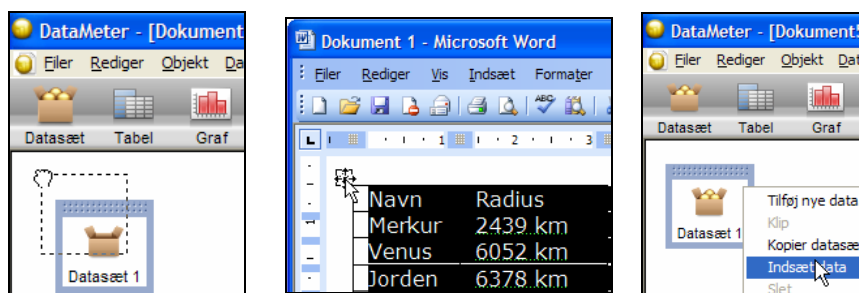
Her dukker ikonerne op i samme størrelse. Men det kunne være rart at få dem vist i et mere realistisk indbyrdes størrelsesforhold. Det kan ske ved at regulere **Bredde** og **Højde**, som regner i pixelstørrelser, ved hjælp af en formel, der fx kan bygge på **Radius**. Her har vi brugt $16\sqrt{\text{Radius} / \text{jordradius}}$, med *jordradius* som indbygget naturkonstant



1.2 Fremskaffelse af data

I det foregående tastede vi simpelthen vore data direkte ind i en tabel. Men med mange data kan det godt blive uoverskueligt. Vi kan da i stedet **kopiere** dataene fra en tabel i fx et regneark.

- 1) Først åbner man et nyt **DataMeter** dokument og trækker et tomt datasæt ned i dokumentet (en tom skattekiste). Data-sættet markeres.
- 2) Derefter markerer man og kopierer den ønskede tabel (fx med **CTRL-C**).
- 3) Endelig indsættes tabellen i det nye **DataMeter** dokument (fx med **CTRL-V** eller menupunktet **Indsæt data** fra **Rediger**-menuen), hvorved skattekisten fyldes op med guldugler. Man kan se dataene ved efterfølgende at trække en tabel ned.



Der er to ting man skal være speciel opmærksom på: Man *skal* huske at oprette et tomt datasæt i sit **DataMeter** dokument først. Ellers får man kun mulighed for at indsætte tabellen som et billede. Og så skal man være opmærksom på at tabellens dataformat *skal* være kompatibelt med dataformatet i **DataMeter**. Det gælder selvfølgelig detaljer som at **DataMeter** arbejder med søjler. Men det gælder også talformatet: Hvis fx tabellen bruger decimalpunktum skal **DataMeter** også arbejde med decimalpunktum. Ellers opfatter den ikke dataene som tal. Her følger **DataMeter** simpelthen Windows opsætning af talformatet med henblik på decimalkilletegnet, ciffergruppeseparatoren (tusindtal-separatoren) og listeseparatoren (fx mellem to koordinater i et koordinatsæt). Hvis der er uoverensstemmelse mellem tabellens dataformat og **DataMeters** dataformat må man derfor enten – hvis det er muligt – reparere på tabellens dataformat (ved fx at udskifte decimalpunktum med decimalkomma), eller man må gå ind i kontrolpanelet under **Internationale og sproglige indstillinger** og skifte dataformatet for tal. Lyder dette som sort tale, skal man snakke med sin datavejleder. Hvis man skifter Windowsindstillingerne skal man huske at genstarte **DataMeter** for at ændringerne også bliver registreret af **DataMeter**.

NB! Her i bogen arbejdes der konsekvent med **decimalpunktum** som decimalkilletegn og med **semikolon** som listeseparator.

Vi fortsætter med at se på, hvordan man henter data fra **Internet-tet**. Som et konkret eksempel vil vi øve os på Jupiters måner. Jupiter udgør et minisolsystem med fire store måner opdaget af Galilei. Der findes efemerider for disse Jupitermåner, dvs. tabeller over deres positioner. Her vil vi bruge hjemmesiden

http://pds-rings.seti.org/tools/ephem2_jup.html

Skulle den af en eller anden grund ikke være tilgængelig kan du finde tilsvarende sider ved fx at søge på 'JPL ephemeris Jupiter moons' (hvor JPL er en forkortelse for Jet Propulsion Laboratory, der er ansvarlig for efemerideberegningerne for NASA)

Jupiter Ephemeris Generator 2.2

Click [here](#) for help, or click on individual highlighted items.

Time Limits

Start time: 2006-07-13
 Stop time: 2006-08-13
 Interval: 4 hours

Ephemeris Options

- ☐ Pre-RPX #1 (JUP071 + JUP066 + JUP067)
1996-Jan-01 to 1999-Dec-31
- ☐ Pre-RPX #2 (JUP100 + JUP104)
1973-Nov-01 to 2023-Nov-01 (Galilean satellites)
1996-Sep-04 to 2002-Jan-02 (small satellites)
- ☐ Post-Galileo #1 (JUP100 + JUP120 + DE405)
1973-Nov-01 to 2023-Nov-01 (Galilean satellites)
1996-Jan-01 to 2010-Jan-01 (small satellites)
- ☒ Post-Galileo #2 (JUP166 + JUP120)
2000-Jan-01 to 2023-Oct-31 (Galilean satellites)
1996-Jan-01 to 2010-Jan-01 (small satellites)

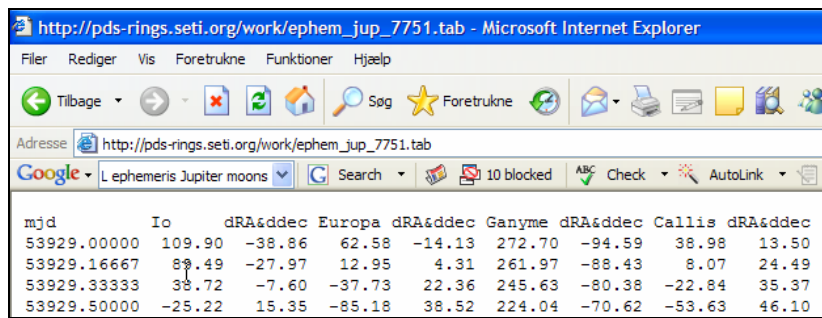
Vi skal selv udfylde **Time Limits**. Som vist vælger vi at se data for en hel måned med 4 timers mellemrum. Derudover skal vi vælge de fire Jupitermåner med *forskudte* data i forhold til Jupiter

Moon Columns:

- ☐ RA (hours) & Dec (deg)
- ☒ Offset RA & Dec from Jupiter (arcsec)
- ☐ Sub-solar latitude & rotating longitude (deg)
- ☐ Sub-Earth latitude & rotating longitude (deg)
- ☒ Io (J1)
- ☒ Europa (J2)
- ☒ Ganymede (J3)
- ☒ Callisto (J4)

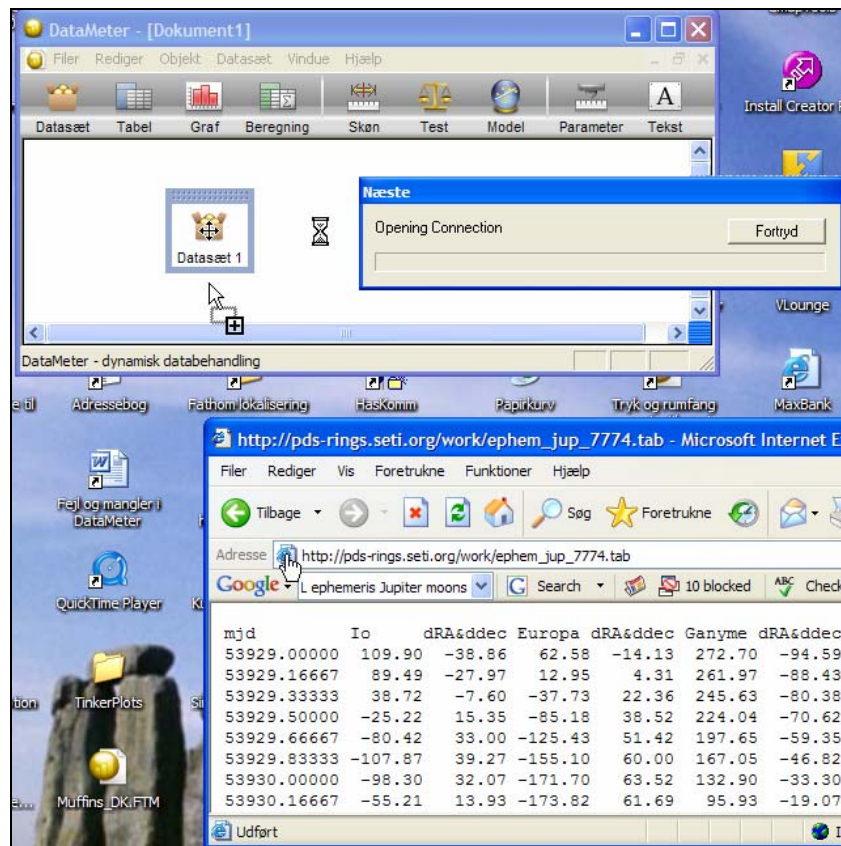
Generate table Reset form

Vi frembringer nu en tabel, der kommer frem på en ny hjemmeside (efter vi har klikket på 'Click here to download table')



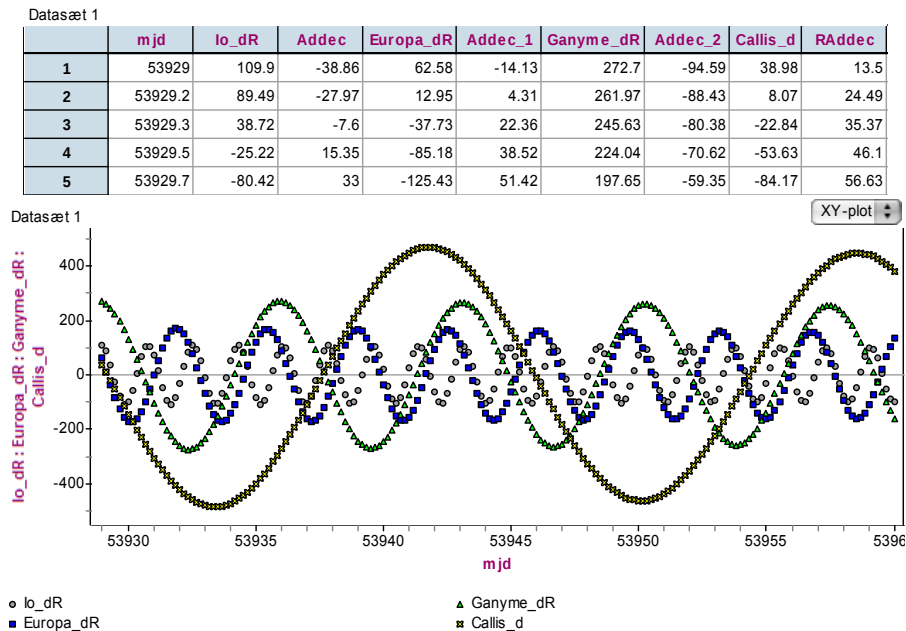
mjd	Io	dRA&ddec	Europa	dRA&ddec	Ganyme	dRA&ddec	Callis	dRA&ddec
53929.00000	109.90	-38.86	62.58	-14.13	272.70	-94.59	38.98	13.50
53929.16667	89.49	-27.97	12.95	4.31	261.97	-88.43	8.07	24.49
53929.33333	38.72	-7.60	-37.73	22.36	245.63	-80.38	-22.84	35.37
53929.50000	-25.22	15.35	-85.18	38.52	224.04	-70.62	-53.63	46.10

Det er denne tabel vi vil have overført til **DataMeter**. Det gør vi ved at sørge for at både **DataMeter** og hjemmesiden vises på en gang og derefter trække webikonet  ind i **DataMeter** dokumentet



DataMeter åbner da selv en forbindelse til webstedet og henter tabellen ind i dokumentet (idet eventuel overskydende tekst fra hjemmesiden samtidigt gemmes i datasætinspektørens kommentarside). Det sker altså alt sammen automatisk, men husk at dataformatet på hjemmesiden skal være kompatibelt med dataformatet i **DataMeter**!

Vi er nu klar til selve øvelsen. Ved at trække tidsvariablen **mjd** (der faktisk måler i dage) ind på førsteaksen og positionsvariablene for de fire Jupitermåner **Io_dR**, **Europa_dR**, **Ganyme_dR** og **Callis_d** (der faktisk måler i buesekunder) ind på andenaksen, kan vi se det tidlige forløb af planeternes bevægelse rundt om Jupiter. Læg mærke til at titlerne er blevet en anelse forvrængede, om end det er ret klart hvilke data, der hører til hvilke måner



Man kan tydeligt se at der er forskel på hvor hurtigt de bevæger sig rundt om Jupiter og hvor langt væk de kommer fra Jupiter. Det er sværere at tolke enhederne. Men der kan vi nu bruge det samme trick som ved Solsystemet. Vi behøver ikke de absolutte mål for omløbstider og baneradier, men kan i stedet benytte relative mål. Da det er Europa, der er den Jordlignende måne med chance for liv, bruger vi Europa som standard.

Øvelse 1.2.1: Jupiters måner

Hent efemeriden for Jupitermånerne på Internettet som beskrevet i det foregående og læg den ind i et datasæt. Træk variable for tid og sted (Offset RA & Dec from Jupiter) for de fire Jupitermåner ind i et grafvindue. Husk at bruge relative mål!

Udmål nu på figuren omløbstider og baneradier for de fire Jupitermåner.

Omsæt de fundne mål til relative mål ved at dividere igennem med omløbstiden henholdsvis baneradien for Europa.

Undersøg nu sammenhængen mellem de relative omløbstider og de relative baneradier for Jupitermånerne.

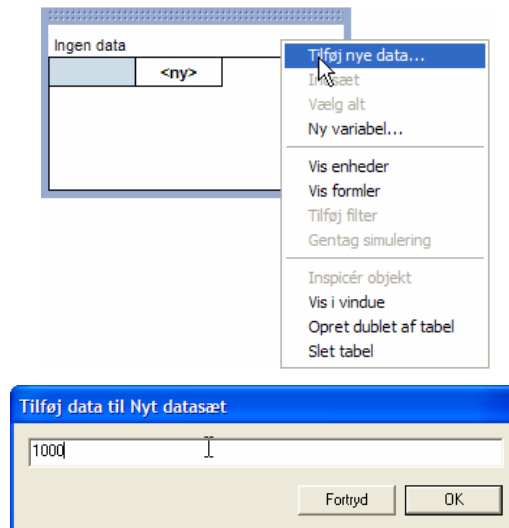
Passer Keplers tredje lov?

Øvelse 1.2.2: Tilfældige rektangler

Vi slutter med en klassisk øvelse om rektangler og de variable, der karakteriserer rektangler, hvor data konstrueres ved hjælp af *formler*. For overskuelighedens skyld baseres øvelsen på følgende variable

bredde – højde – omkreds – areal

For at kunne undersøge sammenhængene mellem disse variable skal vi nu have opbygget en database over et stort antal rektangler. Det gør vi ved at åbne en tabel og vælge menupunktet **Tilføj nye data**, hvor vi fx kan bede om 1000 nye data, svarende til at vi vil arbejde med en database indeholdende 1000 rektangler



Vi vælger nu at tilskrive rektanglerne tilfældige bredder og højder mellem 0 og 1. Det gøres ved at indskrive formelen **Tilfældig(0;1)** i tilknytning til variablene **bredde** og **højde** (faktisk behøver vi ikke en gang tilføje parametrene 0 og 1, da det netop er standardværdierne)

Tilfældige rektangler

	bredde	højde	<ny>
=	tilfældig (0; 1)	tilfældig (0; 1)	
1	0.240727	0.0776033	
2	0.489813	0.988961	
3	0.406076	0.747103	
4	0.23847	0.00639032	
5	0.927043	0.998266	

Tilføj nu selv formlerne for de andre variable, dvs. **omkreds** og **areal** og undersøg sammenhængene mellem de fire variable ved at trække dem ind i passende grafrum to af gangen. Overvej hvor mange kombinationer det er nødvendigt at undersøge.

Klassificér de figurer, der fremkommer (trekanter, firkanter osv.) og prøv specielt at finde ligningerne for de randkurver, der afgrænser disse figurer.

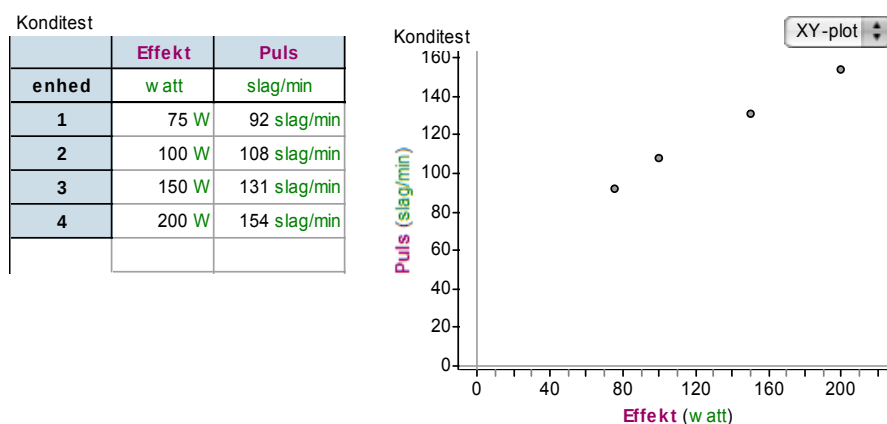
2. Variabelsammenhænge

2.1 Lineære sammenhænge

Den vigtigste sammenhæng mellem to variable er den lineære sammenhæng, hvor grafen som bekendt er en ret linje. Så lad os kigge lidt nærmere på hvordan **DataMeter** håndterer lineære sammenhænge. Vi tager udgangspunkt i følgende eksempel: En elev gennemfører en test på en ergometercykel (kondicykel). Skemaet viser sammenhørende værdier af den effekt eleven yder og elevens puls

Effekt/Watt	75	100	150	200
Puls	92	108	131	154

Vi skal undersøge sammenhængen mellem pulsen og den ydede effekt. Når man kører kondicykel kan man typisk regulere belastningen og samtidigt aflæse effekten. Derved kan man selv vælge hvilken effekt man vil køre med. Samtidigt kan man måle pulsen med en puls-måler. Det er derfor nærliggende at opfatte effekten som den uafhængige variabel som vi selv kan bestemme og pulsen som den afhængige variabel som vi måler. Vi indskrifter derfor dataene i en tabel og trækker **Effekt** ind på førsteaksen og **Puls** ind på andenaksen

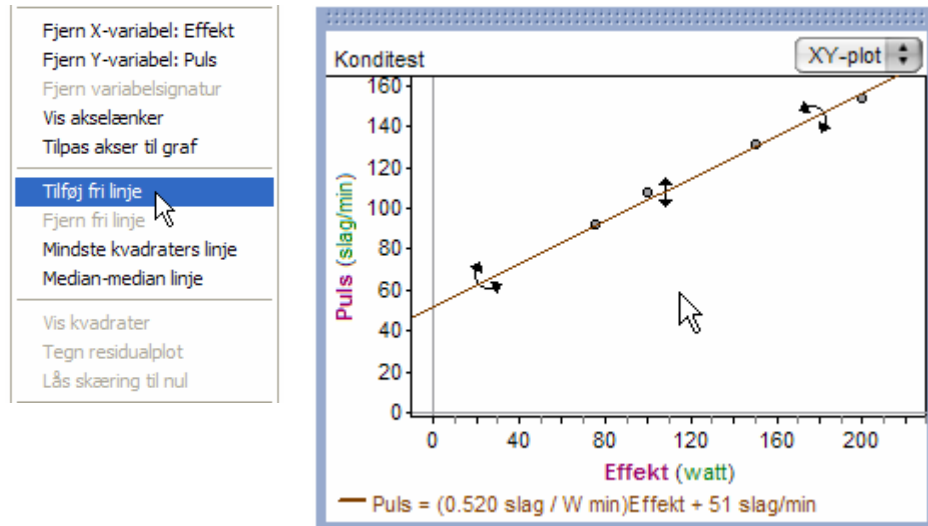


Læg mærke til enhederne: **Slag** (dvs. egentlig antal hjerteslag) er ikke en standardenhed, så før man indtaster værdierne for **Puls** skriver man **slag** i enhedsfeltet og bliver spurgt om den skal tilføjes som brugerdefineret enhed. Derefter kan man ændre den til slag/min og begynde indskrivningen af værdierne. I grafrummet har vi også justeret på andenaksen, så man kan se hele første kvadrant. Det kunne godt ligne en lineær sammenhæng da punkterne ser ud til med tilnærmelse at ligge på ret linje. Vi skal så have tilføjet denne rette linje. Her er der nu tre muligheder for at tegne rette linjer i **DataMeter**

Fri linje – Mindste kvadraters linje – Median-median linje

Vi vil ikke her komme nærmere ind på median-median linjen.

Men den frie linje er en interessant mulighed

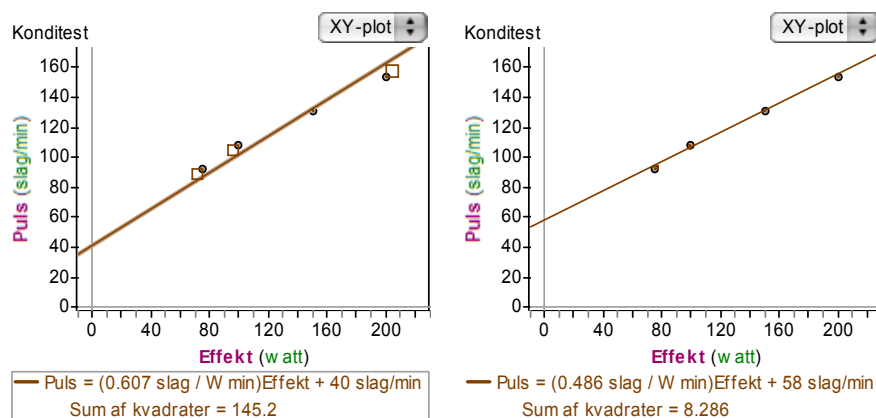


Den styres med markøren, der vipper linjen, hvis man griber linjen ude i en af siderne, mens linjen forskydes lodret, hvis man griber den på midten. Det gør det forbavsende nemt at placere linjen fornuftigt i forhold til datapunkterne og samtidigt kan man aflæse linjens ligning for neden. Læg dog mærke til at enheden for hældningskoefficienten er skrevet lidt sjusket: Der skal både divideres med W og min, så der burde have været en parentes. Det er lidt flovt, men når bare vi er opmærksomme på det, burde det ikke volde store problemer i praksis. Alene ud fra denne graf kan vi altså slutte, at der med god tilnærmelse er tale om en lineær sammenhæng: Hældningen er givet ved 0.520 slag per minut per Watt, dvs. for hver gang effekten stiger med 1 Watt stiger pulsen med 0.520 slag i minuttet. Startværdien er givet ved 51 slag i minuttet. Det er den såkaldte hvilepuls, idet det er værdien for pulsen, når effekten er 0 Watt, dvs. når vi ikke yder noget arbejde. For en utrænnet person er hvilepuls typisk ca. 60 slag i minuttet, mens den for en toptrænnet person kan komme ned på 40 slag i minuttet. I dette tilfælde ligger hvilepulsens tilsyneladende midt i mellem.

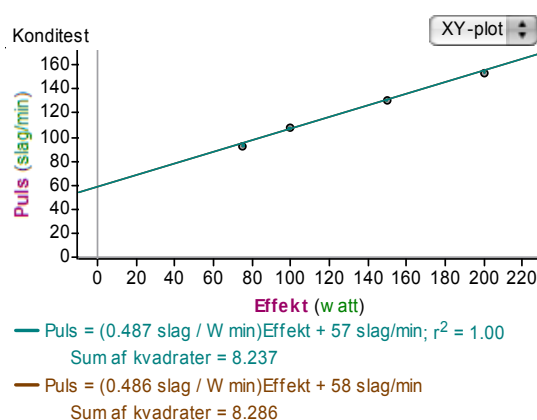
Her har vi justeret linjen ind efter øjemål. Men der findes også mere objektive kriterier man kan justere den ind efter. Den mest benyttede metode er den såkaldte *mindste kvadraters metode*. Her tilføjer man de kvadratiske afvigelser ved at vælge menupunktet **Vis kvadrater**. Samtidigt bliver man oplyst om summen af de kvadratiske afvigelser (dvs. det samlede areal af de små kvadrater). Man prøver nu så godt man kan at justere linjen ind til denne kvadratsum er så lille som muligt. Det kræver lidt fedteri, men som vist på figuren næste side kan man altså godt få kvadratsummen ned omkring 8.3 og en tilhørende linje med ligningen

$$\text{puls} = 0.486 \frac{\text{slag}}{\text{min} \cdot \text{Watt}} \cdot \text{Effekt} + 58 \frac{\text{slag}}{\text{min}}$$

2.1 Variabelsammenhænge: Lineære sammenhænge



Nu er hvilepulsen pludselig knap så imponerende. Heldigvis skal vi ikke selv fedte med mindste kvadraters linje. Den er indbygget som et menupunkt man får adgang til ved at højreklikke i grafrummet



Den ligger altså ret tæt på den linje vi selv lige har fedtet os frem til, men den mindste kvadratsum er altså givet ved 8.237. Samtidigt får vi tilføjet en forklaringsgrad på $r^2 = 1.00$, dvs. på 100%. Det betyder dog ikke at forklaringsgraden er præcis 100%, men blot at den afrundet til et helt tal er 100%. Det understøtter dog klart at den lineære model giver en god beskrivelse af variationerne i data, idet vi kan forklare tæt ved 100% af variationen med den lineære model.

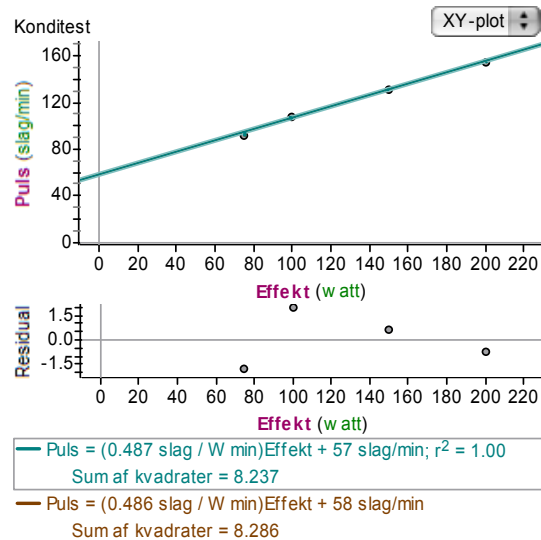
Vores bedste bud på en ligning for den lineære sammenhæng er altså

$$\text{Puls} = 0.487 \frac{\text{slag}}{\text{min-Watt}} \cdot \text{Effekt} + 57 \frac{\text{slag}}{\text{min}}$$

Hvilepuls er 57 slag i minuttet og pulsen stiger med ca. 0.487 slag i minuttet for hver gang effekten stiger med 1 Watt.

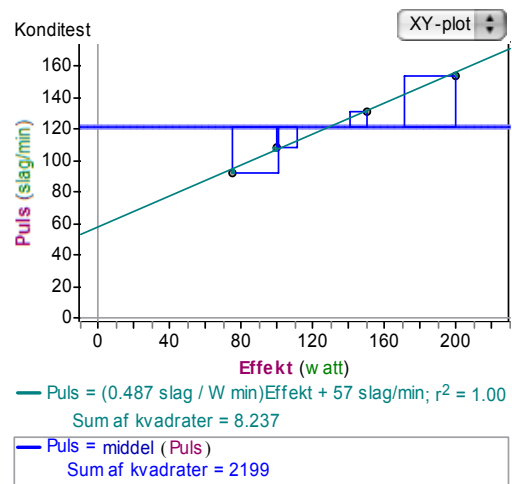
Man kan godt kaste yderligere lys over den lineære model ved at tilføje et *residualplot*. Det viser tydeligt forskellen mellem de observerede værdier og modelværdierne, idet man så at sige har sat et forstørrelsesglas på fejlene

2.1 Variabelsammenhænge: Lineære sammenhænge



Vi ser da at den typiske fejl på pulsmålingerne er op til 1.5 slag i minuttet. Det virker ikke urimeligt for et forsøg af denne type.

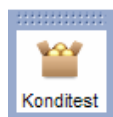
Man kan også kaste mere lys over **forklaringsgraden**. Den er et mål for hvor lille kvadratsummen 8.237 i grunden er. Men for at kunne vurdere om kvadratsummen er lille eller stor må man have et sammenligningsgrundlag: En standard for kvadratsummer. Her bruger man som sædvanligt i statistik en konkurrerende model, den såkaldte **nulhypotese**, i følge hvilken, der slet ikke er nogen sammenhæng mellem de to variable.



Ifølge nulhypotesen svinger den afhængige variabel, her pulsen, altså i virkeligheden tilfældigt omkring sin middelværdi. Men den kan vi tilføje som graf for en funktion med ligningen **Puls = middel(puls)**. Vi ser da, at hvis alt er tilfældigt, fås kvadratsummen 2199. Men anvendes den lineære model fås kun 8.237, hvilket er under 0.5 % af 2199. Den lineære model har altså fjernet mere end 99.5% af kvadratsummen i forhold til nulhypotesen. Derfor kan man forklare mere end 99.5% af variationen i datapunkterne ved hjælp af den lineære model, dvs. forklaringsgraden er over 99.5%, hvilket i den grafiske illustration afrundes til 100%.

Teknisk bemærkning: Funktioner for lineær regression

Man kan godt trække flere decimaler ud for såvel hældningen og skæringen med andenaksen som forklaringsgraden. De er nemlig alle indbyggede som funktioner i lommeregneren. Man kan derfor nemt finde de 'præcise' værdier i en beregningsboks. For at kunne regne på datasættet skal man dog først skabe en tilknytning mellem datasættet og beregningsboksen. Det sker ved at trække titlen fra datasættet ind i beregningsboksen



Ingen data

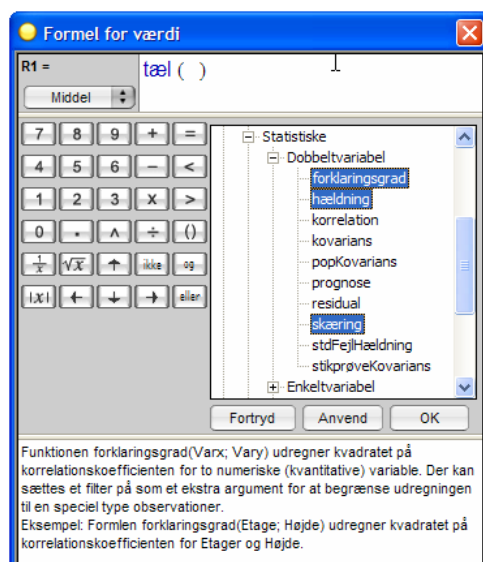
Indtast en variabel her	

Konditest

	4

R1 = tæl ()

Man får da som standard optalt antallet af observationer i datasættet, dvs. i dette tilfælde 4. Men denne formel kan vi nu ændre efter for-godtbefindende, ligesom vi kan tilføje flere formler. De relevante formler findes da i afdelingen for **Statistiske funktioner** under **Dobbeltvariabel**



Konditest

	0.48745763 slag / W min
	57.271186 slag/min
	0.99625365
	0.021136891 slag / W min

R1 = hældning (Effekt; Puls)

R2 = skæring (Effekt; Puls)

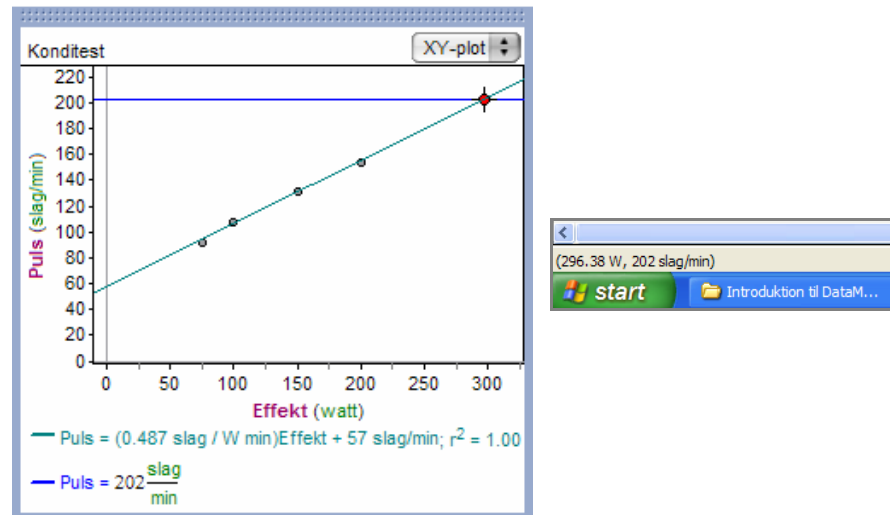
R3 = forklaringsgrad (Effekt; Puls)

R4 = stdFejlHældning (Effekt; Puls)

Forklaringsgraden er altså mere præcist givet ved 99.63%.

Det ser selvfølgelig imponerende ud med alle decimalerne, men i betragtning af den usikkerhed som datapunkterne i øvrigt udviser, er der i almindelighed slet ikke belæg for flere decimaler end dem vi får oplyst på grafen. Faktisk kan man som vist også godt få udregnet et skøn over standardfejlen for hældningen af mindste kvadraters linje. Den viser sig at være på 0.02 slag i minuttet per Watt, så derfor giver det reelt ikke mening at oplyse den med mere end 2 decimaler som 0.49 slag i minuttet per Watt.

Men lad os lige gøre eksemplet med kondicyklen færdig. Vi har allerede fundet elevens hvilepuls, men i virkeligheden har eleven også en maksimal puls. Der findes her en empirisk regel, der siger at summen af den maksimale puls og elevens alder (i år) er 220. I dette tilfælde er eleven 18 år gammel og har derfor typisk en maksimal puls på 202 slag i minuttet. Dette kan vi bruge til at finde den maksimale effekt.



Vi tilføjer som vist den vandrette linje med ligningen **Puls = 202** $\frac{\text{slag}}{\text{min}}$.

Der, hvor den skærer grafen for pulsen som funktion af effekten, finder vi den maksimale effekt. Ved at placere markøren tæt på skæringspunktet kan koordinaterne som vist aflæses på statuslinjen.

De kan være svære at aflæse nøjagtigt, og man kan derfor også i stedet for forsøge sig med en beregning, idet man vender linjens ligning om

$$y = a \cdot x + b \Leftrightarrow \frac{y - b}{a} = x$$

Ingen data

Slip en variabel her	
	297.74127 W

$$R1 = \frac{(202 \frac{\text{slag}}{\text{min}} - 57 \frac{\text{slag}}{\text{min}})}{0.487 \frac{\text{slag}}{\text{min} \cdot \text{W}}}$$

Konklusionen er altså at den maksimale effekt ligger på ca. 298 W. Ud fra den maksimale effekt kan man nu beregne den maksimale ilt-optagelse, idet der også her gælder en empirisk formel

$$\text{Maksimal_Iltoptagelse} = 13.3 \frac{\text{mL}}{\text{min} \cdot \text{W}} \cdot \text{Maksimal_Effekt}$$

Ved at dividere med elevens vægt, som i dette tilfælde er oplyst til at være 83 kg, fås endelig konditallet, som viser sig at være 48

Ingen data

Slip en variabel her	
	47.751807 mL / kg min

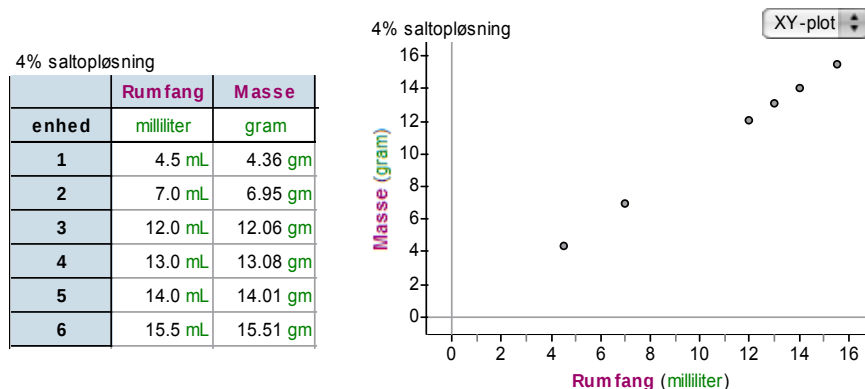
$$R1 = \frac{13.3 \frac{\text{mL}}{\text{min} \cdot \text{W}} \cdot 298 \text{ W}}{83 \text{ kg}}$$

Dette kondital ligger indenfor middelloområdet for en ung mand.

2.2 Simple proportionaliteter

Ligefrem proportionalitet

Næst efter den lineære sammenhæng er *ligefrem proportionalitet* den vigtigste sammenhæng mellem to variable. Dens vigtigste kendetegn er at den *bevarer proportioner*, dvs. hvis den uafhængige variabel fx bliver dobbelt så stort, gælder det samme for den afhængige variabel. Det gør det ret nemt at genkende en ligefrem proportionalitet. Fx er prisen for en vare ligefrem proportional med mængden: Køber vi dobbelt så meget skal vi også betale dobbelt så meget. Tilsvarende er massen af en genstand ligefrem proportional med rumfanget: Fylder genstanden dobbelt så meget får den også den dobbelte masse. Grafen for en ligefrem proportionalitet er en ret linje gennem (0,0), dvs. ligningen er på formen $y = k \cdot x$, hvor hældningen k også kaldes proportionalitetskonstanten. Den har ofte en simpel betydning. Lad os fx se på et forsøg, hvor vi vejer en væske, der hældes op i et måleglas, samtidigt med at vi aflæser rumfanget på siden af måleglasset. Vi kan da afsætte sammenhørende værdier af **Rumfang** og **Masse** i en tabel og afbilde sammenhængen i et XY-plot. Her ser vi på nogle typiske elevdata fra et elevforsøg med en 4% saltopløsning



Grafen kunne godt se ud som om det var en ret linje gennem (0,0), ligesom vi fx kan se at når rumfanget fordobles fra 7 til 14 mL fordobles massen tilsvarende fra ca. 7 gram til ca. 14 gram. Vi kan tilføje den rette linje ved enten at vælge menupunktet **Tilføj fri linje** eller menupunktet **Mindste kvadraters** linje og derefter vælge menupunktet **Lås skæring til nul**. Tilsvarende kan vi finde proportionalitetskonstanten k i tabellen ved at tilføje en ny variabel, **k_værdi**, der i dette tilfælde står for væskens massefylde og som udregnes som forholdet (proportionen) mellem variablene **Masse** og **Rumfang**

$$k_værdi = \frac{\text{Masse}}{\text{Rumfang}}$$

Den bør så med god tilnærmelse være konstant, hvorfor vi kan bruge middelværdien som et bud på opløsningens massefylde.

2.2 Variabelsammenhænge: Simple proportionaliteter

4% saltopløsning

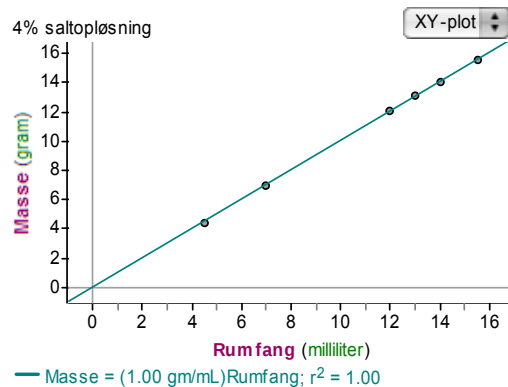
	Rumfang	Masse	k_værdi
enhed	milliliter	gram	gm/mL
=			$\frac{\text{Masse}}{\text{Rumfang}}$
1	4.5 mL	4.36 gm	0.968889 gm/mL
2	7.0 mL	6.95 gm	0.992857 gm/mL
3	12.0 mL	12.06 gm	1.005 gm/mL
4	13.0 mL	13.08 gm	1.00615 gm/mL
5	14.0 mL	14.01 gm	1.00071 gm/mL
6	15.5 mL	15.51 gm	1.00065 gm/mL

4% saltopløsning

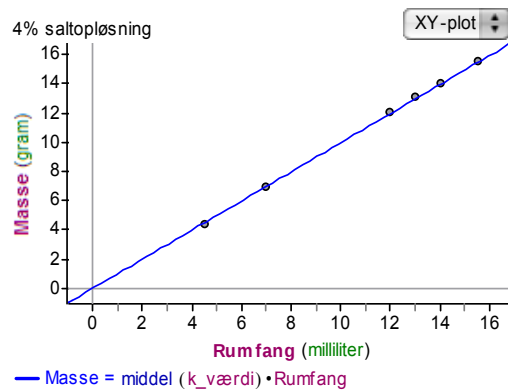
	0.99570989 gm/mL

R1 = middel (k_værdi)

Vi kan som vist udregne denne middelværdi i en beregningsboks, idet vi skal huske først at trække titlen **4 % saltopløsning** ind i beregningsboksen for at få adgang til datavariablene. Massefyllden ligger meget tæt på 1.00 g/mL i overensstemmelse med grafen

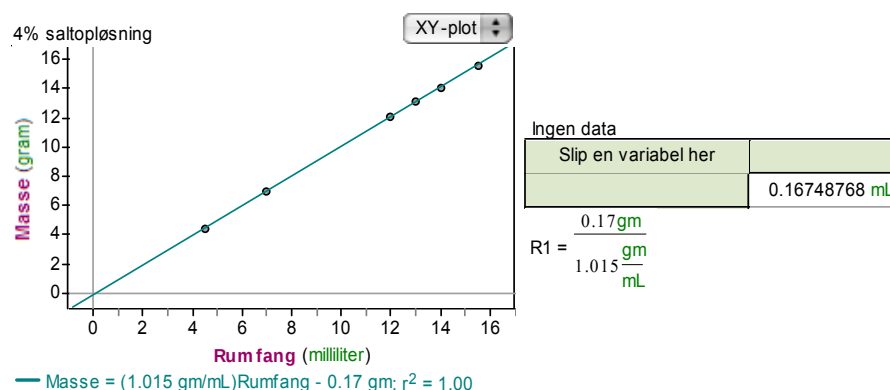


Så langt er alt altså godt og tilsvarende teknikker kan vi benytte til andre ligefremme proportionaliteter. Vi kan endda også bruge den fundne **k_værdi** til at indskrive ligningen for grafen direkte ved at vælge menupunktet **Plot funktion**. Det gør os uafhængig af mindste kvadraters linje, når vi skal arbejde med andre modeller, hvor graferne ikke nødvendigvis er rette linjer



Teknisk bemærkning: Nulpunktsfejl

Der er dog et subtilt problem i den ovenstående model, som det kan være værd at være opmærksom på. Standardværdien for vands massefylde er 1.00 g/mL. Den er dog svagt temperaturafhængig og ved stuetemperatur vil den typisk være faldet til 0.998 g/mL. Det kan imidlertid ikke forklare, hvorfor vi får en middelværdi så lav som 0.996 g/mL, idet massefylden jo skulle blive højere, når vi opløser salt i vandet. Vi kan vælge at slå det hen som et resultat af den uundgåelige usikkerhed ved elevforsøg, dvs. reelt kan vi slet ikke se forskel på rent vand og saltopløsninger i et elevforsøg af denne type. Men der er jo trods alt brugt en præcisionsvægt, så vi kan også se om fejlen kunne ligge et andet sted. Her kan vi hæfte os ved at det er svært at undgå nulpunktsfejl, specielt i rumfangsmålingen. Det er tricket at aflæse rumfanget præcist, bl.a. fordi væsken buer. Men hvis der er nulpunktsfejl i målingerne af rumfang og masse betyder det jo på den anden side, at grafen ikke længere bliver en ret linje gennem (0,0), idet de fysiske nulpunkter for rumfang og masse nu har nogle andre værdier. Grafen er derfor nok stadigvæk en ret linje, men ligger forskudt. Den svarer derfor ikke længere til en ligefrem proportionalitet, men nok til en lineær sammenhæng. En sådan forskydning ændrer *ikke* ved hældningen af den rette linje, og det er jo hældningen vi er interesseret i. I praksis skal vi derfor *ikke* låse mindste kvadraters linje til at gå gennem (0,0) hvis vi vil undgå nulpunktsfejl!



Et bedre bud på hældningen, og dermed saltopløsningens densitet, er altså 1.015 g/mL. Samtidigt skyder vi altså skylden for den 'forkerte' skæring med andenaksen, dvs. de -0.17 g, på nulpunktsfejlen. Da nulpunktsfejlen formentlig skyldes problemer med rumfangsmålingen skal vi så i virkeligheden finde skæringen med førsteaksen. Den bliver som vist 0.17 mL, hvilket altså betyder at rumfangsmålingerne systematisk ligger 0.17 mL for højt. Det er ikke nogen stor systematisk fejl og kan derfor ikke udelukkes. Men moralen er altså den overraskende, at den bedste bestemmelse af hældningen for en ligefrem proportionalitet fås ved at håndtere den som en lineær sammenhæng for at undgå eventuelle nulpunktsfejl.

Omvendt proportionalitet

Vi går derefter over til at se nærmere på den omvendte proportionalitet. Dens vigtigste kendetegn er at den *vender proportioner* (reciprok proportion), dvs. hvis den uafhængige variabel fx bliver dobbelt så stort, bliver den afhængige variabel til gengæld kun halvt så stor. Det gør der også nemt at genkende en omvendt proportionalitet. Fx er den tid man bruger for at tilbagelægge en bestemt strækning – fx mellem huset og skolen – omvendt proportional med den hastighed man kører: Kører vi dobbelt så hurtigt tager det kun den halve tid. Tilsvarende er trykket af en indespærret gas omvendt proportional med rumfanget: Spærres gassen inde på den halve plads stiger trykket til det dobbelte.

Grafen for en omvendt proportionalitet er en ligesidet hyperbel, dvs. ligningen er på formen $y = k/x$ ($y = k \cdot x^{-1}$), hvor k kaldes proportionalitetskonstanten. Lad os som et eksempel se på hvor lang tid en mand, som drikker alkohol, er om at forbrænde en genstand

Så lang tid er du om at forbrænde en genstand		
Tid om at forbrænde 1 genstand		
Vægt	Mænd	Kvinder
50 kg	1 time og 36 minutter	2 timer 24 minutter
60 kg	1 time og 20 minutter	2 timer
70 kg	1 time og 9 minutter	1 time 43 minutter
80 kg	1 time	1 time 30 minutter
90 kg	53 minutter	1 time 20 minutter
100 kg	48 minutter	1 time 12 minutter

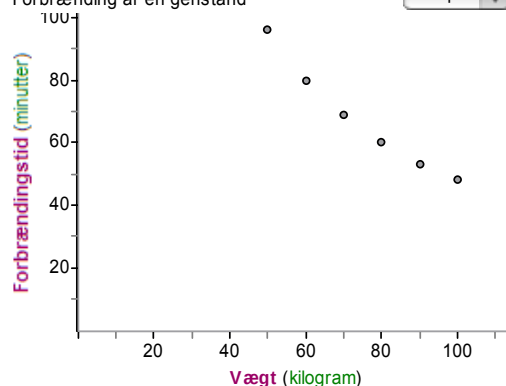
Kilde: Pjece fra Sundhedsstyrelsen: Fakta om alkohol, 1997.

Vi indskriver derfor data for forbrændingstiden for mænd, idet vi først omsætter alle tiderne til minutter og trækker variabelen **Vægt** ind som uafhængig variabel og variabelen **Forbrændingstid** ind som afhængig variabel

Forbrænding af en genstand

	Vægt	Forbrændingstid
enhed	kilogram	minutter
1	50 kg	96 min
2	60 kg	80 min
3	70 kg	69 min
4	80 kg	60 min
5	90 kg	53 min
6	100 kg	48 min

Forbrænding af en genstand



Sammenhængen er tydeligt aftagende. Grafen kunne godt se ud som om det var en ligesidet hyperbel, ligesom vi fx kan se at når vægten fordobles fra 50 kg til 100 kg falder forbrændingstiden til det halve, idet den falder fra 96 minutter til 48 minutter. Vi kan finde proportionalitetskonstanten **k** i tabellen ved at tilføje en ny variabel, **k_værdi**, der udregnes som produktet mellem variablene **Vægt** og **Forbrændingstid**

$$k_værdi = Vægt \cdot Forbrændingstid$$

Den bør så med god tilnærmelse være konstant, hvorfor vi kan bruge middelværdien som et bud på værdien af proportionalitetskonstanten.

Forbrænding af en genstand

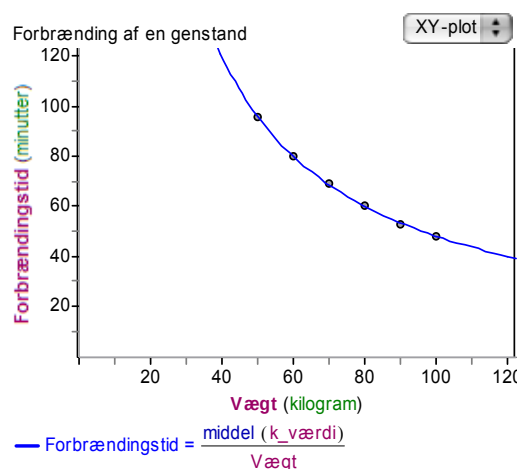
	Vægt	Forbrændingstid	k_værdi
enhed	kilogram	minutter	kg min
=			Vægt·Forbrændingstid
1	50 kg	96 min	4800 kg min
2	60 kg	80 min	4800 kg min
3	70 kg	69 min	4830 kg min
4	80 kg	60 min	4800 kg min
5	90 kg	53 min	4770 kg min
6	100 kg	48 min	4800 kg min

Forbrænding af en genstand

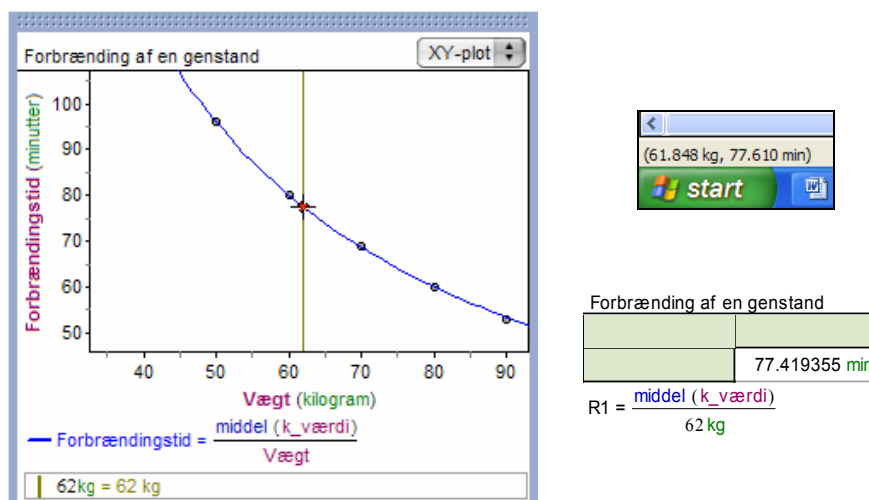
	4800 kg min

R1 = middel (k_værdi)

Vi kan som vist udregne denne middelværdi i en beregningsboks, idet vi skal huske først at trække titlen **Forbrænding af en genstand** ind i beregningsboksen for at få adgang til datavariablene. Proportionalitetskonstanten **k** ligger meget tæt på 4800 kg·min. Vi kan også bruge den fundne **k_værdi** til at indskrive ligningen for grafen ved at vælge menupunktet **Plot funktion**.



Ønsker vi nu fx at finde forbrændingstiden for en mand med vægten 62 kg kan vi dels finde den ved en beregning ud fra den fundne ligning **Forbrændingstid** = 4800 kg·min/**Vægt** eller vi kan aflæse den grafisk, idet vi tilføjer den lodrette linje Vægt = 62 kg ved hjælp af menupunktet **Plot værdi**



I begge tilfælde finder man en forbrændingstid på ca. 77.5 minutter.

Men nu handlede de givne oplysninger også om kvinders forbrændingstid. Det er derfor nærtliggende at tilføje kvindernes data til tabellen. Det kan gøres på to væsensforskellige måder:

- 1) Vi kan tilføje en ny søjle med kvindernes forbrændingstider og fx kalde den **Forbrændingstid_kvinde** (og tilsvarende ændre den forrige til **Forbrændingstid_mænd**). Men læg mærke til, at det kan vi kun gøre så nemt fordi det er de samme vægtklasser for de to køn.
- 2) Vi kan også tilføje en ny variabel **Køn**, hvorefter vi tilføjer de resterende oplysninger om kvinder til tabellen. Dette er langt mere i **DataMeters** ånd

Forbrænding af en genstand

	Vægt	Forbrændingstid	Køn	k_værdi
enhed	kilogram	minutter		kg min
=				Vægt•Forbrændingstid
1	50 kg	96 min	Mand	4800 kg min
2	60 kg	80 min	Mand	4800 kg min
3	70 kg	69 min	Mand	4830 kg min
4	80 kg	60 min	Mand	4800 kg min
5	90 kg	53 min	Mand	4770 kg min
6	100 kg	48 min	Mand	4800 kg min
7	50 kg	144 min	Kvinde	7200 kg min
8	60 kg	120 min	Kvinde	7200 kg min
9	70 kg	103 min	Kvinde	7210 kg min
10	80 kg	90 min	Kvinde	7200 kg min
11	90 kg	80 min	Kvinde	7200 kg min
12	100 kg	72 min	Kvinde	7200 kg min

I tabellen ses da, at kvindernes konstant er ca. 7200 kg·min.

Vi kan nu håndtere de to køn separat i grafvinduet ved dels at trække variablen **Køn** ind i grafrummet, dels udregne middelværdierne for variablen **k_værdi** separat for de to køn. Det kan fx ske ved at tilføje betingelserne (filtrene)

Køn = "Mand" henholdsvis **Køn** = "Kvinde" til den statistiske funktion **mid-del()**, dvs. ved at skrive kommandoen på formen **middel(k_værdi; Køn = "Mand")** og tilsvarende for kvinder. Læg mærke til at vi har sat kategorien "Mand" i anførselstegn for at markere, at det er navnet på en kategori og *ikke* navnet på en variabel.

Forbrænding af en genstand

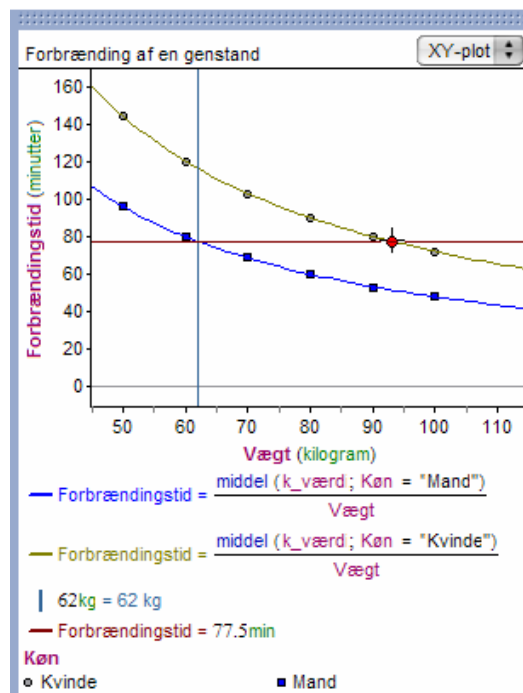
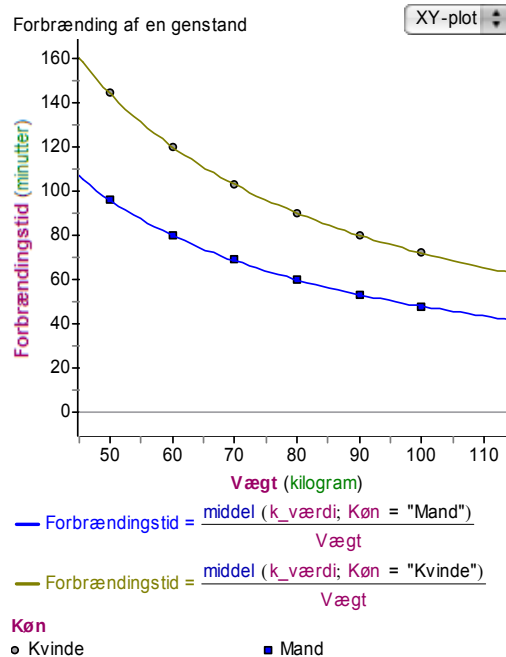
Køn	Kvinde	7201.6667 kg min
	Mand	4800 kg min
Søjle total		6000.8333 kg min

R1 = **middel(k_værdi)**

Men vi kan også som vist få de to køn adskilt i en beregningsboks.

Hvis vi så fx vil finde ud af hvor meget en kvinde skal veje for at have den samme forbrændingstid som en mand på 62 kg, kan vi finde ud af det ved først at tilføje den lodrette linje **Vægt** = 62 kg og aflæse skæring med mændenes graf (77.5 min). Derefter tilføjes den vandrette linje **Forbrændingstid** = 77.5 min og vi aflæser skæring med kvindernes graf. Vi finder da på statuslinjen at den korresponderende vægt for kvinder er ca. 93 kg

(92.938 kg, 77.489 min)



Sammenhængen mellem bølgelængde og frekvens for en lydgiver er et andet typisk eksempel på en omvendt proportionalitet. Bølgelængden og frekvensen er nemlig omvendt proportionale og proportionalitetskonstanten er netop lydens fart

$$\text{bølgelængde} \cdot \text{frekvens} = \text{lydens fart}$$

Her følger nu data fra et eleveksperiment med lydresonanser

Fotoet viser en opstilling til bestemmelse af lydens fart i luft. Den ene ende af et åbent rør er stukket ned i vand. Over røret er anbragt en højttaler, som udsender en tone med en valgt frekvens f . Røret flyttes op og ned for at finde de positioner af røret, hvor lyden forstærkes. For hver frekvens bestemmes afstanden ΔL mellem to nabopositioner, hvor lyden forstærkes. Lydbølgerne i røret har bølgelængden $2 \cdot \Delta L$.

Tabellen viser sammenhørende værdier for frekvensen f og afstanden ΔL .

Bestem ud fra oplysningerne i tabellen en værdi for lydens fart i luft.



FOTO: ELSEBETH PETERSEN

f / kHz	1,2	1,6	2,0	2,4	2,8	3,2	3,6	4,0
$\Delta L / \text{m}$	0,144	0,110	0,083	0,070	0,060	0,053	0,050	0,043

Da bølgelængden er det dobbelte af ΔL i følge opgaveteksten kan vi nu nemt konstruere den fornødne tabel i **DataMeter**

Lydens fart

	f	Delta_L	Lambda	k_værdi
enhed	kilohertz	meter	meter	m/s
=			$2 \cdot \text{Delta_L}$	$f \cdot \text{Lambda}$
1	1.2 kHz	0.144 m	0.288 m	345.6 m/s
2	1.6 kHz	0.110 m	0.22 m	352 m/s
3	2.0 kHz	0.083 m	0.166 m	332 m/s
4	2.4 kHz	0.070 m	0.14 m	336 m/s
5	2.8 kHz	0.060 m	0.12 m	336 m/s
6	3.2 kHz	0.053 m	0.106 m	339.2 m/s
7	3.6 kHz	0.050 m	0.1 m	360 m/s
8	4.0 kHz	0.043 m	0.086 m	344 m/s

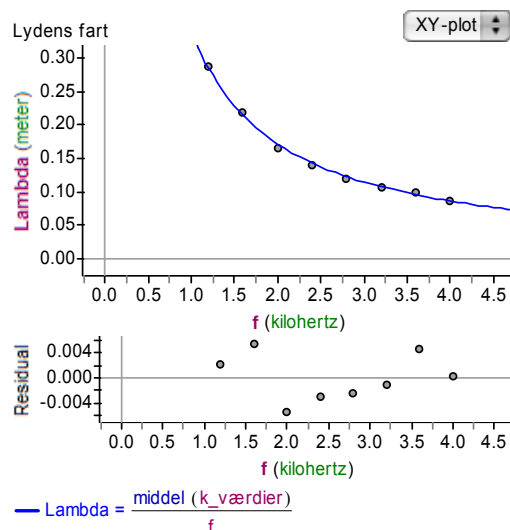
Lydens fart

	343.1 m/s

$R1 = \text{middel} (k_værdi)$

Lydens fart, dvs. **k_værdi**, er altså med rimelighed den samme for alle datapunkterne. Et godt bud på lydens fart i luft er derfor middelværdien, som er givet ved 343.1 m/s. Læg mærke til enheden for variabelen **k_værdi**, der i første omgang anføres som kiloHertz-meter. Ved at gå ind i enhedsfeltet og overskrive den med den compatible enhed m/s, omsættes resultatet automatisk som vist på figuren.

Vi kan også checke grafen for sammenhængen mellem frekvensen og bølgelængden for at se om den med rimelighed kan beskrives med en omvendt proportionalitet



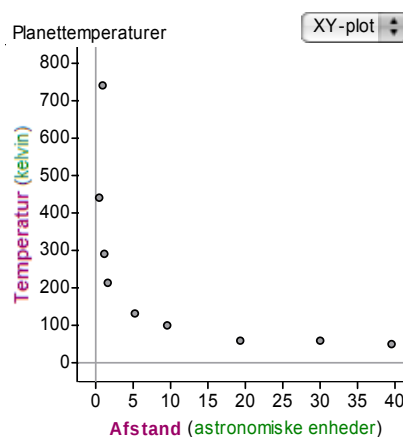
Vi har da tilføjet et residualplot for at kunne vurdere afvigelserne mellem modellen og de faktiske målinger. Vi ser dels, at afvigelserne ligger rimeligt tilfældigt, dels at de holder sig under 0.005 m, dvs. under 5 mm, og det er ikke urimeligt for et elevforsøg af denne type.

Omvendt kvadratrodsproportionalitet

Som et eksempel på en lidt mere kompliceret proportionalitetsmodel slutter vi med at se nærmere på sammenhængen mellem en planets afstand til Solen (i astronomiske enheder) og dens middeltemperatur (i kelvin, der er standardenheden for temperatur i **DataMeter** idet temperaturen i Kelvinskalaen netop har værdien nul i det absolutte nulpunkt)

Planettemperaturer

	Planet	Afstand	Tempe...
enhed		astronomisk...	kelvin
1	Merkur	0.387 æ	440 K
2	Venus	0.723 æ	737 K
3	Jorden	1 æ	288 K
4	Mars	1.524 æ	210 K
5	Jupiter	5.203 æ	129 K
6	Saturn	9.555 æ	97 K
7	Uranus	19.218 æ	58 K
8	Neptun	30.11 æ	58 K
9	Pluto	39.545 æ	50 K



Bortset fra planeten Venus, der har taget et ordentligt hop opad i temperaturen, kunne det tyde på en pæn aftagende sammenhæng, hvor temperaturen falder systematisk med den voksende afstand til Solen. Ser vi fx på parret Merkur og Mars eller på parret Jupiter og Uranus kunne det godt se ud til, at når afstanden bliver ca. fire gange så stor bliver temperaturen ca. halveret. Det tyder på en omvendt kvadratrodsproportionalitet $y = k / \sqrt{x} = \sqrt{k^2 / x} = (k \cdot x^{-\frac{1}{2}})$. Men det kan vi jo godt gøre ved at undersøge om $y^2 \cdot x$ med tilnærmelse er konstant. Vi tilføjer derfor en ny sammensat variabel, der i dette tilfælde korresponderer med kvadratet på k-værdien

Planettemperaturer

	Planet	Afstand	Tempe...	k_kvadratværdi
enhed		astronomisk...	kelvin	ae K^2
=				AfstandTemperatur ²
1	Merkur	0.387 ae	440 K	74923 ae K^2
2	Venus	0.723 ae	737 K	392711 ae K^2
3	Jorden	1 ae	288 K	82944 ae K^2
4	Mars	1.524 ae	210 K	67208 ae K^2
5	Jupiter	5.203 ae	129 K	86583 ae K^2
6	Saturn	9.555 ae	97 K	89903 ae K^2
7	Uranus	19.218 ae	58 K	64649 ae K^2
8	Neptun	30.11 ae	58 K	101290 ae K^2
9	Pluto	39.545 ae	50 K	98862 ae K^2

Planettemperaturer

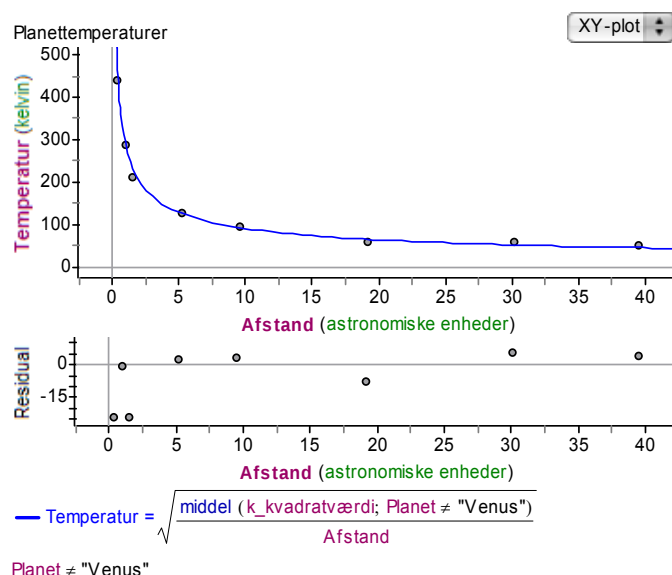
	83295.451 ae K^2

R1 = middel (k_kvadratværdi; Planet ≠ "Venus")

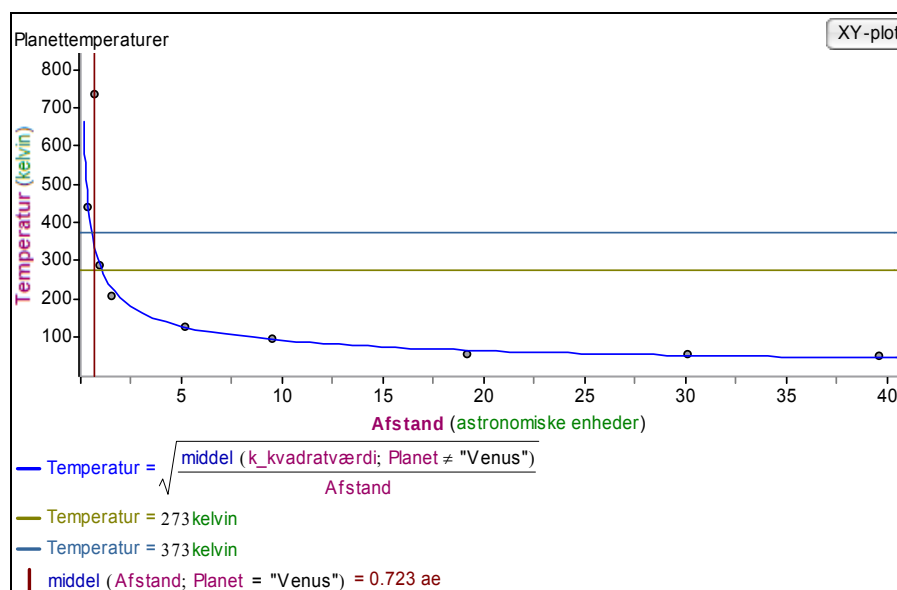
Med planeten Venus som en bemærkelsesværdig undtagelse ser det ud som om k-kvadratet er rimeligt konstant. Derfor har vi trukket en middelværdi ud, hvor vi har fjernet Venus ved hjælp af en ekstra parameter (filter) til **middel()**-funktionen, dvs. skrevet beregningen på formen **middel(k_kvadratværdi; Planet ≠ "Venus")** med "Venus" i anførselstegn for at markere, at det er navnet på en kategori. Man får fat i 'forskellig fra'-tegnet ≠ (der ligger samme sted på lommeregnetastaturet som lighedstegnet =) ved at holde **CTRL**-tasten nede. Vi kan så checke sammenhængen grafisk idet vi igen filtrerer Venus ud som en bemærkelsesværdig undtagelse. For at kunne vurdere modellen nøjere tilføjer vi ydermere et residualplot over afvigelserne mellem modellen og de faktiske temperaturer (se figur næste side). Vi ser da, at Merkur og Mars ligger lidt lavt, men ellers passer modellen rimeligt. Der gælder altså følgende standardlov for sammenhængen mellem en planets afstand og dens middeltemperatur

$$\text{Temperatur} = \sqrt{\frac{83300 \text{ ae} \cdot \text{K}^2}{\text{Afstand}}}$$

2.2 Variabelsammenhænge: Simple proportionaliteter



En sådan model kan bl.a. bruges til at undersøge den såkaldte *habilitetszone* omkring Solen, dvs. indenfor hvilke grænser afstanden kan ligge for at der er chance for at finde liv. Et rimeligt kriterium er da at forlange at der skal kunne være flydende vand på planeten. Men så skal temperaturen jo ligge mellem 273 K og 373 K. Vi tilføjer derfor disse vandrette linjer til grafen



Vi ser at Jorden lige netop ligger indenfor denne zone. Venus derimod ligger alt for højt (idet noget er gået grueligt galt for Venus, der burde ligge indenfor). Endelig ligger Mars for lavt. Blandt planeterne er det altså kun Jorden, der har fået mulighed for at udvikle stabilt liv.

2.3 Fibonaccital og eksponentiel vækst

Fibonaccitalle sættes ofte i forbindelse med formeringsmønstre hos kaniner eller bier. Her vil vi se på dem fra en kombinatorisk synsvinkel. Vi arbejder da med flisedækninger bestående af en enkelt række kvadratiske felter, hvor vi kun må anvende enkeltbrikker (der dækker et enkelt kvadratisk felt: ) og dominobrikker (der dækker to kvadratiske felter: ). Det kan fx se således ud



Denne flisedækning dækker i alt 7 felter ved hjælp af 5 fliser, idet vi har anvendt 3 enkeltbrikker og 2 dominobrikker.

Det kombinatoriske problem består nu i at undersøge på hvor mange måder man kan dække n felter med enkeltbrikker og dominobrikker.

Vi kan da starte med at udfylde en simpel tabel

Antal felter n	Flisedækninger	Fibonaccital F_n
1		1
2	 	2
3	   	3
4	      	5
5		
6		
7		
8		

Prøv selv at udfylde resten af tabellen og finde derved de første 8 fibonaccital. Kan du se noget mønster i tallene?

Teknisk bemærkning: Den rekursive formel for fibonaccital

Det ser ud som om hvert nyt fibonaccital er summen af de to foregående, dvs.

$$F_n = F_{n-1} + F_{n-2}$$

At dette virkeligt er rigtigt kan vi nu argumentere for på følgende måde: Vi ser på en flisedækning med n felter



Den må da som vist enten slutte med en enkeltbrik eller med en dominobrik:

- 1) Hvis den slutter med en enkeltbrik må de resterende brikker dække de $n-1$ foregående felter. Det kan de gøre på F_{n-1} forskellige måder.
- 2) Hvis den slutter med en dominobrik må de resterende brikker dække de $n-2$ foregående felter. Det kan de gøre på F_{n-2} forskellige måder.

Men det viser netop at der er $F_{n-1} + F_{n-2}$ forskellige måder man kan dække n felter på. Altså gælder den følgende rekursive formel for Fibonaccitalene

$$F_n = F_{n-1} + F_{n-2}$$

Formlen kaldes rekursiv fordi den i stedet for at tillade en direkte beregning af Fibonaccitalene viser tilbage til tidligere Fibonaccital. På et eller andet tidspunkt når man da tilbage til startværdierne $F_1 = 1$ og $F_2 = 2$, og man kan da endeligt fastlægge slutværdien. Hvis vi fx vil beregne F_5 ser det således ud

$$\begin{aligned} F_5 &= F_4 + F_3 \\ &= (F_3 + F_2) + (F_2 + F_1) \\ &= ((F_2 + F_1) + F_2) + (F_2 + F_1) \\ &= 3F_2 + 2F_1 = 3 \cdot 2 + 2 \cdot 1 = 8 \end{aligned}$$

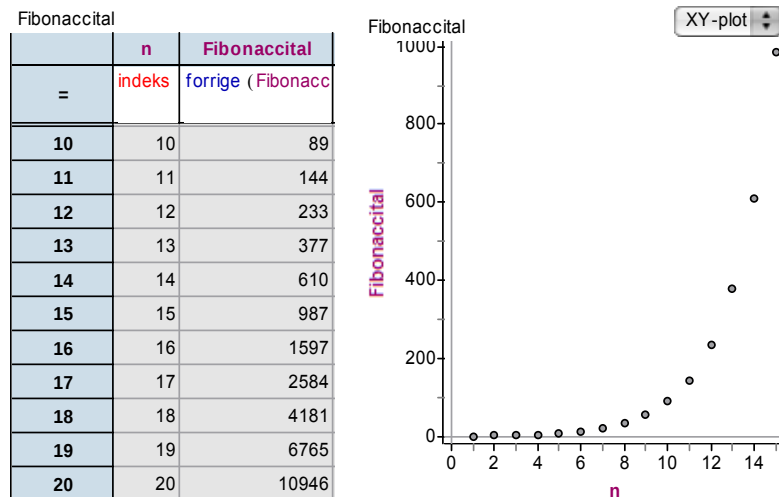
Hvis vi nu vil lege videre med Fibonaccitalene skal de skrives ind i en tabel. Det kan godt være træls, hvis de skal skrives ind i hånden. Heldigvis understøtter **DataMeter** rekursive formler for variable. I dette tilfælde er hvert nyt fibonaccital summen af de to foregående. I **DataMeters** formulering ser det således ud

Fibonaccital = forrige(Fibonaccital) + forrige(forrige(Fibonaccital))

Men dertil kommer en ekstra parameter i **forrige()**-funktionen, der fortæller hvilken værdi man skal bruge, når man når til det første Fibonaccital $F_1 = 1$ i tabellen. Vi udnytter da at der forud for det første Fibonaccital må gå værdierne $F_0 = 1$ og $F_{-1} = 0$, da de ellers ikke vil passe med mønstret. Den færdige formel ser derfor således ud

Fibonaccital = `forrige (Fibonaccital; 1) + forrige (forrige (Fibonaccital; 1); 0)`

Det er nu ikke svært at oprette en tabel med de første 20 fibonaccital og konstruere den tilsvarende graf

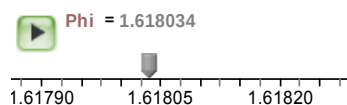


Fibonaccitalene vokser meget hurtigt og vi kan nemt få en fornemmelse af at væksten foregår eksponentielt. I så fald skal fremskrivningsfaktoren med tilnærmelse være konstant. Men den kan vi jo udregne som forholdet mellem et fibonaccital og det næste fibonaccital (hvor den ekstra tomme parameter "" skal sikre at **næste()**-funktionen er udefineret, når vi når slutningen af tabellen).

Fremskrivningsfaktor =	$\frac{\text{næste}(\text{Fibonaccital}; "")}{\text{Fibonaccital}}$
------------------------	---

Det ser ret så lovende ud idet fremskrivningsfaktoren hurtigt stabiliserer sig på tallet 1.618034... Måske kan du genkende dette tal som *det gyldne snit*, hvis du har leget med det gyldne snit i andre sammenhænge. I hvert fald indfører vi nu en parameter ved at trække parameterværktøjet ned fra værktøjshylden. Denne parameter giver vi værdien det gyldne snit Φ , som vi skriver som **Phi**

Fibonaccital			
	n	Fibonaccital	Fremskriv...
=	indeks	forrige (Fibonacc	næste (Fibonacc
			Fibonacc
10	10	89	1.61797753
11	11	144	1.61805556
12	12	233	1.61802575
13	13	377	1.61803714
14	14	610	1.61803279
15	15	987	1.61803445
16	16	1597	1.61803381
17	17	2584	1.61803406
18	18	4181	1.61803396
19	19	6765	1.61803400
20	20	10946	



Ved at højreklikke på parameteren **Phi** får vi adgang til at indskrive formelen for det gyldne snit som vist på figuren.

Phi =	$\frac{(1 + \sqrt{5})}{2}$
-------	----------------------------

Vi kan nu også fremskaffe et bud på startværdien b for den eksponentielle vækst $y = b \cdot a^x$, idet vi vender formelen om: $b = y/a^x$. Også startværdien stabiliseres hurtigt på en fast værdi 0.72360680... som det kan være sværere at genkende. Men får man det lykkelige indfald at kigge på multipla af startværdien opdager man ret hurtigt en forbindelse til det gyldne snit. Fem gange startværdien b er nemlig det samme som det gyldne snit lagt sammen med to

$$5b = \Phi + 2 \Leftrightarrow b = \frac{\Phi + 2}{5}$$

Den eksponentielle vækstmodel

for fibonaccitalle er altså givet ved $F_n = \frac{\Phi+2}{5} \cdot \Phi^n$.

Fibonaccital

	n	Fibonaccital	Startværdi
=	indeks	forrige (Fibonacc)	Fibonaccital Φ^n
11	11	144	0.72359982
12	12	233	0.72360946
13	13	377	0.72360578
14	14	610	0.72360719
15	15	987	0.72360665
16	16	1597	0.72360685
17	17	2584	0.72360678
18	18	4181	0.72360681
19	19	6765	0.72360679
20	20	10946	0.72360680

Fibonaccital

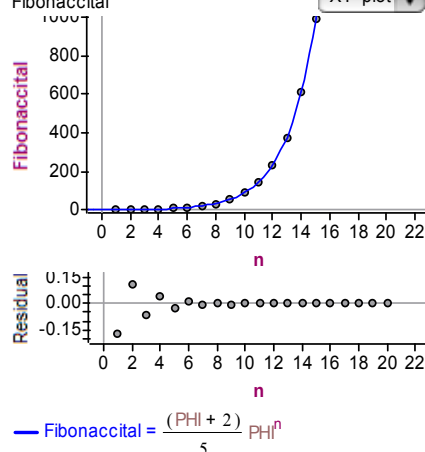
	3.618034

R1 = 5 sidste (startværdi)

Fibonaccital

	n	Fibonaccital	Eksponent...
=	indeks	forrige (Fibonacc)	$\frac{(\Phi + 2)}{5} \Phi^n$
1	1	1	1.17
2	2	2	1.89
3	3	3	3.07
4	4	5	4.96
5	5	8	8.02
6	6	13	12.98
7	7	21	21.01
8	8	34	33.99
9	9	55	55.00
10	10	89	89.00

Fibonaccital



Formlen passer utroligt flot. Faktisk kan man finde den præcise værdi af det n 'te fibonaccital ved at afrunde modelværdien til et helt tal.

Øvelse 2.4.1: Kvadratet på fibonaccital

Undersøg summen af to på hinanden følgende kvadrater på Fibonaccital: $F_{n-1}^2 + F_n^2$. Hvad er mønstret?

Udfordring: Kan du vise den fundne formel kombinatorisk?

Øvelse 2.4.2: Den direkte formel for fibonaccital

Undersøg afvigelserne for den eksponentielle vækstmodel:

$F_n - \frac{\Phi+2}{5} \cdot \Phi^n$. Find derved en direkte eksakt formel for først afvigelserne og dernæst fibonaccitalle selv.

2.4 Variabelkontrol

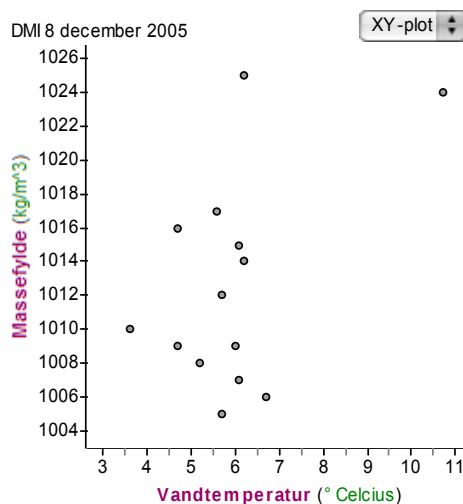
Når man skal undersøge sammenhængen mellem to variable er det meget vigtigt, at man har kontrol over variablene, dvs. har sikret sig, at der ikke er andre – skjulte – baggrundvariable, der blander sig i undersøgelsen. Ikke blot kan de sløre sammenhængen mellem de to variable, men de kan endda vende en sammenhæng helt om. Lad os fx tænke os at vi vil undersøge massefylden af vand og at vi vil indhente nogle empiriske data. Dem kan vi fx finde på DMI's hjemmeside, hvor man løbende rapporterer massefylden og temperaturen af havvand for en række udvalgte lokaliteter langs de danske kyster



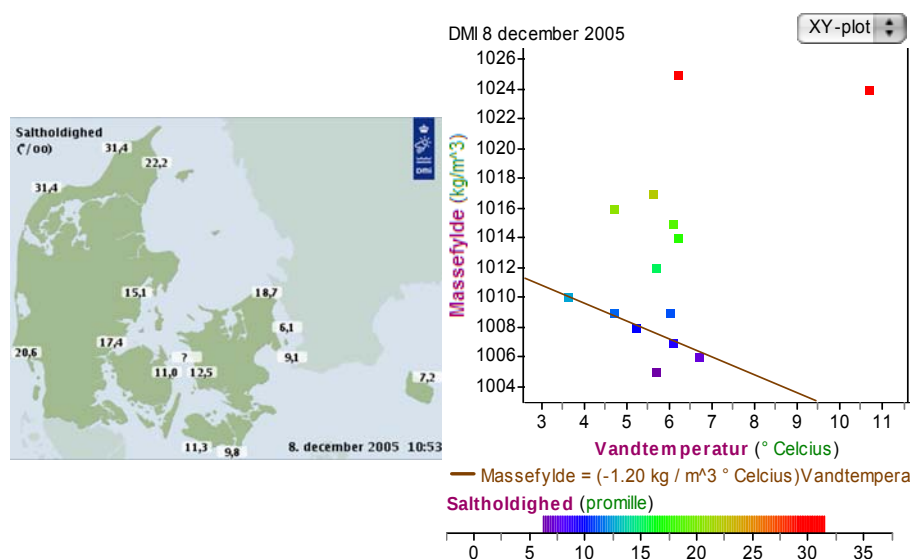
Men det kunne jo så give anledning til en undersøgelse af sammenhængen mellem temperaturen og massefylden for dansk havvand. Vi taster derfor tallene ind i en tabel (for de målestationer, hvor begge variable er oplyst) og konstruerer en graf over sammenhørende værdier af variablene **Vandtemperatur** (med Celcius som en brugerdefineret enhed og gradtegnet ° indskrevet som *grader*) og **Massefylde**

DMI 8 december 2005

	Vandtemperatur	Massefylde
enhed	° Celcius	kg/m ³
1	6.7 ° Celcius	1006 kg/m ³
2	5.2 ° Celcius	1008 kg/m ³
3	4.7 ° Celcius	1009 kg/m ³
4	6.1 ° Celcius	1007 kg/m ³
5	5.7 ° Celcius	1005 kg/m ³
6	6.1 ° Celcius	1015 kg/m ³
7	3.6 ° Celcius	1010 kg/m ³
8	6.0 ° Celcius	1009 kg/m ³
9	6.2 ° Celcius	1014 kg/m ³
10	5.7 ° Celcius	1012 kg/m ³
11	5.6 ° Celcius	1017 kg/m ³
12	10.7 ° Celcius	1024 kg/m ³
13	6.2 ° Celcius	1025 kg/m ³
14	4.7 ° Celcius	1016 kg/m ³



Det ser unægteligt lidt mudret ud – og hvad værre er, så ser det nærmest ud som om sammenhængen er stigende. Men alt andet lige forventer vi jo at massefylden falder med stigende temperaturer, fordi væsker ligesom faste stoffer udvider sig når de varmes op (med vand som en bemærkelsesværdig undtagelse når vi nærmer os frysepunktet, men så koldt er det ikke i den ovennævnte graf). Så hvad er der gået galt? Hvis vi skal undersøge sammenhængen mellem vandtemperatur og massefylde skal vi sikre os at vi kun varierer variabelen **Vandtemperatur**, dvs. vi skal holde alle andre variable konstante i undersøgelsen. Her skal vi altså sikre os, at det danske havvand i alle andre henseender end lige netop temperaturen har den samme beskaffenhed. Men det er langt fra tilfældet. Der er stor forskel på sammensætningen af det ferske vand fra Østersøen og det salte vand fra Vesterhavet. Så hvis nu saltholdigheden havde en væsentlig indflydelse på massefylden kunne den gå ind og ødelægge hele vores undersøgelse. Heldigvis har DMI også taget målinger af saltholdigheden

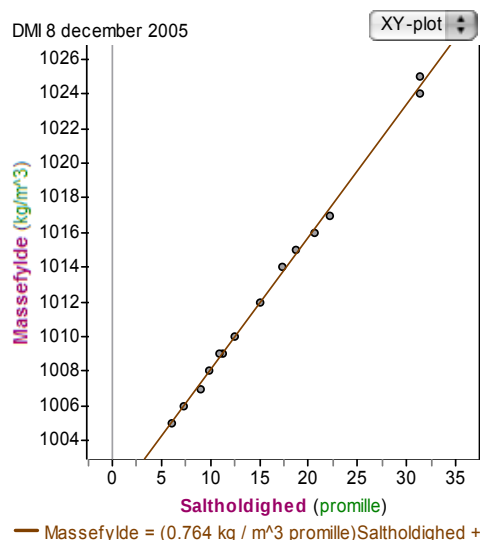


Tilføjes disse data til vores tabel kan vi som vist trække variabelen **Saltholdighed** ind i grafrummet for at se hvordan den påvirker sammenhængen. Den viser da klart at hvis vi indskrænker os til kun at se på datapunkter med tilnærmelsesvis samme saltholdighed er sammenhængen som forventet aftagende. På grafen har vi fx tilføjet en tendenslinje for det nederste bånd bestående af 7 datapunkter med saltholdigheder tæt på 10 promille. De høje massefylder kommer altså ikke fra en stigning i temperaturen men fra en stigning i saltholdigheden. Faktisk har saltholdigheden langt større indflydelse på massefylden end temperaturen. Det kan vi se ved i stedet at afbilde sammenhængen mellem saltholdighed og massefylde. Vi får da en meget klar lineær sammenhæng, der viser at massefylden stiger med ca. 0.75 kg/m^3 for hver gang saltholdigheden stiger med 1 promille (og endvidere er der en startværdi på 1000.4 kg/m^3)

DMI 8 december 2005

	Massefylde	Saltholdighed
enhed	kg/m ³	promille
1	1006 kg/m ³	7.2 promille
2	1008 kg/m ³	9.8 promille
3	1009 kg/m ³	11.3 promille
4	1007 kg/m ³	9.1 promille
5	1005 kg/m ³	6.1 promille
6	1015 kg/m ³	18.7 promille
7	1010 kg/m ³	12.5 promille
8	1009 kg/m ³	11.0 promille
9	1014 kg/m ³	17.4 promille
10	1012 kg/m ³	15.1 promille
11	1017 kg/m ³	22.2 promille
12	1024 kg/m ³	31.4 promille
13	1025 kg/m ³	31.4 promille
14	1016 kg/m ³	20.6 promille

DMI 8 december 2005



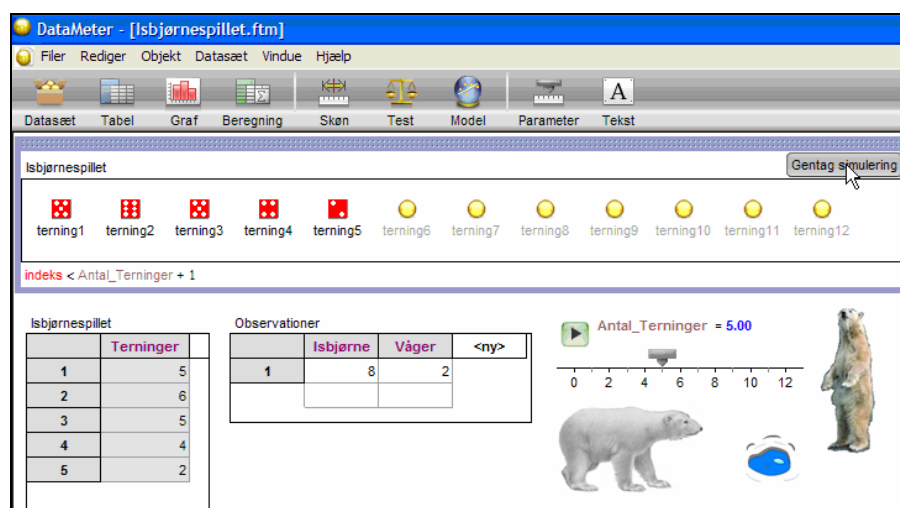
Når man overser en skjult variabel, der har afgørende indflydelse på undersøgelsen af et fænomen, siger man at undersøgelsen er *konfunderet* (dvs. rodet sammen, jævnfør det beslægtede ord konfus) og man kalder også den skjulte variabel for en *konfunderet variabel*. I daglig tale betyder ordet konfunderet det samme som forvirret (jævnfør Engelsk: confuse). Hvis man ikke har styr på de relevante variable er undersøgelsen altså noget rod. Det forhold at en skjult variabel ikke blot kan sløre en sammenhæng, men i uheldige tilfælde endda vende en sammenhæng, kaldes for *Simpsons paradoks*.

Man udfører variabelkontrol, dvs. kontrollerer variablene, når man dels har styr på alle de relevante variable for det pågældende fænomen, dels kan fokusere på de uafhængige variable én for én mens de øvrige holdes konstant. Det er nemt nok, hvis man selv kan tilskrive de uafhængige variable de værdier man ønsker. Men i mange sammenhænge er den sidste betingelse desværre ikke opfyldt. Når man fx indhenter oplysninger om havvand rundt omkring i Danmark kan man jo ikke holde saltholdigheden konstant, mens man varierer temperaturen. Man må tage de kombinationer af de uafhængige variable man nu får stillet til rådighed og så efterfølgende prøve at fokusere på passende udsnit af observationer, hvor den ene variabel (med tilnærmelse) er konstant.

I de følgende to spil kan du selv lege med variabelkontrol i situationer, hvor du ikke har fuldstændigt styr på de uafhængige variable, fordi de styres stokastisk, dvs. udvælges tilfældigt. Spillene minder altså mere om betingelserne for feltstudier end om kontrollerede laboratorieforsøg. Det gør det selvfølgelig sværere at se sammenhænge, men derved bliver spillene også mere udfordrende og forhåbentligt derved også mere interessante.

Øvelse 2.4.1: Isbjørnespillet

For Erik den Røde, der opdagede Grønland var den hvide bjørn et hel- ligt dyr. Han bad derfor sine folk være meget opmærksomme på is- bjørnenes gøren og laden. I isbjørnespillet skal du nu være med til at holde udkig efter isbjørne og våger i isen. Det er et terningspil, hvor du kaster med et vist antal terninger svarende til de områder du hol- der øje med. Når du har kastet terningerne får du at vide hvor mange våger, der er i områderne, samt hvor mange isbjørne, der er i områ- derne. Du skal altså spille med en ekspert, der kan forsyne dig med disse fornødne oplysninger. I dette tilfælde er det **DataMeter**, der er eksperten, så for at spille isbjørnespillet skal du først hente det på **DataMeters** hjemmeside. Når du åbner spillet er det allerede sat i gang, så du kan få en fornemmelse for spillet

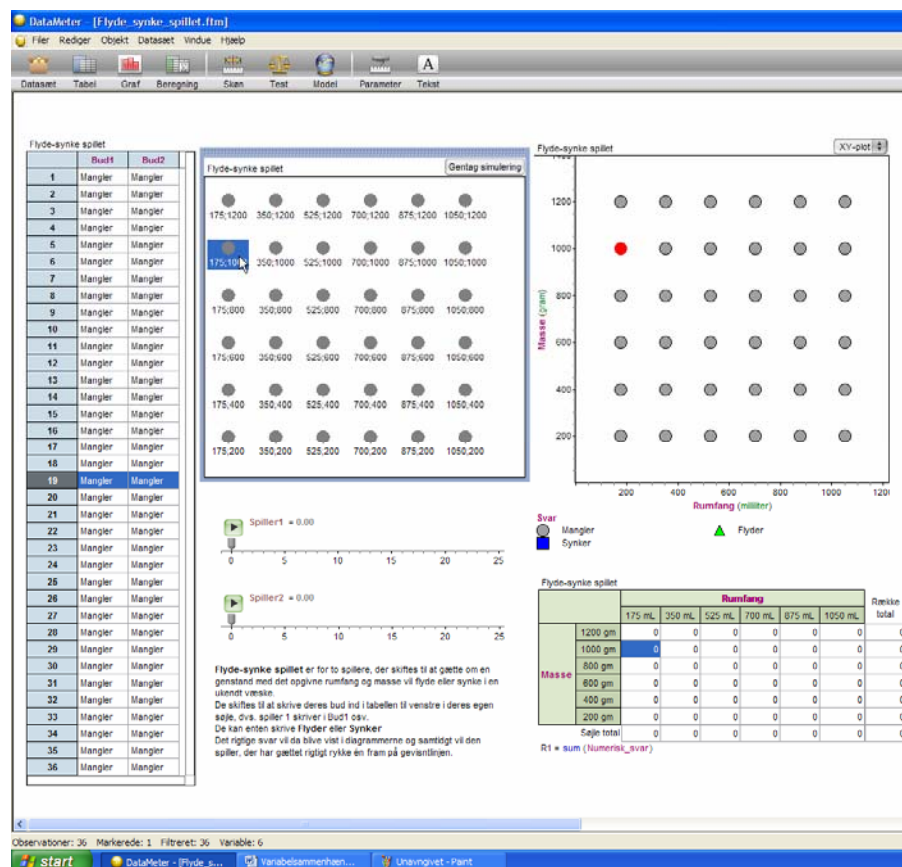


Ved hjælp af parameteren **Antal_terninger** kan du selv styre hvor mange områder du vil observere. Som udgangspunkt får du tildelt 5 områder, dvs. du kaster med 5 terninger. Du kan se terningerne i det øverste vindue, og du kan samtidigt se deres talværdier i tabellen. I dette tilfælde har du altså slået en femmer, en sekser, en femmer, en firer og en toer. Endelig får du oplyst det vigtigste, nemlig at der i disse fem områder er observeret 2 våger og 8 isbjørne. Du kaster terningerne for at indhente nye observationer ved som vist at klikke på **Gentag simulering**.

Din opgave er nu et finde ud af hvilken regel der ligger bag antallet af våger og isbjørne. Når du har fundet reglen bør du teste den indtil du er helt sikker på at du virkeligt har en gyldig regel. Desuden bør du tolke reglen, dvs. du bør give en overbevisende for- klaring på, hvad der er våger og hvad der er isbjørne i dette terninge- spil. Du er ikke tvivl når du først har fået din aha-oplevelse. Indtil da ønskes du god fornøjelse med spillet.

Øvelse 2.4.2: Flyde-synke-spillet

Flyde-synke-spillet handler om et virtuelt eksperiment, hvor en genstand sættes ned i et kar med en ukendt væske, hvorefter man observerer om genstanden flyder eller synker. Denne gang er det et spil for to personer, der skiftes til at gætte på om genstanden flyder eller synker. Gætter man rigtigt får man selv et point, gætter man forkert får modstanderen et point. Dertil har man brug for en ekspert, der kan afgøre om ens bud er korrekt eller forkert. I dette tilfælde er det altså **DataMeter**, der er eksperten, så for at spille flyde-synke-spillet skal du først hente det på **DataMeters** hjemmeside



Spillepladen består af

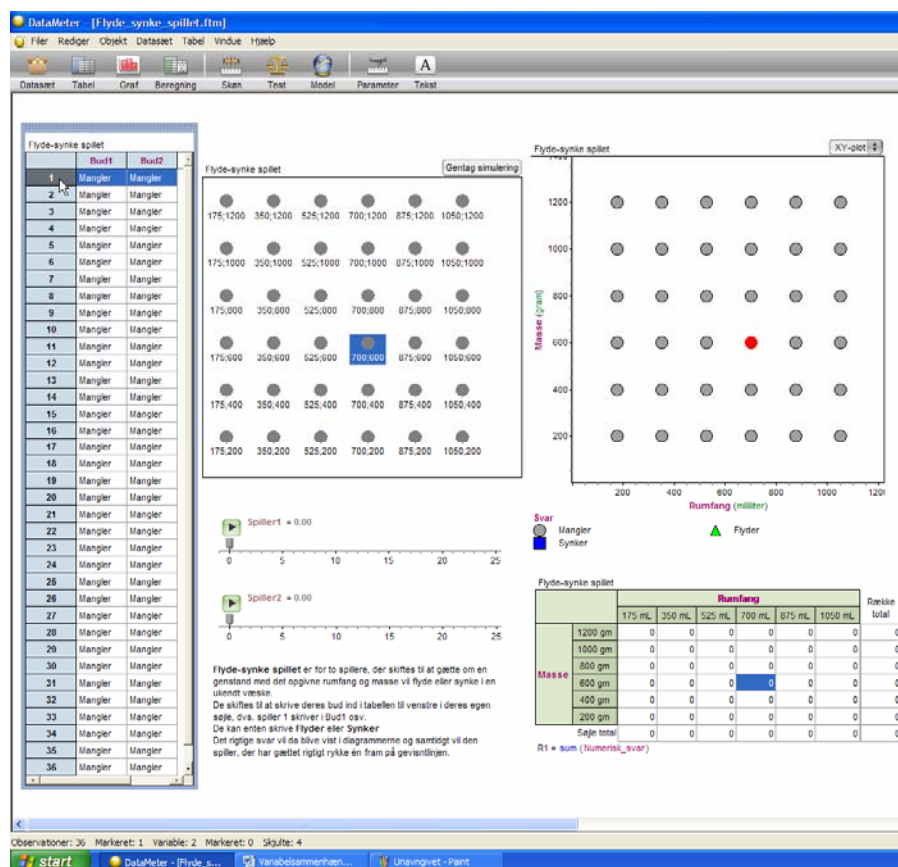
- 1) en tabel med en oversigt over de to spilleres bud
- 2) et diagram over de 36 mulige genstande med deres rumfang og masse noteret
- 3) et grafvindue, der viser sammenhængen mellem rumfang og masse for de 36 genstande
- 4) en krydstabel over sammenhængen mellem rumfang og masse for de 36 genstande

Hvis man som vist klikker på en genstand i et af vinduerne bliver det markeret samtidigt i alle fire vinduer.

I diagrammet, grafvinduet og krydstabellen kan man endvidere se hvilke eksperimenter der allerede er udført. Hvis der mangler et eksperiment bliver det noteret som en grå cirkel, henholdsvis et 0 i krydstabellen. Når eksperimentet er udført viser diagrammet, grafvinduet og krydstabellen yderligere om det blev en flyder (markeret som grøn firkant/trekant, henholdsvis et 1-tal) eller en synker (markeret som en blå firkant, henholdsvis et -1-tal). Diagrammet, grafvinduet og krydstabellen er altså tre ækvivalente fremstillinger af de samme 36 genstande med rumfang fra 175 mL til 1050 mL og masser fra 200 g til 1200 g. Man kan selvfølgelig frit vælge at bruge den af fremstillingerne, der giver det bedste overblik.

Men nok så vigtigt er de to skydere, der holder rede på pointene for de to spillere, idet en spiller får et point, hvis vedkommende enten selv gætter rigtigt eller modstanderen gætter forkert.

Spillet findes i to varianter: En forholdsvis triviell variant hvor de to spillere efter tur vælger hvilken genstand de vil sætte ned i karret ved som vist at klikke på genstanden i enten diagrammet, grafvinduet eller krydstabellen. Men denne variant er faktisk for nem. I stedet bør man skiftevis klikke i tabellen i den rækkefølge buddene er opstillet inde i tabellen, dvs. man klikker først som vist i række 1 ude i indeksfeltet: Derved lyser et tilfældigt eksperiment op



Spiller nr. 1 afgiver derefter sit bud, idet vedkommende enten skriver *Flyder* eller *Synker* i feltet hørende til variabelen **Bud1**. Man kan med det samme se om man har gættet rigtigt, idet det korrekte svar offentliggøres i diagrammet, grafvinduet og krydstabelen

Eksperimentets Tilstand	Diagram	Grafvindue	Krydstabel
<i>Mangler</i>			
<i>Flyder</i>			
<i>Synker</i>			

Samtidigt bliver der tilskrevet points til den af de to spillere, der vandt den første runde. Den aktuelle pointstilling ses i de to dynamiske parametre.

Derefter klikker man i række 2 ude i indekset og nu er det spiller 2's tur til at komme med et bud – *Flyder* eller *Synker* – der skrives i feltet hørende til **Bud2**. Igen kan man umiddelbart se om man har gættet rigtigt, samtidigt med at der bliver tilskrevet points til den af de to spillere, der vandt den anden runde.

Sådan fortsætter spiller 1 med at komme med buddene for alle rækkerne med ulige indeks, 1, 3, 5, ..., mens spiller 2 kommer med buddene for alle rækkerne med lige indeks, 2, 4, 6, På et tidspunkt vil det selvfølgelig være klart, hvem af de to, der har vundet, men fortsæt alligevel med at spille til pladen er fyldt helt ud, idet konkurrencen mellem de to spillere kun er første del af spillet.

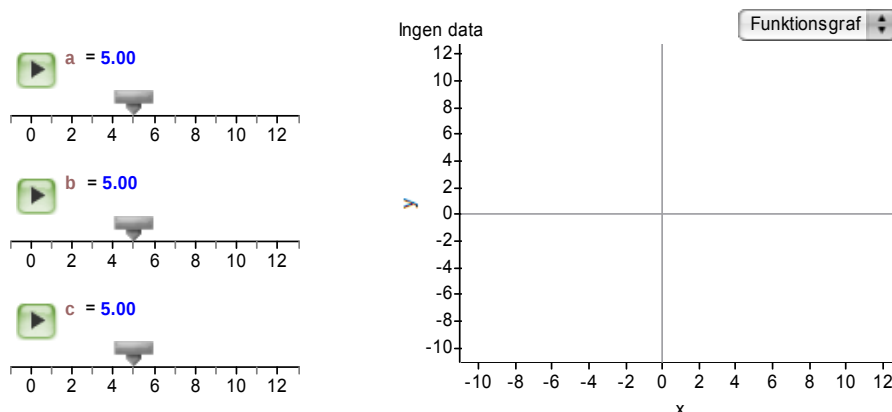
Sidste del af spillet handler om mekanismen bag flyde-synke-spillet:

Hvilken sammenhæng skal der være mellem rumfang og masse for at en genstand flyder henholdsvis synker? Når første del er færdigt er rumfang-masse grafvinduet blevet udfyldt med flydere og synkere, der deler første kvadrant i to områder for henholdsvis flydere og synkere. Hvilken form har disse to områder? Kan du konstruere en enkelt variabel, der er afgørende for om genstanden flyder eller synker?

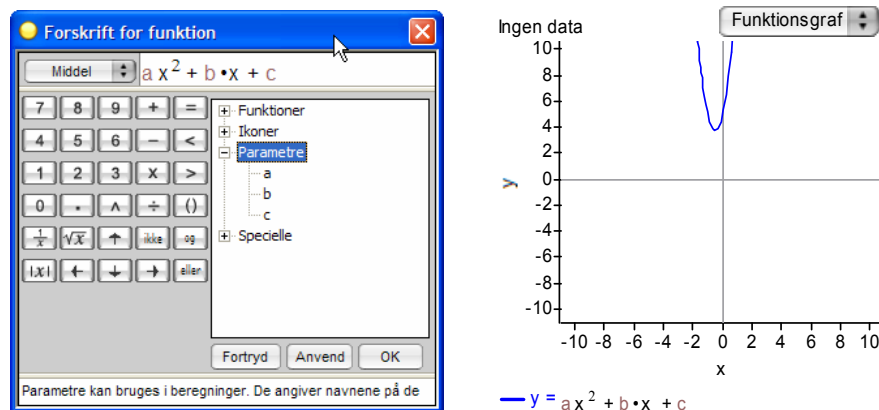
Udfordring: Hvad kan du sige om den ukendte væske? Kunne det fx være vand? Hvis det ikke kan være vand, hvad kan du så i givet fald slutte om væsken i forhold til vand?

Øvelse 2.4.3: variabelkontrol med andengradspolynomier

Parameter-værktøjet egner sig godt til variabelkontrol, hvor man har et fænomen, der afhænger af flere parametre, og derefter undersøger deres indflydelse på fænomenet ved at variere parametrene systematisk én efter én. Vi starter med et eksempel hvor vi bruger **DataMeter** som en almindelig grafregner (dvs. helt uden et datasæt). Grafen for et andengradspolynomium $y = a \cdot x^2 + b \cdot x + c$ er en parabel, hvis form og beliggenhed afhænger af parametrene a , b og c . Men hvordan afhænger formen og beliggenheden af parametrene? For at undersøge dette, trækker vi tre **Parameter**-værktøjer ned fra værktøjshylden og navngiver dem a , b og c . Tilsvarende trækker vi et **Graf**-værktøj ned og skifter graftypen til **Funktionsgraf**

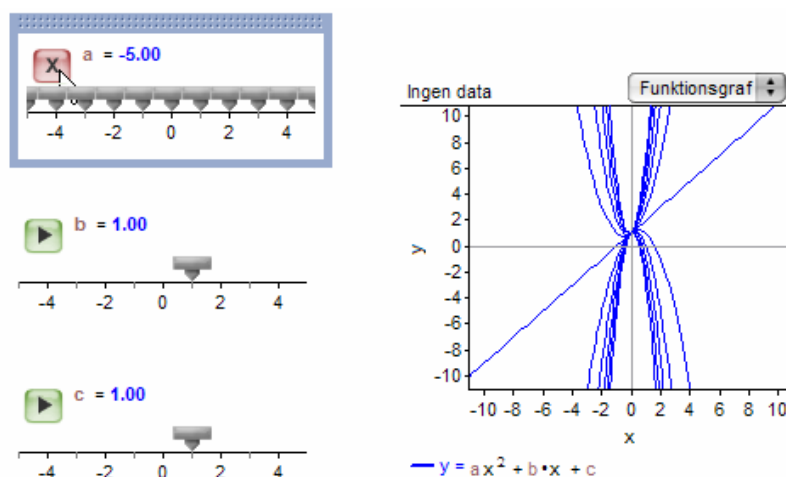


Der dukker da et x-y-kordinatsystem op med et standardvindue, der går fra -10 til 12 langs begge akser. Højreklikker vi i grafrummet kan vi nu vælge menupunktet **Plot funktion** og indskrive ligningen for andengradspolynomiet



Inden vi går i gang med at lege med grafen kan det betale sig at sætte vinduesgrænserne for parametrene til at gå fra -5 til 5 . Det gøres nemmest ved at dobbeltklikke i parameteren og sætte grænserne ved hjælp af inspektøren.

Derefter sættes alle parametrene som udgangspunkt til 1 (klik i parameterværdien og ret den til det ønskede)



Her er det vist som en montage af billeder, der viser hvordan grafbilledet ændres når parameteren a **animeres** ved at trykke på animationsknappen . Hvis det går for hurtigt kan man i stedet trække i skyderen. Når man er færdig med at overveje a 's indflydelse sættes parameterværdien tilbage til 1, hvorefter man på samme måde kan undersøge, hvad der sker når b varieres og derefter når c varieres.

Øvelse 2.4.3a: Hvilken indflydelse har a , b og c så på parablens form og beliggenhed?

Øvelse 2.4.3b: Prøv også at tilføje graferne for såvel den lineære del $y = b \cdot x + c$ som den konstante del $y = c$. Kaster det yderligere lys over den rolle som parametrene b og c spiller for parablens form og beliggenhed?

Der findes også en anden standardform for et andengradspolynomium

$$y = a \cdot (x - h)^2 + k$$

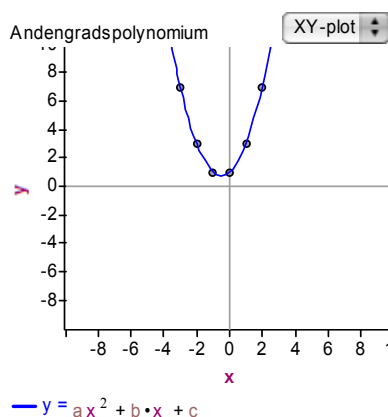
Man siger da at man har omskrevet andengradspolynomiet ved hjælp af en kvadratkomplettering.

Øvelse 2.4.3c: Undersøg på samme måde hvilken indflydelse parametrene h og k har på parablens form og beliggenhed. Det kan betale sig at tilføje graferne med ligningerne $x = h$ og $y = k$, idet grafen for den lodrette linje $x = h$ tilføjes ved at vælge menupunktet **Plot værdi**.

Vi vil nu prøve at kaste yderligere lys over andengradspolynomiets graf med ligningen $y = a \cdot x^2 + b \cdot x + c$, herunder ikke mindst parablens toppunkt, som vi vil prøve at opdage en formel for.

Det kan da betale sig at supplere med et datasæt for andengradspolynomiet. Vi trækker derfor en tabel ned i dokumentet og tilføjer en variabel for successive x -værdier, der spiller rollen af den uafhængige variabel, mens y -værdien givet ved ligningen $y = a \cdot x^2 + b \cdot x + c$ spiller rollen som den afhængige variabel

Andengradspolynomium		
	x	y
=		$a \cdot x^2 + b \cdot x + c$
1	-5	21
2	-4	13
3	-3	7
4	-2	3
5	-1	1
6	0	1
7	1	3
8	2	7
9	3	13
10	4	21
11	5	31



Inspicér graf hørende til A...	
Data	Egenskaber
Parameter	Værdi
punktstør...	5
xNedre	-10
xØvre	10
yNedre	-10
yØvre	10
xModsats...	falsk
yModsats...	falsk
xAutotilpas	falsk
yAutotilpas	falsk

Som vist optræder datapunkterne da som perler på en snor langs parabelen. Det kan nu betale sig at låse koordinatsystemet, så det ikke hele tiden tilpasses datapunkterne i takt med at parametrene a , b og c varieres. Det sker ved at dobbeltklikke i grafrummet og derved vække grafinspektøren, hvorved man får adgang til at slå grafparametrene **xAutotilpas** og **yAutotilpas** fra ved at sætte dem til **falsk**.

Vi vil nu knytte et førstegradspolynomium til parabelen, der angiver hvor hurtigt andengradspolynomiet vokser. Det gør vi ved at udregne den *symmetriske differenskvotient*, dvs. hældningen for sekanten der forbinder det foregående datapunkt med det efterfølgende datapunkt (se **GeoMeter**-figuren den følgende side). Den symmetriske differenskvotient har formen

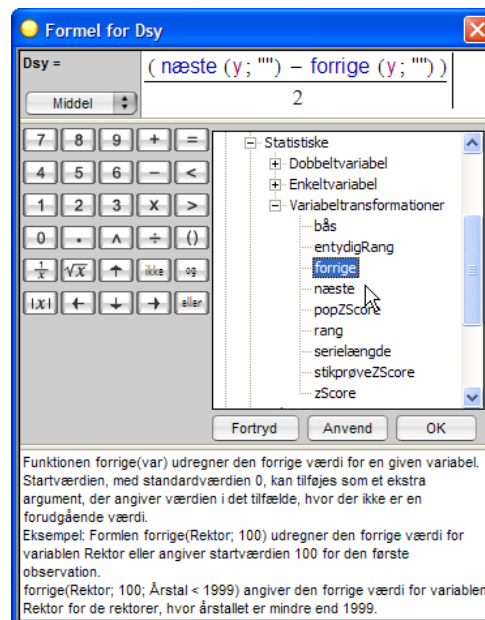
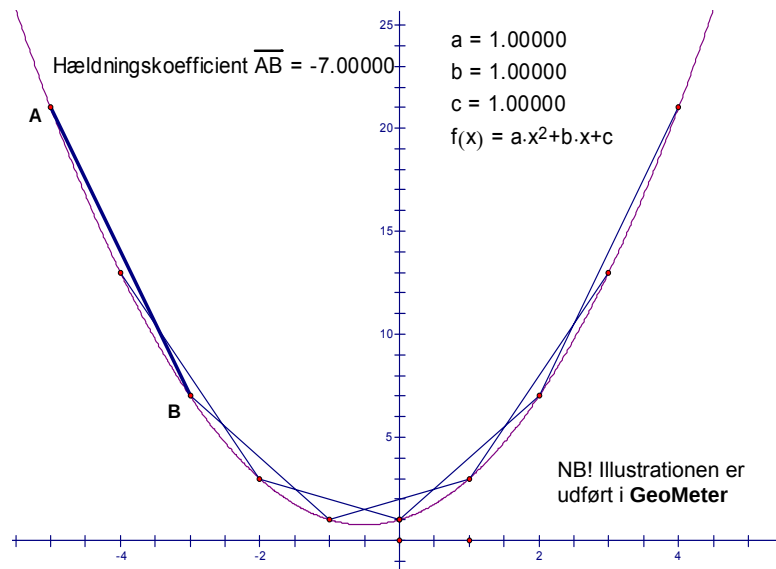
$$D_s y = \frac{næste(y) - forrige(y)}{2}$$

Den udregnes ved hjælp af variabelfunktionerne **næste()** og **forrige()**, der netop henter værdien fra det efterfølgende datapunkt henholdsvis det foregående datapunkt. Da disse funktioner omdanner den givne variabel til en ny variabel, hvor alle værdier er skiftet én plads frem eller én plads tilbage, står de i lommeregneren under

Statistiske funktioner > Variabeltransformationer

(se figuren næste side)

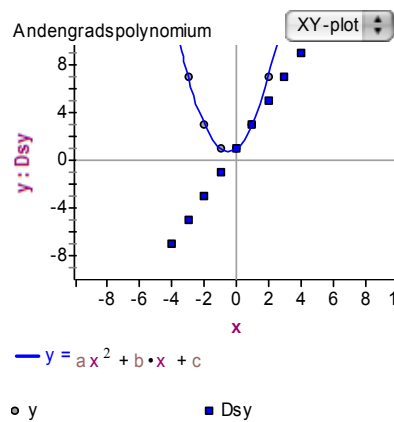
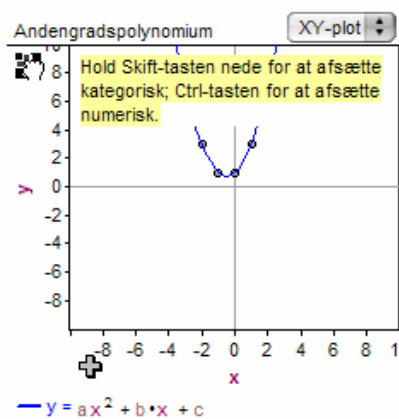
2.4 Variabelsammenhænge: Variabelkontrol



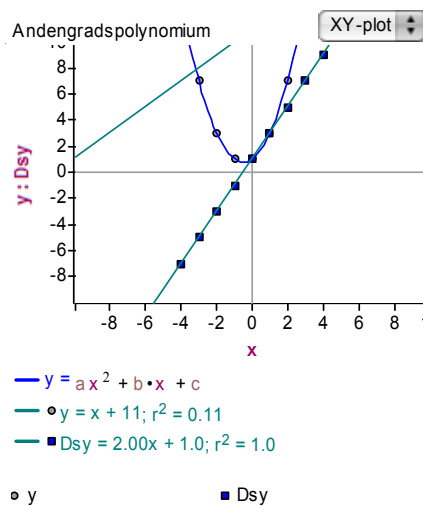
Læg mærke til den ekstra parameter `""'`. Den skal sikre at variabelen `næste(y;""')` er udefineret når vi når frem til det sidste datapunkt, hvor der jo ikke er en efterfølgende værdi. Undlades `""'` får variabelen standardværdien 0, hvilket i dette tilfælde er noget kludder.

Vi tilføjer nu grafen for den symmetriske differenskvotient sammen med grafen for andengradspolynomiet ved at trække variabelen **Dsy** ind på andenaksen og slippe den på $\div$ -tegnet, så den ikke smider grafen for andengradspolynomiet ud

Andengradspolynomium			
	x	y	Dsy
=		$ax^2 + b \cdot x + c$	$\frac{(\text{næste}(y; "") - \text{forrige}(y; ""))}{2}$
1	-5	21	
2	-4	13	-7
3	-3	7	-5
4	-2	3	-3
5	-1	1	-1
6	0	1	1
7	1	3	3
8	2	7	5
9	3	13	7
10	4	21	9
11	5	31	



Resultatet er en ny perlekæde der tydeligvis ligger på ret linje. Den symmetriske differenskvotient er altså grafen for en ret linje. Men den kan vi jo finde ligningen for ved at vælge menupunktet **Mindste kvadraters linjer** i grafrummet. I dette tilfælde finder vi altså hældningen 2 og skæringen med andenaksen 1. Men hældningen og skæringen for denne rette linje afhænger jo af parablen, dvs. af parametrene a , b og c . Det er disse sammenhænge vi vil prøve at finde.

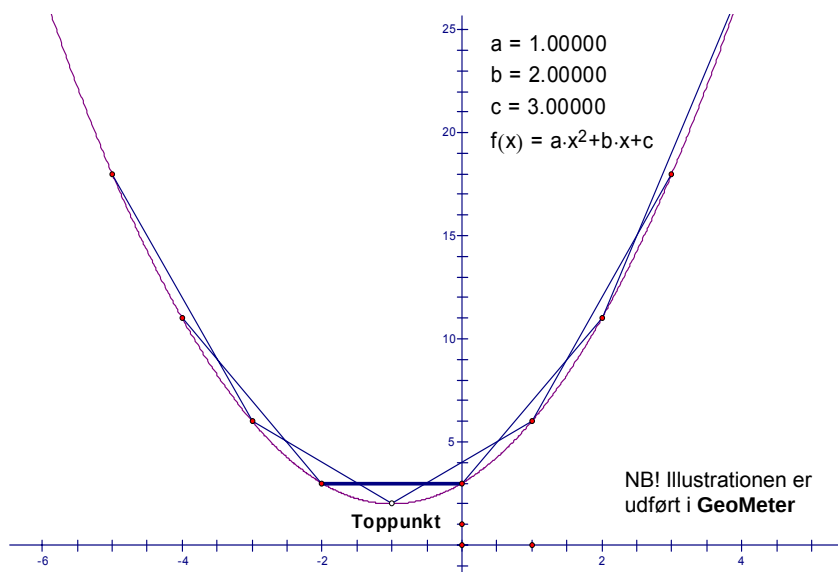


Øvelse 2.4.3e: Prøv ved hjælp af variabelkontrol at finde ud af hvordan hældningen for den rette linje (x , **Dsy**) og skæringen med andenaksen for den rette linje (x , **Dsy**) afhænger af parametrene a , b og c . Opstil herved en ligning for denne rette linje, hvori indgår parametrene for andengradspolynomiet.

Vink: Sørg for at parametrene a , b og c kun antager heltallige værdier. Det kan fx gøres elegant ved at vække parameterinspektørerne for a , b og c efter tur og sætte værdien af inspektørparameteren **BegrænsTilMultiplaAf** til 1.

Inspicér parameter		
Egenskaber		
Parameter	Værdi	For...
a	1	
Opdateringsfrekvens		
Nedre_	-5	
Øvre_	5	
BegrænsTilMultiplaAf	1	
VendAksen	falsk	

Øvelse 2.4.3f: Kig nu igen på parablen med dens toppunkt



Hvad bliver hældningen for det linjestykke, der forbinder det foregående grafpunkt med det efterfølgende grafpunkt, hvis vi netop tager udgangspunkt i toppunktet for parablen? Hvordan kan du nu udnytte ligningen fra den foregående øvelse til at finde en formel for toppunktets x -værdi?

Øvelse 2.4.3g: Kig nu igen på parablen med ligningen

$$y = a \cdot x^2 + b \cdot x + c.$$

Tilføj såvel den vandrette som den lodrette linje gennem toppunktet, idet du udnytter den ligning du har fundet i øvelse f til den lodrette linje (som tegnes med menupunktet **Plot værdi**). Hvilken forbindelse er der mellem øvelse g og øvelse c?

3. Regressionsmodeller

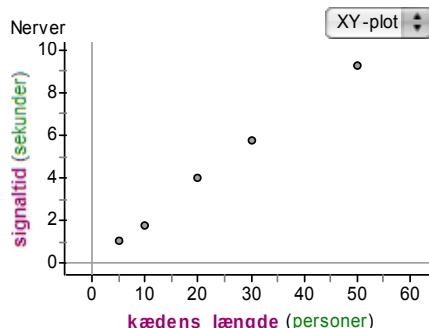
3.1 Lineær regression

Vi vil starte med at prøve at kaste nyt lys over mindste kvadraters linje med ligningen $y = a \cdot x + b$, idet vi vil præsentere metoden, så den nemt kan generaliseres til andre nærtbeslægtede modeller med to parametre a og b . Som eksempel vil vi se på et simpelt fornøjeligt eksperiment, hvor man undersøger hastigheden af et nervesignal. Vi stiller derfor deltagerne op i en kæde (som gerne må være en rundkreds), så de holder hinanden i hænderne – med lukkede øjne!



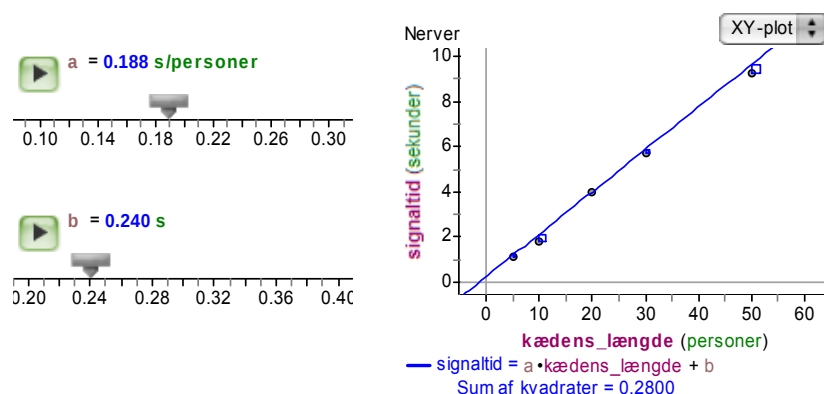
På et givet tegn startes signalet ved, at den første deltager giver den anden deltager et forsigtigt klem med hånden, som så giver klemmet videre til den tredje deltager osv. Der tages tid på hvor lang tid signalet er om at nå sidste person i kæden. Her er nogle måledata fra en sådan rundkreds, hvor vi har opfattet kædens længde som den uafhængige variabel (idet den kan vi jo selv bestemme) og signaltiden som den afhængige variabel (som vi fx måler med en mobiltelefon)

Nerver		
	kædens_længde	signaltid
enhed	personer	sekunder
1	5 personer	1.09 s
2	10 personer	1.80 s
3	20 personer	4.01 s
4	30 personer	5.72 s
5	50 personer	9.27 s



Det kunne jo godt ligne en lineær sammenhæng, så vi vil nu prøve om vi systematisk fra bunden af kan finde den bedste rette linje. Her-til indfører vi to dynamiske parametre a og b svarende til linjens hældning og skæringen med andenaksen. Her svarer hældningen a til den tid signalet er om at passere en enkelt person i kæden, mens startværdien b svarer til noget i retning af den reaktionstid, der er involveret for at opfange startsignalet og afslutte tidtagningen.

Vi indskrives ligningen $\text{signaltid} = a \cdot \text{Kædens_længde} + b$ ved hjælp af menupunktet **Plot funktion** i grafrummet. Tilsvarende vælger vi menupunktet **Vis kvadrater** i grafrummet. Ved at fedte med parametrene kan vi nu prøve at gøre kvadratsummen så lille som mulig



Men det er klart at det nemt bliver noget fedteværk, når man skal styre to parametre på én gang. Vi vil derfor prøve om vi kan eliminere den ene af dem. Her er det nemmest at eliminere startværdien b . Har vi først fundet et bud på hældningen a kan vi nemt finde den tilhørende værdi af startværdien b ved hjælp af omskrivningen

$$\text{signaltid} - a \cdot \text{kædens_længde} = b$$

Vi indfører nu en ny afhængig variabel, som vi kalder **b_værdi** og som udregner startværdien ud fra de enkelte datapunkter svarende til at vi trækker linjer med hældningen a gennem datapunkterne

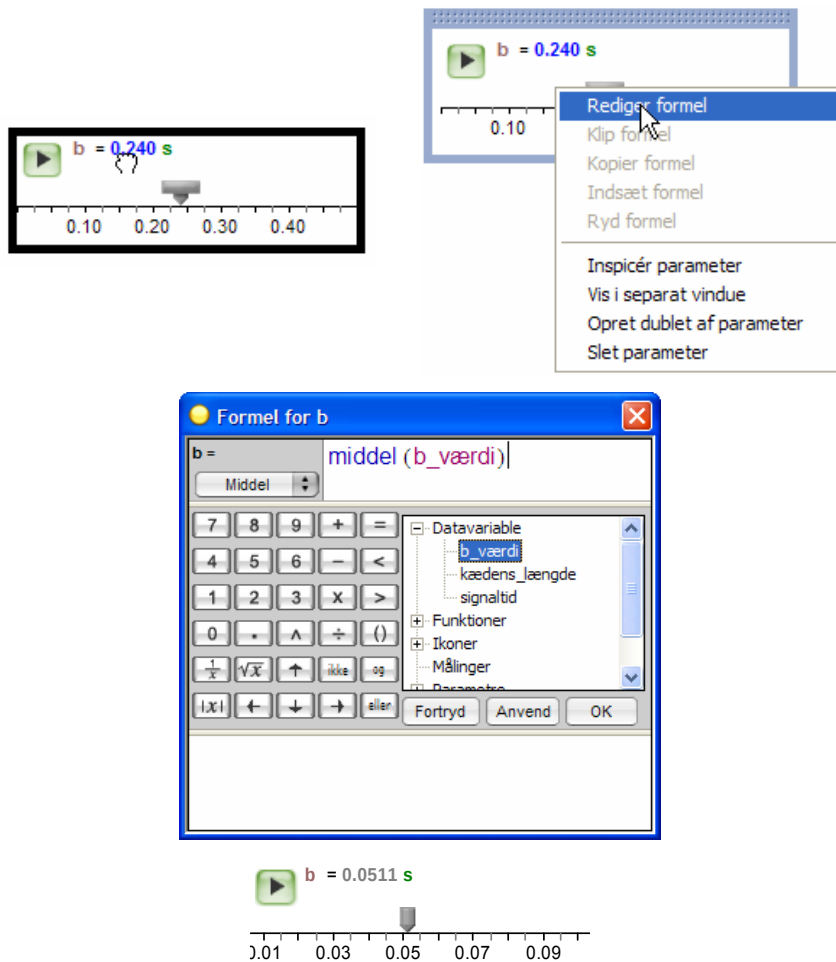
Nerver			
	kædens_længde	signaltid	b_værdi
enhed	personer	sekunder	sekunder
=			signaltid - a · kædens_længde
1	5 personer	1.09 s	0.149362 s
2	10 personer	1.80 s	-0.0812766 s
3	20 personer	4.01 s	0.247447 s
4	30 personer	5.72 s	0.0761702 s
5	50 personer	9.27 s	-0.136383 s

Det kan godt se forvirrende ud med fem bud på startværdien, men hvis **b_værdi** med tilnærmelse er konstant kan vi jo trække en fælles værdi ud som middelværdien af de fem bud. Det er på denne måde vi eliminerer b : Først trækker vi datasættet ind i parameteren ved at gribe titlen og slippe den inde i parameterområdet. Dernæst vælger vi menupunktet **Rediger formel** og indskrives formelen

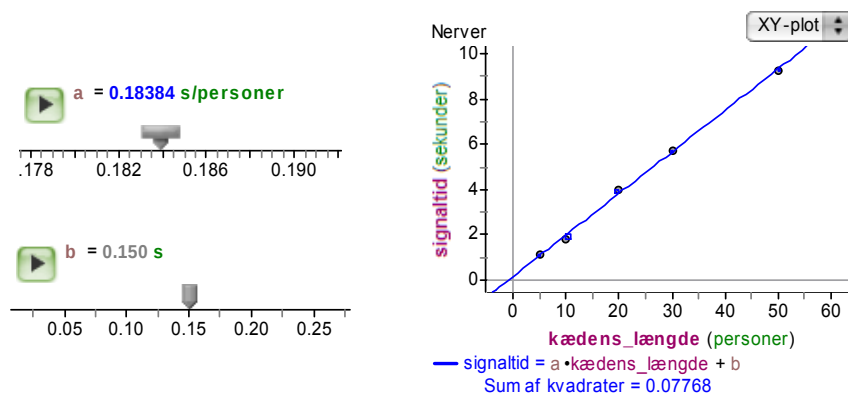
$$b = \text{middel}(\text{b_værdi})$$

i lommeregneren. Derved binder vi værdien af parameteren b til den middelværdi, der igen afhænger af parameteren a , som jo indgår i formelen for **b_værdi**

3.1 Regressionsmodeller: Lineær regression

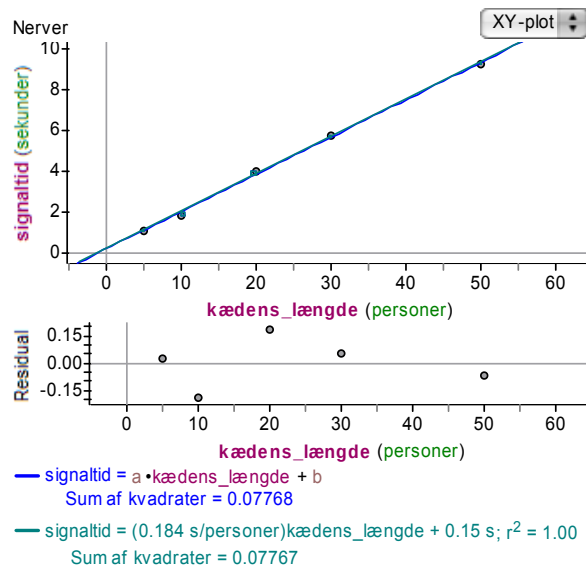


Men så er der kun hældningen a tilbage, som vi kan variere på. Vi kan zoome ind på parameteren a ved at holde **CTRL**-tasten nede, hvorved markøren omdannes til et forstørrelsesglas med et plus. Herefter kan vi regulere på hældningen a til kvadratsummen er gjort mindst mulig



Vi ser altså at den mindste kvadratsum er 0.07768 og at den tilhørende hældning er givet ved $a = 0.18384$ s/person, dvs. i gennemsnit bruger nervesignalet 0.18 s på at passere en person i kæden. Tilsvarende er startværdien b givet ved 0.150 s.

Det svarer rimeligt godt til den indbyggede mindste kvadraters linje, hvor vi også har tilføjet et residualplot, så man kan se at afvigelserne ligger tilfældigt fordelt omkring 0



Godt nok er det lykkedes for mindste kvadraters linje at komme helt ned på kvadratsummen 0.07767, men hvis vi fedter videre med indzoomninger kan vi såmænd også få denne lille forbedring med. Så vi har ramt værdien af a og b rimeligt præcist. Vi kan få oplyst mere præcise værdier for a , b og forklaringsgraden, hvis vi åbner en beregningsboks og anvender de detaljerede funktioner for mindste kvadraters linje

Nerver	
	0.1838125 s/personer
	0.1503125 s
	0.99820716
R1 = hældning (kædens_længde; signaltid)	
R2 = skæring (kædens_længde; signaltid)	
R3 = forklaringsgrad (kædens_længde; signaltid)	

De samme værdier kan i princippet findes ved fortsat indzoomning, men det er ikke besværet værd. Vi har tidligere set på forklaringsgraden i detaljer, så den vil vi ikke kommentere yderligere. Men vi kan altså konkludere at vi nu har fundet en metode til i detaljer at genskabe mindste kvadraters linje. Denne metode vil vi nu anvende på andre tilsvarende modeller – først og fremmest potensproportionaliteter og eksponentielle vækstmodeller.

3.2 Potensregression

Linen med brudstyrke-garanti

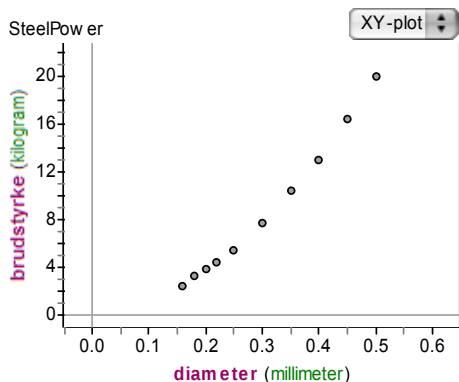
Stærk og smidig!

SteelPower er linen for dig, der *hader* »overraskelser«, når de store hugger...

Kilde: Fisk & Fri, nr. 1, februar 1998.

Vi tager denne gang udgangspunkt i det viste eksempel, hvor man får oplyst en tabel over sammenhængen mellem fiskelinens diameter i millimeter og dens brudstyrke i kg. Man hænger altså tungere og tungere lodder i enden af linerne indtil de sprænger, hvorefter man noterer massen af det kritiske lod. Disse data kan vi som vist uden videre skrive ind i en tabel.

Tilsvarende kan vi få tegnet en graf ved at trække variablene ind på akserne. Vi trækker variabelen **diameter** ind som uafhængig variabel på førsteaksen (da vi selv bestemmer hvilken type vi vil teste) og **brudstyrke** som afhængig variabel på andenaksen (da vi måler brudstyrken)



Damyl SteelPower
– stærk som stål!

Diameter (i mm)	Brudstyrke (i kg)
0,16	2,40
0,18	3,15
0,20	3,80
0,22	4,30
0,25	5,30
0,30	7,70
0,35	10,40
0,40	13,00
0,45	16,30
0,50	20,00

SteelPower

	diameter	brudstyrke
enhed	millimeter	kilogram
=		
1	0.16 mm	2.40 kg
2	0.18 mm	3.15 kg
3	0.20 mm	3.80 kg
4	0.22 mm	4.30 kg
5	0.25 mm	5.30 kg
6	0.30 mm	7.70 kg
7	0.35 mm	10.40 kg
8	0.40 mm	13.00 kg
9	0.45 mm	16.30 kg
10	0.50 mm	20.00 kg

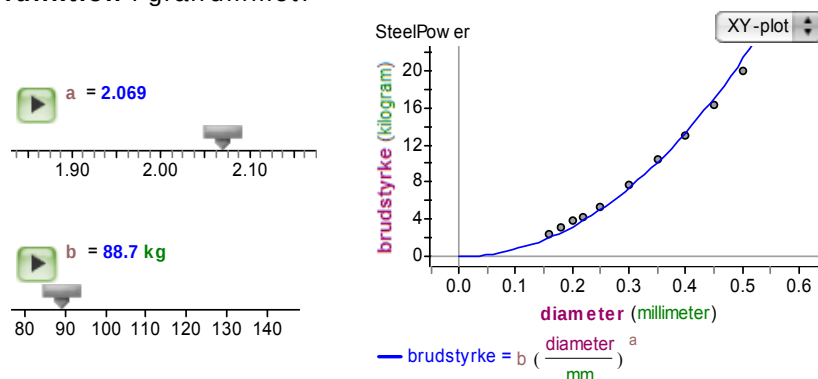
Vi vil altså undersøge om denne sammenhæng kan beskrives ved en potensproportionalitet på formen

$$y = b \cdot x^a, \text{ dvs. } \text{brudstyrke} = b \cdot \left(\frac{\text{diameter}}{\text{mm}} \right)^a$$

Læg mærke til at vi dividerer den uafhængige variabel **diameter** med enheden *mm*. Det er prisen for at bruge enheder i ligningen.

Første fase: Grafisk tilpasning af parametre

Vi indfører nu to dynamiske parametre: Én for eksponenten a og én for startværdien b og indskrives variabelsammenhængen med ligningen $\text{brudstyrke} = b \cdot (\text{diameter} / \text{mm})^a$ ved at vælge menupunktet **Plot funktion** i grafrummet:



Men selv med en del øvelse er det svært at tilpasse to parametre på én gang, så det er her vi skal forbedre processen. Det gør vi ved hjælp af følgende simple betragtning: For en given værdi af eksponenten a kan vi finde startværdien b ved hjælp af følgende simple omskrivning:

$$b = \frac{y}{x^a}, \text{ dvs. } b = \frac{\text{brudstyrke}}{\left(\frac{\text{diameter}}{\text{mm}}\right)^a}$$

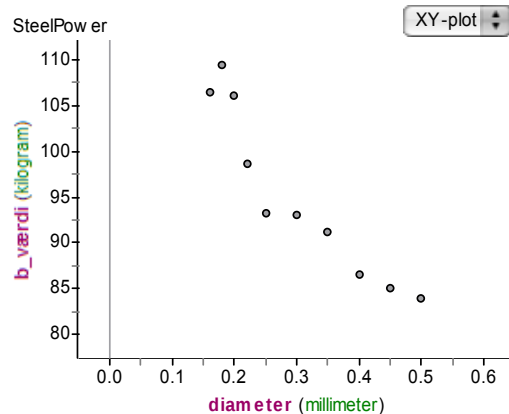
Vi indfører derfor en ny variabel, **b_værdi**, der netop udregner startværdien for hvert enkelt datapunkt

SteelPower

	diameter	brudstyrke	b_værdi
enhed	millimeter	kilogram	kilogram
=			$\frac{\text{brudstyrke}}{\left(\frac{\text{diameter}}{\text{mm}}\right)^a}$
1	0.16 mm	2.40 kg	106.387 kg
2	0.18 mm	3.15 kg	109.434 kg
3	0.20 mm	3.80 kg	106.158 kg
4	0.22 mm	4.30 kg	98.627 kg
5	0.25 mm	5.30 kg	93.3121 kg
6	0.30 mm	7.70 kg	92.9666 kg
7	0.35 mm	10.40 kg	91.276 kg
8	0.40 mm	13.00 kg	86.5528 kg
9	0.45 mm	16.30 kg	85.0533 kg
10	0.50 mm	20.00 kg	83.9191 kg

Har vi fundet et godt bud på eksponenten a vil dette give et godt bud på startværdien b , dvs. **b_værdi** vil med tilnærmelse være konstant.

I dette tilfælde er vores bud på a fx *ikke* særligt overbevisende, da der er en tydelig tendens til at **b_værdi** falder fra ca. 110 kg til ca. 85 kg hen gennem datasættet



Men uanset om vi har et godt bud på a eller ej kan vi trække et rimeligt bud på værdien af b ud af tabellen som gennemsnittet af variabelen **b_værdi**, dvs. middelværdien af alle disse bud på startværdien b . Denne værdi kan vi så bruge til at binde parameteren b , så vi kun har en fri parameter tilbage som skal varieres.

Teknisk bemærkning: Formlen for a

Hvis man kender formelen for a , dvs.

$$a = \frac{\log(y_2) - \log(y_1)}{\log(x_2) - \log(x_1)} = \frac{\log\left(\frac{y_2}{y_1}\right)}{\log\left(\frac{x_2}{x_1}\right)}$$

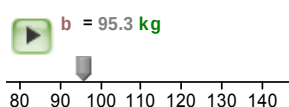
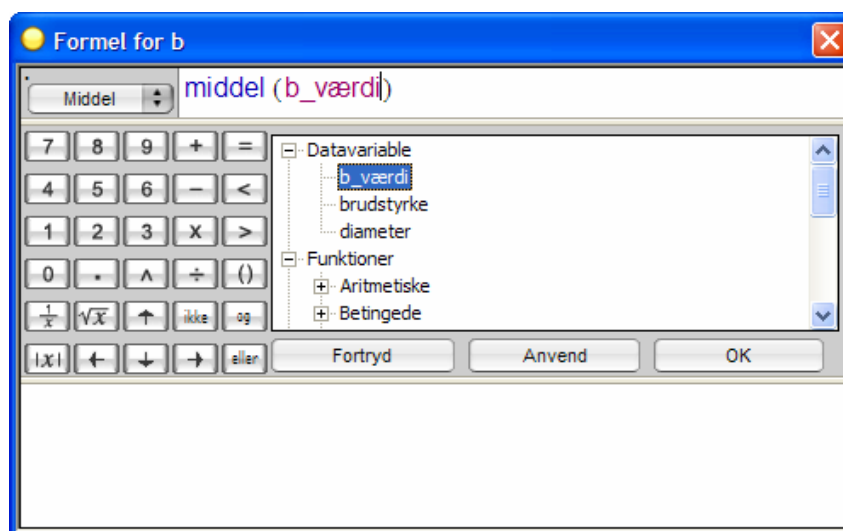
kan vi bruge datameter til at finde et rimeligt udgangspunkt for a , fx ved at bruge det første og det sidste datapunkt. Ved at trække datasættet ind i en beregningsboks får vi adgang til variablene fra datasættet og kan indskrive formelen for eksponenten a

SteelPower	
	1.8608037

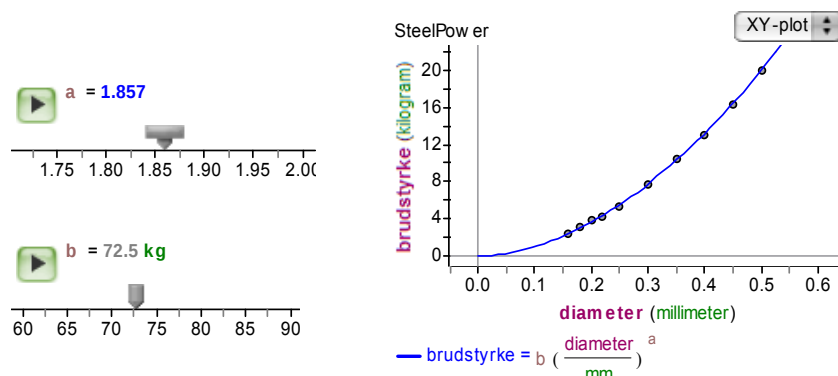
$$R1 = \frac{\log\left(\frac{\text{sidste (brudstyrke)}}{\text{første (brudstyrke)}}\right)}{\log\left(\frac{\text{sidste (diameter)}}{\text{første (diameter)}}\right)}$$

Men det er altså ikke noget, der er en forudsætning for det følgende.

Vi binder altså nu parameteren b til værdien **middel(b_værdi)**. Det kræver at vi først trækker datasættet ind i parameterboksen ved at gribe titlen og slippe den inde i parameterområdet. Derefter kan formelen skrives direkte ind



Læg mærke til at skyderen mistede sit vandrette håndtag. Den kan ikke længere rykkes med håndkraft, idet parameteren b nu er *bundet*: Den afhænger af den frie parameter a og kan derfor kun ændres i takt med a . Det er nu en smal sag at tilpasse den frie parameter a indtil grafen for potensproportionalitetsmodellen ser ud til at gå gennem datapunkterne

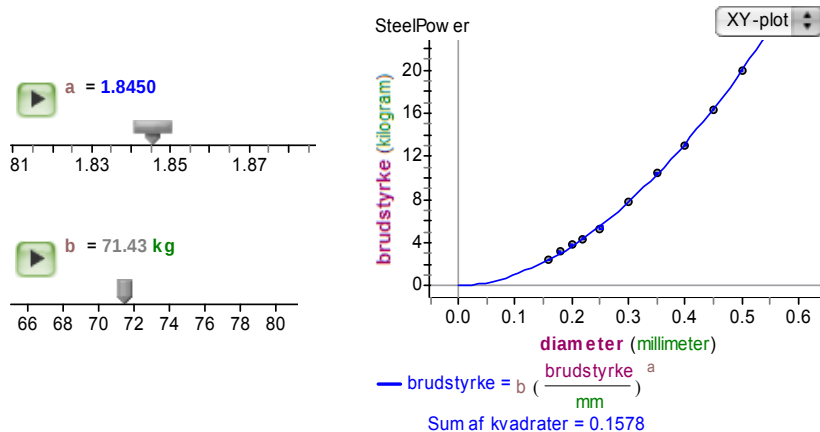


Undervejs kan man passende zoome ind på de to parametre ved at holde **CTRL**-tasten nede (hvorved markøren ændres til et plus i en cirkel).

Dermed er vi igennem den første fase i modellen og har fundet et rimeligt bud på værdien af parametrene a og b og har samtidigt fået et godt visuelt check på at potensproportionalitetsmodellen er rimelig.

Anden fase: Mindste kvadraters metode

Vi går så i gang med finjusteringen af parameteren a ved hjælp af *mindste kvadraters metode*. Hertil tilføjer vi kvadraterne på afvigelserne ved at vælge menupunktet **Vis kvadrater** i grafrummet

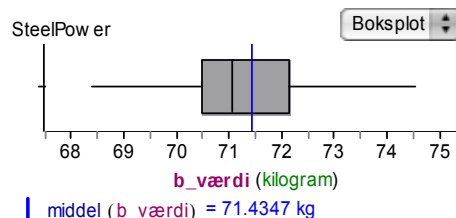
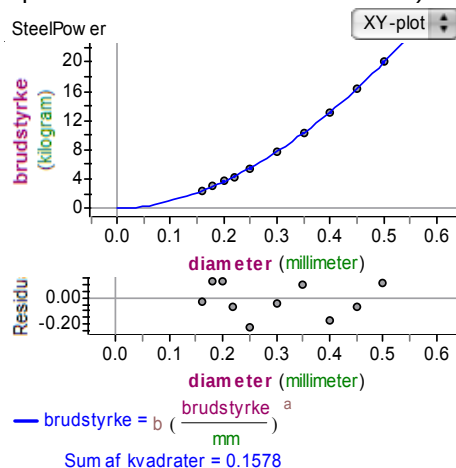


Vi får da automatisk oplyst summen af kvadraterne og efter lidt ind-zoomning kan vi gøre summen mindst muligt. Den mindste kvadratsum er da som vist 0.1578. Vi har nu fastlagt parametrene a og b så præcist som muligt. Vi ser altså at b har værdien 71.43 kg, som svarer til brudstyrken for et tovværk med diameteren 1 mm. Tilsvarende har eksponenten a værdien 1.8450 (altså rimeligt tæt ved 2, dvs. brudstyrken er tilnærmelsesvis proportional med tværsnitsarealet).

Vi kan også tilføje et *residual-plot* for at se om de resterende afvigelser ligger tilfældigt fordelt. Det sker ved at vælge menupunktet **Tegn residualplot** i grafrummet.

Vi ser da at residualerne varierer rimeligt tilfældigt omkring nul. Der er ikke noget tydeligt mønster og residualerne ligger typisk mellem ± 0.10 kg, og det angiver dermed den typiske usikkerhed i modellen.

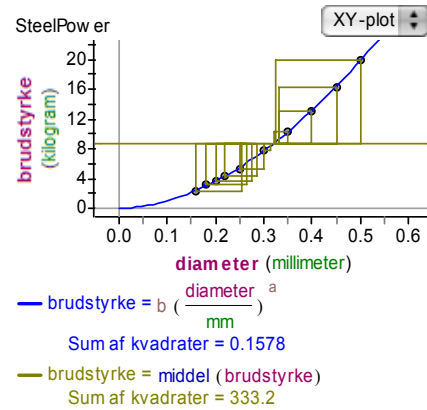
Et boksplot for startværdien viser tilsvarende at b -værdierne ligger rimeligt samlet med en usikkerhed af størrelsesordenen 1.0 kg. Læg dog mærke til at fordelingen er højreskæv, hvilket trækker middelværdien en anelse op over medianen.



Tredje fase: Forklaringsgraden

Vi slutter med at finde forklaringsgraden for proportionalitetsmodellen. Hertil sammenligner vi den med den mest modsatte model, **nulhypotesen**, ifølge hvilken variationerne i brudstyrken er rent tilfældige og derfor i virkeligheden slet ikke afhænger af diameteren af fiskelinen. I så fald svinger brudstyrkerne tilfældigt omkring deres middelværdi og vi tilføjer derfor denne ved hjælp af menupunktet **Plot funktion** i graf rummet. Som forventet er summen af kvadraterne meget større under nulhypotesen. Vi har en god model. Vi kan se at de residualer der er tilbage i potensproportionalitetsmodellen

kun udgør 0.05% af afvigelserne under nulhypotesen. Vi kan derfor forklare 99.95% af dataenes variation ved hjælp af vores potensproportionalitetsmodel. Resten skyldes tilfældige variationer (jfr. residualplottet). Forklaringsgraden for vores potensproportionalitetsmodel er derfor $R^2 = 99.95\%$.



Ingen data

Slip en variabel her	
	0.00047358944
	0.99952641

$$R1 = \frac{0.1578}{333.2}$$

$$R2 = 1 - \frac{0.1578}{333.2}$$

Teknisk bemærkning: Om enheder og potensproportionalitet

I det foregående har vi arbejdet konsekvent med enheder, idet vi har skrevet ligningen for potensproportionaliteten på formen

$$y = b \cdot \left(\frac{x}{\text{enhed}} \right)^a$$

Derved sikres at eksponenten a ikke får nogen enhed (dvs. a er dimensionsløs) og at b netop får samme enhed som y , hvilket er i overensstemmelse med b 's rolle som startværdi for y , dvs. b angiver netop værdien for y , når x har værdien 1 enhed. Mens eksponenten a er uafhængig af valget af enheder, bliver værdien for b derfor stærkt afhængig af valget af enheder. Der er selvfølgelig stor forskel på brudstyrken for et tov med diameteren 1 mm sammenlignet med brudstyrken for et tov med diameteren 1 cm.

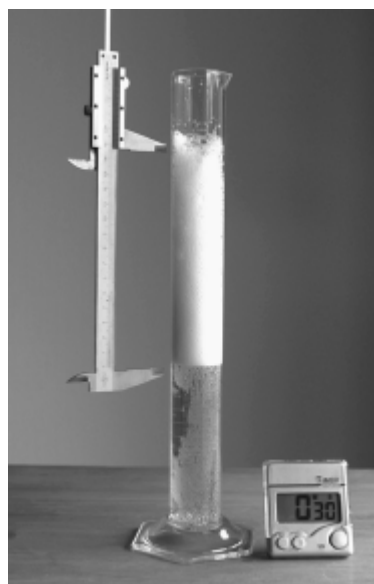
Denne måde at skrive ligningen på stemmer også godt overens med den generelle ligning for en potensproportionalitet gennem et givet punkt (x_0, y_0) , der kan skrives på formen

$$y = y_0 \cdot \left(\frac{x}{x_0} \right)^a$$

3.3 Eksponentiel regression

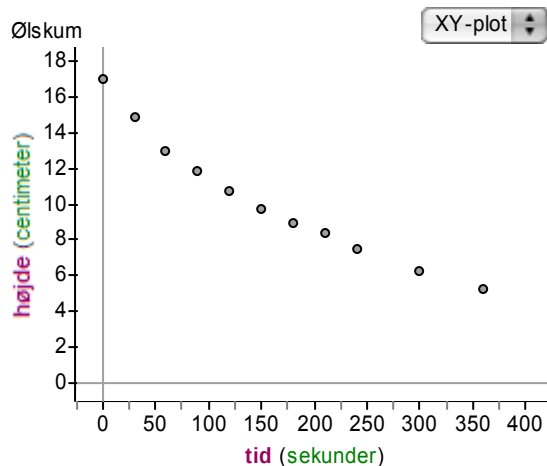
t	0	30	60	90	120	150	180	210	240	300	360
h	17,0	14,9	13,2	11,9	10,7	9,7	8,9	8,4	7,5	6,3	5,2

Vi tager denne gang udgangspunkt i et eksempel med ølskum, hvor man får oplyst højden h af ølskummet i cm som funktion af den tid t i sekunder, der er gået siden den velskummende øl blev hældt op i et måleglas. Vi vil nu undersøge om sammenhængen mellem tiden og højden med rimelighed kan beskrives ved en eksponentiel vækstmodel. Vi skriver derfor de oplyste data ind i en tabel som vist. Derefter trækkes **tid** ind som uafhængig variabel på førsteaksen (idet vi som vist nemt kan styre tiden med et stopur) og **højde** ind som afhængig variabel på andenaksen (som vi måler med den viste skydelære)



Ølskum

	tid	højde
enhed	sekunder	centimeter
=		
1	0 s	17.0 cm
2	30 s	14.9 cm
3	60 s	13.0 cm
4	90 s	11.9 cm
5	120 s	10.7 cm
6	150 s	9.7 cm
7	180 s	8.9 cm
8	210 s	8.4 cm
9	240 s	7.5 cm
10	300 s	6.3 cm
11	360 s	5.2 cm



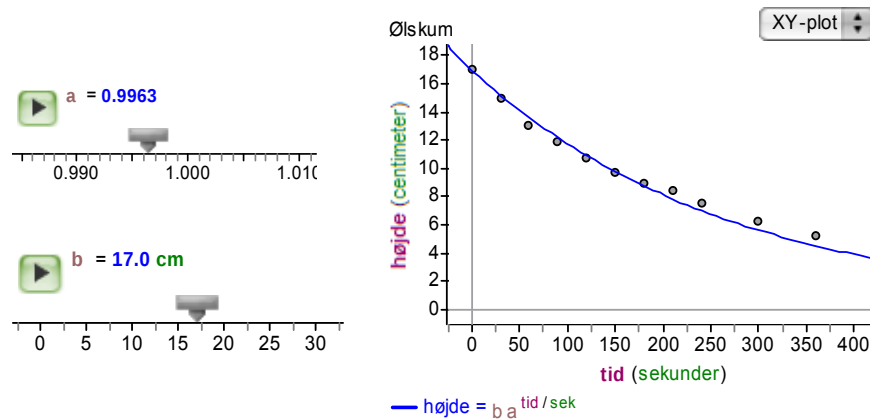
Vi skal derfor bestemme ligningen for denne eksponentielle vækstmodel. Den er på formen

$$y = b \cdot a^x, \text{ dvs. } \text{højde} = b \cdot a^{\frac{\text{tid}}{\text{sek}}}$$

Læg mærke til at vi dividerer den uafhængige variabel **tid** med enheden sek. Det er prisen for at bruge enheder i den eksponentielle vækstmodel.

Første fase: Grafisk tilpasning af parametre

Vi indfører nu to dynamiske parametre: Én for fremskrivningsfaktoren a og én for startværdien b og indskrives variabelsammenhængen med ligningen **højde** = $b \cdot a^{\text{tid} / \text{sek}}$ ved at vælge menupunktet **Plot funktion** i grafrummet



Her finder vi et rimeligt bud på b ved at indsætte startværdien fra tabellen, dvs. 17.0 cm. Værdien kan så efterfølgende justeres.

Teknisk bemærkning: Formlen for a

Hvis man kender formelen for a , dvs.

$$a = x_2 - x_1 \sqrt{\frac{y_2}{y_1}} = \left(\frac{y_2}{y_1} \right)^{\frac{1}{x_2 - x_1}}$$

kan vi bruge datameter til at finde et rimeligt udgangspunkt for a , fx ved at bruge det første og det sidste datapunkt. Ved at trække datasættet *Ølskum* ind i en beregningsboks får vi adgang til de variable og kan skrive formelen ind

Ølskum	
	0.99671498

$$R1 = \frac{\text{sidste (højde)}}{\text{første (højde)}}^{\frac{1}{\left(\text{sidste } \left(\frac{\text{tid}}{\text{sek}} \right) - \text{første } \left(\frac{\text{tid}}{\text{sek}} \right) \right)}}$$

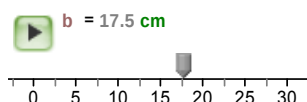
Men det er altså ikke noget, der er en forudsætning for det følgende.

Men som sædvanlig er det svært at tilpasse to parametre på en gang, så igen forbedrer vi processen kraftigt ved hjælp af følgende simple betragtning: For en given værdi af parameteren a kan vi finde b ved hjælp af følgende simple omskrivning

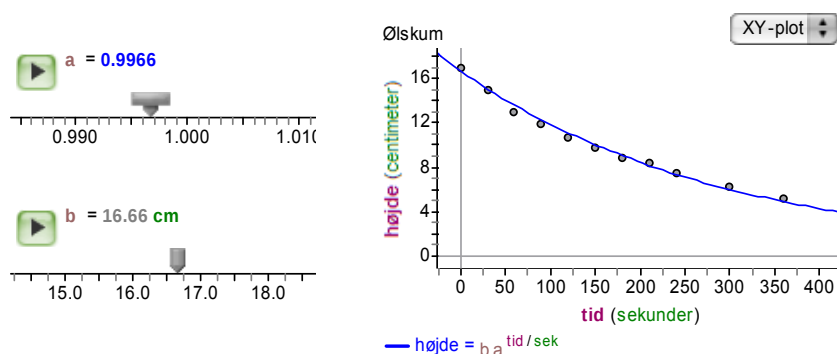
$$b = \frac{y}{a^x}, \text{ dvs. } b = \frac{\text{højde}}{a^{\text{tid} / \text{sek}}}$$

Vi indfører derfor en ny variabel, **b_værdi**, (eller startværdi), der netop udregner b -værdien for hvert enkelt datapunkt. Hvis vi har fundet et godt bud på fremskrivningsfaktoren a vil dette give et godt bud på b , dvs. **b_værdi** vil med tilnærmelse være konstant. Men uanset om vi har et godt bud på a eller ej kan vi trække et rimeligt bud på værdien af b ud som gennemsnittet, dvs. middelværdien af alle disse bud på b . Denne b -værdi kan vi så bruge til at binde parameteren b , så vi kun har én fri parameter tilbage. Vi binder derfor parameteren b til **middel(b_værdi)**. Det kræver at vi trækker datasættet *Ølskum* ind i parameteren ved at gribe titlen og slippe den inde i parameterområdet. Derefter kan formelen skrives direkte ind ved at højreklikke og vælge menupunktet rediger formel.

Ølskum			
	tid	højde	b_værdi
enhed	sekunder	centimeter	centimeter
=			$\frac{\text{højde}}{\text{tid/sek}}$
1	0 s	17.0 cm	17.00 cm
2	30 s	14.9 cm	16.65 cm
3	60 s	13.0 cm	16.23 cm
4	90 s	11.9 cm	16.59 cm
5	120 s	10.7 cm	16.67 cm
6	150 s	9.7 cm	16.88 cm
7	180 s	8.9 cm	17.31 cm
8	210 s	8.4 cm	18.25 cm
9	240 s	7.5 cm	18.20 cm
10	300 s	6.3 cm	19.08 cm
11	360 s	5.2 cm	19.66 cm



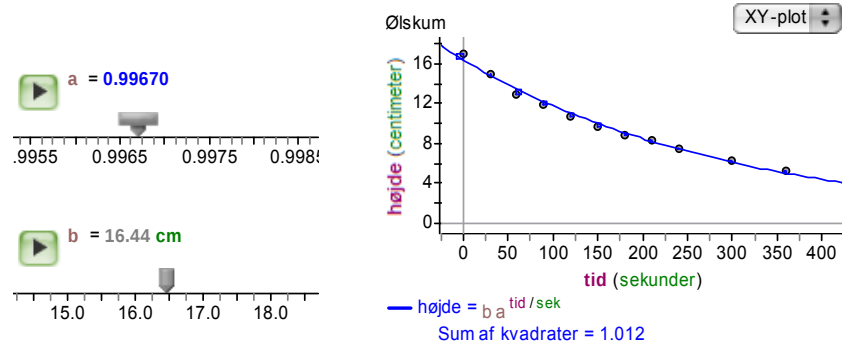
Læg mærke til at skyderen mistede sit vandrette håndtag. Den kan ikke længere rykkes med håndkraft, idet parameteren b nu er *bundet*, idet den afhænger af den frie parameter a og derfor kun ændres i takt med a . Det er nu en smal sag at tilpasse den frie parameter a indtil grafen for den eksponentielle vækstmodel går gennem datapunkterne. Undervejs kan man passende zoome ind på de to parametre ved at holde **CTRL**-tasten nede.



Dermed er vi kommet igennem den første fase i modellen og har fundet et rimeligt bud på værdien af parametrene a og b . Samtidigt har vi fået et godt visuelt check på at den eksponentielle model er i rimelig overensstemmelse med datasættet.

Anden fase: Mindste kvadraters metode

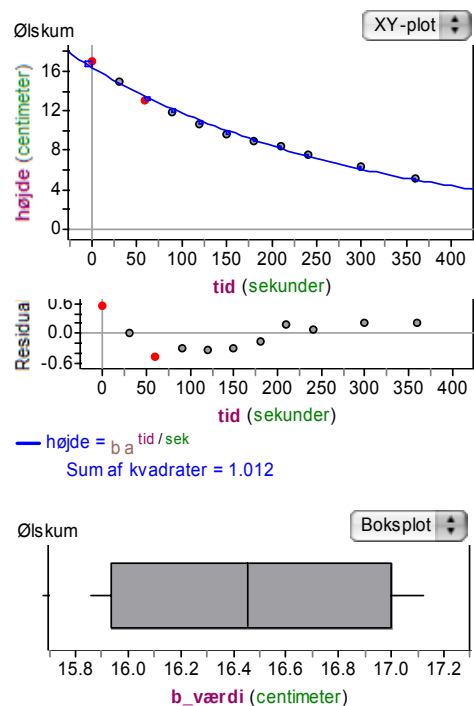
Derefter kommer finjusteringen af parameteren a ved hjælp af mindste kvadraters metode. Hertil tilføjer vi kvadraterne på afvigelserne ved at vælge menupunktet **Vis kvadrater** i grafrummet



Vi får da automatisk oplyst summen af kvadraterne og efter lidt ind-zoomning kan vi gøre den mindst muligt. Den mindste kvadratsum er da som vist 1.012. Vi har nu fastlagt parametrene a og b så præcist som muligt. Vi ser altså at b har værdien 16.44 cm, svarende til højden af ølskummet til tiden $t = 0$ sek. Tilsvarende har fremskrivningsfaktoren a værdien 0.9967, dvs. hvert sekund falder ølskumets højde med 0.23%.

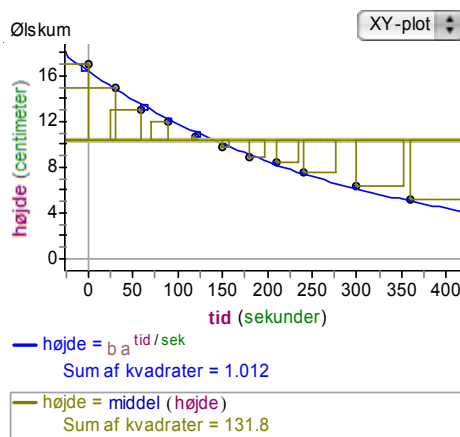
Vi kan også tilføje et residualplot for at se om de resterende afvigelser ligger tilfældigt fordelt. Det sker ved at vælge menupunktet **Tegn residualplot** i grafrummet. Vi ser da, at residualerne varierer rimeligt tilfældigt omkring nul med de største afvigelser i starten. Bortset fra starten er afvigelserne typisk nogle få millimeter, så det virker ikke urimeligt.

Vi kan også se at b -værdien er rimeligt samlet. Et boksplot viser tilsvarende at fordelingen er pænt symmetrisk og der er ikke nogen perifere observationer (afvigere). Endelig er usikkerheden på startværdien b af størrelsesordenen 0.5 cm.



Tredje fase: Forklaringsgraden

Vi slutter med at finde forklaringsgraden for den eksponentielle vækstmodel. Hertil sammenligner vi den med den mest modsatte model, **nulhypotesen**, ifølge hvilken variationerne i højden er rent tilfældige, dvs. de afhænger slet ikke af tiden. I så fald svinger højderne tilfældigt omkring deres middelværdi og vi tilføjer derfor denne ved hjælp af menupunktet **Plot funktion** i grafrummet. Som forventet er summen af kvadraterne meget større under nulhypotesen. Vi har en god model. Vi kan se at de residualer der er tilbage i den eksponentielle model kun udgør 0.8% af afvigelserne under nulhypotesen. Vi kan derfor forklare 99.2% af dataenes variation ved hjælp af vores eksponentielle vækstmodel. Resten skyldes tilfældige variationer (jfr. residualplottet). Forklaringsgraden for vores eksponentielle vækstmodel er altså $R^2 = 99.23\%$.



Ingen data

Slip en variabel her	
	0.0076783005
	0.9923217

$$R1 = \frac{1.012}{131.8}$$

$$R2 = 1 - \frac{1.012}{131.8}$$

Teknisk bemærkning: Om enheder og eksponentiel vækst

I det foregående har vi arbejdet konsekvent med enheder, idet vi har skrevet ligningen for den eksponentielle vækstmodel på formen

$$y = b \cdot a^{x / \text{enhed}}$$

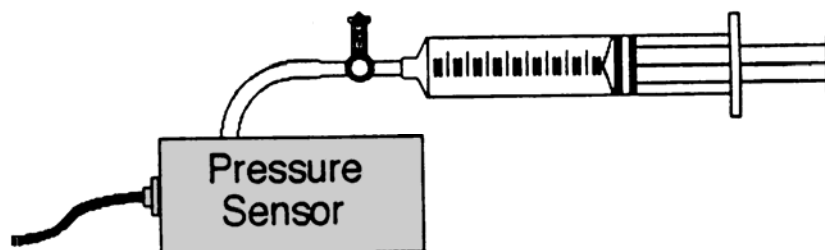
Derved sikres at grundtallet a ikke får nogen enhed (dvs. a er dimensionsløs). Samtidigt får b netop samme enhed som y , hvilket er i overensstemmelse med b 's rolle som startværdi for y , dvs. b angiver netop værdien af y , når x har værdien 0. Mens startværdien b er uafhængig af enheden for x , kommer værdien af grundtallet a til at afhænge stærkt af enheden for x . Det er helt i den eksponentielle vækstmodels ånd, idet a jo netop angiver fremskrivningsfaktoren for y -værdien, når den uafhængige variabel x vokser med 1 enhed. Der er selvfølgelig stor forskel på fremskrivningsfaktorerne når vi venter 1 sekund henholdsvis 1 minut.

Denne måde at skrive ligningen på stemmer også godt overens med den form for ligning, hvor vi som enhed vælger fordoblings/halveringskonstanten som enhed

$$y = b \cdot 2^{x / X_2} \quad \text{henholdsvis} \quad y = b \cdot \left(\frac{1}{2}\right)^{x / X_{1/2}}$$

3.4 Dynamisk parametertilpasning

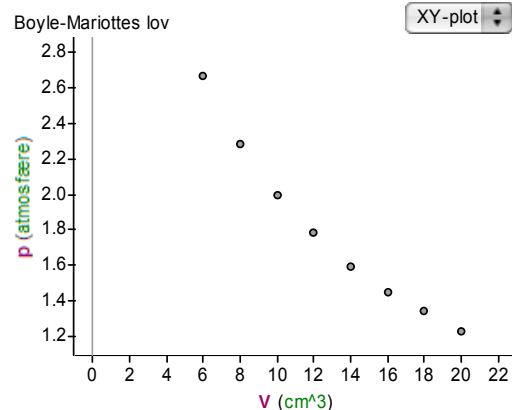
I de foregående afsnit har vi set på de vigtigste standardmodeller: lineær, potens og eksponentiel vækst. Vi vil nu se på en mere almen modellering, hvor vi skal fitte en parameter ved hjælp af mindste kvadraters metode. Nedenstående skema indeholder data fra et forsøg, hvor vi har målt lufttrykket p i en gassprøjte med stempel som funktion af rumfanget V . Der er anvendt standard CBL udstyr.



Rumfang/cm ³	6	8	10	12	14	16	18	20
Tryk/atm	2.67	2.28	2.00	1.78	1.59	1.45	1.34	1.23

Boyle-Mariottes lov

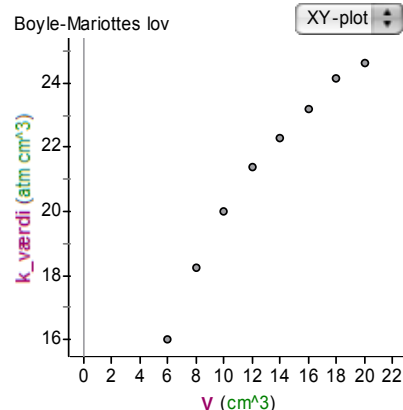
	V	p
enhed	cm ³	atmosfære
1	6 cm ³	2.67 atm
2	8 cm ³	2.28 atm
3	10 cm ³	2.00 atm
4	12 cm ³	1.78 atm
5	14 cm ³	1.59 atm
6	16 cm ³	1.45 atm
7	18 cm ³	1.34 atm
8	20 cm ³	1.23 atm



Det er nærliggende at gætte på en omvendt proportionalitet, hvorfor vi checker om produktet $p \cdot V$ (dvs. **k_værdi**) er konstant

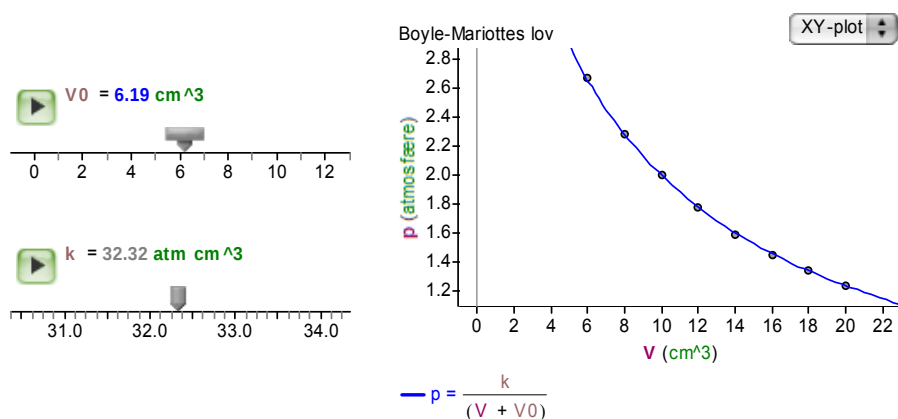
Boyle-Mariottes lov

	V	p	k_værdi
enhed	cm ³	atmosfære	atm cm ³
=			$p \cdot V$
1	6 cm ³	2.67 atm	16.02 atm cm ³
2	8 cm ³	2.28 atm	18.24 atm cm ³
3	10 cm ³	2.00 atm	20.00 atm cm ³
4	12 cm ³	1.78 atm	21.36 atm cm ³
5	14 cm ³	1.59 atm	22.26 atm cm ³
6	16 cm ³	1.45 atm	23.20 atm cm ³
7	18 cm ³	1.34 atm	24.12 atm cm ³
8	20 cm ³	1.23 atm	24.60 atm cm ³

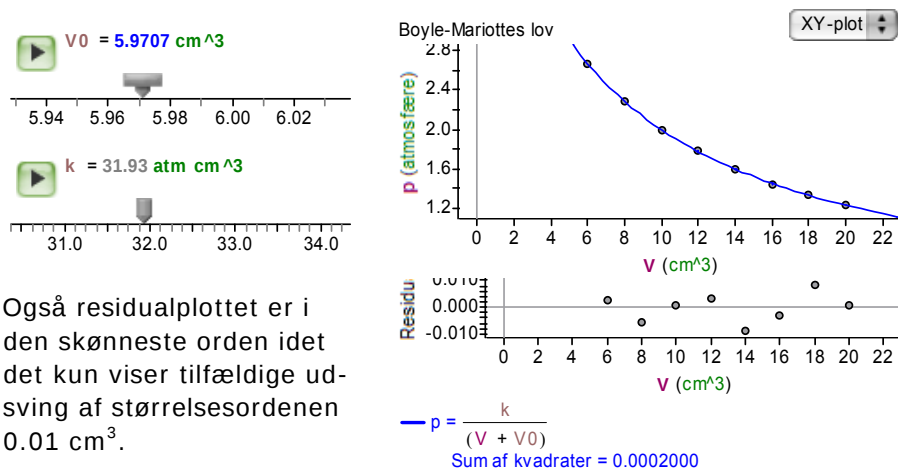


Der er dog tydeligt, at der noget galt. Produktet vokser fra ca. 16 i den første måling til ca. 24 i den sidste måling, altså en stigning på 50% og et plot over produktet $p \cdot V$ som funktion af rumfanget V viser tydeligt, at produktet varierer systematisk hen gennem måleserien. Fejlen ligger i måleudstyret. Det målte rumfang er aflæst på *gas-sprøjten*, dvs. den luft, der er i slanger og tryksonde, måles ikke med. Det aflæste rumfang skal altså korrigeres med et bidrag for det overskydende rumfang V_0 , og vi forventer derfor ifølge Boyle-Mariottes lov sammenhængen $p \cdot (V + V_0) = k$, hvor k er en konstant. Problemet er nu, hvordan vi kan finde parametrene k og V_0 ?

Vi indfører nu dynamiske parametre for proportionalitetskonstanten k og det ukendte overskydende rumfang V_0 . Derved kan k bindes til middelværdien af **k_værdi** (dvs. produktet $p \cdot (V + V_0)$), ligesom vi kan tilføje grafen for sammenhængen $p = k / (V + V_0)$

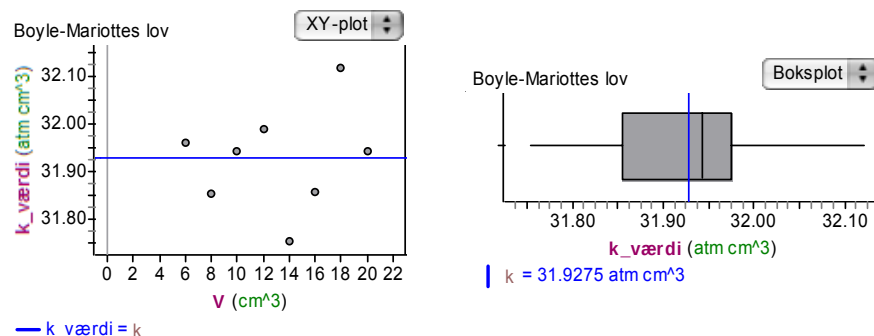


Grafisk hænger det altså fint sammen og det overskydende rumfang V_0 er altså givet ved ca. 6.2 cm^3 , mens proportionalitetskonstanten k er givet ved ca. $32.3 \text{ atm} \cdot \text{cm}^3$. Vi kan selvfølgelig underbygge det endnu mere præcist ved hjælp af mindste kvadraters metode

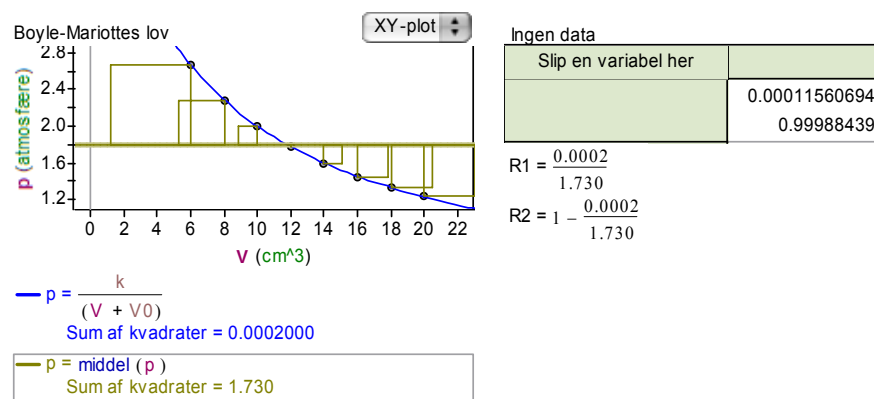


Også residualplottet er i den skønneste orden idet det kun viser tilfældige udsving af størrelsesordenen 0.01 cm^3 .

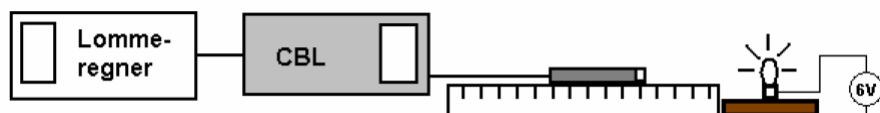
Vi kan tilsvarende se på **k_værdi**, som med tilnærmelse bør være konstant. Grafen for **k_værdi** som funktion af rumfanget afspejler det samme forløb som residualplottet og viser altså kun tilfældige udsving omkring middelværdien. Tilsvarende afspejler boksplottet en relativ snæver fordeling, der dog er trukket venstreskæv, hvorfor middelværdien ligger en anelse under medianen. Endelig ses det, at den typiske usikkerhed på proportionalitetskonstanten er af størrelsesordenen $0.5 \text{ atm}\cdot\text{cm}^3$



Til slut kan vi finde forklaringsgraden for modellen



Hertil sammenligner vi den med den mest modsatte model, **nulhypotesen**, ifølge hvilken variationerne i trykket er rent tilfældige, dvs. de afhænger slet ikke af rumfanget. I så fald svinger trykket tilfældigt omkring sin middelværdi og vi tilføjer derfor denne ved hjælp af menupunktet **Plot funktion** i grafrummet. Som forventet er summen af kvadraterne meget større under nulhypotesen. Vi har en god model. Vi kan se at de residualer der er tilbage i den teoretiske model kun udgør 0.01% af afvigelserne under nulhypotesen. Vi kan derfor forklare 99.99% af dataenes variation ved hjælp af vores teoretiske model. Resten skyldes tilfældige variationer (jfr. residualplottet). Forklaringsgraden for vores teoretiske model er altså $R^2 = 99.99\%$. Vi ser altså, at der er meget fin overensstemmelse mellem vores forsøgsdata og Boyle-Mariottes lov, når først vi har fået korigeret for det ukendte overskydende rumfang i slange og sonde.

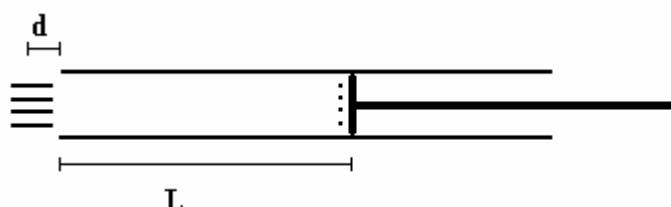
Øvelse 3.4.1: Lysintensitetens afstandsafhængighed

Da det areal lyset spredes over vokser med kvadratet på afstanden til lyskilden, forventer vi at lysintensiteten er omvendt proportional med kvadratet på afstanden. I et konkret forsøg fik man følgende data

Afstand (cm)	10	15	20	25	30	35	40
Lysintensitet (mW/cm^2)	0.631	0.399	0.255	0.178	0.134	0.107	0.095

Undersøg om lysintensiteten er omvendt proportional med kvadratet på afstanden.

I virkeligheden ligger lysmåleren et stykke inde i sonden, og pæren står et stykke inde på brættet, hvorfor vi skal lægge en overskydende længde d til afstanden til pæren. Udnyt dette til at vise at lysintensiteten faktisk er omvendt proportional med kvadratet på afstanden.

Øvelse 3.4.2: Lydens fart i luft

I et klassisk forsøg til fastlæggelse af lydens fart anvender man en tonegiver og et resonansrør, der kan indstilles i forskellige dybder. Ved at variere på frekvensen f af tonegiveren kan man fastlægge den første resonans, der i praksis ligger når rørets længde L er ca. $\frac{1}{4}$ bølgelængde. I dette forsøg har vi altså umiddelbart adgang til sammenhørende værdier af frekvens og bølgelængde og kan dermed bestemme lydens fart ud fra relationen

$$\text{lydens fart} = \text{frekvens} \cdot \text{bølgelængde}$$

I et konkret forsøg fandt man følgende resultater

f/Hz	200	250	300	400	500
L/mm	402	318	260	190	149

Undersøg om frekvens og rørlængde er omvendt proportionale.

I virkeligheden ligger den yderste svingningsbug som vist et lille stykke d uden for resonansrøret. Vi skal altså i virkeligheden lægge en overskydende længde d , den såkaldte randeffekt, til rørlængden L for at finde en kvart bølgelængde. Udnyt dette til at finde en værdi for randeffekten, såvel som en mere præcis værdi for lydens fart.

Indeks

afviger 74
andengradspolynomier 55-60
animation (parameter) 56
astronomisk enhed 9, 41
begrænsTilMultiplaAf 60
beregningsskema 31-32, 34, 37, 64, 67, 72
boksplot 69, 74, 78
Boyle-Mariottes lov 76-78
brugerdefineret enhed 27, 48
Celcius 48
data (indsæt) 22
data (kopier) 22
data fra internettet 23
data (kompatibel) 22, 24
datasæt 7, 19, 22
decimalpunktum 22
decimalskilletegn 22
densitet 14-16, 35
dublet 18
dynamisk parametertilpasning 76-79
eksponentiel regression 71-75
eksponentiel vækst 46-47
enhed 8-9, 15-17
enhed (brugerdefineret) 27, 48
enhed (eksponentiel vækst) 75
enhed (inkompatibel) 17
enhed (kompatibel) 40
enhed (potensproportionalitet) 70
enhed (ændring) 15, 40
fibonacci 44-47
filter 39, 42, 58
flyde-synke-spillet 52-54
forbrænding af genstand 36-39
forklaringsgrad 29-31, 64, 70, 75, 78
forklaringsgrad() 31, 64
formatér variabel 15
formel for a (potensproportionalitet) 67
formel for a (eksponentiel vækst) 72
formel for fibonacci 47
forrige() 45, 57
fremkræftsfaktor 46, 72-75
fri linje 27-28
funktionsgraf 55-60
første() 67, 72
grafisk tilpasning af parametre 66, 72
grafværktøj 9-12
gyldent snit 46-47
habitabilitetszone 43
heltallige værdier 60
hældning() 31, 64
ikon 20-21

indeks 8, 53-54
indsæt data 22
inkompatible enheder 17
inspektør for datasæt 13, 20, 24
inspektør for parameter 55-56, 60
inspektør for grafrum 18, 57
internationale og sproglige indstillinger 22
internettet (hent data) 23
isbjørnespillet 51
Jupiters måner 23-25
kartotekskort (datasæt) 13, 19
kategorilister 13
kategoriserede variable 8, 10, 13
kelvin 41
Keplers tredjelov 16-19, 25
kompatibel data 22, 24
kompatible enheder 40
kondital 32
konfunderet variabel 50
kopier data 22
krydstabel 52-54
kvadratet på fibonacci 47
kvadratsum 28-30, 62-64, 69
ligefrem proportionalitet 33-35
lineær regression 31, 61-64
lineær sammenhæng 27-32, 35
listeseparator 22
lodret linje 37, 39, 56, 60
lommeregner 10-11, 14, 17, 20, 31, 57, 62
lydens fart 40, 79
lys afstandsafhængighed 79
lås koordinatsystem 57
Lås skæring til nul 33
massefylde 14, 16, 33-35, 48-49
median 69
middel() 30, 34, 37, 39, 40, 42, 62, 67, 70, 73, 75, 78
mindste kvadraters linje 27-29, 31, 33, 59, 61, 64
mindste kvadraters metode 28, 69, 74, 76-77
modermærkekræft 50
nervesignalet fart 61-64
nulhypotese 30, 70, 75, 78
nulpunktsfejl 35
numeriske variable 8, 12
næste() 46, 57, 58
ombyt() 21
omvendt kvadratproportionalitet 41-43
omvendt proportionalitet 36-41, 76
parabel 55-60
parameterværktøj 55

perifer observation 74
planetikoner 20-21
planetsystemet 7-21, 41-43
Plot funktion 17, 34, 37, 55, 62, 66, 70, 72, 75, 78
Plot værdi 37, 56, 60
potensproportionalitet 65-70
potensregression 65-70
proportionalitetskonstant 33, 36, 37, 40, 77, 78
proportioner 33, 36
puls 27, 32
rekursiv formel 45
residualplot 29-30, 41-42, 64, 69-70, 74-75, 77-78
saltopløsning 4% 33-34
sammensat variabel 14, 42
siderisk år 9
sidste() 67, 72
signatur (variabel) 9, 12
simpsons paradoks 50
skjult variabel 48, 50
skæring() 31, 64
standardfejl for hældning 31
startværdi (eksponentiel vækst) 47, 72-73, 75
startværdi (lineær vækst) 28, 49, 61-62, 64
startværdi (potensvækst) 66-67, 69-70
statuslinje 32, 39
stdFejlHældning() 31
symmetrisk differenskvotient 57-59
søjlediagram 10-12
tabelværktøj 7-9
talvariable 8
tekstvariable 8
tilfældig() 26
tilfældige rektangler 26
Tilføj nye data 26
tæl() 10
udseende 20
usikkerhed startværdi 69, 74, 78
vandret linje 32, 39, 43, 60
variabel (konfunderet) 50
variabel (skjult) 48, 50
variabelkontrol 48-60
variabelsignatur 9, 12
variabeltransformationer 57
videnskabeligt talformat 8
vinduesgrænser (graf) 18
Vis kvadrater 28, 62, 69, 74
webikon 24
xAutotilpas 57
xy-plot 12, 33
yAutotilpas 57
zoom 18, 63, 68, 73
ølskum 71-73
år (siderisk) 9