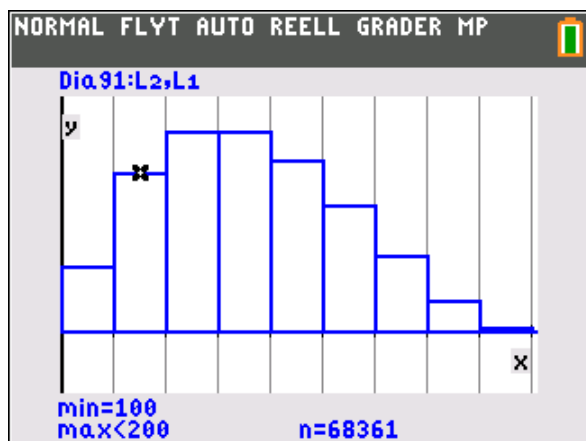


Integraler och analys av inkomstfördelning

Övningen passar bäst för både för Kurs 3 och kurs 4. Här får man använda kunskaper från statistik i tidigare kurser. Statistiska begrepp som kommer in i denna aktivitet är klassindelning, normalfördelning standardavvikelse och regression. Det är bra om du har en viss vana vid att plotta statistiska diagram med räknaren.

I statistikeditorn har vi matat in data över inkomstfördelningen för personer över 18 år i en kommun med ca en halv miljon invånare.

På räknaren kan vi naturligtvis inte lagra inkomsten för alla dessa personer utan vi har gjort en sammanställning i ett histogram med antalet personer i olika inkomstkick.



På x-axeln har vi inkomsterna och vi har staplar med bredden 100. Den andra stapeln visar att antalet inkomsttagare med inkomster mellan 100 och 200 kkr, dvs. 100 000 till 200 000 kr, är ca 68 000.

Detta är naturligtvis inte den exakta bilden av inkomstfördelningen. Vi har ju klumpat ihop det hela och antagit att alla i ett inkomstkick har lika stor inkomst. För den andra stapeln ovan så har vi antagit att alla tjänar 150 kkr och vi använder detta värde när vi beräknar de totala inkomsterna och medelinkomsten.

I nästa spalt ser du i statistikeditorn de värden vi har utgått ifrån när vi plottat histogrammet.

L1	L2	L3	L4	L5	3
28560	50	-----	-----	-----	
68361	150				
85909	250				
86173	350				
74118	450				
54713	550				
32925	650				
13721	750				
2068	850				
-----	-----				

L3(1)=medel(L2,L1)

Vi kan nu också beräkna medelvärde och standardavvikelse för inkomstfördelningen och det exakta antalet personer. Placera markören där du vill få beräkningen utförd och tryck på **2nd**[list] och välj matematiska beräkningar (MA). Här finns instruktioner för beräkning av olika statistiska mått:

NORMAL FLYT AUTO REELL GRADER MP					
NAMN OPS MA					
1:min(
2:max(
3:medel(
4:median(
5:sum(
6:prod(
7:stdAv(
8:varians(

L1	L2	L3	L4	L5	3
28560	50	-----	-----	-----	
68361	150				
85909	250				
86173	350				
74118	450				
54713	550				
32925	650				
13721	750				
2068	850				
-----	-----				

L3(1)=medel(L2,L1)

Formeln ovan betyder att vi beräknar medelvärdet för personer med inkomsterna 50, 100 osv där frekvenserna är 28560, 68361 osv.

Om du trycker på **enter** får du beräkningen utförd. Gör likadant med beräkningen av stan-

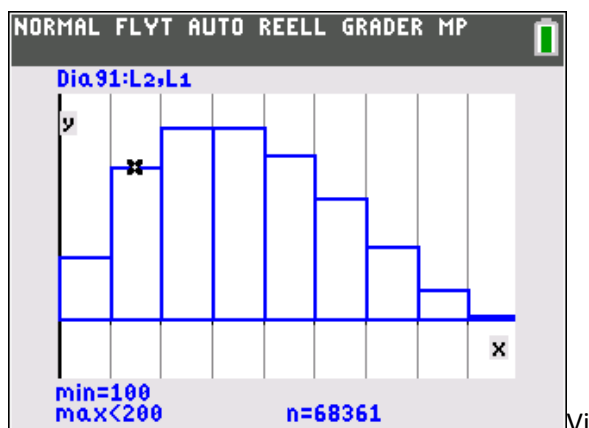
dardavvikelsen. Antalet personer beräknas med formeln $\text{sum}(L1)$.

L1	L2	L3	L4	L5	3
28560	50	358.79	-----	-----	
68361	150	181.65			
85909	250	446548			
86173	350				
74118	450				
54713	550				
32925	650				
13721	750				
2068	850				
-----	-----				

L3(4)=

Medelinkomsten är ca 359 000 kr, standardavvikelsen ca 182 000 kr och totala antalet personer 446548. Den siffran är ju korrekt eftersom vi har exakta data i lista L1.

Diskutera utseendet på fördelningen. Påminner det om en normalfördelning? Knappast. Det är ganska många i de lägre inkomstskikten och betydligt färre i de högre. Vi återkommer till detta.



Vi ska nu titta lite närmare på dessa data och göra en matematisk modellering med analysverktyg. Man kan nämligen med utgångspunkt från ett enda funktionsuttryck göra ett stort antal beräkningar. Visar något av den matematiska analysens kraft.

Nu ska först vi göra en matematisk modellering av inkomstfördelningen med räknarens regressionsverktyg.

Det visar sig att man kan få en bra modell med en tredjegradsfunktion. Vi går här inte igenom varför just en tredjegradsfunktion passar bra.

Tryck på $\boxed{\text{stat}}$ och välj beräkningar (BERÄK). Välj sedan alternativ 6 KubikReg.

NORMAL FLYT AUTO REELL GRADER MP					
REDIGERA BERÄK TESTER					
1: 1-Var-stat					
2: 2-Var-stat					
3: Med-Med					
4: LinReg(ax+b)					
5: KvadReg					
6: KubikReg					
7: 4-gradsReg					
8: LinReg(a+bx)					
9: LnReg					

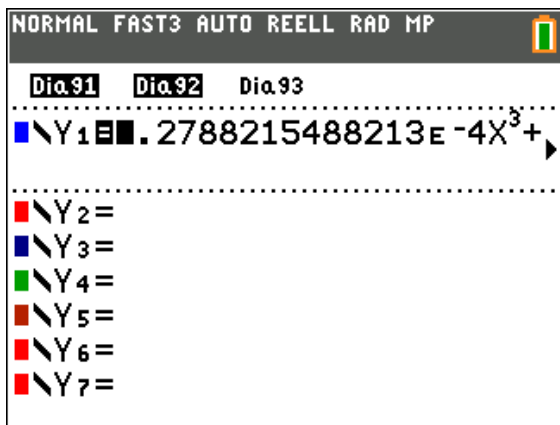
Tryck nu på $\boxed{\text{enter}}$ och fyll i enligt nedan. Vi lagrar den beräknade funktionsmodellen i Y1. L1 och L2 knappar du in från tangentbordet. Y1 kan du infoga genom att trycka på $\boxed{\alpha}$ $\boxed{f4}$ och sedan välja Y1. Detta är en genväg.

NORMAL FLYT AUTO REELL GRADER MP					
KubikReg					
Xlista:L2					
Ylista:L1					
FrekvLista:					
Lagra RegEkv:Y1					
Beräkna					

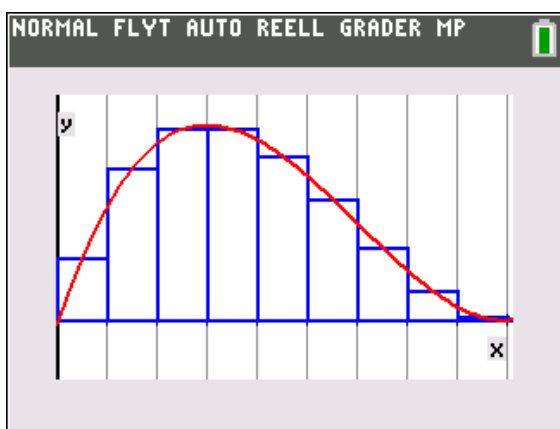
Markera **Beräkna** och tryck på $\boxed{\text{enter}}$ igen. Nu får vi resultatet.

NORMAL FLYT AUTO REELL GRADER MP					
KubikReg					
$y=ax^3+bx^2+cx+d$					
a=8.278821549E-4					
b=-1.48516645					
c=668.1353018					
d=-1237.293831					

Funktionen är nu lagrad i Y1 med en massa decimaler.

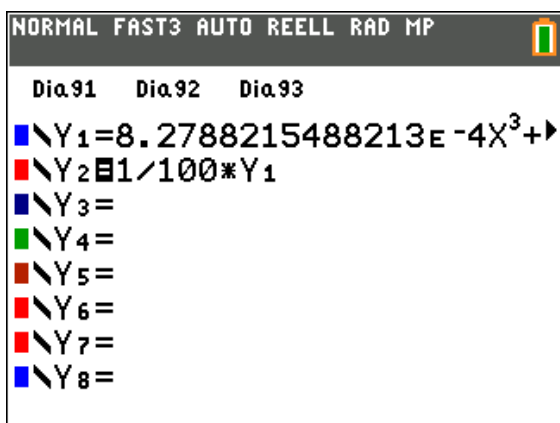


Eftersom den beräknade modellen finns lagrad i Y1 så kan vi nu direkt plotta resultatet tillsammans med histogrammet.

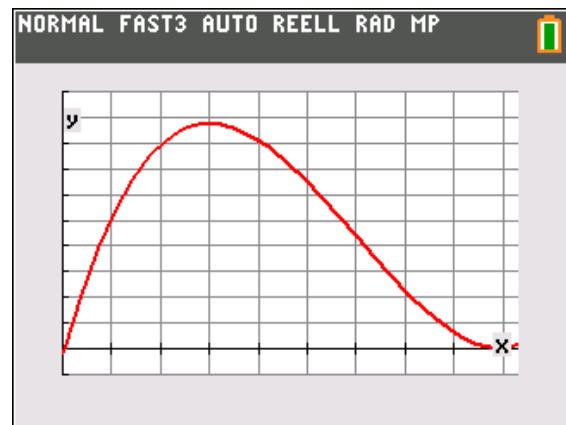


Vi får en väldigt bra anpassning. Den röda kurvan följer data väldigt väl.

För att kunna beräkna arean under kurvan på ett korrekt sätt så ska vi dividera tredjegradsuttrycket med 100, som är klassbredden för våra data. Enklaste sättet att göra detta är att skriva in en ny funktion Y2 enligt nedan och avmarkera Y1.



Nu plottar vi denna kurva i ett bra fönster och vi stänger också av diagramritningen.

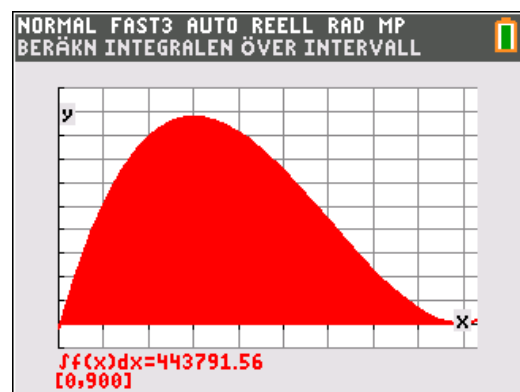


Den vågräta axeln visar alltså årsinkomsten och den lodräta axeln frekvensen. Detta är en s.k. *frekvensfunktion*, precis som normalfördelningen.

Vad händer om vi beräknar integralen av denna funktion från 0 till 900 med räknarens analysverktyg? Vi prövar!

Tryck på 2^{nd} [calc] och välj 7: $\int f(x)dx$. Nu får man först välja nedre gräns. Skriv in 0 från tangentbordet och tryck på enter . Välj sedan övre gränsen 900 och tryck på enter igen.

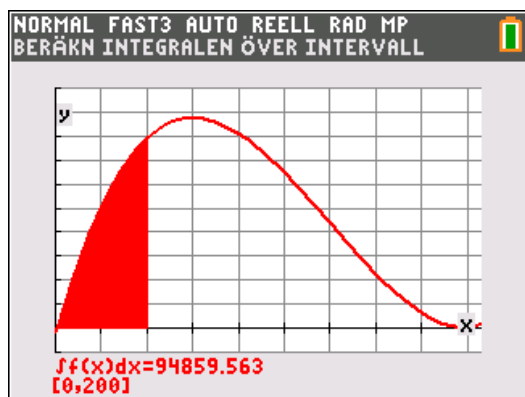
Nu får fylls området under kurvan från 0 till 900 och integralen beräknas. Vi får värdet 443791. Detta svar är alltså totala antalet inkomsttagare. Stämmer väldigt bra med beräknade data på sid 2.



Du kan göra andra beräkningar genom att sätta andra gränser. T.ex. hur många tjänar mer än 500 000 kr? Hur många tjänar mindre än 200 000 kr osv.

Rensa först det röda området genom att trycka på 2^{nd} [draw] och välj alternativ 1:RensaRit. Det måste man göra varje gång

man gjort en integralberäkning och ska göra en ny sådan.



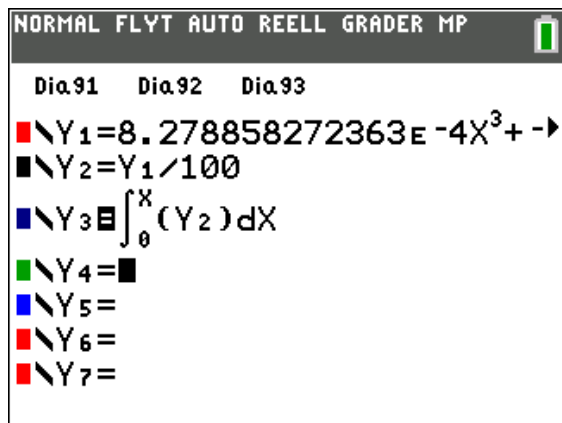
Ca 95 000 personer tjänar mindre än 200 000 kr.

Antalet inkomsttagare med en inkomst mindre eller lika med x kr är

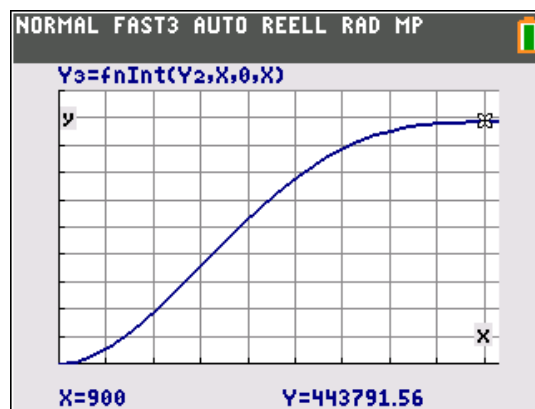
$$\int_0^x Y_2 dx$$

Där Y_2 är frekvensfunktionen på förra sidan. Vi kan nu rita denna funktion. Det är ju frekvensfunktionens *primitiva funktion*.

Här har vi matat in denna funktion och avmarkerat plottningen av Y_1 och Y_2 .



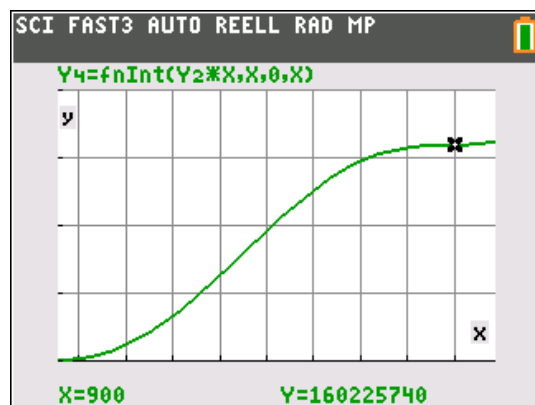
I grafen i nästa spalt kan vi återigen se att det finns ca 444 000 inkomsttagare i kommunen. Värdena (y -koordinaten) står ju för antalet inkomsttagare med inkomster från noll och fram till värdet på x . Innan plottning, se till att du ställer in ett bra fönster.



Om vi nu istället plottar funktionen

$$\int_0^x (f_2(x) \cdot x) dx$$

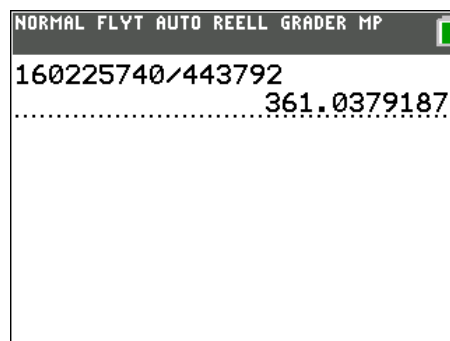
får vi den *totala inkomsten* från inkomsten 0 fram till x , där x är inkomsten. Diskutera gärna hur det kan stämma.



$\int_0^{900} (Y_2 \cdot x) dx$ ger totalinkomsten 160225740 kkr., vilket kan skrivas som $1,602 \cdot 10^8$ kkr.

Eftersom vi har beräkningar i kkr blir det då ca $1,6 \cdot 10^{11}$ kr eller 160 miljarder kr.

Då kan vi beräkna medelinkomsten som totala inkomsterna dividerat med antal inkomsttagare. Stämmer bra med tidigare beräkningar.



Nu jämför vi vår fördelning med en normalfördelning, där vi använt värdena för medelvärde och standardavvikelse från statistikeditorn.

L1	L2	L3	L4	L5	Σ
28560	50	358.79	-----	-----	
68361	150	181.65			
85909	250	446548			
86173	350	-----			
74118	450				
54713	550				
32925	650				
13721	750				
2068	850				
-----	-----				

L3(4)=

För att plotta normalfördelningen, placera markören på ledig plats i Y=-editorn och skriv först in 446548*.

Di.91	Di.92	Di.93
Y1=	446548*	
Y2=		
Y3=		
Y4=		
Y5=		
Y6=		
Y7=		

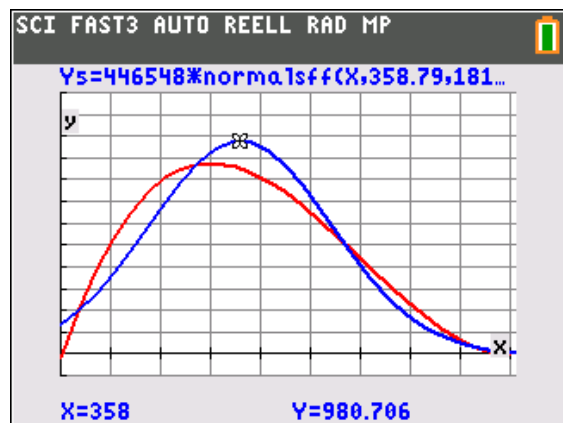
Tryck på 2nd [distr] och välj normalsff

FÖRDELN RITA
1: normalsff(
2: normalkff(
3: invNorm(
4: invT(
5: tsff(
6: tkff(
7: χ^2 sff(
8: χ^2 kff(
9: F χ^2 sff(

Då kommer en dialogruta upp som du fyller i enligt nedan

SCI FAST3 AUTO REELL RAD MP
normalsff
x-värde: X
μ : 358.79
σ : 181.65
Klistra in

Markera Klistra in och tryck på enter . Funktionen är nu på plats. Med en bra fönsterinställning (man kan få prova sig fram) blir det så här. Du ser funktionsuttrycket överst på skärmen.



Diskutera skillnaderna mellan fördelningarna utifrån plottningarna. Den röda normalfördelningskurvan verkar till exempel inte börja vid 0 på den lodräta axeln. Varför är det så?

Det exempel vi valt är naturligtvis anpassat så att man kan anpassa till en någorlunda enkel funktion. Så här ser inkomstfördelningen för personer folkbokförda i Sverige hela år 2019 och i åldrarna 20–64 år.

Vi har här klassbredd på 50 kkr. Dessutom ska man lägga till att ca 300 000 personer saknar inkomst.

