Opinionsundersökning

Vi ska nu titta lite närmare på osäkerheten bakom stickprov på andelar. Man stöter ofta på detta i samband undersökningar om partisympatier.

Vi antar att stickprovsstorleken är 1000 personer, ungefär vad man brukar ha i opinionsundersökningar. Frågan är vad man kan förvänta sig om hur nära andelen i stickprovet kommer andelen i hela befolkningen. Vi kallar den för *p*.

Vi börjar här med teorin och genomför sedan några undersökningar med slumpen.

Ma<n kan visa att i 19 fall av 20 (95 % konfidensintervall) så hamnar stickprovets andel i intervallet

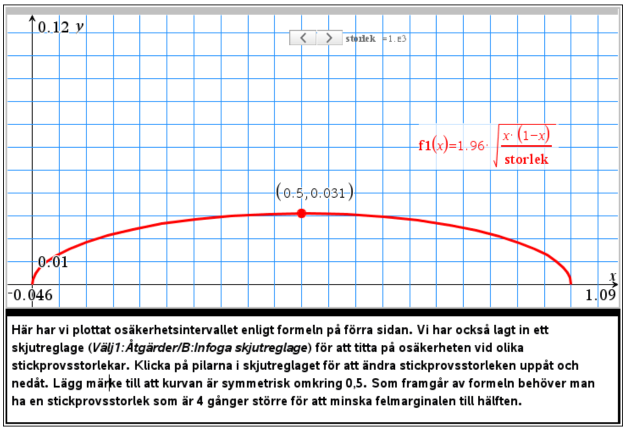
$$p \pm 1{,}96 \cdot \sqrt{\frac{p(1-p)}{n}}$$ där *n* är stickprovsstorleken.

Vi antar den *sanna* proportionen i populationen är 50 % och stickprovsstorleken 1000. Då får vi

$$0{,}50 \pm 1{,}96 \cdot \sqrt{\frac{0{,}50 \cdot (1-0{,}50)}{1000}} \approx 0{,}031$$

Vi får alltså då en felmarginal på ca 3,1 %. Av uttrycket ser vi att felmarginalen beror både på andelen och stickprovsstorleken. Vi visar detta grafiskt på sid 2.

Sid 2: Här har vi plottat osäkerhetsintervallet enligt formeln på förra sidan. Vi har också lagt in ett skjutreglage (Välj1:Åtgärder/B:Infoga skjutreglage) för att titta på osäkerheten vid olika stickprovsstorlekar. Klicka på pilarna i skjutreglaget för att ändra stickprovsstorleken uppåt och nedåt. Lägg märke till att kurvan är symmetrisk omkring 0,5. Osäkerheten är alltså lika stor om andelen är 80 % resp. 20 %. Som framgår av formeln behöver man ha en stickprovsstorlek som är 4 gånger större för att minska felmarginalen till hälften.



Sid 3: Om det sanna värdet är 50 % för ett parti och vi drar ett stickprov på 1000 personer och upprepar detta 1000 gånger får vi ett resultat ungefär som i diagrammet. Vi har här simulerat detta med instruktionen

randbin(1000,0.5,1000)

Instruktionen *randbin* kan du hitta i verktygslådan under Data/Slump/Binomial.

Vi simulerar alltså slumptal från en binomialfördelning. Det ligger antagligen utanför kursmålen att ta upp detta men man kan nämna att om man upprepar ett försök, som har en viss sannolikhet, ett antal gånger så kan man beräkna sannolikheten att man lyckas 0, 1, 2 … *n* gånger.

Att beräkna sannolikheten att man lyckas *precis* 500 gånger kan beräknas i en cell med den inbyggda instruktionen så här

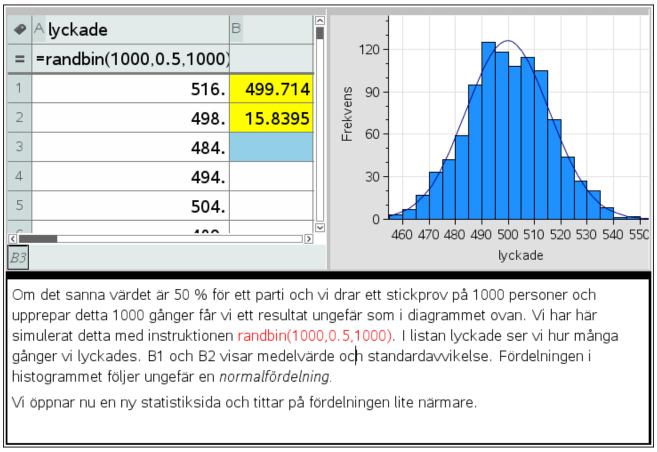
=Binompdf(1000,0.5,500)

Vi får resultatet 0.025225, ca 2,5 % alltså

Om vi använder formeln för binomialfördelningen (ncr är antalet kombinationer) kan vi skriva

ncr(1000,500)·0.5^{500}·0.5^{500}

Vi får samma resultat. TI-Nspire kan hantera beräkningen av ncr(1000,500), som är ett förfärligt stort tal (ung 2,7·10^{299}).

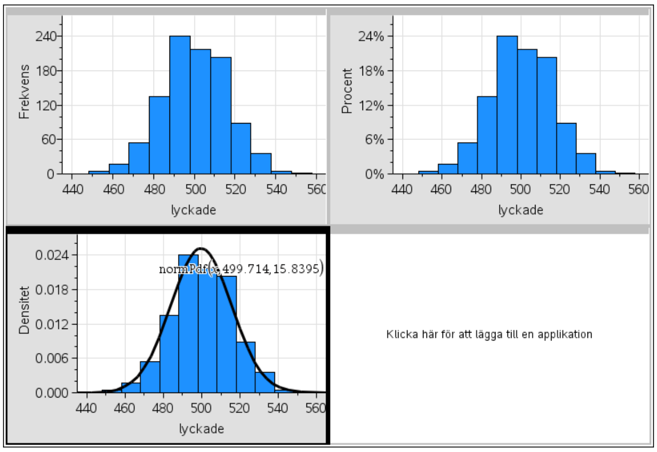


I listan *lyckade* ser vi hur många gånger vi lyckades. B1 och B2 visar medelvärde och standardavvikelse. Fördelningen i histogrammet följer ungefär en *normalfördelning*. Det är utifrån detta som statistikerna kan beräkna en s.k. felmarginal för osäkerheten i mätningen. Det var det vi visade på sid 1-2.

Läs mer om felkällor på
http://www.scb.se/sv_/Dokumentation/Statistikguiden/Kvalitet-i-statistiken/Felmarginaler/

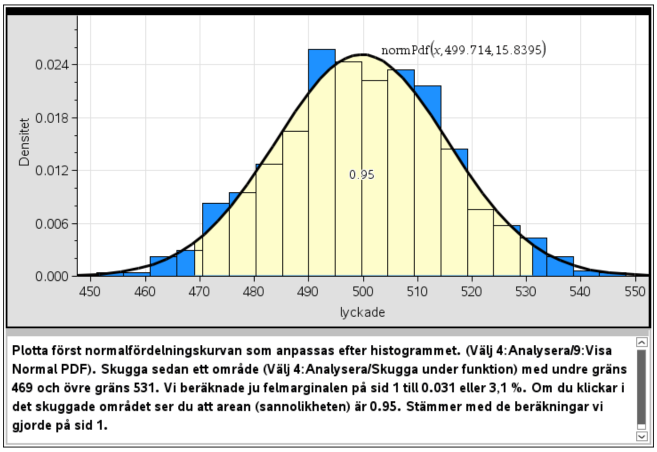
När man har sitt histogram kan man ställa in *skalan* på frekvensaxeln på tre olika sätt. Vi visar här alla. Vi får ju exakt samma diagram men med oliks skalor. Skalan densitet betyder att arean under staplarna är 1 area-enhet. Procent är ju den relativa frekvensen.

Du kan ställa in dessa detta och en del annat, bl.a. stapelbredd, genom att högerklicka i diagrammet.



Om man i verktygslådan väljer Analysera kan man välja att rita en normalfördelningskurva till sitt histogram. Välj Visa Normal PDF. Om du klickar på kurvan ser du parametrarna för funktionen.

Sid 4: Vi öppnar nu en ny statistiksida och tittar på fördelningen för ”lyckade” lite närmare.



Plotta först normalfördelningskurvan som anpassas efter histogrammet. (Välj 4:Analysera och sedan 9:Visa Normal PDF). Skugga sedan ett område (Välj 4:Analysera/Skugga under funktion) med undre gräns 469 och övre gräns 531. Vi beräknade ju fel-marginalen på sid 1 till 0.031 eller 3,1. %.

Om du klickar i det skuggade området ser du att arean (sannolikheten) är 0.95. Stämmer med de beräkningar vi gjorde på sid 1. 95 % av arean av normalfördelnings-kurvan ligger inom felgränserna. Elegant!

Man kan göra detta försök ***utan*** binomialfördelningen genom att simulera slantsingling. Där är ju sannolik-heten för Krona t.ex. .50 %.

Börja med att skapa en slumptalslista, 0 eller 1, i första kolumnen. I cell B! skriver du sedan

antal_krona:=1000·mean(krona_klave)

Då beräknas antal krona vi fick i slumptalsaltringen i variabeln antal_krona.
I kolumnen med variabelnamnet ”fångade” skriver du sedan i formelfältet =capture(antal_krona,1).
Då infångas alla värden i cell B1 och läggs in en lista.
Tryck upprepade gånger på Ctrl r så fylls denna lista på.

	A krona_klave	B	C fångade	D	E
=	=randint(0,1,1000)		=capture(antal_krona,1)		
1	1	510	481		
2	0		484		
3	0		480		
4	1		498		
5	1		528		
6	0		508		
7	1		487		
8	0		525		
9	1		499		
10	1		505		
11	0		495		

B1 antal_krona:=1000· mean(krona_klave)

Efter några hundra tryckningar på Ctrl r så kan fördelningen se ut så här. Vi får här värdet 0.96.