

Hur kan vi få reda på medelvärdet?

I denna övning ska vi titta på hur man genom stickprov kan dra slutsatser om hela populationen. Vi visar med ett par exempel hur det går till när man undersöker medelvärden. I verkligheten drar man kanske bara ett stickprov men statistikerna vet hur det blir med fördelningen om man drar många stickprov.

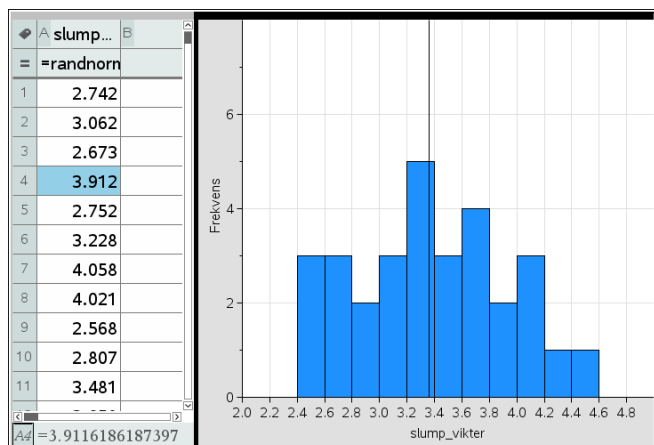
Övningen passar att genomföra i kurs 2 när man har gått igenom normalfördelningen. Vi har även en övning där vi undersöker proportioner (exempel med opinionsundersökningar) genom stickprov.

Första exemplet handlar om födelsevikter där vi vet att medelvikten är ca 3,5 kg och standardavvikelsen ca 0,5 kg.

Sid 2: Först öppnar vi en kalkylbladssida och namnger kolumnen "slump_vikt" t.ex. Placera markören i formelfältet och välj från verktygsmenyn Data/Slump/Normal. Skriv sedan

`=randnorm(3.5,0.5,30)`

Då får vi en lista med 30 slumpade värden från en normalfördelning. Om vi trycker på Ctrl r får vi en ny slumpvalslista.



Frågan är hur man ska göra för att få en bättre skattning av medelvärdet. Jo, man kan göra ett stort antal stickprov. Låt oss säga att vi gör så och tittar på hur medellängden fördelar sig för de olika stickproven. Vi ska nu göra detta i kalkylbladet och titta närmare på detta.

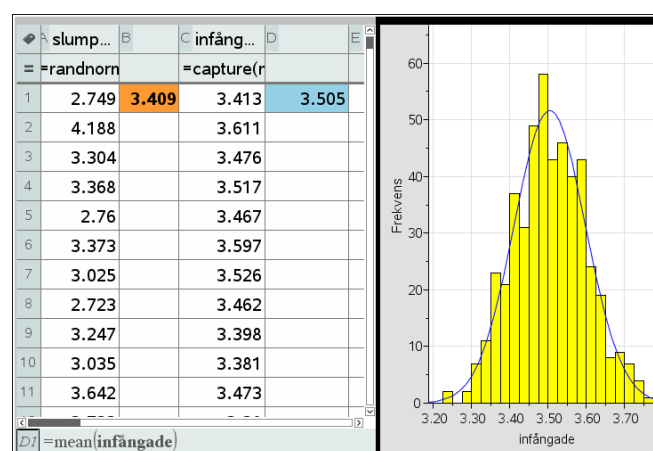
Vi ska nu "fånga in" medelvärdet när vi gör många slumpvalsstringar. Sedan ska vi titta på fördelningen av medelvärdena. Vi ska också lägga till en del nya saker i vårt kalkylark.

Sid 3-4: I cell B1 har vi aktuellt värde på medelvärdet från slumpvalslistan. Uppdateras varje gång vi trycker på Ctrl r.

I kolumn C (infångade) har vi med instruktionen capture fångat in medelvärdena från slumpvalslistan. Vi får ett nytt värde varje gång vi trycker på Ctrl r.

I cell D1 har vi sedan medelvärdet av de infångade medelvärdena. Titta på formlerna i de nämnda cellerna. I diagrammet har vi sedan ett histogram över fördelningen av infångade medelvärdena. Som ni ser har vi redan gjort en del simuleringar. Fyll nu på listan genom att upprepade gånger tryckas på Ctrl r.

Det går också att rensa listan med infångade värden och börja om igen.



Vi såg att vi fick en fördelning som liknade en normalfördelning när vi tittade på stickprovsmedelvärdena.

Man kan visa att i ungefär 19 fall av tjugo, dvs. i 95 % av alla stickprov kommer stickprovsmedelvärdet att ligga i intervallet

$$\bar{x} \pm 2 \cdot \frac{s}{\sqrt{n}}$$

där \bar{x} är stickprovsmedelvärdet, s är standardavvikelsen och n är stickprovsstorleken.

Problem 2

Sid 1-2: I en större undersökning med 2000 personer i åldrarna 18-64 år undersökte man antalet skickade sms personerna gjorde under en viss bestämd tidsperiod. Det visade sig att man i medeltal skickade 177 sms men att spridningen var mycket stor. Standardavvikelsen var 89. Detta är vår population.

Vi har nu gjort samma undersökning som i problem 1. Enda skillnaden är att vi i kolumn C gjort en slumpmässigt urval på 50 data från den stora listan i kolumn 1, som är vår population.

Vad som är remarkabelt är att vi även denna gång får stickprovsmedelvärden som liknar en normalfördelning trots att populationen inte alls har denna fördelning. Så är det alltid!

A	B	C	D	E	F	G
sms_alla		urval		fångade		
=		=randsam		=capture(
1	0.	medelvärde	472.		226.7	275.6
2	0.	276.6	869.	medelvärde	284.8	
3	0.		151.	urval	232.7	
4	0.		274.	259.2	280.6	
5	0.		167.		245.	
6	0.		22.		257.8	
7	0.		279.		267.4	
8	0.		38.		212.9	
9	0.		225.		241.7	
10	0.		316.		335.7	
11	0.		262.		202.8	

Sid 3: På denna sida har vi nu tre diagram. Till vänster fördelningen för hela populationen. Överst till höger aktuell fördelning för medelvärdet för urvalet och längst ner fördelningen av stickprovsmedelvärdena. Mycket information på en sida.

