
16. UPPSKATTNINGSPROBLEMET

16.1 Uppskattning vid oberoende observationer

Det här avsnittet ägnas åt problemet att uppskatta populationsmedelvärdet μ då observationerna är stokastiskt oberoende. Som vi har sett tidigare har vi observationer på stokastiskt oberoende variabler då vi slumpmässigt väljer individer ur en oändlig population och mäter en egenskap på de valda individerna. En annan situation då vi har oberoende observationer är då vi gör dragningar med återläggning ur en ändligt stor population. Det kanske mest naturliga sättet att uppskatta populationsmedelvärdet är att använda urvalsmedelvärdet som uppskattning. Förutom att det verkar intuitivt rimligt att använda urvalsmedelvärdet som uppskattning av populationsmedelvärdet, vet vi från kapitel 14 att urvalsmedelvärdet både är minstakvadratuppskattningen och, för en normalfördelad population, maximum likelihood uppskattningen av populationsmedelvärdet.

Vi har tidigare noterat att ett observerat urvalsmedelvärde $\bar{x}$ kan betraktas som en observation på den stokastiska variabeln $\bar{X}$, där

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

och $X_1, X_2, \dots, X_n$ är urvalets observationer. Vi vet då från kapitel 15 att

$$E(\bar{X}) = \mu$$

En tolkning av detta är att urvalsmedelvärdet i medeltal tenderar att vara lika med populationsmedelvärdet, dvs det finns ingen systematisk skillnad mellan mellan $\bar{X}$ och μ . Vi säger då att $\bar{X}$ är en *väntevärdesriktig* uppskattning av μ . Vidare vet vi att

$$V(\bar{X}) = \sigma^2/n.$$

Om populationen är normalfördelad vet vi slutligen att

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

eller, om vi så vill

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n}).$$

I de fall då populationsvariansen är okänd, kan den uppskattas med urvalsvariansen s^2 . Genom att betrakta s^2 som en observation på den stokastiska variabeln S^2 , där

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

får vi från kapitel 15.4 att

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Om populationen inte är normalfördelad måste vi förlita oss på centrala gränsvärdessatsen vilken förutsätter att antalet observationer, n , är stort. Tydligt gäller det då approximativt att

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

om σ är känd och n är stort, eller om σ är okänd approximativt att

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0, 1)$$

då n är stort.

Proportioner kan betraktas som medelvärden av variabler som endast kan anta värdena 0 och 1,

$$P = \frac{1}{n} \sum_{i=1}^n X_i,$$

där $X_i = 0$ om individ nr i saknar den studerade egenskapen och $X_i = 1$ om individ nr i har den studerade egenskapen. Om proportionen individer i populationen som har den studerade egenskapen är π vet vi att $P(X_i = 1) = \pi$ och $P(X_i = 0) = 1 - \pi$. På samma sätt som för medelvärden i allmänhet följer under samma förutsättningar som tidigare att

$$E(P) = \pi,$$

dvs att urvalsproportionen är en väntevärdesriktig uppskattning av populationsproportionen och att

$$V(P) = \frac{\pi(1 - \pi)}{n}.$$

Från centrala gränsvärdessatsen har vi slutligen att

$$\frac{P - \pi}{\sqrt{P(1 - P)/n}} \sim N(0, 1)$$

approximativt för stora urval.

De approximationer vi anger här baserar sig alla på centrala gränsvärdessatsen. Tyvärr säger inte den satsen någonting om hur många observationer vi måste göra för att approximationen skall bli bra. Den säger bara att ju fler observationer vi gör, ju bättre blir approximationen. För de flesta tillämpningar och om populationen inte är alltför snedfördelad brukar man anse att $n = 30$ observationer ger en tillfredsställande god approximation.

Exempel 16.1

Herr A's restid till arbetet är en normalfördelad stokastisk variabel med (populations-) variansen $\sigma^2 = 4$. Under en fyraveckorsperiod ($n = 20$ arbetsdagar) mäter han restiden och beräknar medelvärdet till $\bar{x} = 26,5$ minuter. Vi använder därför $\bar{x} = 26,5$ minuter som en uppskattning av herr A's restid till arbetet. Om vi antar att observationerna på herr A's restid är oberoende av varandra får vi att urvalsmedelvärdet har variansen $\sigma_{\bar{X}}^2 = 4/20 = 0,2$ och följaktligen standardavvikelsen $\sigma_{\bar{X}} = 2/\sqrt{20} = \sqrt{0,2} \approx 0,45$. ■

16.2 Uppskattning vid ändlig population

Om populationen är ändlig och urvalet görs utan återläggning gäller det fortfarande att $E(\bar{X}) = \mu$ men variansen måste modifieras till

$$V(\bar{X}) = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$$

där N är populationsstorleken. I övrigt gäller alla resultat från föregående avsnitt. Skillnaden mot fallet med oberoende observationen är att variansuttrycket nu också innehåller faktorn $(N-n)/(N-1)$. Denna faktor kallas *ändlighetskorrektur*. Om urvalsstorleken n är liten i förhållande till populationsstorleken N är ändlighetskorrekturen nära 1 och kan normalt bortses ifrån. I det fallet kan vi således approximera $V(\bar{X})$ med det enklare uttrycket σ^2/n som vi använder vid oberoende observationer. Som tumregel brukar man bortse från ändlighetskorrekturen om den är större än 0,95.

Exempel 16.2

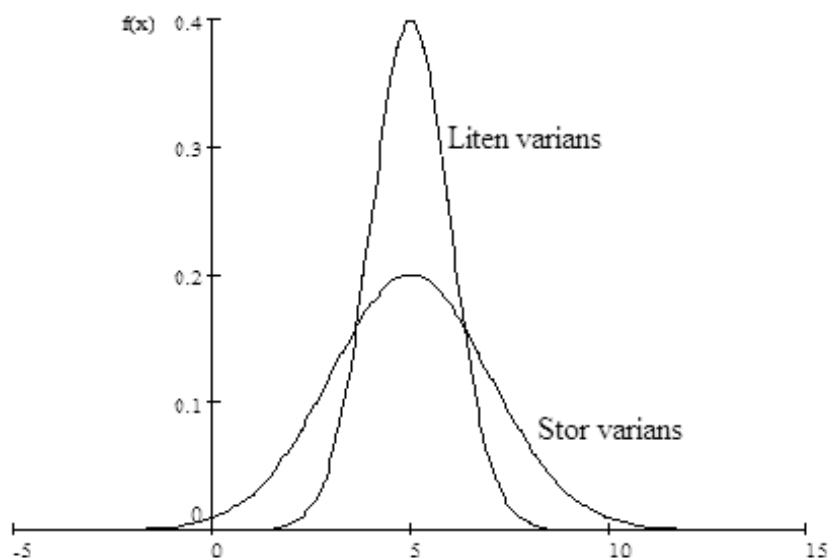
En skola har $N = 500$ elever. Ett urval om $n = 100$ elever görs för att undersöka elevernas åsikter om trivsel i skolans matsal. Bland annat visade det sig att 42 elever var nöjda med matsalens salladsbord. Vi använder därför $p = 42/100 = 42\%$ som en uppskattning av andelen elever som är nöjda med salladsbordet. En uppskattning av variansen för denna proportion är

$$s^2 = \frac{0,42(1-0,42)}{100} \left(\frac{500-100}{500-1} \right) \approx 0,001953$$

och på motsvarande sätt blir en uppskattning av standardavvikelsen för proportionen nöjda $s = \sqrt{0,001953} \approx 0,04419$. ■

16.3 Osäkerhet i uppskattningar

Begreppet osäkerhet i uppskattningar kan definieras på många olika sätt. En något "luddig" definition som förmodligen de flesta kan acceptera är följande: Osäkerheten är stor om sannolikheten är stor att det uppskattade värdet avviker mycket från det sanna värdet. Och omvänt, osäkerheten är liten om det är liten sannolikhet att vi skall få en uppskattning som avviker kraftigt från det sanna värdet. Genom att studera samplingfördelningen kan vi avgöra om sannolikheten är stor eller liten att en uppskattning avviker mycket från det sanna värdet.



Figur 16.1 Två samplingfördelningar med samma väntevärde, μ , men med olika varians.

Samplingfördelningen för urvalsmedelvärdet har väntevärdet lika med det sanna värdet (populationsmedelvärdet), μ . Figur 16.1 visar två samplingfördelningar med väntevärdet $\mu = 5$. Den ena är mer koncentrerad kring μ och har alltså mindre varians. Den andra fördelningen är mer utbredd och har

därför större varians. Tydligt är det mindre sannolikhet att vi skall få ett urvalsmedelvärde som ligger långt från μ i den fördelning som har liten varians. Detta innebär att variansen för urvalsmedelvärdet är ett mått på osäkerheten i uppskattningen. Liten varians hos urvalsmedelvärdet ger koncentrerad samplingfördelning, som i sin tur ger liten osäkerhet i uppskattningen. En stor varians hos urvalsmedelvärdet ger en utbredd, ”tjocksvansad”, samplingfördelning och en stor osäkerhet i uppskattningen.

Vi har tidigare noterat att variansen för urvalsmedelvärdet minskar då urvalsstorleken ökar. Om vi använder variansen för urvalsmedelvärdet som mått på osäkerheten, ser vi att osäkerheten minskar då urvalsstorleken ökar. Detta stämmer väl överens med vår intuition att osäkerheten bör minska då vi baserar våra slutsatser på fler observationer.

16.4 Konfidensintervall

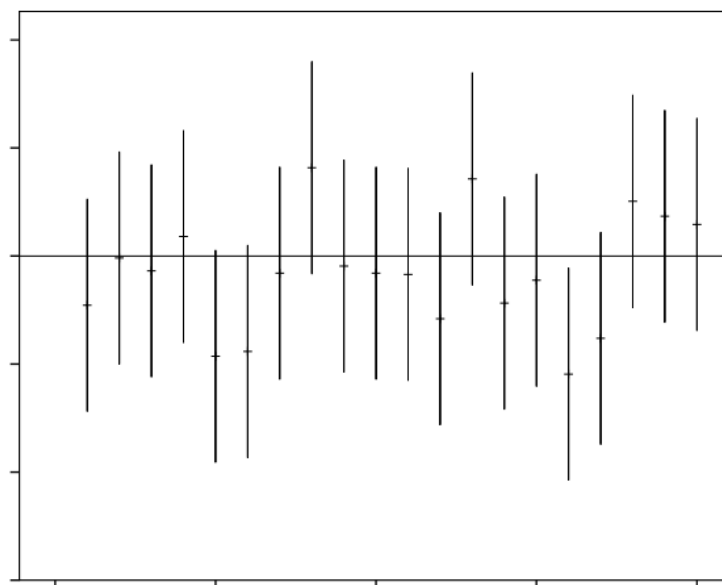
Vi ska nu fördjupa diskussionen om osäkerhet i uppskattningar och införa ett mer sofistikerat sätt att ange osäkerhet. Från centrala gränsvärdessatsen vet vi att om vi har ett tillräckligt stort urval kommer urvalsmedelvärdet $\bar{X}$ att bli approximativt normalfördelat. Därav följer att sannolikheten att $\bar{X}$ skall hamna utanför gränserna $\mu \pm 1.96\sigma_{\bar{X}}$ är 0.05, där $\sigma_{\bar{X}}$ är standardavvikelsen för urvalsmedelvärdet. Skrivet i symboler är det

$$\begin{aligned} 0,05 &= P\left(-1,96 \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq 1,96\right) \\ &= P(-1,96 \cdot \sigma_{\bar{X}} \leq \bar{X} - \mu \leq 1,96 \cdot \sigma_{\bar{X}}) \\ &= P(\bar{X} - 1,96 \cdot \sigma_{\bar{X}} \leq \mu \leq \bar{X} + 1,96 \cdot \sigma_{\bar{X}}) \end{aligned}$$

dvs händelsen att intervallet $\bar{X} \pm 1,96 \cdot \sigma_{\bar{X}}$ innehåller det sanna värdet μ inträffar med sannolikheten 0,95. Vi kan därför skriva $\bar{x} \pm 1,96 \cdot \sigma_{\bar{x}}$ och därigenom ange en ”osäkerhetsmarginal” eller ett intervall som i 95 fall av 100 innehåller det sanna värdet. Ett sådant intervall brukar kallas ett 95 %-igt *konfidensintervall*.

Vi kan alltså tolka konfidensintervallet som en intervallskattning av μ , dvs i stället för att ange ett enda värde, en punktskattning i form av $\bar{x}$, anger vi ett helt intervall med rimliga värden på μ .

Observera att ett konfidensintervall bestäms av det observerade värdet $\bar{x}$. Ett konfidensintervall bestäms alltså inte av någon stokastisk variabel. Om ett givet intervall innehåller eller inte innehåller det sanna populationsvärdet vet vi inte. Det är därför fel att säga att ett 95 %-igt konfidensintervall innehåller det sanna värdet med sannolikheten 0,95. Ett konfidensintervall kan antingen innehålla eller inte innehålla det sanna värdet. Däremot kan vi säga att om vi utför vårt försök (urval) väldigt många gånger och för varje gång beräknar ett 95 %-igt konfidensintervall, kommer ungefär 95 % av alla intervall att innehålla det sanna värdet. Denna tolkning kommer ur följande resonemang: Antag att vi upprepar urvalsdragningen ett antal gånger och för varje urval vi drar beräknar vi ett 95%-igt konfidensintervall. Vi kommer då att få ett antal sådana intervall, ett för varje urval. Intervallen kommer att bli något olika eftersom urvalsmedelvärdet och därmed intervallens mittpunkt, kommer att bli olika för de olika urvalen. Figur 16.2 visar ett exempel med konfidensintervall beräknade från 20 olika urval. Urvalen har samma storlek och är dragna ur samma population. Linjen mitt i figuren visar populationsmedelvärdet. De flesta intervallen innehåller det sanna populationsmedelvärdet men ett av intervallen innehåller inte populationsmedelvärdet.



Figur 16.2 Konfidensintervall från 20 olika urval av samma storlek från samma population.

Oftast kan man dra ett mycket stort antal urval av en viss storlek från en population och varje urval har sitt eget urvalsmedelvärde och sitt eget konfidensintervall. De intervall som visas i figur 16.2 är bara ett litet antal av alla intervall som kan beräknas från alla möjliga urval. Det urval och det intervall vi faktiskt fått i en undersökning är endast ett av alla intervall som är möjliga. Tyvärr vet vi inte om det intervall vi fått är ett intervall som innehåller det sanna värdet eller inte. Å andra sidan vet vi att 95 % av alla möjliga 95 %-iga konfidensintervall verkligen innehåller populationsmedelvärdet. Vi säger därför att konfidensgraden är 95 %.

Tydligt har vi att ett 95 %-igt konfidensintervall har formen

$$\bar{x} \pm 1,96 \cdot \sigma_{\bar{x}}.$$

Längden av konfidensintervallet på var sida om $\bar{x}$ brukar kallas den *statistiska felmarginalen*.

På motsvarande sätt som vi bildar ett 95 %-igt konfidensintervall kan vi konstruera konfidensintervall med andra konfidensgrader. Vi får t ex ett 99 %-igt konfidensintervall genom att använda kunskapen att

$$P\left(-2,58 \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq 2,58\right) = 0,99$$

för att komma fram till att $\bar{x} \pm 2,58 \cdot \sigma_{\bar{X}}$ är ett 99 %-igt konfidensintervall.

Exempel 16.1 (*fortsättning*)

Ett 95 %-igt konfidensintervall för den sanna restiden för herr A är

$$26.5 \pm 1.96 \cdot 2/\sqrt{20}$$

dvs ungefär 26.5 ± 0.9 eller $25.6 - 27.4$ minuter. ■

Exempel 16.2 (*fortsättning*)

Ett 95 %-igt konfidensintervall för andelen elever som är nöjda med matsalens salladsbord är

$$0,42 \pm 1,96 \cdot 0,04419$$

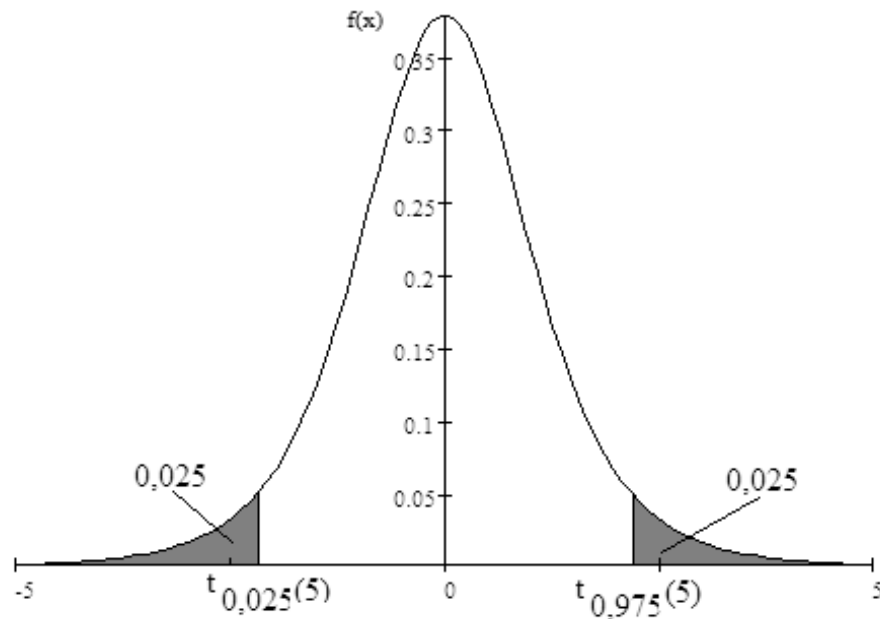
dvs ungefär 33 % – 51 %. ■

Tyvärr känner vi inte alltid populationsvariansen σ^2 och därmed heller inte urvalsmedelvärdets varians. Detta gör att vi inte kan beräkna ett konfidensintervall om vi inte kan använda en uppskattning av σ^2 i stället. Vi har tidigare nämnt att populationsvariansen kan uppskattas med urvalsvariansen s^2 . Om observationerna är normalfördelade bygger konstruktionen av ett konfidensintervall för μ då σ^2 är okänd på det faktum att den stokastiska variabeln

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

är t -fördelad med $n-1$ frihetsgrader. Vi skall nu använda detta och genomföra precis samma resonemang som tidigare för att konstruera ett 95 %-igt konfidensintervall. Låt därför $t_{0,975}(n-1)$ vara det tal som är sådant att sannolikheten är 0,025 att en t -fördelad variabel med $n-1$ frihetsgrader är större

än $t_{0,975}(n-1)$. Följaktligen är sannolikheten 0,975 att en sådan stokastisk variabel är mindre än $t_{0,975}(n-1)$. Ett exempel med $n-1 = 5$ visas i figur 16.2.



Figur 16.3 t -fördelningen med 5 frihetsgrader och talen $t_{0,025}(5)$ och $t_{0,975}(5)$ markerade

Det gäller då att

$$P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{0,975}(n-1)\right) = 0,975.$$

Eftersom t -fördelningen är symmetrisk kring noll, gäller det att $t_{0,025}(n-1) = -t_{0,975}(n-1)$, och

$$P\left(-t_{0,975}(n-1) \leq \frac{\bar{X} - \mu}{S/\sqrt{n}}\right) = 0,025.$$

Sätter vi ihop detta får vi enligt additionssatsen

$$P\left(-t_{0,975}(n-1) \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{0,975}(n-1)\right) = 0,95$$

vilket är detsamma som

$$P\left(\bar{X} - t_{0,975}(n-1) \cdot S/\sqrt{n} \leq \mu \leq \bar{X} + t_{0,975}(n-1) \cdot S/\sqrt{n}\right) = 0,95.$$

Vi bildar därför vårt konfidensintervall genom

$$\bar{x} \pm t_{0,975}(n-1) \cdot s/\sqrt{n}.$$

På motsvarande sätt bildas ett 99 %-igt konfidensintervall genom

$$\bar{x} \pm t_{,995}(n-1) \cdot s/\sqrt{n},$$

där $t_{0,995}(n-1)$ är det tal som är sådant att sannolikheten är 0,005 att en t -fördelad stokastisk variabel med $n-1$ frihetsgrader är större än $t_{0,995}(n-1)$.

Exempel 16.1 (fortsättning)

Antag att (populations-) standardavvikelsen σ är okänd men uppskattas med den observerade urvalsstandardavvikelsen $s = 2$. Eftersom $n = 20$ och $t_{0,975}(19) = 2,093$ får vi ett 95 %-igt konfidensintervall för den sanna restiden för herr A genom

$$26,5 \pm 2,093 \cdot 2/\sqrt{20}$$

dvs ungefär $26,5 \pm 0,936$ eller $25,564 - 27,436$. I det fall då vi uppskattar σ får vi alltså ett något större konfidensintervall än då σ är känd. Detta stämmer väl överens med vår intuition att våra utsagor är mer precisa om vi har mer kunskap om populationen (σ känd) och mindre precisa om vi har mindre kunskap om populationen (σ okänd). ■

16.5 Dimensionering av urvalsstorleken

Eftersom osäkerheten i en induktiv slutledning (som t ex uppskattningen av ett populationsmedelvärde) minskar då vi gör fler observationer, är det önskvärt att göra många observationer. Det är dock ofta relativt kostsamt att göra många observationer. I detta avsnitt skall vi visa hur man kan räkna ut erforderligt antal observationer för att uppnå en på förhand given noggrannhet i uppskattningarna.

Tidigare i detta kapitel har vi sett att ett 95 %-igt konfidensintervall för ett populationsmedelvärde μ ges av

$$\bar{x} \pm B$$

där

$$B = 1,96 \cdot \sigma_{\bar{X}}$$

är den statistiska felmarginalen. Om observationerna är stokastiskt oberoende så är

$$\sigma_{\bar{X}}^2 = \sigma^2/n$$

Antag att vi nu bestämmer oss för hur stort konfidensintervall vi vill ha. Eftersom B är halva konfidensintervallets bredd, får vi därigenom ett visst värde på B . Tillsammans med variansformeln får vi ekvationen

$$B = 1,96 \cdot \sigma/\sqrt{n},$$

där urvalsstorleken n är obekant. Lösningen av denna ekvation är

$$n = 1.96^2 \cdot \frac{\sigma^2}{B^2} \approx 4 \cdot \frac{\sigma^2}{B^2},$$

som alltså anger det antal observationer vi måste göra för att ett 95 %-igt konfidensintervall skall ha bredden $2 \cdot B$ (eller för att den statistiska felmarginalen får vara högst B).

Exempel 16.1 (*fortsättning*)

Antag att vi vill att halva bredden av ett 95 %-igt konfidensintervall för herr A's medelrestid till arbetet skall vara $B = 0,8$, dvs vi vill få ett konfidensintervall av typen

$$\bar{x} \pm 0,8$$

Vi måste då göra

$$n \approx 4 \cdot \frac{2^2}{0.8^2} = 25$$

observationer. ■

Vi arbetar ofta med ändliga populationer och gör urval utan återläggning. I sådana situationer tillkommer ändlighetskorrektionen $(N - n) / (n - 1)$ i variansformeln, dvs

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \left(\frac{N - n}{N - 1} \right),$$

där N är antalet individer i hela populationen. Vi kan nu på samma sätt som tidigare beräkna erforderlig urvalsstorlek, dvs med hjälp av variansformeln och ett värde på B får vi ekvationen

$$B = 1.96 \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}.$$

Löser vi ut n ur denna ekvation får vi erforderligt antal observationer för att den statistiska felmarginalen skall vara B . Vi får lösningen

$$n = \frac{N \cdot \sigma^2}{\frac{B^2}{1.96^2} (N - 1) + \sigma^2} \approx \frac{\sigma^2}{\frac{B^2}{4} + \frac{\sigma^2}{N}}.$$

Detta kan ytterligare förenklas om beräkningarna genomförs i två steg. I det första steget gör vi en "approximativ" beräkning enligt formeln för oberoende observationer. Beteckna svaret med n_0 . I det andra steget beräknar vi den exakta lösningen genom

$$n = \frac{n_0}{1 + \frac{n_0}{N}}.$$

Observera att om N är stort i förhållande till n (eller n_0) blir de två formlerna för urvalsstorleken ungefär lika.

Exempel 16.3

Ett företag har 5000 anställda. Vi är intresserade av att bilda ett 95 %-igt konfidensintervall för medelinkomsten hos företagets anställda. Intervallets totala längd skall vara 200 kr. Om standardavvikelsen är 2000 kr måste vi göra

$$n = \frac{2000^2}{\frac{100^2}{4} + \frac{2000^2}{5000}} \approx 1213$$

observationer. ■

När vi skall dimensionera storleken på ett urval måste vi känna populationsvariansen σ^2 . Tyvärr är populationsvariansen oftast okänd, men om vi på förhand kan uppskatta den kan vi få en approximativ beräkning av erforderlig urvalsstorlek.

Ett sätt att uppskatta populationsvariansen är att genomföra en förundersökning med ett mindre antal observationer. Urvalsvariansen s^2 från förundersökningen kan då användas som en uppskattning av populationsvariansen. Sätter vi sedan in s^2 i stället för σ^2 i uttrycket för beräkning av urvalsstorleken får vi en uppskattning av erforderligt antal observationer.

Ett annat sätt att uppskatta populationsvariansen bygger på den så kallade "empirical rule", som innebär att

$$\sigma \approx \frac{\text{största värdet} - \text{minsta värdet}}{4}.$$

Om vi har en uppfattning om vad det största och det minsta värdet är i den undersökta populationen, och inget av dessa värden är oändligt stort, kan vi få ett approximativt värde på σ . Detta värde används sedan för att uppskatta erforderlig urvalsstorlek.

När det gäller proportioner, vet vi att populationsvariansen är $\sigma^2 = \pi(1 - \pi)$, där π är populationsproportionen. Eftersom kvantiteten $\pi(1 - \pi)$ aldrig kan bli större än 0,25 kan populationsvariansen heller aldrig bli större än 0,25.

Sätter vi in $\sigma^2 = 0,25$ i uttrycket för urvalsstorleken och beräknar n , kan vi därför inte få för litet n .

Exempel 16.4

I en kommun finns 80 000 röstberättigade. Hur stort urval måste vi dra för att uppskatta proportionen sympatisörer med x-partiet, om vi vill att längden på ett konfidensintervall skall vara högst fyra procentenheter?

Det gäller att $N = 80000$, $B = 0,04/2 = 0,02$ och eftersom vi undersöker en proportion sätter vi $\sigma^2 = 0,25$. Med hjälp av uttrycket för urvalsstorleken får vi

$$n \approx \frac{0,25}{\frac{0,02^2}{4} + \frac{0,25}{80000}} \approx 2424,27$$

Urvalet bör omfatta 2425 individer för att vi skall vara säkra på att uppnå den givna precisionen. ■

Övningar

16.1 För att uppskatta medelvärdet μ för en normalfördelad stokastisk variabel med känd varians $\sigma^2 = 36$ gjorde man 25 observationer på den aktuella variabeln varvid man erhöll $\bar{x} = 14,25$. Bilda ett 95 %-igt konfidensintervall för variabelns medelvärde μ .

16.2 Antag att maxhastigheten för bilar av ett visst bilmärke är en stokastisk variabel med variansen 30. För att uppskatta medelvärdet μ togs ett urval om 50 bilar ut. Summan av de observerade maxhastigheterna blev $\sum_{i=1}^n x_i = 8168$. Beräkna en uppskattning och ett 95 %-igt konfidensintervall för variabelns väntevärde.

16.3 Ett postorderföretag lovar i sin reklam att order som görs via Internet levereras inom tre dagar. Ett slumpmässigt urval av gjorda ordrar visar att ett 95 %-igt konfidensintervall för proportionen ordrar levererade inom tre dagar är $91\% \pm 5\%$. Förklara vad detta innebär och avgör om nedanstående påståenden är sanna eller falska.

- a) 95 % av alla slumpmässiga urval av ordrar visar att 91 % av orderarna levereras inom tre dagar.
- b) I 95 % av alla dagar kommer mellan 86 % och 96 % av orderarna att levereras inom tre dagar.
- c) Mellan 86 % och 96 % av alla ordrar levereras inom tre dagar.
- d) 95 % av alla slumpmässiga urval av ordrar visar att leverans sker inom tre dagar för mellan 86 % och 96 % av alla ordrar.
- e) Vi är 95 % säkra på att mellan 86 % och 96 % av orderarna i detta urval levereras inom tre dagar.

16.4 Urvalsstorlek, konfidsensgrad och statistisk felmarginal är några faktorer som är inblandade då man bildar ett konfidensintervall. Avgör om följande påståenden är korrekta.

- a) Färre observationer ger en mindre konfidsensgrad då den statistiska felmarginalen hålls konstant.
- b) Att minska den statistiska felmarginalen innebär att man minskar konfidensgraden då urvalsstorleken hålls konstant.
- c) Ett urval som är fyra gånger större halverar den statistiska felmarginalen då konfidensgraden hålls konstant.
- d) Man får en mindre statistisk felmarginal genom att öka antalet observationer då konfidensgraden hålls konstant.

16.5 I en undersökning av livslängden hos transistorer framställda enligt en viss metod valdes 390 transistorer slumpmässigt ut. Dessa utsattes för en viss behandling och livslängden hos de utvalda transistorerna observerades, se nedanstående tabell.

Livslängd i timmar	Frekvens
1-10	91
11-20	70
21-30	54
31-40	43
41-50	33
51-60	26
61-70	20
71-80	16
81-90	12
91-100	9
101-110	5
111-120	3
121-130	4
131-140	3
141-150	1

Ange ett 95 %-igt konfidensintervall för förväntade livslängden hos transistorer framställda enligt metoden.

16.6 I ett slumpmässigt urval om 1224 personer över 18 år bosatta i Sverige uppgav 6 % att de hade svårigheter att få pengarna att räcka till för sina utgifter varje månad. Bestäm ett 95 %-igt konfidensintervall för proportionen personer över 18 år som har svårigheter att få pengarna att räcka till alla utgifter varje månad.

16.7 Vi har fått i uppdrag att genomföra en enkätundersökning bland värnpliktiga vid ett regemente. Undersökningens syfte är att utröna de värnpliktigas inställning till vissa trivselfrämjande åtgärder. Bredden av ett 95 %-igt konfidensintervall för proportionen positiva till en viss åtgärd får högst vara 6 procentenheter. Hur stort urval måste vi dra om antalet värnpliktiga är 1775 vid undersökningstillfället?

16.8 Inför ett riksdagsmannaval vill man i en kommun uppskatta hur väljarnas sympatier för de olika politiska partierna fördelar sig. Avsikten är att jämföra resultaten från denna undersökning med valresultatet, för att utröna hur

den politiska debatten under tiden före valet påverkar partisympatierna i kommunen. För den skull planerar man att dra ett urval från en komplett röstlängd omfattande 30 000 personer. De olika partiernas andel av väljarkåren förmodas ligga inom nedanstående intervall:

Moderaterna	20 - 30 %
Folkpartiet	5 - 15 %
Kristdemokraterna	4 - 12 %
Centerpartiet	5 - 15 %
Socialdemokraterna	30 - 45 %
Miljöpartiet de gröna	3 - 6 %
Vänsterpartiet	3 - 10 %

- a) Hur stort är det minsta urval som erfordras om man vill bestämma den relativa andelen av väljarkåren för vart och ett av de sju partierna med sådan precision att längden av ett 95 %-igt konfidensintervall ej kommer att överstiga 0,04 för något av partierna?
- b) Diskutera undersökningens målsättning.
- c) Vilka intervalllängder erhålles för socialdemokrater respektive kristdemokrater om observerade proportioner är 0,35 respektive 0,07?

16.9 För det senaste räkenskapsåret uppvisar företaget Soling AB en bruttovinst som i förhållande till företagets storlek är betydligt lägre än branchen i genomsnitt. Bland annat bestäms att en revision skall göras hos företaget. Speciellt intresse ägnas åt granskning av företagets försäljning. Under räkenskapsåret har företaget bokfört 1100 fakturor på försäljningskontot. Ett slumpmässigt urval om 100 fakturor väljs och granskas. För 11 av dessa har det fakturerade beloppet bokförts felaktigt. Sammanlagt är det bokförda värdet 22 000 kr lägre än de fakturerade beloppen. Den statistiskt intresserade revisorn beräknar standardavvikelsen för bokföringsfelen (= skillnaden mellan fakturerat och bokfört belopp) till 700 kr för hela urvalet.

- a) Uppskatta proportionen felaktigt bokförda försäljningsfakturor, dels med en punktskattning och dels med ett 95 %-igt konfidensintervall.
- b) Uppskatta genomsnittligt bokföringsfel för de bokförda fakturorna. Ange dels en punktskattning och dels ett 95 %-igt konfidensintervall.

c) Ge en verbal tolkning av ett konfidensintervall. Gör det i formuleringar begripliga för revisorer. Exemplifiera med ett av de båda intervallen du beräknat ovan.

16.10 Efter omfattande studier av en förpackningsmaskin för potatis vet man att den producerar potatispåsar vars vikter kan betraktas som oberoende observationer på en stokastisk variabel. Vidare vet vi att en normalfördelning med väntevärdet μ och standardavvikelsen 0,12 kg tjänar som en god modell för variabeln.

a) I samband med en service på maskinen gjordes kontrollpåfyllningar av 20 påsar. Det visade sig att genomsnittsvikten blev 2,53 kg. Beräkna ett 95 %-igt konfidensintervall för väntevärdet μ av vikten av en påse.

b) Hur många påsar bör du kontrollväga om du vill att din skattning av μ inte ska ha en större statistisk felmarginal (halva konfidensintervallet) än 0,02 kg?

16.11 För att bestämma en reaktionstid använder en psykolog en mätmetod som ger normalfördelade tider med standardavvikelsen 0,05 sekunder. Antag att psykologen vid ett tillfälle erhållit följande fem reaktionstider:

1,05 sek 0,97 sek 1,02 sek 1,01 sek 0,96 sek

a) Bestäm konfidensintervall för väntevärdet med 95 %, 99 % och 99,9 % konfidensgrad.

b) Antag att psykologen vill göra så många mätningar av en reaktionstiden att det är möjligt att beräkna ett 95 %-igt konfidensintervall för väntevärdet med totala längden 0,01 sekunder. Hur många mätningar måste psykologen minst göra?

16.12 Av en årsmodell finns 600 inregistrerade personbilar av ett visst märke. Man ville uppskatta den årliga genomsnittliga körsträckan för dessa bilar. Ett slumpmässigt urval om 200 bilar drogs med hjälp av bilregistret. För de 200 bilarna fann man en genomsnittlig körsträcka på 1373 mil med standardavvikelsen 264 mil.

a) Beräkna ett 95 %-igt konfidensintervall för den genomsnittliga årliga körsträckan i den studerade populationen.

b) På Trafiksäkerhetsverket fick man se resultatet av undersökningen. Man tyckte det skulle vara intressant att göra en motsvarande undersökning som gällde alla personbilar (oavsett märke) av just den studerade årsmodellen. Om standardavvikelsen för denna utvidgade population kan antas vara ungefär samma som för den speciella modellen, hur stort urval krävs för att den statistiska felmarginalen skall bli detsamma?

16.13 Med uppgifter ur en enkät till 300 villaägare i landet med hus byggda efter år 2000 kunde man beräkna den årliga genomsnittliga belåningen av villorna till 2 236 200 kr med en standardavvikelse på 346 000 kr.

a) Beräkna ett 95 %-igt konfidensintervall för den genomsnittliga belåningen av villorna i den studerade populationen.

b) I en kommun vill man göra en liknande undersökning. I kommunen finns totalt 75 villor byggda efter år 2000. Man vill dock ha en dubbelt så stor noggrannhet som i undersökningen i a) (hälften så stor statistisk felmarginal). Man kan anta att standardavvikelsen för villabelåningen är ungefär densamma som gällde för undersökningen i uppgift a. Vilken är den minsta urvalsstorleken som krävs?

16.14 I en artikel i *Journal of Accounting Research* studerade Ashton, Willingham och Elliot tidsfördröjningen mellan verksamhetsårets slut och publiceringen av revisionsrapporten för industriföretag och företag inom finansbranchen. Ett slumpmässigt urval om 250 industriföretag gav ett medelvärde om 68,04 dagar och en standardavvikelse om 35,72 dagar i tidsfördröjning. Ett slumpmässigt urval om 238 finansföretag gav en fördröjning som i medeltal var 56,74 dagar med en standardavvikelse om 34,87 dagar.

a) Bestäm ett 95 %-igt konfidensintervall för medelfördröjningstiden för alla industriföretag.

b) Bestäm ett 95 %-igt konfidensintervall för medelfördröjningstiden för alla finansföretag.

c) Undersök om det finns belägg för att tidsfördröjningarna är olika för industriföretag och finansföretag genom att jämföra konfidensintervallen i uppgift a och uppgift b.