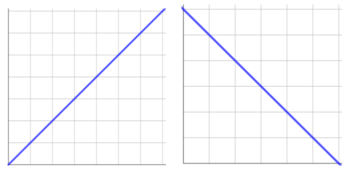
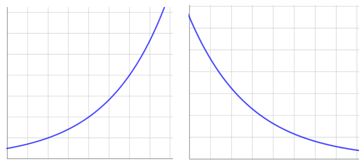
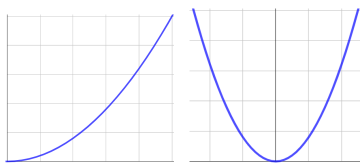
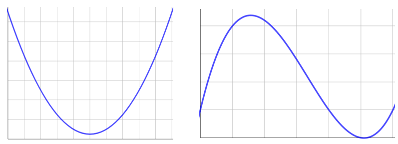
Olika regressionsmodeller

I förra genomgången såg vi hur vi kunde ta fram ett linjärt samband genom att anpassa en rät linje efter punkter. Det finns andra samband som inte är linjära och vi kan ta fram många av dessa på liknande sätt.

För att veta vilken modell som passar bäst utifrån givna punkter kan det vara bra att känna till några kännetecken för utseendet hos olika funktioner.

Funktion	**Form**	**Kännetecken**	**Utseende**
Linjär	$y = kx + m$	Rät linje. Ökningen eller minskningen är lika stor överallt.	
Exponentiell	$y = Ca^x$	Liten ökning i början och stor ökning i slutet ELLER stor minskning i början och liten minskning i slutet.	
Potens	$y = Cx^a$	Kurva som vänder och går genom origo.	
Polynom	$y = ax^2 + bx + c$ (grad 2)	Kurva som vänder, inte nödvändigtvis genom origo.	

Ibland är det svårt att avgöra vilken modell som passar bäst. Det kan t.ex. vara svårt att se om punkterna anpassas bäst av en exponential- eller en potensfunktion (se likheten mellan de två graferna till vänster i tabellen ovan). Potens- och polynomfunktioner kan vara ännu svårare att skilja åt eftersom de dessutom ibland kan sammanfalla ($y = x^2$ är både en potensfunktion och en polynomfunktion).

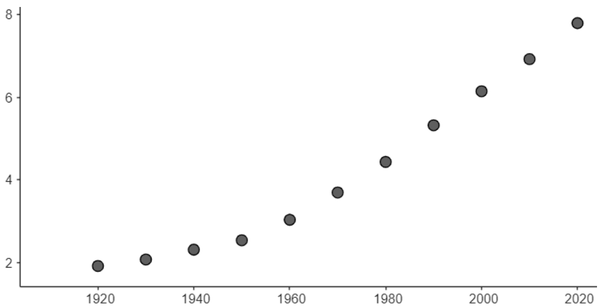
För att avgöra vilken kurvanpassning som är bäst kan vi utnyttja korrelationskoefficientens kvadrat, r^2. r^2-värdet ligger mellan 0 (ingen korrelation alls) och 1 (perfekt korrelation). Så ju närmare 1 som r^2-värdet ligger desto bättre anpassad är kurvan. För att ta fram r^2-värdet matar vi, som vanligt, in värdena i Geogebras kalkylblad, klickar på knappen "Visa statistik" Σx och läser av r^2-värdet i tabellen.

Ex. Världens befolkning vid några olika årtal redovisas i tabellen nedan.

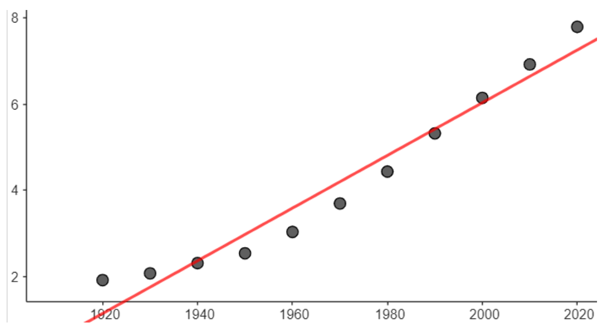
År	**Befolkning (miljarder)**		1970	3,692
1920	1,912		1980	4,434
1930	2,070		1990	5,321
1940	2,307		2000	6,145
1950	2,536		2010	6,922
1960	3,032		2020	7,794

Skapa ett spridningsdiagram och ta fram en regressionsmodell som är
a) linjär   b) exponentiell
c) Vilken av modellerna beskriver sambandet bäst?

a) Vi matar in värdena i Geogebras kalkylblad (tänk på att decimaltal skrivs med punkt), markerar värdena, klickar på "blå staplar" och väljer *Tvåvariabels regressionsanalys*. Nedanstående spridningsdiagram erhålls.



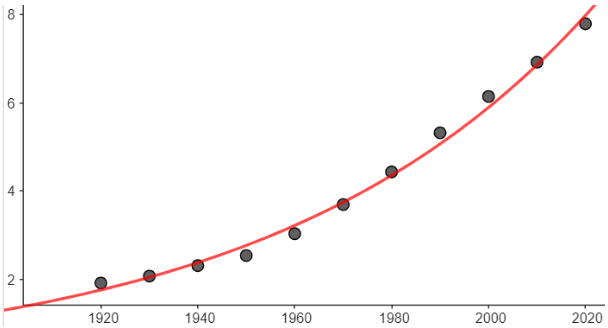
Vi väljer regressionsmodell *Linjär* och läser av ekvationen.



$y = 0.0612\,x - 116.3385$

Svar: $y = 0{,}061x - 120$

b) Vi skiftar regressionsmodell i rullistan till *Exponentiell* $C \cdot a$^x och läser av ekvationen.



$$y = 0 \cdot 1.0153^x$$

Svar: $y = 0 \cdot 1{,}0153^x$

(Startvärdet C = 0 i modellen är orealistiskt, men Geogebra skriver inte ut alla decimaler för tal nära 0.)

c) För att avgöra vilken av modellerna som beskriver sambandet bäst tittar vi på vilken kurva som ligger bäst anpassad efter kurvorna. I detta fall är det lätt att se att exponentialfunktionen passar punkterna bättre och därmed är bäst.

Om vi vill vara säkra kan vi jämföra modellernas r^2-värden. Vi klickar på knappen "Visa statistik" Σx och läser av vid R^2 för respektive modell.

Linjär: $r^2 = 0{,}954$
Exponentiell: $r^2 = 0{,}9933$

Eftersom r^2-värdet ligger närmast 1 för exponentialfunktionen är denna bäst.

Svar: Exponentiella modellen är bäst.