

Projekt 3.3 Rygtespredning - modellering af logistisk vækst

(Projektet anvender værktøjsprogrammet TI Nspire)

Alle de tilstedeværende i klassen tildeles et nummer, så med 28 elever i klassen uddeles numrene 1-29.

Derefter trækkes en startperson ved hjælp af kommandoen

$\text{RandInt}(1,29)$.

Lad os sige vi finder tallet 17. Person nr. 17 rejser sig op, og spillet er i gang. Der trækkes et nyt nummer for at se hvem der første gang hører rygtet fra person nr. 17. Det foregår nemmest i grafregner-værkstedet, hvor man kan pile op og hente den forrige kommando ved at taste ENTER.

Lad os sige, at vi denne gang finder tallet 13. Person nr. 13 rejser sig op. Nu kender både 17 og 13 rygtet, så i næste runde trækker vi **to** numre ved at ændre kommandoen til

$\text{RandInt}(1,29,2)$.

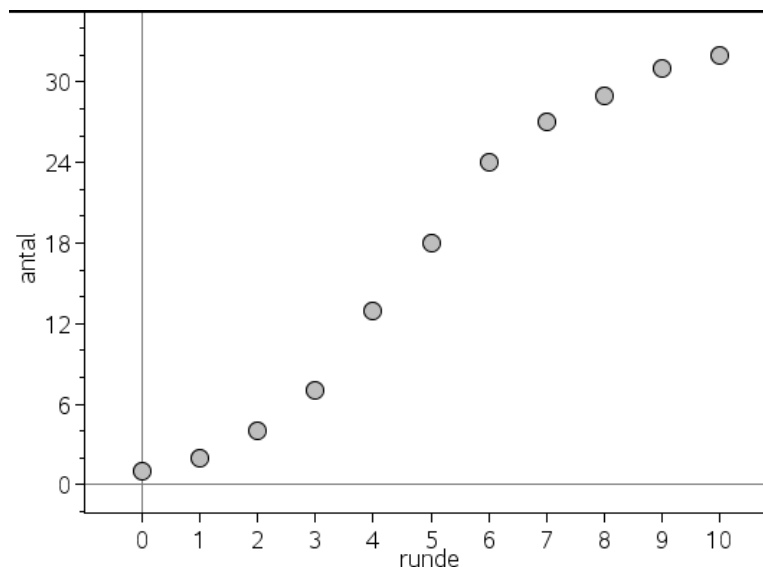
De inficerede personer, som har hørt rygtet fra person nr. 17 og person nr.13 rejser sig op. Vi trækker så igen numre svarende til det *totale antal inficerede personer* osv. indtil alle har hørt rygtet. Antallet af inficerede personer som funktion af antallet af rygtespredninger kan nu undersøges nærmere.

I et konkret eksperiment udført på et T³-kursus med 32 tilstedeværende udviklede det sig således:

0	1	2	3	4	5	6	7	8	9	10
1	2	4	7	13	18	24	27	29	31	32

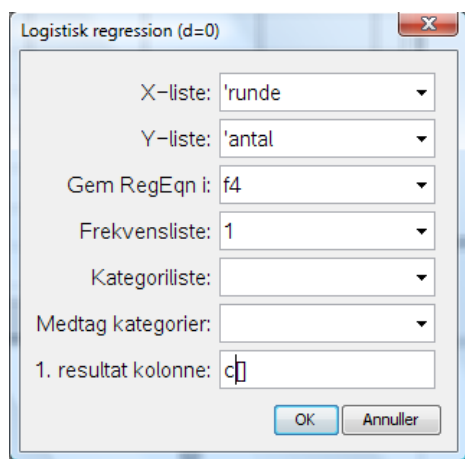
Tallene tages ind i listerne **runde** og **antal** og vi er klar til at kigge på et statistisk plot.

	A runde	B antal
1	0	1
2	1	2
3	2	4
4	3	7
5	4	13
6	5	18
7	6	24
8	7	27
9	8	29
10	9	31
11	10	32



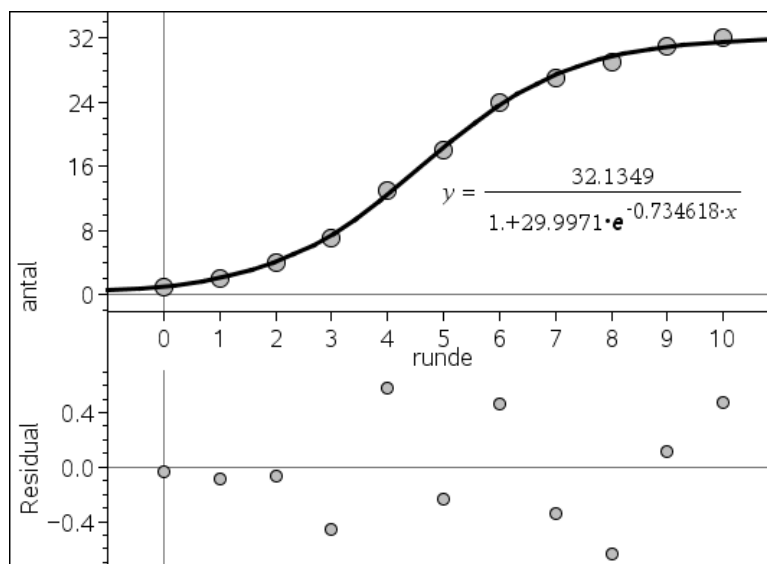
Logistisk regression: Maskinens tur

Det kunne godt ligne en logistisk vækst. For at undersøge det nærmere kan vi nu gå frem på flere forskellige måder. Vi kunne for eksempel starte med at tænke os om, men vi kan også bruge maskinens indbyggede logistiske vækstmodel. Lad os derfor vente lidt med at tænke og først kigge på maskinens indbyggede regressions model **Logistisk vækst (d=0)**. Som alle andre regressionsmodeller ligger den under menuen **Statistik > stat beregning**. Vi udfører derfor kommandoen i lister og regnearkværkstedet, så vi også kan ligningen for regressionsfunktionen at se. Der er tale om en ikke-lineær regressionsmodel, så maskinen finder parametrene iterativt ved at lave løbende justeringer, hvorved summen af de kvadratiske afvigelse hele tiden søges nedbragt.



	A	B	C	D
	runde	antal		
◆				=Logistic('runde','antal,1): Co
1	0	1	Titel	Logistisk regression (d=0)
2	1	2	RegEqn	$c/(1+a \cdot e^{(-b \cdot x)})$
3	2	4	a	29.9971
4	3	7	b	0.734618
5	4	13	c	32.1349
6	5	18	Resid	{-0.036708030738902,-0.088...
7	6	24		

Læg mærke til, at vi *ikke* får oplyst en forklaringsgrad. Til gengæld er grafen rimeligt overbevisende. Men selvfølgelig bør vi ikke bare lade os stille tilfredse med en pæn graf. Vi bør i det mindste checke residualerne for at se om vi skulle have overset systematiske variationer, der går udover den logistiske model:



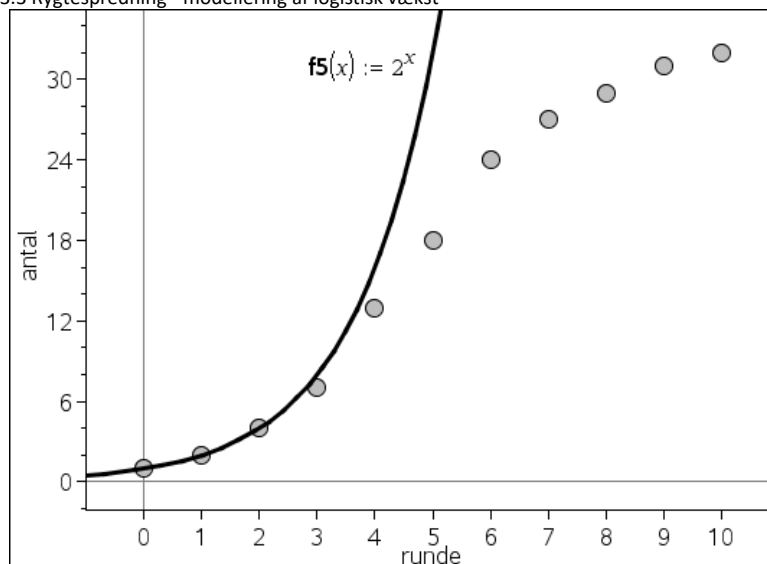
Som det ses svinger residualerne tilfældigt op og ned, og da ydermere den største afvigelse er helt nede på ca. 1/2, må det siges at være et særdeles tilfredsstillende residualplot, i betragtning af at alle de observerede y-værdier er hele tal.

Manuel fitting til en logistisk model

Så er det vist på tide at vi også prøver at tænke os lidt om! Til at begynde med vil sandsynligheden for at vi rammer en genganger være meget lille, da antallet af inficerede elever er meget lille i forhold til det samlede antal. Den første må derfor forventes at smitte en ny. De to første må så forventes at smitte to nye. Disse fire må så forventes at smitte fire nye osv. Til at begynde med fordobles antallet derfor hver gang, indtil gengangerne begynder at optræde. I den ovenstående serie ser vi da også at antallet vokser som 1 – 2 – 4 – 7, så først i tredje runde dukker den første genganger op.

Den eksponentielle vækst i den indledende fase er derfor simpelthen givet ved den eksponentielle vækstfunktion 2^x . En graf over fordoblingsfunktionen $y = 2^x$ sammen med det statistiske plot for rygtespredningerne bekræfter da også dette:

Projekter: Kapitel 3. Projekt 3.3 Rygtespredning - modellering af logistisk vækst



Dernæst bemærker vi, at mæthedsniveauet også er givet på forhånd idet der er 32 deltagere i alt, hvorfor $M = 32$. Vi forventer altså en logistisk vækst af formen:

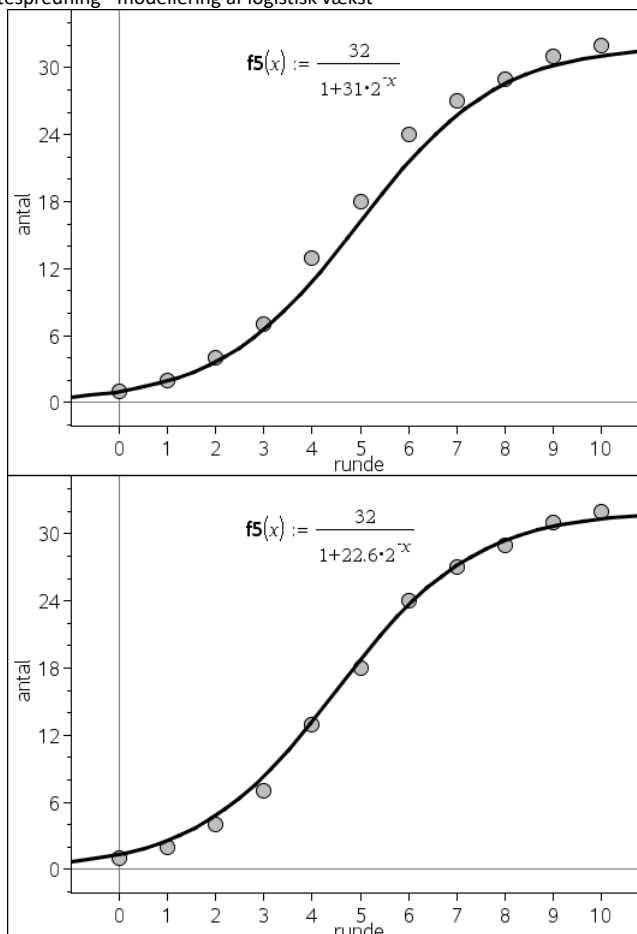
$$y = \frac{32}{1 + c \cdot 2^{-x}}$$

Bemærkning: Læg mærke til overensstemmelsen med det resultat vi fandt ved den logistiske regression. Her blev mætningsniveauet M estimeret til 32.13, altså en smule over de forventede 32. Tilsvarende blev vækstraten b estimeret til 0.7346, altså igen en smule over de forventede $\ln(2) = 0.6931$. Men alt i alt peger det i den rigtige retning.

Der er altså reelt kun 1 parameter, der skal estimeres i vores teoretiske model, nemlig konstanten c , som kontrollerer den *vandrette forskydning* af grafen. Den er til gengæld sværere at forudsige teoretisk. Vi vil derfor estimere den ud fra det konkrete eksperiment. Hvis vi fx insisterer på startværdien 1, ser vi, at vi er nødt til at sætte $c = 31$. Hvis vi i stedet benytter vendepunktet, ser vi at antallet af smittede elever passerer halvdelen af de tilstedeværende, dvs. 16, et sted mellem 4 og 5 runder henne. Altså er konstanten givet ved $c \approx 2^{4.5} \approx 22.6$.

Ingen af estimerterne er særligt præcise, men ved at se på graferne kan vi dog godt se, at de følger observationerne rimeligt, om end de ikke gengiver forskydningen i x -aksens retning helt præcist.

Projekter: Kapitel 3. Projekt 3.3 Rygtespredning - modellering af logistisk vækst



Det ses dog klart ud fra graferne, at estimatet baseret på vendepunktet er væsentligt mere præcist end estimatet baseret på startpunktet! Da der kun er én fri parameter i vores model kan vi imidlertid nemt finde en rimelig værdi for denne. Fx kan vi udnytte at c er en konstant, men c er givet ved udtrykket:

$$c = \left(\frac{32}{y} - 1 \right) \cdot 2^x$$

Listen givet ved udtrykket

$$\left(\frac{32}{\mathit{antal}} - 1 \right) \cdot 2^{\mathit{runde}}$$

skal derfor med tilnærmelse være konstant. Ved indtastning i **lister- og regneark-værkstedet** fås da også:

Projekter: Kapitel 3. Projekt 3.3 Rygtespredning - modellering af logistisk vækst

	A	B	C
	runde	antal	c_data
♦			=(32/antal
1	0	1	31.
2	1	2	30.
3	2	4	28.
4	3	7	28.5714
5	4	13	23.3846
6	5	18	24.8889
7	6	24	21.3333
8	7	27	23.7037
9	8	29	26.4828
10	9	31	16.5161
C	c_data := $\left(\frac{32}{\text{antal}} - 1\right) \cdot 2^{\text{runde}}$		

Specielt inde i midten giver det rimelige værdier for c, men stadig med en betydelig variation. Vi står os derfor ved at anvende medianen

$$\text{median}(c_data) \rightarrow 24.888888889$$

dvs. i dette tilfælde 24.9:

