

# Statistisk acceptansk kontroll

## BILAGA 2 - Grundbegrepp i statistiken

Metodbeskrivning 908:1994

B2 Grundbegrepp i statistiken.....	3
B2.1 Syften och fundament.....	3
B2.1.1 Statistik - ett hjälpmedel för verklighetsbeskrivning.....	3
B2.1.2 Slumpmässighet.....	4
B2.2 Variabel, sannolikhet och fördelning.....	5
B2.2.1 Diskreta och kontinuerliga variabler .....	5
B2.2.2 Sannolikhetsbegreppet.....	6
B2.3 Fördelningars karaktäristika .....	12
B2.3.1 Central tendens - "medelvärden" .....	12
B2.3.2 Variabilitet - "spridning" .....	13
B2.3.3 Några fördelningstyper .....	15
B2.4 Skattning .....	23
B2.4.1 Skattning = estimation .....	23
B2.4.2 Punkt- och intervallskattning.....	23
B2.4.3 Vänteriktighet .....	23
B2.4.4 Effektivitet (efficiens).....	24
B2.4.5 Samplingfördelning .....	26
B2.4.6 Intervallskattning - intervallestimat - konfidensintervall .....	27
B2.4.7 "Enkelsidigt" (asymmetriskt) konfidensintervall .....	28
B2.5 Hypotesprövning .....	29
B2.5.1 Den generella teorin.....	29
B2.5.2 Exempel, dubbelsidigt test.....	31
B2.5.3 Fel av typ I och fel av typ II, power och OC-kurvor .....	32
B2.5.4 Enkelsidigt test .....	34
B2.6 Acceptansintervall .....	36

## B2 Grundbegrepp i statistiken

Metodbeskrivningars syfte är i första hand att garantera, att ingenting viktigt blir förbiset i ett arbetssätt som man bestämt sig för att tillämpa - inte att *motivera* ett visst arbetssätt. Emellertid har det ofta visat sig, att just tillämpning av statistisk metodik utan insikt om varför metoden är uppbyggd som den är både inbjuder till allvarliga felgrepp och leder till att det statistiska arbetssättet uppfattas som konstigt, tråkigt och "onödigt". Denna bilaga har tagits med i metodbeskrivningen med syftet att ge alla, som använder sig av de i tidigare avsnitt presenterade och i [VÄG 94](#) föreskrivna statistiska arbetssätten, möjligheter att skaffa sig en åtminstone översiktlig metodinsikt - förstå *varför* man gör som man gör och därför göra det bra.

Texten är skriven "från början", dvs avsikten är att även den som inte kan något nämnvärt om statistik skall kunna förstå framställningen, under förutsättning att vederbörande läser textstyckena i den ordning de kommer. Detta hindrar dock inte, att den som redan kan en del men vill friska upp gamla kunskaper också kan ha nytta av texten genom att endast läsa valda delar utan att trötta sig med sådant som redan är välbekant. Rubrikerna bör kunna vara till ledning för sådant urval.

### B2.1 Syften och fundament

#### B2.1.1 Statistik - ett hjälpmedel för verklighetsbeskrivning

"Statistik" är ett ord som väcker blandade känslor. Många tror, att statistik är enbart torrt och tråkigt. Den som skaffar sig en smula kunskap på området inser snart, att så är det inte alls. Statistik är ett hjälpmedel för att ta sig fram i och fatta beslut avseende den stundom kaotiska verkligheten. Statistik är dels ett sätt att beskriva verkligheten kvantitativt - och dels en metodarsenal för att utnyttja de kvantitativa beskrivningarna för att dra slutsatser om hur verkligheten är beskaffad. I en på sin tid berömd lärobok i statistik (Wallis-Roberts) beskrivs statistik som " ... a bunch of methods to make wise decisions in the face of uncertainty." En något fri översättning skulle vara: "Statistik är en knippe metoder att använda för att fatta kloka beslut i ovissa lägen".

Man talar om *beskrivande statistik* (deskriptiv statistik) respektive *statistisk slutsatsdragning* (statistisk inferens).

De flesta känner till "basverktygen" i den beskrivande statistiken. För ett observationsmaterial - t ex ett antal mätningar av en terrassytenivå - kan man beräkna medelvärdet. Man kan beskriva mätningarna i en tabell eller i diagramform (med t ex ett stapeldiagram, histogram). Man kan ange, hur pass samlade de olika mätvärdena är, t ex genom att ange största och minsta mätvärdet (eller differensen dessa emellan, den s k variationsvidden). Andra mått på variabiliteten i mätvärdena finns, t ex den s k standardavvikelsen.

Lite förenklat kan man säga, att beskrivande statistik tjänar till att *åskådliggöra konkreta observationsmaterial* med hjälp av tabeller, diagram och olika sammanfattningsmått.

Statistisk slutsatsdragning kan sägas kretsa kring grundfrågan: "När vi nu har det här faktiska observationsmaterialet, vad kan vi då våga säga om hur verkligheten är beskaffad - och med vilken säkerhet vågar vi påstå detta?"

Den statistiska slutsatsdragningen utnyttjar ofta förutsättningen att ett visst konkret observationsmaterial tillkommit genom att man "slumpmässigt" valt ut de observerade värdena ur en större (ofta oändligt stor) mängd tänkbara värden. Terrassytenivån, som vi nyss tog som exempel, skulle vi ju i och för sig kunna göra hur många mätningar som helst av. Det gör vi inte i praktiken. Vi genomför ett begränsat antal observationer, men gör dessa på ett sådant sätt, att de blir "representativa" för den oändliga mängd observationer som teoretiskt sett skulle kunna göras. Då kan vi enligt vissa regler använda den kunskap vi mätt oss fram till i vårt begränsade observationsmaterial för utsagor om hela den teoretiska, oändliga mängden. Vi kan "överföra" en utsaga om ett stickprov till en utsaga om "hela verkligheten", dvs göra *inference* (av latinets *inferre* ≈ bära med sig).

Ett sätt att åstadkomma representativitet är att göra ett "slumpmässigt" urval. För slumpmässiga urval gäller ett antal "naturlagar" som kan formuleras matematiskt och som kan hjälpa oss formulera slutsatser om "hela verkligheten" utifrån vad vi direkt kan mäta på vårt observationsmaterial.

### B2.1.2 Slumpmässighet

Vad är då slumpmässighet? Alla har vi väl en intuitiv förståelse av begreppet, men låt oss studera ett exempel för att precisera innebörden:

Om man har en helt korrekt, symmetrisk sexsidig tärning och står i begrepp att kasta den en gång så finns det några utsagor man kan göra innan kastet:

- Det kommer att bli en etta, tvåa, trea, fyra, femma eller sexa.
- Det är omöjligt att förutsäga, vilket av de sex möjliga värdena som kommer upp, men...
- det är lika troligt, att det blir en etta som att det blir en tvåa etc.

När tärningen kastas är det - rent "mekaniskt" sett - ett stort antal olika faktorer som samverkar till hur den slutligen kommer att falla. Den vinkel tärningen har vid utkastet, kastets kraft och riktning, vilket "spinn" tärningen får, vilket underlag den hamnar på, friktionen... Med modern industrirobotteknik vore det kanske möjligt att konstruera en maskin som kunde kasta tärning på ett så precist och oföränderligt upprepbart sätt att man alltid fick samma utfall, t ex bara treor. Detta skulle i så fall vara *motsatsen* till slumpmässighet - förutbestämda (deterministiska) utfall. Men vid vanligt tärningkastande är orsaksfaktorerna så många, så oöverskådliga, så varierande och så omöjliga att beskriva preciserat, att vi måste betrakta effekten, dvs utfallet av det enskilda kastet, som något som varierar "slumpmässigt" från försök till försök.

Men som exemplet visar, gäller även för slumpmässigheten vissa begränsningar och regler:

- Vi kan ange, vilka de *möjliga utfallen* är.
- Vi kan med 100 % säkerhet säga, att det *inte går* att i förväg ange, vilket av de möjliga utfallen vi kommer att få i ett enskilt kast, men...
- vi kan ange, hur *troligt* ett visst utfall är, *utifrån teoretisk utgångspunkt*.

Detta är typiskt för händelser, vars utfall varierar slumpmässigt. När orsakerna till variationen är mångfaldiga, utan att någon orsaksfaktor dominerar kraftigt över de andra och när faktorerna samverkar på ett sätt som inte kan beskrivas i funktionssamband - då är variationen slumpmässig. Eller, mer korrekt, då *hanterar* vi variationen som slumpmässig.

## **B2.2 Variabel, sannolikhet och fördelning**

### **B2.2.1 Diskreta och kontinuerliga variabler**

"En variabel är något som varierar", skulle man något naivt kunna säga. Inom statistiken är ofta "variabel" en kortform för "slumpvariabel", dvs en variabel ( $X$ ) som antar olika värden ( $x_1, x_2, x_3 \dots x_n$ ) på ett slumpmässigt sätt.

(I det följande betecknar vi variabel med stor bokstav, medan det eller de värden variabeln antar (eller kan anta) betecknas med små bokstäver.)

Man skiljer mellan slumpvariabler som endast kan anta *vissa* värden (inom ett intervall) - diskontinuerliga eller *diskreta* slumpvariabler - och slumpvariabler som kan anta *alla* värden (inom ett visst intervall) - *kontinuerliga* slumpvariabler. Att man behöver skilja situationerna åt beror på att vissa sannolikhetsteoretiska resonemang genomförs på olika sätt för de två fallen - se nedan.

I exemplet med tärningkast ovan hade vi en diskret slumpvariabel:  $X =$  "Antal ögon upp vid kast med tärning". De värden slumpvariabeln kan anta är givetvis  $x_1 = 1, x_2 = 2, x_3 = 3, \dots x_6 = 6$ . Med en mer "matematisk" formulering kan alltså slumpvariabeln anta värden som är naturliga tal inom intervallet  $1 \leq x_i \leq 6$ .

Ett exempel på en kontinuerlig slumpvariabel kunde vara:  $X =$  "Avvikelse i mm från riktvärde för terrassytenivå" vid mätning genom avvägning. Denna variabel kan anta vilket värde som helst inom ett intervall som förhoppningsvis inte är det teoretiskt tänkbara  $-\infty \leq x_i \leq +\infty$  utan något snävare.

Normalt avläser vi inte mätvärdena med oändlig noggrannhet, troligen inte mer än på hela mm när. Hantering av kontinuerliga variabler i praktiken innebär sålunda alltid en viss "diskretisering" - men vi behandlar ändå variabeln som vad den teoretiskt sett är, dvs kontinuerlig.

Ett annat exempel på diskret slumpvariabel kan vara:  $X = \text{"Antal godkända av fem uttagna prover avseende hålrums halt"}$ . Det inses lätt att  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$ ,  $x_4 = 3$ ,  $x_5 = 4$  och  $x_6 = 5$ .

Inom statistisk acceptansk kontroll talar man ibland om "variabelmetoden" och "attributmetoden". Variabelmetoden förutsätter, att man kan använda acceptanskriterier för *kontinuerliga* slumpvariabler. Om det inte går att hitta lämpliga kontinuerliga kriterievariabler får man konstruera *diskreta*, dvs tillämpa attributmetoden. Exemplet med hålrums halt ovan skulle leda till användning av attributmetoden.

Ibland förekommer (särskilt i ursprungligen engelskspråkig text) ordet "stokastisk" (stochastic på engelska). Det talas om stokastiska variabler och stokastiska processer. Ordet betyder helt enkelt slumpmässig. En stokastisk variabel är en slumpvariabel. En stokastisk process är en situation där en slumpvariabel antar värden. Mätningar av terassytenivån är alltså en stokastisk process - märkvärdigare är det inte.

### B2.2.2 Sannolikhetsbegreppet

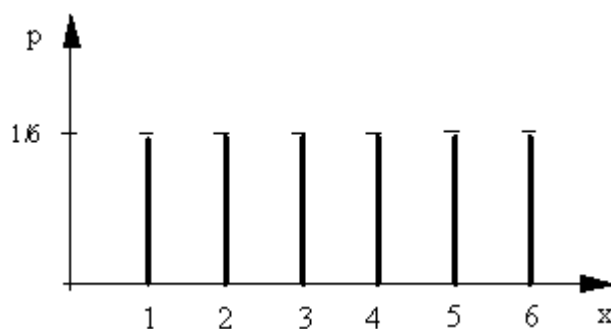
Flertalet nu levande svenskar kommer förmodligen att tänka på Tage Danielssons Harrisburg-sketch när de hör ordet "sannolikhet". Trots hans i och för sig briljanta förklaring är det kanske på sin plats att säga några ord till om detta fundamentala begrepp.

I formler betecknas sannolikhet ofta med  $P$  (efter engelska probability). Vi använder här stor bokstav för sannolikhet i allmänhet [ $P(X=x_i)$ ], liten indicerad bokstav [ $p_i$ ] för att ange sannolikhet för visst utfall [dvs  $p_i = P(X=x_i)$ ]. (Jfr beteckningssättet avseende slumpvariabler).

Man kan "förstå" sannolikhet på ett empiriskt sätt: Vi återgår till exemplet med tärningen och frågar oss: Vad är sannolikheten att få en etta [ $p_1 = P(X=x_1=1)$ ]? För att få svaret kastar vi tärningen ett mycket stort antal gånger ( $n$ ) och noterar frekvensen ettor, dvs hur många ettor vi får,  $f(x_1)$ . Kvoten  $f(x_1)/n$  kommer alltmer att närma sig  $1/6$  ju fler kast vi gör. Vi kan uppfatta sannolikheten för etta [ $p_1 = P(X=x_1)$ ] som *gränsvärdet* för kvoten  $f(x_1)/n$  när vi låter  $n$  gå mot oändligheten.

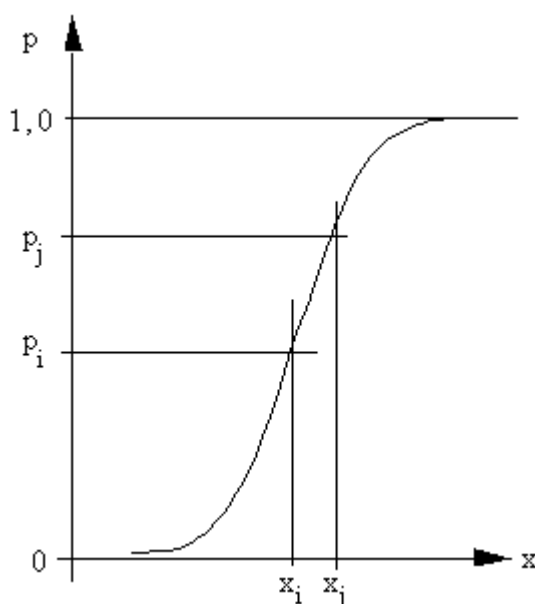
Mer generellt gäller, att sannolikheten för ett visst utfall kan uppfattas som gränsvärdet för kvoten (gynnsamma utfall)/(totalantalet försök) när totalantalet försök går mot oändligheten. Och om vi undersöker detta för *alla* tänkbara utfall  $X=x_i$  kommer vi att finna, att *summan* av de olika gränsvärdena är 1. Naturligtvis. Sannolikheten för att  $X$  ska anta *något* av sina möjliga värden är förstuds 100 %, dvs 1.

Ett mer "teoretiskt" sätt att definiera sannolikhet är att säga, att sannolikheten  $p_i$  för ett visst utfall  $X=x_i$  (där  $X$  är en slumpvariabel med möjliga utfall  $x_1, x_2, x_3, \dots, x_i, \dots, x_n$ ) är ett tal *förknippat* med utfallet  $X=x_i$  sådant att  $\sum p_i = 1$ .



Figur B2.2-1 Sannolikhet, diskret slumpvariabel

För kontinuerliga slumpvariabler kan vi inte definiera sannolikheten för att  $X$  skall anta ett visst värde. En kontinuerlig variabel kan ju anta *alla* värden (inom ett visst intervall) vilket innebär en oändlig mängd möjliga värden, hur litet intervallet än är. Sannolikheten för att  $X$  antar exakt värdet  $x_i$  blir därför noll. Här får definitionen formuleras t ex så att  $p_i$  anger sannolikheten för att (den kontinuerliga) slumpvariabeln  $X$  antar ett värde  $x_i$  *eller mindre*, dvs antar värden inom ett visst intervall. Sannolikheten för att  $X$  antar värden mellan två gränser,  $x_i$  och  $x_j$  (där  $x_j > x_i$ ) beräknas som  $p_j - p_i$ .

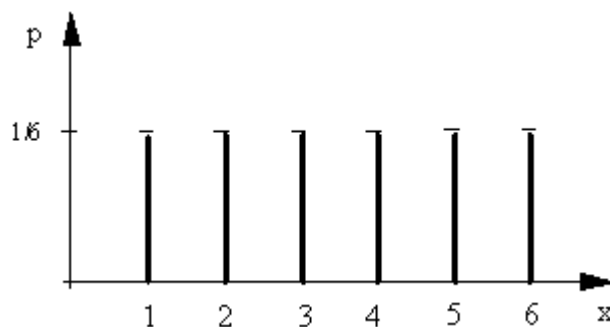


Figur B2.2-2 Sannolikhet, kontinuerlig slumpvariabel

Särskilt för kontinuerliga slumpvariabler brukar man också tala om "sannolikhetstäthet" - se vidare nedan.

### B2.2.3 Fördelningsbegreppet

Med fördelning menar man det "sätt" på vilket en variabels olika värden "uppträder". Låt oss betrakta tärningsexemplet igen. De olika utfallen har samma sannolikhet och en bild av det "sätt" på vilket värdena för slumpvariabeln  $X =$  Antal ögon upp vid kast med tärning "uppträder" är följande:



Figur B2.2-3 Sannolikhetsfördelning för  $X$ =Antal ögon upp vid kast med tärning

Bilden av fördelningen illustrerar ett antal av de *egenskaper* som slumpvariabeln ifråga har - möjliga utfall, utfallens utbredning, symmetrin kring värdet 3,5, alla värden lika sannolika mm.

"På statistiska" säger man att en slumpvariabel *si* och så har (eller kan antas ha) den eller den fördelningen. Därmed ger man en karaktäristik av variabeln med avseende på en mängd av dess egenskaper.

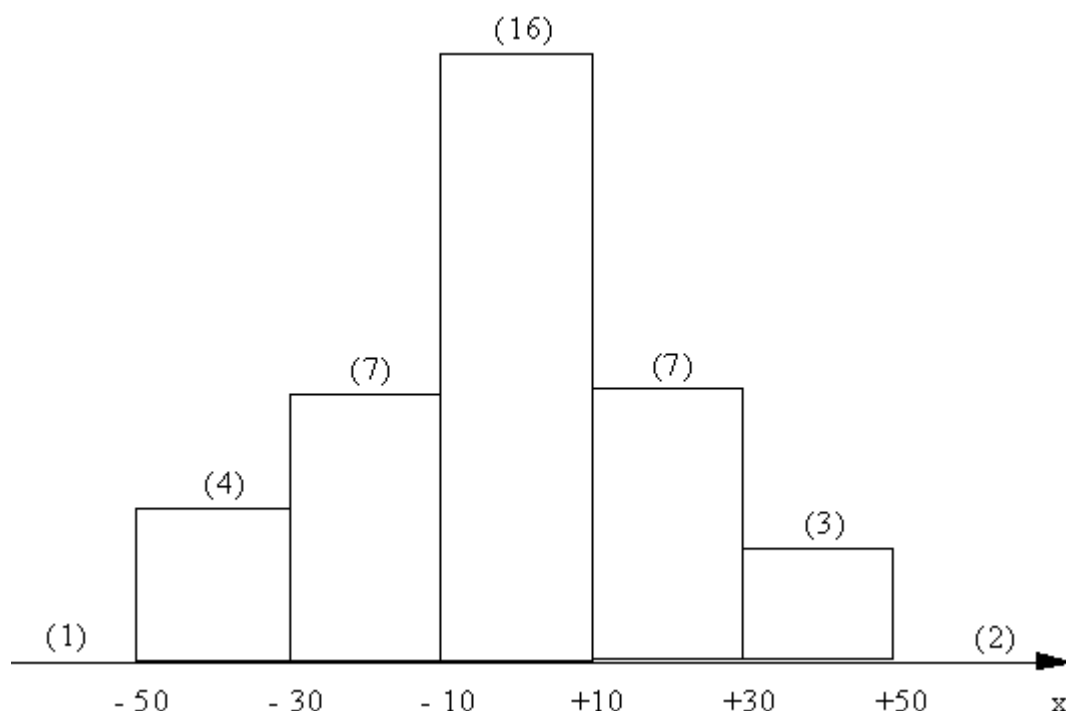
Låt oss ta ett annat exempel: Vi har gjort 40 mätningar av en lagerytenivå för att kontrollera eventuella avvikelser från riktvärdet  $[x_i]$  och sammanfattar observationerna i följande tabell:

Avvikelse ( $x_i$ ), mm	Antal observationer ( $f_i$ )
$x_i < - 50$	1
$- 50 \leq x_i < - 30$	4
$- 30 \leq x_i < - 10$	7
$- 10 \leq x_i < + 10$	16
$+ 10 \leq x_i < + 30$	7
$+ 30 \leq x_i \leq + 50$	3
$x_i > + 50$	2

Tabell B2.2-4 40 mätningar av lagerytenivåns avvikelse från riktvärde

Ett stapeldiagram (histogram, dvs celldiagram, där stapelytorna - inte höjderna! - är proportionella mot frekvenserna) får följande utseende:

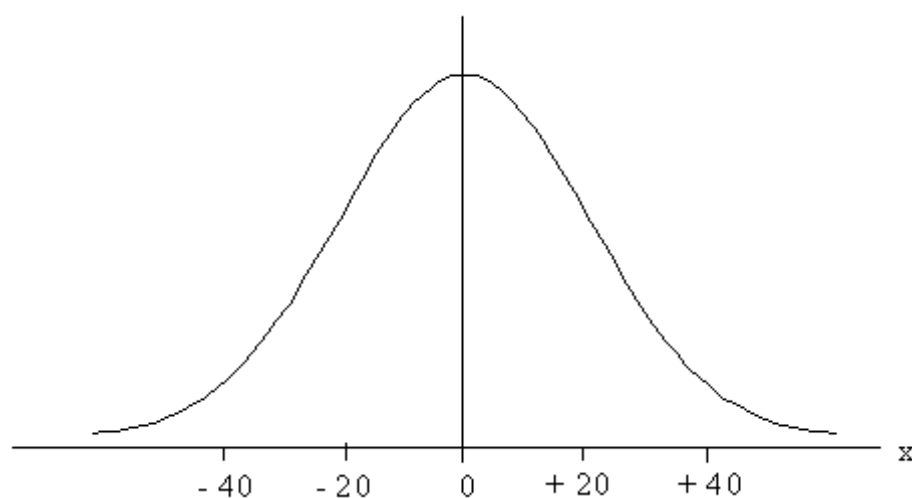




Figur B2.2-5 40 mätningar av lagerytenivåns avvikelser från riktvärde

Tabellen såväl som figuren ovan visar just "det sätt på vilket variabelvärdena uppträder" i vårt *stickprov* av 40 värden avseende slumpvariabeln  $X =$  "Mätning av lagerytenivåns avvikelse från riktvärde". Tabellen och figuren visar *fördelningen i ett stickprov*. Vi ser, t ex, att värdena i mätserien *centreras* kring noll (vilket förefaller vara det ungefärliga medelvärdet), att de ligger någorlunda väl *samlade* kring noll och att de ligger ganska *symmetriskt* kring noll (dvs avvikelser uppåt respektive nedåt är ungefär lika vanliga).

Om vi gör tankeexperimentet att vi fortsätter mäta samma lagerytenivå i all oändlighet och representerar våra mätningar i ett histogram där vi successivt låter klassbredderna bli mindre och mindre (i all oändlighet) skulle vi - om byggaren gjort sitt jobb enligt överenskommelserna - få en figur av följande typ:



*Figur B2.2-6 Ett oändligt antal mätningar av lagerytenivåns avvikelser från riktvärde representerat i histogram med oändligt små klassbredder (så små att gränserna mellan klasser inte kan markeras i figuren)*

Figuren visar (på ett empiriskt sätt) hur *fördelningen* av "hela verkligheten", dvs "alla tänkbara mätningar av lagerytenivåns avvikelser från riktvärde" ser ut. Om byggaren gjort rätt ska avvikelserna centrera kring noll och ligga symmetriskt väl samlade kring noll. *Ytan* under "kurvan" representerar "vanligheten" (frekvensen) av avvikelser av visst slag. Vi ser t ex, att hälften av alla avvikelser är negativa, hälften positiva. Vi ser också, att betydligt mer än hälften av avvikelserna ligger inom intervallet  $-20 \leq x_i \leq +20$ .

Många känner igen "kurvans" form i figuren ovan. Den brukar kallas "normalfördelningskurva" eller "Gauss-kurva" eller ibland "klockkurva". En slumpvariabel som antar sina värden på ett sätt som kan beskrivas med denna kan kallas "normalfördelad slumpvariabel".

Normalfördelningen är ett mycket viktigt "verktyg" inom statistiken. Detta beror på, att många slumpvariabler låter sig hanteras *som om* de vore normalfördelade, dvs deras utfallssannolikheter kan med god approximation beräknas med hjälp av normalfördelningen. Eftersom normalfördelningens matematiska funktion är känd kan (t ex) ytor under kurvan (som representerar sannolikheter) beräknas genom integrering. Se vidare nedan.

Man kan matematiskt visa, att slumpvariabler som i sig är *summor* av andra slumpvariabler tenderar att bli alltmer normalfördelade, ju *fler* de ingående slumpvariablerna är (den s k centrala gränsvärdessatsen). Det är vanligt, att mätvärden (av t ex lagerytenivå) kan betraktas som värden på en någorlunda normalfördelad slumpvariabel. De olika orsakerna till att mätvärdena faktiskt varierar är ju oftast både många och var för sig slumpvariabler (avläsningsfel, ojämn packning och avjämning, variationer i materialtäthet mm mm - alltså *ett antal* slumpvariabler vars *summa* blir till mätvärdet).

Ännu större möjligheter att använda normalfördelningen får man om man i stället för *enskilda* mätvärden använder *summan* av ett antal mätvärden. Om varje mätvärde kan uppfattas som en summa av slumpvariabler så omfattar givetvis *summan av mätvärdena* ännu flera slumpvariabler - och är ännu mera normalfördelad enligt centrala gränsvärdessatsen (under vissa förutsättningar om s k oberoende, som vi hoppar över här). Det är detta som är huvudskälet till att vi strävar efter att använda stickprovsmedelvärden vid statistisk acceptanskontroll. Ett stickprovsmedelvärde är ju just en summa av slumpvariabler (observation 1 + observation 2 + .... + observation n). Att summan sedan divideras med antalet observationer påverkar inte dess fördelningskaraktär, enbart placeringen längs tallinjen.

#### **B2.2.4 Population och stickprov**

Vad ett stickprov är har de flesta klart för sig, och vi har redan använt begreppet ovan. Stickprov innebär ju bara, att man tar ut "en del ur en större mängd". Däremot är det i statistiksammanhang viktigt att observera *hur* stickprovet tas ut, eftersom vi i regel vill använda mätningar av stickprovet för att dra slutsatser om den större

mängden. Vi vill, att stickprovet skall vara "representativt" för den större mängden och tillämpar därför urvalsförfaranden, som med beräkningsbar säkerhet ger representativa stickprov.

För "den större mängden", den helhet ur vilken vi drar stickprov och till vilken vi så småningom vill överföra våra slutsatser, används i statistiken benämningen *population*. Just för att veta, hur långt vi kan generalisera våra slutsatser baserade på stickprovsmätning, är det viktigt att ha synnerligen klart för sig vilken population stickprovet egentligen dragits ur.

Låt oss återigen exemplifiera med nivåmätning. Vi vill undersöka, om ett visst byggnadsobjekt - säg, underbyggnaden för 4 km ny riksväg - utförts enligt överenskommen specifikation. Vi vet, att den första kilometern byggs av X-man & söner medan sträckan 1-3 km utförs av Schakt & C:o och den resterande kilometern av Jord & Berg AB. Ingen beställare skulle komma på idén att godkänna samtliga entreprenörers arbete enbart på grundval av mätningar av tio aldrig så slumpmässigt valda punkter inom X-man & Söners arbetsområde. Och knappast skulle Schakt & C:os arbete godkännas på grundval av mätningar inom enbart de första 200 längdmeterna av deras vägavsnitt.

Det gäller givetvis att välja kontrollobjekt - populationer - med förstånd och insikt. Det kanske finns grund att tro, att varje entreprenör för sig presterar en någorlunda jämn kvalitet inom sitt arbetsområde medan däremot skillnaderna kan vara stora emellan beroende på erfarenhet, resurser etc. Det kanske finns grund att tro, att en viss entreprenörs arbete inom ett område av enhetlig geologisk karaktär håller jämn kvalitet - däremot att kvaliteten kan växla om kringförutsättningarna växlar (geologi, utförandesäsong etc).

I de fall som statistisk acceptansk kontroll används enligt [VÄG 94](#) anges alltid, hur s k *kontrollobjekt* skall väljas. Ofta anges storleken av en yta, inom vilket ett stickprov omfattande ett visst antal mätpunkter skall tas ut på visst sätt - och till vilken slutsatserna från stickprovsmätningarna kan dras. Population i sammanhanget är "alla tänkbara mätningar" av aktuellt slag som *kan* göras inom den yta som är kontrollobjekt - alltså en oändlig mängd.

För att åstadkomma representativa stickprov tillämpas olika former av slumpmässigt urval av mätpunkter inom kontrollobjektet. Några manuella förfaranden beskrivs i avsnitt 4 tidigare i denna metodbeskrivning och ytterligare (datorstödda) kommer med all säkerhet att utvecklas. Gemensamt för de olika metoderna är att de är utformade så att *varje punkt* inom kontrollobjektet har *samma* à-priori-sannolikhet att bli utvald till stickprovet (dvs *innan* urvalet av mätpunkter tekniskt påbörjas har alla punkter samma urvalssannolikhet). Därmed garanteras, att dragna stickprov blir "i genomsnitt representativa". Men därmed garanteras *inte*, att ett visst, faktiskt erhållet stickprov verkligen *är* representativt för kontrollobjektet - man kan ju ha "otur" i den ena eller andra riktningen.

Det är viktigt att observera, att det är *urvalsmetoden* som har en viss, beräkningsbar sannolikhet för att "träffa rätt" - om det *faktiskt erhållna* stickprovets relation till populationen vet man i efterhand *ingenting* - utom att det är åstadkommet med en metod, som ger representativitet - med en viss felrisk.

## B2.3 Fördelningars karaktäristika

### B2.3.1 Central tendens - "medelvärden"

Det man i regel i första hand vill veta om en fördelning är var den "i stort sett" ligger längs tallinjen - är fördelningens värden av storleksordningen 1, 100 eller 10 000? Mått som anger detta brukar med ett gemensamt namn kallas mått på central tendens. Det mest välkända torde vara det "vanliga" medelvärdet, något mer precist benämnt det *aritmetiska medelvärdet*. På engelska heter medelvärde "mean value", ofta bara "mean".

Som nämnts i avsnitt 3 betecknas (här) medelvärdet i fördelningen av ett stickprovs mätvärden med  $\bar{x}$ , i regel uttalat "x-streck" på svenska (och "x-bar" på engelska). Ibland används beteckningen  $M$  liktydigt med  $\bar{x}$ . Och, som nästan alla vet, definieras det aritmetiska medelvärdet (för ett stickprov) som summan av observationsvärdena dividerad med antalet värden.

Om stickprovet innehåller flera ( $f_i$  st) observationer med samma variabelvärde  $x_i$  kan det vara praktiskt att beräkna medelvärdet som

$$\bar{x} = 1/n \cdot \sum f_i \cdot x_i = \sum f_i/n \cdot x_i$$

I det sistnämnda sättet att skriva medelvärdesberäkningsformeln anger den s k relativa frekvensen  $f_i/n$  "vanligheten" av det aktuella  $x_i$ -värdet som en andel  $0 \leq f_i/n \leq 1$ . Detta ansluter till ett sätt att definiera medelvärdet "teoretiskt":

Om slumpvariabeln  $X$  antar sina värden  $x_i$  med sannolikheterna  $p_i$  är det aritmetiska medelvärdet för slumpvariabeln:

$$\mu_X = \sum p_i \cdot x_i$$

där den grekiska bokstaven  $\mu$  (uttalas "my" på svenska och "mjo" på engelska) är den generellt använda beteckningen för medelvärdet i en population.

Man kan visa, att ett stickprovsmedelvärde  $\bar{x}$  är en s k *vänsterriktig skattning* av populationsmedelvärdet  $\mu_X$  - om stickprovet dragits slumpmässigt ur populationen. Detta betyder, att ett stickprovsmedelvärde  $\bar{x}$  "i genomsnitt" ger en korrekt uppfattning om värdet av  $\mu_X$ . Se vidare avsnitt B2.4.

Det finns också *andra* medelvärden än det aritmetiska: Det *geometriska* medelvärdet (ibland kallat medelproportionalen) av  $n$  st värden definieras som  $n$ :te roten ur värdenas produkt. Geometriska medelvärdet av  $a$  och  $b$  är alltså kvadratroten ur  $a \cdot b$ , geometriska medelvärdet av  $a$ ,  $b$  och  $c$  tredjeteroten ur  $a \cdot b \cdot c$ , etc. Det *harmoniska* medelvärdet bygger på inverterade variabelvärden. Ytterligare speciella medelvärden utnyttjas ibland, men är inte av intresse för statistisk acceptansk kontroll. Emellertid bör man observera, att ordet "medelvärde" inte är entydigt - om det finns risk för missförstånd bör den längre formen utnyttjas.

Utöver medelvärden finns ytterligare mått på central tendens, som ibland används för att beskriva observationsmaterial (stickprov):

*Medianvärdet* i en fördelning är "det mittersta värdet", det värde, som har lika många observationer "till vänster" (dvs på lägre värden) som "till höger" (dvs på högre värden). Medianen ger ibland en "sannare" bild av fördelningens centrala tendens än det aritmetiska medelvärdet, särskilt när fördelningen är osymmetrisk (sned). I t ex en inkomstfördelning brukar median användas som mått på central tendens - de relativt få som har *mycket* höga inkomster "drar upp" medelvärdet så att det inte uppfattas som ett mått på den "naturliga" centralpunkten i inkomstfördelningen.

Median heter "median value" eller bara "median" på engelska.

*Typvärdet* i en fördelning är "det vanligaste värdet", det värde, som har flest observationer. Om man t ex undersökt åldern hos tio mopedförare och fått resultatet:

14, 15, 15, 16, 16, 15, 15, 54, 13, 17

är typvärdet 15 eftersom det förekommer 4 gånger medan 16 resp 14 förekommer 2 gånger och övriga värden en gång vardera.

Typvärde heter på engelska "mode" eller "modal value".

### **B2.3.2 Variabilitet - "spridning"**

Variabiliteten hos en fördelning är i regel nästan lika intressant som den centrala tendensen. Ordet "spridning" används på svenska som en allmän benämning på fenomenet att värdena i en fördelning ligger "utbredda" över tallinjen. "Stor spridning" innebär givetvis, att värdena fördelar sig över ett ganska stort intervall, "liten spridning" motsatsen.

Det finns också personer, som med "spridning" menar det speciella spridnings*måttet* standardavvikelse. Detta språkbruk bör undvikas, eftersom det är otydligt och kan orsaka missförstånd.

Spridning heter på engelska "dispersion".

Precis som för central tendens finns ett antal olika mått som används för att mäta spridning. Det allra enklaste är *variationsvidden*, dvs distansen mellan den lägsta och största observationen. I exemplet med mopedförarna ovan är variationsvidden  $54 - 14 = 40$ . Härav framgår också, att variationsvidden är ett ganska "oskarpt" spridningsmått. Det är ju bara en enda observation av de 10 som avsevärt skiljer sig från de andra, vilka ligger tätt samlade i intervallet 14 - 17.

Variationsvidd heter på engelska "range".

Ett annat spridningsmått är den s k kvartilavvikelsen. Kvartilerna är "släktingar" till medianen. Den första kvartilen i en fördelning har 25 % av observationerna "till vänster" om sig (dvs på lägre värden), den andra 50 % och den tredje 75 %. (Den andra kvartilen är alltså samma värde som medianen.) Kvartilavvikelsen definieras som halva det s k kvartilavståndet, som i sin tur är avståndet mellan den första och den tredje kvartilen.

Kvartilavvikelse heter "quartile deviation" på engelska. Kvartilavstånd heter "quartile distance", ibland "quartile difference".

Variationsvidd och kvartilavvikelse är spridningsmått som i regel används för snabba uppskattningar av variabiliteten i ett observationsmaterial, när man inte har annat än papper och penna till hands. För att noggrannare beskriva variabiliteten beräknar man oftast standardavvikelsen, som definieras som (positiva) roten ur den s k variansen, vilken i sin tur är aritmetiska medelvärdet av observationsvärdenas kvadrerade avvikelser från sitt medelvärde. För ett stickprovsmaterial gäller alltså:

$$s = \sqrt{s^2}, \text{ där:}$$

$$s^2 = [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] / n ; \text{ dvs}$$

$$s^2 = 1/n \cdot \sum (x_i - \bar{x})^2$$

(i är ett index som genomlöper alla värden 1, 2, 3 .... n)

För "handräkning" kan det vara praktiskt att utnyttja följande:

$$\begin{aligned} s^2 &= 1/n \cdot \sum (x_i - \bar{x})^2 = 1/n \cdot \sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \\ &= 1/n \cdot \{\sum x_i^2 - 2\bar{x} \cdot \sum x_i + n \cdot \bar{x}^2\} = 1/n \cdot \{\sum x_i^2 - 2\bar{x} \cdot n \cdot \bar{x} + n \cdot \bar{x}^2\} = \\ &= 1/n \cdot \{\sum x_i^2 - n \cdot \bar{x}^2\} = 1/n \cdot \sum x_i^2 - \bar{x}^2 \end{aligned}$$

Variansen kan alltså uttryckas som medelvärdet av kvadraterna på observationerna minskat med kvadraten av observationernas aritmetiska medelvärde. (Jfr dock alternativ definition nedan).

Om stickprovet innehåller flera ( $f_i$  st) observationer med samma variabelvärde  $x_i$  kan det vara praktiskt att beräkna variansen som

$$s^2 = 1/n \cdot \sum f_i \cdot (x_i - \bar{x})^2 = \sum f_i / n \cdot (x_i - \bar{x})^2$$

I det sistnämnda sättet att skriva variansberäkningsformeln anger den s k relativa frekvensen  $f_i/n$  "vanligheten" av det aktuella  $x_i$ -värdet som en andel  $0 \leq f_i/n \leq 1$ . Detta ansluter till ett sätt att definiera variansen "teoretiskt":

Om slumpvariabeln  $X$  antar sina värden  $x_i$  med sannolikheterna  $p_i$  är variansen för slumpvariabeln:

$$\sigma^2_X = \sum p_i \cdot (x_i - \mu_X)^2$$

Standardavvikelsen är roten ur variansen:

$$\sigma_X = \sqrt{\sigma^2_X}$$

där den grekiska bokstaven  $\sigma$  (som uttalas "sigma" på såväl svenska som engelska) generellt betecknar standardavvikelse i population. På motsvarande sätt betecknar  $\sigma^2$  populationsvariens.

Variansen är alltså egentligen ett speciellt aritmetiskt medelvärde (av kvadrerade avvikelser) och standardavvikelsen roten härur. Kvadreringen innebär bl a, att variansens och standardavvikelsens storlek påverkas särskilt kraftigt av värden som ligger "långt ut" i fördelningen jämfört med hur sådana värden påverkar t ex kvartilavvikelsen (jfr aritmetiskt medelvärde och median). Det finns situationer, där standardavvikelsen kan uppfattas som ett "överdrivet" spridningsmått, på motsvarande sätt som att det aritmetiska medelvärdet inte alltid mäter vad som uppfattas som en "den korrekta" centrala tendensen hos en sned fördelning.

Man kan visa, att stickprovsvariansen  $s^2$  definierad på ovanstående sätt *inte* är en helt vänteriktig skattning av populationsvariansen  $\sigma_{x_2}$  - den hamnar i genomsnitt lite "snett", nämligen vid  $\sigma_{x_2} \cdot (n-1)/n$ . Om man sålunda "justerar"  $s^2$  genom att multiplicera med faktorn  $n/(n-1)$  får man en stickprovskaraktäristika som är en vänteriktig skattning av populationsvariansen,  $\sigma_{x_2}$ . Se vidare avsnitt B2.4.

Eftersom skälet till att beräkna stickprovsvariansen  $s^2$  mycket ofta är att man vill utnyttja den för att uppskatta populationsvariansen brukar man av praktiska skäl "justera  $s^2$  redan från början", dvs *definiera*  $s^2$  så att  $s^2$  i sig själv är en vänteriktig skattning för populationsvariansen  $\sigma_{x_2}$ . Alltså:

$\sigma_{x_2}$  skattas med stickprovsvariansen  $s^2$  som beräknas enligt:

$$s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$$

$$s = \sqrt{s^2}$$

Det är stickprovsstandardavvikelsen  $s$  beräknad på detta sätt (dvs med  $s^2$  definierad som  $s$  k vänteriktig estimator för  $\sigma_{x_2}$ ) som avses i de olika krav som anges i VÄG 94. Även med denna definition kan man härleda en formelvariant som ger lite enklare hantering vid "handräkning":

$$\begin{aligned} s^2 &= 1/(n-1) \cdot \sum (x_i - \bar{x})^2 = 1/(n-1) \cdot \sum (x_i^2 - 2\bar{x} x_i + \bar{x}^2) = \\ &= 1/(n-1) \cdot \{\sum x_i^2 - 2\bar{x} \cdot \sum x_i + n \cdot \bar{x}^2\} = \\ &= 1/(n-1) \cdot \{\sum x_i^2 - 2\bar{x} \cdot n \cdot \bar{x} + n \cdot \bar{x}^2\} = \\ &= 1/(n-1) \cdot \{\sum x_i^2 - n \cdot \bar{x}^2\} = 1/(n-1) \cdot \sum x_i^2 - n/(n-1) \cdot \bar{x}^2 \end{aligned}$$

### B2.3.3 Några fördelningstyper

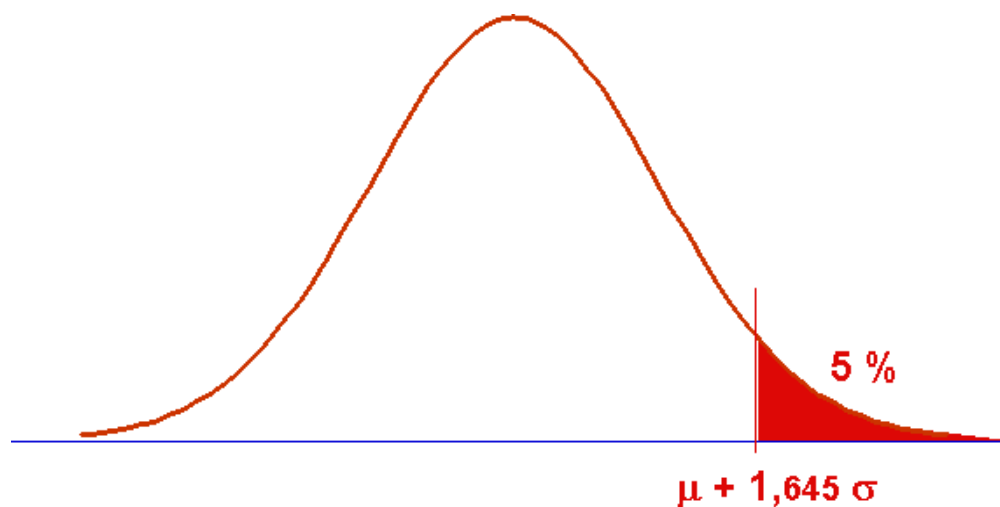
Vi har konstaterat, att det finns kontinuerliga och diskreta slumpvariabler, samt att fördelningen anger slumpvariabelns "sätt att uppträda". Fördelningen kan beskrivas med avseende på (bl a) central tendens och variabilitet. Fördelningens frekvensfunktion och/eller fördelningsfunktion kan (om den är känd) användas för att beräkna sannolikheter för olika utfall hos slumpvariabeln.

Statistisk inferens - dvs uttalanden om "hur verkligheten är" baserade på stickprovsmässiga mätningar - kan inte göras med någon noggrannhet utan kunskap om den fördelning som den i sammanhanget aktuella slumpvariabeln (t ex

stickprovsmedelvärdet  $\bar{x}$  eller stickprovsandelen "godkända prov",  $f/n$ ) har eller kan anses ha. Först när man kan klara ut vilken s k *referensfördelning* som kan vara tillämplig går det att göra inferens, som då utnyttjar sannolikhetsberäkningar baserade på (den matematiskt hanterbara) referensfördelningen.

### B2.3.3.1 Normalfördelning

För kontinuerliga slumpvariabler är den s k normalfördelningen den utan jämförelse mest utnyttjade referensfördelningen. Dess matematiska funktion skall vi inte ta upp här; vi nöjer oss med att konstatera att den finns och medger att man kan beräkna utfallssannolikheter t ex enligt följande:

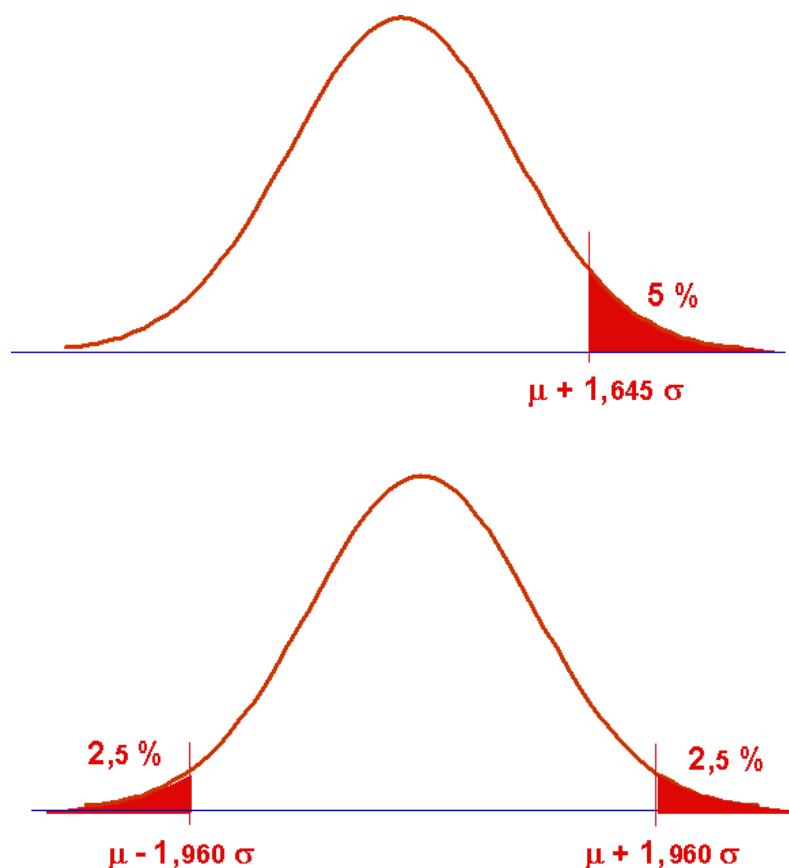


Figur B2.3-1 Sannolikhetsberäkning med hjälp av normalfördelningstabell

Ur tabeller kan man avläsa de gränser under respektive över vilka en viss andel av fördelningen ligger. Så kan man t ex finna, att i normalfördelningar ligger alltid 5 % av fördelningen (sannolikhetsmassan, "populationsvärdena") ovanför gränsen  $\mu + 1,645 \sigma$ . Detta innebär således, att sannolikheten för att vid slumpmässig dragning av ett värde ur en normalfördelad population få ett värde som överstiger medelvärdet med mer än 1,645 standardavvikelseenheter är 5 %.

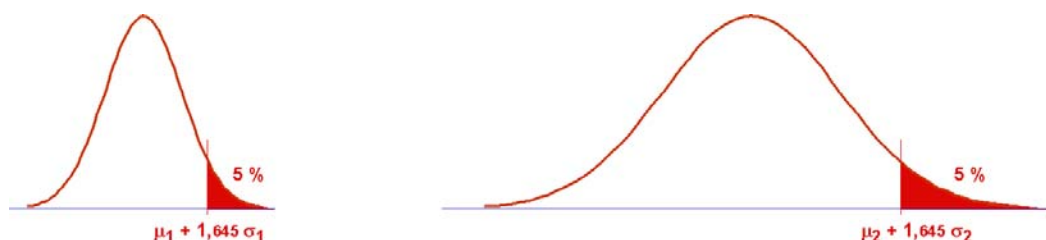
Normalfördelningen är symmetrisk kring sitt medelvärde. Man kan på motsvarande sätt som ovan finna, att sannolikheten att få ett värde som ligger "längre från medelvärdet" än 1,96 standardavvikelseenheter är 5 %. Se figur B2.3-2.





Figur B2.3-2 Enkel- och dubbelsidiga intervall i normalfördelning

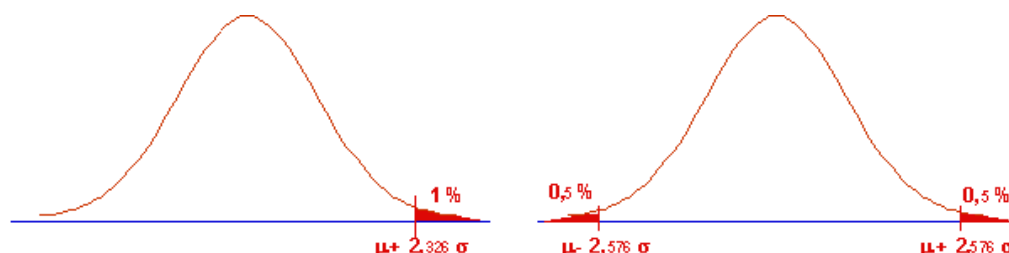
De tabeller som utnyttjas för att bestämma sannolikheter av det här slaget avser så gott som alltid den normalfördelning vars medelvärde ( $\mu$ ) är 0 och vars standardavvikelse ( $\sigma$ ) är 1. En fördelning med dessa värden på  $\mu$  och  $\sigma$  brukar kallas *normerad*. Tabeller för den normerade normalfördelningen kan användas för att beräkna gränser i vilken normalfördelning som helst - eftersom alla normalfördelningar är "likformiga". I den normerade normalfördelningen ligger 5 % av sannolikhetsmassan ovanför gränsen 1,645 - i vilken annan normalfördelning som helst ligger samma andel (5 %) ovanför gränsen  $\mu + 1,645 \sigma$ . Se figur B2.3-3.



Figur B2.3-3 Normalfördelningar med olika  $\mu$  och  $\sigma$

Om man ur en normalfördelning med, säg,  $\mu = 200$  och  $\sigma = 20$  har fått ett värde 250 kan man följaktligen räkna ut, huruvida det erhållna utfallet är "vanligt" eller "ovanligt" genom att jämföra med den "likformiga" normerade normalfördelningen: Man tar reda på, hur många

standardavvikelseenheter från medelvärdet som utfallet 250 ligger och slår upp sannolikheten för ett så extremt värde (eller extremare) i normalfördelningstabellen: Utfallet överstiger medelvärdet med 50, vilket motsvarar 2,5 standardavvikelseenheter. Tabellen visar, att sannolikheten att vid slumpmässig dragning ur en normalfördelning få ett så stort positivt värde (eller ännu större) är mindre än en procent. Annorlunda uttryckt är sannolikheten att få ett värde som ligger så långt från medelvärdet (eller ännu längre ut i endera av fördelningens två "svansar") bara drygt en procent.



Figur B2.3-4 Enkelsidiga och dubbelsidiga utfallsområden

Detta är den grundprincip som kommer till användning t ex vid statistisk acceptansk kontroll. Man undersöker, huruvida det erhållna värdet på en kriterievariabel (t ex ett stickprovsmedelvärde  $\bar{x}$ ) är "rimligt" eller "orimligt" med hänsyn till antagandet, att den (normal-) fördelning som det teoretiskt kan antas komma från har ett visst medelvärde,  $\mu$  (= riktvärdet för den undersökta egenskapen). Om jämförelsen visar, att värdet ligger så långt från fördelningens förmodade medelvärde att utfallet är "osannolikt" drar man slutsatsen, att fördelningens medelvärde i själva verket är något annat, dvs avviker från riktvärdet.

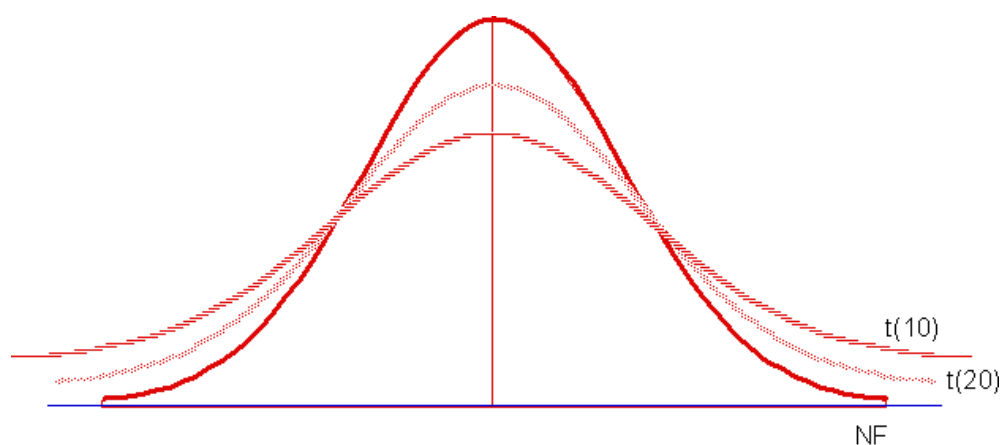
### B2.3.3.2 t-fördelning

Som framgår ovan är det nödvändigt att känna till en normalfördelningens medelvärde  $\mu$  och standardavvikelse  $\sigma$  för att kunna transformera en kriterivariabels värde så att detta kan jämföras med den normerade normalfördelningen. I praktiken tillhör detta undantagen -  $\mu$  och  $\sigma$  är i regel okända. Vad gäller  $\mu$  är detta dock inget problem - vi har i regel en hypotes om vad  $\mu$  borde vara (= riktvärdet) och använder oss därför av detta  $\mu$ -värde för transformationen. Och, som tidigare påpekats,  $\sigma$ -värdet kan generellt skattas med hjälp av den standardavvikelse ( $s$ ) som vi kan mäta upp i vårt stickprov.

Emellertid - och utan att gå in på den teoretiska bakgrunden - finns det vissa komplikationer med detta. Om inte vårt stickprov är rejält stort (åtminstone  $n = 30$  à  $50$ ) så kommer vår transformerade kriterivariabel inte att vara tillräckligt väl normalfördelad för att vi skall kunna använda den normerade normalfördelningen för sannolikhetsberäkning. I vissa fall

finns dock möjligheter att använda en annan referensfördelning, den så kallade t-fördelningen (ibland kallad "Student's t" efter sin anonyme upptäckare). t-fördelningen är strängt taget en "familj" av referensfördelningar, där det exakta utseendet hos respektive t-fördelning bestäms av det s k *frihetsgradsantalet* (som vi här för enkelhets skull kan uppfatta som antalet observationer till stickprovet minus ett).

t-fördelningarna liknar normalfördelningen i många avseenden och utseendet närmar sig normalfördelningens med ökat antal frihetsgrader. Därför brukar man i praktiken sällan bry sig om att använda t-fördelning förrän kriterievariabeln - och variansskattningen - baseras på små stickprov ( $n \leq 30$ ).



Figur B2.3-5 Normalfördelning och t-fördelningar

Som synes är t-fördelningarna litet "plattare" än normalfördelningen och alltmer platta ju mindre antalet frihetsgrader är. En allt större andel av sannolikhetsmassan ligger på distans från medelvärdet ju mindre antalet frihetsgrader är - dvs osäkerheten tilltar när stickprovsstorleken minskar.

På samma sätt som för normalfördelningen finns tabeller över de normerade t-fördelningarnas frekvens- och fördelningsfunktioner. Dessa används på motsvarande sätt som normalfördelningstabellerna - det gäller bara att välja den tabell som avser rätt antal frihetsgrader.

Det skall också framhållas, att användning av t-fördelning som referensfördelning formellt ställer krav på att den population som kriterivariabeln antas komma från skall vara normalfördelad. Den praktiska slutsatsen härav för den som sysslar med statistisk acceptansk kontroll blir: Undvik att arbeta med små stickprov ur populationer med okänd standardavvikelse och fördelningsform!

### B2.3.3.3 Övriga kontinuerliga referensfördelningar

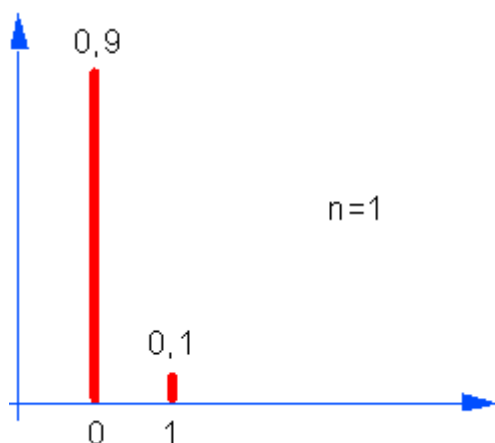
Förutom de ovan beskrivna - och oftast använda - kontinuerliga referensfördelningarna finns ett stort antal övriga, som används i olika

sammanhang inom statistisk inferens. Dessa är dock (f n) inte aktuella vid statistisk acceptansk kontroll enligt [VÄG 94](#). Den intresserade hänvisas till någon lämplig standardlärobok för statistikstudier på högskolenivå.

#### B2.3.3.4 Binomialfördelning

Låt oss betrakta en situation där en population uppfattas som att den består av två slags enheter - "godkända" respektive "ej godkända". Det kan t ex vara populationen av "alla tänkbara prov avseende beläggningstjocklek" som kan tas inom ett kontrollobjekt (dvs ett oändligt antal). Låt oss vidare anta, att den sanna andelen "ej godkända" är så hög som 10 % - beläggningsentreprenören har varit snål.

Den antagna populationen kan beskrivas enligt figur B2.3-6 - sannolikheten att vid slumpmässigt val av ett enda prov få ett som är "ej godkänt" är 10 %. Sannolikheten att vid slumpmässigt val av ett prov få 0 som är "ej godkänt" är 90 %.



Figur B2.3-6 Binomialfördelning - populationsfördelning

Om vi tar två prover kan vi få noll, ett eller två prov som är "ej godkända":

- För att vi skall få 0 "ej godkända" krävs förstås, att det första provet är godkänt ( $P=0,9$ ) och att även det andra är godkänt ( $P=0,9$ ) Eftersom vi tar proverna oberoende av varandra är den kombinerade sannolikheten, dvs sannolikheten att få 0 "ej godkända",  $0,9 \cdot 0,9 = 0,81$ .
- För att få 1 "ej godkänt" får vi det antingen som första prov, för vilket sannolikheten är  $0,1 \cdot 0,9 = 0,09$ ; eller som andra prov, för vilket sannolikheten är  $0,9 \cdot 0,1 = 0,09$ . Totalsannolikheten för 1 "ej godkänt" blir alltså 0,18.
- För att få båda proven "ej godkända" är sannolikheten  $0,1 \cdot 0,1 = 0,01$ .

Summan av sannolikheterna är  $0,81 + 0,18 + 0,01 = 1,00$ . Givetvis. Vi måste ju få antingen 0, 1 eller 2 "ej godkända".

Om vi ökar antalet uttagna prover till t ex 3 - hur är det då? Vad är sannolikheten att få 0, 1, 2 eller alla 3 proverna "ej godkända"? Vi kan naturligtvis räkna vidare på samma sätt som nyss. Sannolikheten för 0 "ej godkända" blir förstås  $0,9 \cdot 0,9 \cdot 0,9 \approx 0,73$  och sannolikheten för 3 "ej godkända"  $0,1 \cdot 0,1 \cdot 0,1 = 0,001 (\approx 0,00)$ . Däremot blir det lite jobbigare att beräkna sannolikheterna för 1 eller 2 "ej godkända" - vi kan ju få dessa antal på flera sätt, dvs i olika sekvenser.

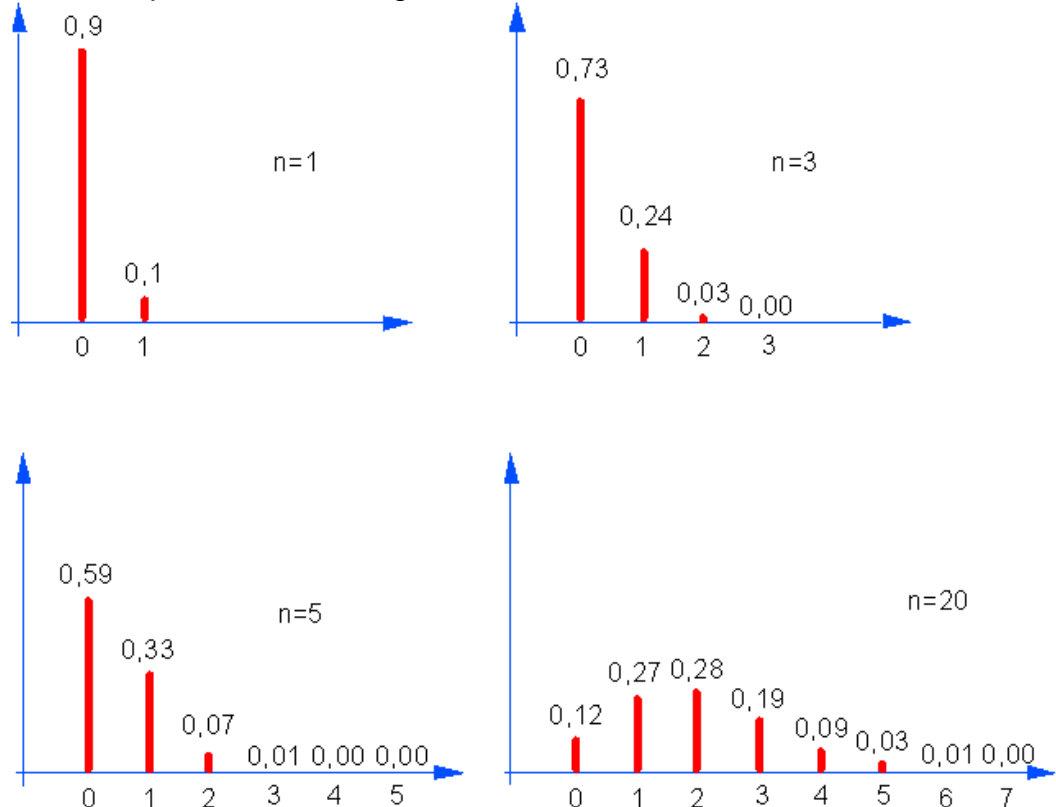
Vissa känner säkert igen situationen från gymnasiets matematik, närmare bestämt kombinatoriken. "På hur många olika sätt kan man få  $f$  st "ej godkänd" bland  $n$  uttagna"? Svaret är: "På  $n$  över  $f$  sätt".

Här har vi att göra med den s k binomialfördelningen, vars frekvensfunktion är:

$$P(X = f) = \binom{n}{f} \cdot P^f \cdot (1 - P)^{(n-f)}$$

där  $n$  = totala antalet uttagna (dvs stickprovsstorleken),  $f$  = antalet (frekvensen) i stickprovet av speciellt slag (t ex "ej godkänd") och  $P$  = sannolikheten att vid slumpmässigt val av en enhet ur populationen få en enhet av detta speciella slag (dvs = populationsandelen "ej godkända").

Utfallssannolikheterna för 0, 1 2 etc "ej godkända" illustreras för några olika stickprovsstorlekar i figur B2.3-7:



*Figur B2.3-7 Binomialfördelningar för olika stickprovsstorlekar*

Man kan visa, att binomialfördelningens medelvärde  $\mu$  motsvaras av  $n \cdot P$  och dess varians  $\sigma^2$  av  $n \cdot P \cdot (1-P)$ .

Som kan anas av de olika fördelningarna i figuren ovan blir även binomialfördelningen "mer och mer normalfördelad" när stickprovsstorleken växer, även om den för små värden på  $n$  och värden på  $P$  nära 0 eller 1 är extremt sned. Även här fungerar nämligen centrala gränsvärdessatsen och för stora värden på  $n$  (och särskilt om  $P$ -värdet ligger kring 0,5) kan även binomialfördelningen med gott resultat hanteras som om den vore en normalfördelning. Detta innebär, att ett erhållet värde på kriterievariabeln  $f$  (antal ej godkända) kan jämföras med den normerade normalfördelningen efter det att det transformerats enligt

$$\frac{f - nP}{\sqrt{nP(1-P)}} \quad (\text{jfr avsnitt B2.5})$$

För måttliga värden på  $n$  i förening med  $P$ -värden i närheten av 0 eller 1 bör dock binomialfördelningen inte approximeras med NF. Det finns standardtabeller för binomialfördelningen liksom för normal- och  $t$ -fördelningarna som kan hjälpa till att avgöra, vad som är ett "extremt" utfall. Vi ser t ex i figur B2.3-7 (nederst till höger) att sannolikheten för att få fem eller flera "ej godkända" när man undersöker 20 prover är så låg som  $0,03 + 0,01 = 4\%$  om populationsandelen defekta är 10%. Skulle man vid ett försök få ett utfall fem eller fler defekta talar detta för att man snarare bör misstänka, att populationsandelen defekta överstiger  $P = 0,10$  än att man haft "väldig otur".

### **B2.3.3.5 Övriga diskreta referensfördelningar**

De situationer där binomialfördelningen är direkt tillämplig kännetecknas av "dragning med återläggning", dvs sannolikheten  $P$  för ett visst utfall (t ex "defekt prov") är konstant under hela urvalsproceduren. Det finns fall där detta inte gäller - dvs utfallssannolikheten förändras successivt, beroende på vilka enheter som råkar väljas till stickprovet. Situationen brukar kallas "dragning utan återläggning" och ofta exemplifieras med bridgegivar. I sådana fall kan den sk hypergeometriska fördelningen utnyttjas som referensfördelning. Denna synes dock inte ha någon tillämpning inom den statistiska acceptansk kontroll som aktualiseras av [VÄG 94](#).

I fall där utfallssannolikheten  $P$  (för t ex "defekt prov") är mycket liten och antalet uttagna prov ( $n$ ) är stort kan den sk Poisson-fördelningen vara aktuell som alternativ till binomialfördelningen (och dessutom enklare att utnyttja rent matematiskt). Inte heller denna förefaller dock tillämplig inom ramen för statistisk acceptansk kontroll enligt [VÄG 94](#).

Ett antal ytterligare diskreta referensfördelningar finns tillgängliga. Den intresserade hänvisas till någon lämplig standardlärobok för statistikstudier på högskolenivå.

## **B2.4 Skattning**

### **B2.4.1 Skattning = estimation**

Skattning innebär, att man med utgångspunkt i en stickprovskaraktäristika gör ett uttalande om förhållanden i den population som stickprovet dragits från. Så kan man t ex använda ett stickprovsmedelvärde  $\bar{x}$  som skattning för populationsmedelvärdet  $\mu$  och använda stickprovsvariansen  $s^2$  som skattning för populationsvariansen  $\sigma^2$ .

Ett annat ord för skattning är estimation, som visserligen är mindre svenskt men har fördelen att man tydligare kan skilja mellan *estimation* - att skatta; *estimat* - den erhållna skattningen, dvs själva värdet och *estimator* - den slumpvariabel som används vid estimationen (t ex  $\bar{X}$  för  $\mu$ ). Man utför alltså en estimation genom att utnyttja en estimator, beräkna dess värde för det erhållna stickprovet och därigenom få ett estimat.

### **B2.4.2 Punkt- och intervallskattning**

Estimation kan göras som punktestimation (punktskattning) eller intervallestimation (intervallskattning). Punktestimation innebär att man beräknar värdet hos en lämplig estimator och får ett värde - ett punktestimat för den populationsstorhet man vill estimeras. Intervallestimation innebär att man beräknar ett punktestimat och därutöver utnyttjar estimatorns varians och fördelningstyp för att lägga ett intervall i anslutning till punktestimatet, inom vilket populationsstorheten anges ligga.

Låt säga, att vi vid en nivåkontroll gjort 49 observationer av avvikelserna från riktvärdet och fått ett stickprovsmedelvärde  $\bar{x} = 2$  mm. Stickprovsstandardavvikelsen har beräknats till 7 (mm). Vi kan på detta underlag göra:

- punktestimat för faktisk avvikelse från riktvärdet,  $\mu$ : " $\mu$  skattas till 2 mm"
- intervallestimat för faktisk avvikelse från riktvärdet,  $\mu$ : " $\mu$  ligger med 95 % konfidens inom intervallet  $2 \pm 1,96$  (mm), dvs inom intervallet 0,04 - 3,96 mm".

Hur vi i det senare fallet bestämmer intervallets bredd skall vi strax återkomma till, liksom till begreppet "konfidens".

### **B2.4.3 Vänteriktighet**

I samband med estimatorer talar man om "vänteriktighet" eller på engelska "unbiasedness". En estimator (stickprovskaraktäristika) är vänteriktig vis-à-vis en

populationsstorhet om den har populationsstorheten som sitt teoretiska medelvärde. Mer formellt brukar detta uttryckas så att en estimator (som är en slumpvariabel) är vänteriktig för en viss populationsstorhet om dess "matematiska förväntan" är lika med populationsstorheten. Så är t ex stickprovsmedelvärdet  $\bar{X}$  vänteriktig estimator för  $\mu$  eftersom medelvärdet av alla tänkbara  $\bar{x}$  man skulle kunna få om man drog stickprov i all oändlighet är just  $\mu$ . Som vi sett i tidigare avsnitt är däremot inte stickprovsvariansen utan vidare vänteriktig estimator för populationsvariansen  $\sigma^2$  - vi har justerat dess definition för att eliminera den "felvisning" (bias) som den skulle ge om vi dividerade  $\sum(x-\bar{x})^2$  med  $n$  i stället för med  $(n-1)$ .

#### B2.4.4 Effektivitet (efficiens)

En estimators effektivitet är beroende av dess varians (standardavvikelse). Om det finns flera olika vänteriktiga estimatorer för en viss populationsstorhet bör den väljas som har den minsta variansen (standardavvikelsen).

Betrakta t ex situationen att vi vill estimeras ett populationsmedelvärde,  $\mu$ . Vi kan göra det på grundval av en enda observation, "ett enda x-värde" ur populationen. Den estimator vi då använder (slumpvariabeln  $X$ ) har variansen  $\sigma^2$ .

Vi kan också ta ett stickprov om flera ( $n$ ) enheter ur populationen, beräkna stickprovsmedelvärdet  $\bar{x}$  och använda detta som skattning för  $\mu$ . Vilken varians (standardavvikelse) har egentligen estimatoren  $\bar{X}$ ?

Man kan visa, att om en slumpvariabel  $Y$  är en summa av ett antal ( $s$  k oberoende) slumpvariabler  $X$ :

$$Y = X + X + X + \dots + X$$

så är variansen för slumpvariabeln  $Y =$  summan av varianserna för de ingående slumpvariablerna  $X$ :

$$\sigma^2_Y = \sigma^2_x + \sigma^2_x + \sigma^2_x + \dots + \sigma^2_x$$

Man kan också visa, att om en slumpvariabel  $Y$  är en linjär funktion av en annan slumpvariabel  $X$ , t ex:

$$Y = C \cdot X + D \text{ (där } C \text{ och } D \text{ är konstanter)}$$

så är variansen för  $Y$

$$\sigma^2_Y = C^2 \cdot \sigma^2_x$$

Låt oss nu betrakta estimatoren  $\bar{X}$ : Den kan uppfattas som summan av  $n$  st slumpvariabler  $X_i$  ( första observation + andra observation + ... +  $n$ :te observation) dividerad med antalet observationer,  $n$ :

$$\bar{X} = 1/n \cdot \sum X_i = 1/n \cdot x_1 + 1/n \cdot X_2 + \dots + 1/n \cdot X_n$$

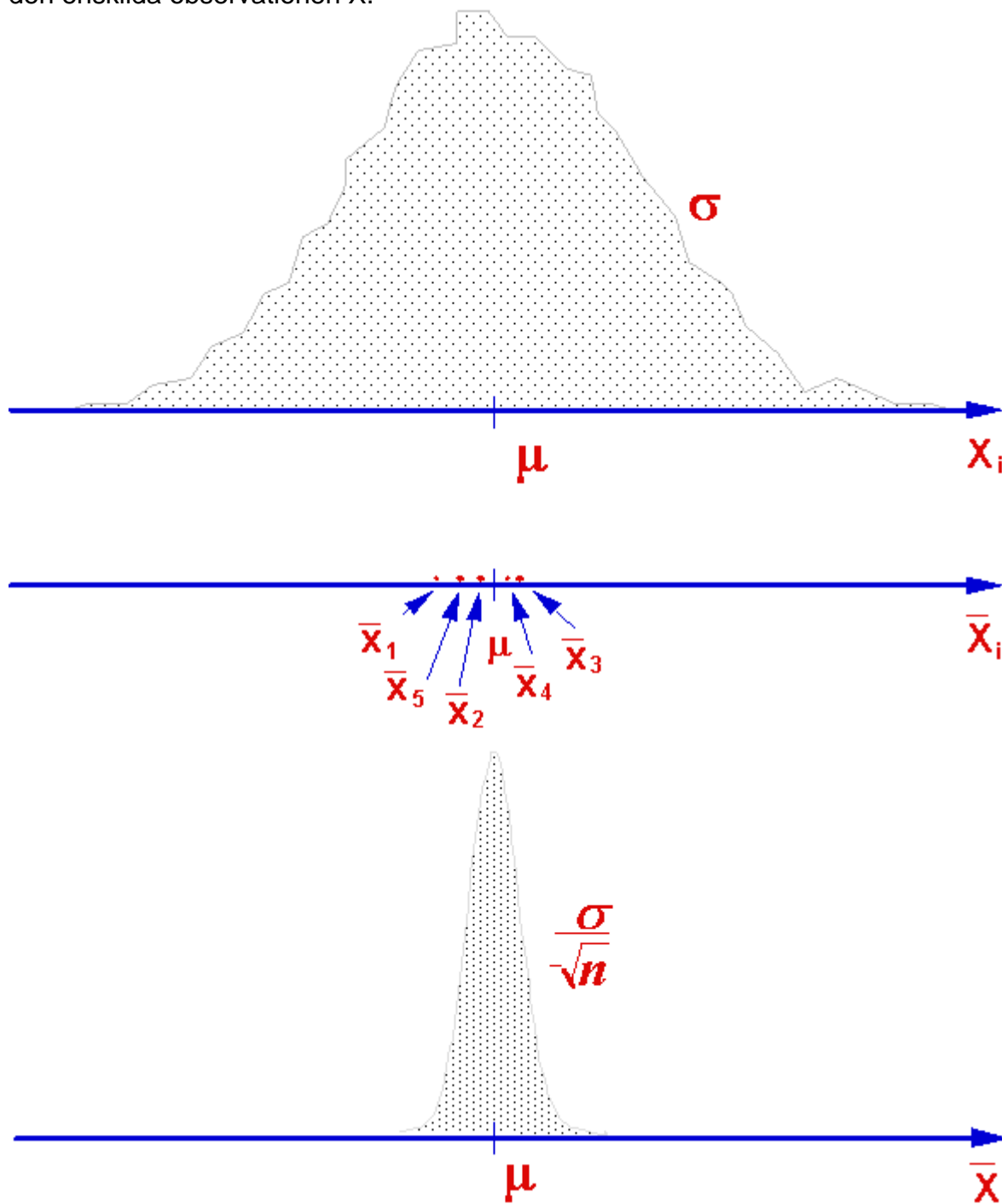


Enligt de refererade räknereglerna för varians är då:

$$\sigma_{\bar{X}}^2 = \left(\frac{1}{n}\right)^2 \cdot \sigma_X^2 + \left(\frac{1}{n}\right)^2 \cdot \sigma_X^2 + \dots + \left(\frac{1}{n}\right)^2 \cdot \sigma_X^2, \text{ dvs}$$

$$\sigma_{\bar{X}}^2 = n \cdot \left(\frac{1}{n}\right)^2 \cdot \sigma_X^2 = \frac{\sigma_X^2}{n} = \frac{\sigma^2}{n}$$

Detta innebär således, att estimatorn  $\bar{X}$  har 1/n gånger mindre varians än estimatorn X. Stickprovsmedelvärdet  $\bar{X}$  är alltså en mycket mer effektiv (efficient) estimator än den enskilda observationen X.



Figur B2.4-1 Populationsfördelning och samplingfördelning

### B2.4.5 Samplingfördelning

"Sampling" är det engelska ordet för "stickprovsdragning" och ordet används också i svenskt fackspråk. Med "samplingfördelning" (sampling distribution på engelska) menas generellt en teoretisk fördelning för en stickprovskaraktäristika - t ex  $\bar{x}$ , dvs någonting som beräknas på grundval av värdena i ett slumpmässigt draget stickprov (sample eller på svenska ofta sampel) och vars värde därmed varierar slumpmässigt från stickprov till stickprov om samplingförfarandet upprepas.

Vi har sett, att stickprovskaraktäristikor ( $\bar{x}$ ,  $s^2$ ) kan användas som estimatorer för motsvarande populationsstorheter ( $\mu$ ,  $\sigma^2$ ). "Samplingfördelning" är därför i de flesta fall liktydigt med "fördelningen av de estimat som kan erhållas med en estimator, när stickprovsdragningen (och estimatberäkningen) upprepas i all oändlighet".

Samplingfördelningar är alltid "mer normalfördelade" än de populationer ur vilka stickproven dras. Detta är en konsekvens av den tidigare nämnda centrala gränsvärdessatsen (CGS), som kort uttryckt anger, att summor av slumpvariabler blir mer normalfördelade ju fler termer (dvs oberoende slumpvariabler) som ingår i summan. Detta är ett av skälen till att vi vill utnyttja samplingfördelningar för inferens - vi kan lättare utnyttja den matematiskt välbestämda normalfördelningen som referensfördelning. Ett annat viktigt skäl är att samplingfördelningens variabilitet alltid är avsevärt mindre än variabiliteten i den population stickproven kommer ifrån (se föregående avsnitt) och minskar, ju mer stickprovsstorleken ökas. Sammanfattningsvis: Om vi drar någorlunda stora, slumpmässiga stickprov ur en kontinuerlig population och beräknar  $\bar{x}$  kan vi räkna med, att samplingfördelningen för  $\bar{x}$  är normalfördelad med  $\mu_{\bar{x}} = \mu_x$  och

$$\sigma^2_{\bar{x}} = 1/n \cdot \sigma^2_x$$

$$\sigma^2_x \text{ kan vi skatta vänteriktigt med } s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$$

Därmed kan vi också skatta samplingfördelningens varians (och standardavvikelse):

$$\hat{\sigma}_{\bar{x}}^2 = \frac{1}{n} \cdot \hat{\sigma}_x^2 = \frac{s_x^2}{n}, \text{ dvs } \hat{\sigma}_{\bar{x}} = \frac{1}{\sqrt{n}} \cdot \hat{\sigma}_x = \frac{s_x}{\sqrt{n}}$$

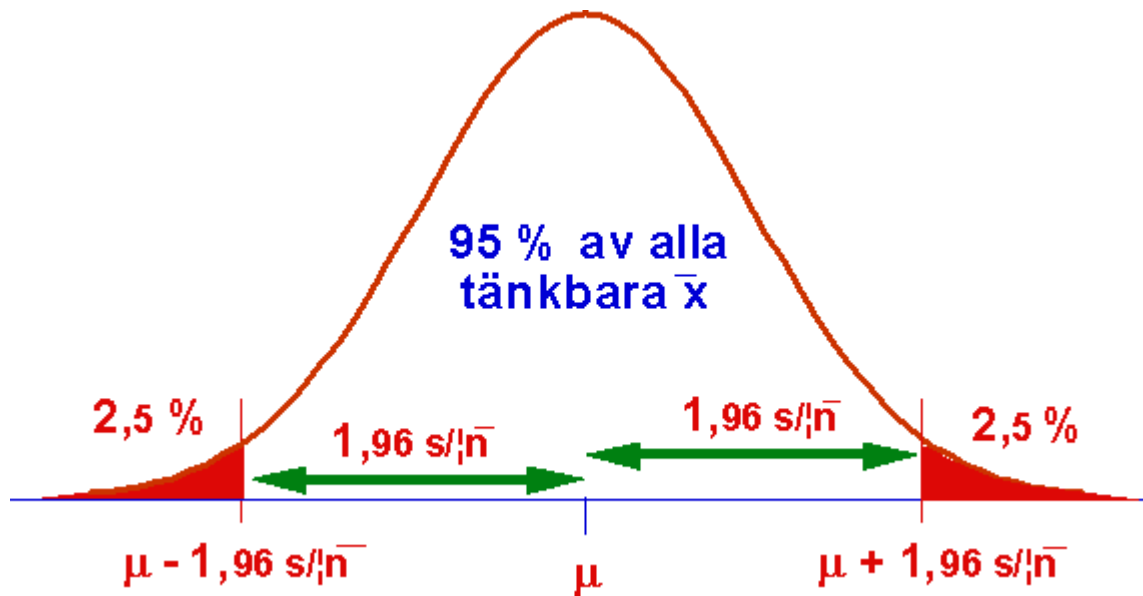
Eftersom samplingfördelningen för  $\bar{x}$  kan hanteras som normalfördelad kan vi t ex konstatera, att 95 % av alla tänkbara  $\bar{x}$  skulle falla inom intervallet:

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

och, om vi skattar populationsvariansen med  $s^2$ :

$$\mu \pm 1,96 \cdot \frac{s_x}{\sqrt{n}}$$

Se figur B2.4-2!

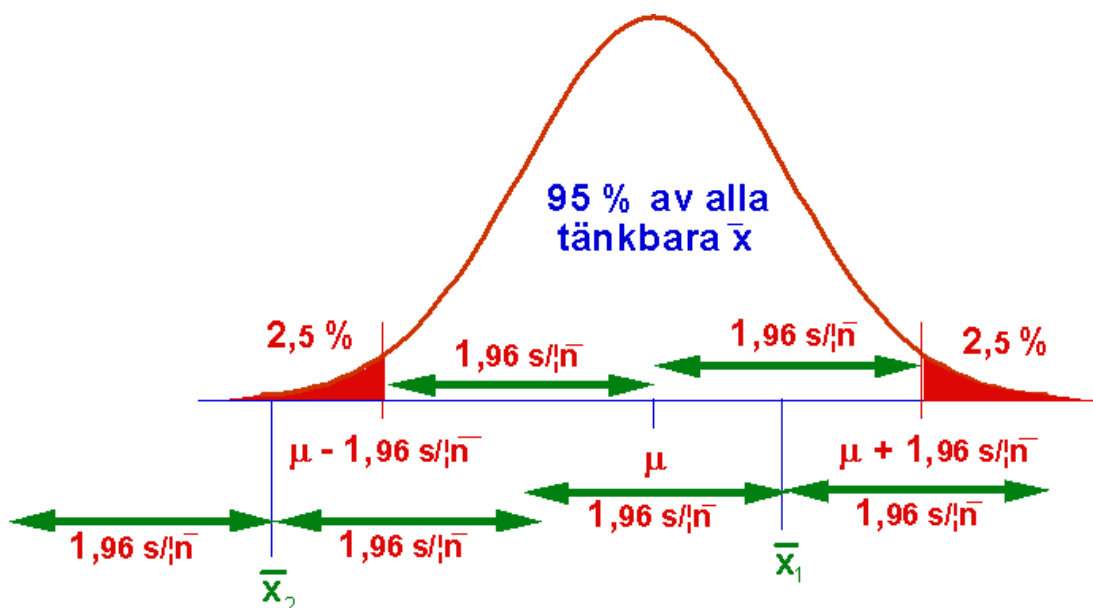


Figur B2.4-2 Det intervall i samplingfördelningen för stickprovsmedelvärdet inom vilket 95 % av alla värden faller är  $2 \cdot 1,96$  standardavvikelseenheter stort

### B2.4.6 Intervallskattning - intervallestimat - konfidensintervall

Om vi tar *ett enda* stickprov och beräknar dess  $\bar{x}$  så är detta teoretiskt sett liktydigt med att vi "tar en observation ur samplingfördelningen för  $\bar{x}$ " (se figur B2.4-2). Vi vet naturligtvis *inte*, om just detta  $\bar{x}$ -värde tillhör de 95 % som ligger mellan gränserna i figur B2.4-2 eller om det är ett värde utanför endera gränsen. Men vi *vet*, att den metod vi använder ger oss ett  $\bar{x}$  mellan gränserna i 95 fall av hundra.

Därför: Om vi symmetriskt kring vårt funna  $\bar{x}$  lägger ett intervall, lika långt som den sträcka inom vilken 95 % av  $\bar{x}$ -värdena i samplingfördelningen faller, så kommer detta intervall att täcka  $\mu$  om vårt  $\bar{x}$  är som  $\bar{x}_1$  i figuren nedan, dvs tillhör de 95 % - och inte nå fram till  $\mu$  om det är som  $\bar{x}_2$ , dvs kommer ifrån någon av "svansarna" utanför 2,5 %-gränserna.



*Figur B2.4-3 Konfidensintervall för  $\mu$* 

Om vi bildar intervall kring  $\bar{x}$  på detta sätt kommer vi att i 95 fall av hundra få intervall - konfidensintervall - som täcker  $\mu$ . Observera dock, att *ett* visst numeriskt bestämt intervall (det enda vi i praktiken beräknar) antingen täcker  $\mu$  eller inte - det täcker inte  $\mu$  med någon sorts "sannolikhet"! Sannolikhetsresonemanget hör till själva metoden, inte till det enskilda intervallet. Det är därför man infört benämningen *konfidens* ( $\approx$  tilltro). Med ett "95 % konfidensintervall" menas alltså ett intervall som bestämts med en metod som i 95 fall av hundra ger korrekta intervall, dvs (i detta exempel) intervall som täcker  $\mu$ . Den s k *konfidensnivån* är 95 %.

Om vi vill öka konfidensnivån, t ex till 99 %, måste vi förlänga intervallet. Medan 95 % av normalfördelningen faller mellan  $\pm 1$ , 96 standardavvikelseenheter från medelvärdet är gränserna för 99 %  $\pm 2,58$ .

Vill vi krympa intervallet (öka precisionen i skattningen) måste vi sänka konfidensnivån, t ex till 90 %, vilket ger gränserna  $\pm 1,65$ .

Hög konfidens innebär sålunda lägre precision hos intervallestimatet (breda intervall) och högre estimatprecision (trängre intervall) lägre konfidens. Det enda sätt på vilket vi kan höja skattningens precision (krympa intervallet) utan att uppoffra konfidens är genom att öka stickprovsstorleken ( $n$ ). Därigenom minskar vi samplingfördelningens varians/standardavvikelse.

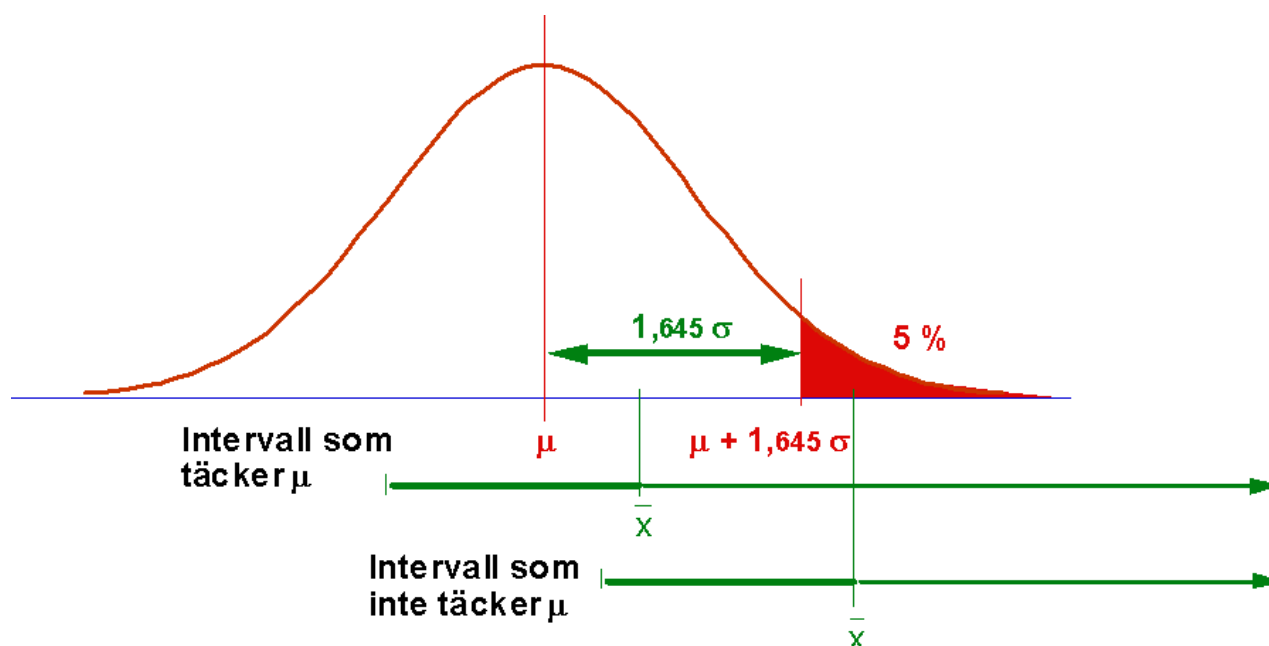
(Teoretiskt sett kunde vi också, för att minska intervallbredden, söka en mer efficient estimator - dvs en estimator vars samplingfördelning hade en lägre varians, jfr avsnitt B2.4.4. Men tyvärr finns det i just det här fallet ingen sådan - den effektivaste estimatoren för  $\mu$  är  $\bar{x}$ .)

**B2.4.7 "Enkelsidigt" (asymmetriskt) konfidensintervall**

I föregående avsnitt bildade vi symmetriska konfidensintervall kring  $\bar{x}$  som med viss konfidens fångade  $\mu$ . Verbalt kan sådana intervall uttryckas: "Medelvärdet ligger inom ett intervall  $a \pm b$ ".

I vissa fall är ett asymmetriskt (s k enkelsidigt) intervall mer intressant än ett symmetriskt (s k dubbelsidigt). Låt säga, att man vill bestämma brottgränsen för en brobalk. Det man då främst är intresserad av är att kunna säga att "brottgränsen ligger vid minst  $a$  kN" - en övre begränsning är sakligt ointressant, balken får gärna vara hur stark som helst.

I en sådan situation vill vi bilda ett intervall med (t ex) 95 % konfidens för brottgränsen  $\mu$  och som inte behöver ha begränsning uppåt. Vi är därmed beredda att ta en risk på 5 % för att få intervall som överskattar  $\mu$  - vilket blir fallet om det stickprov vi tar får ett så högt  $\bar{x}$  att vårt konfidensintervall inte "når ned" till  $\mu$ .



Figur B2.4-4 Enkelsidigt konfidensintervall

Vår utsaga får formen " $\mu$  är åtminstone  $\bar{x} - 1,645 \cdot \text{standardavvikelsen för } \bar{x}$ ", dvs " $\mu$  är åtminstone  $a$ ".

På analogt sätt utformas intervall för " $\mu$  är högst  $a$ " när detta är den intressanta utsagan.

## B2.5 Hypotesprövning

### B2.5.1 Den generella teorin

På i princip samma sätt som vid intervall estimation kan vi utnyttja en estimator och vad vi vet om dess fördelning för att testa *hypoteser* om värdet på estimatorns väntevärde (dvs den populationsstorhet som motsvarar estimatorn/stickprovskaraktäristikan).

Det första steget är att formulera två hypoteser om populationsstorhetens värde - en s k nollhypotes (betecknas  $H_0$ ) under vars villkor vi genomför testet och en mothypotes (betecknas  $H_M$  eller  $H_1$ ) som anger "motsatsen till  $H_0$ "; dvs  $H_0$  och  $H_M$  skall tillsammans täcka alla tänkbara giltighetsfall.

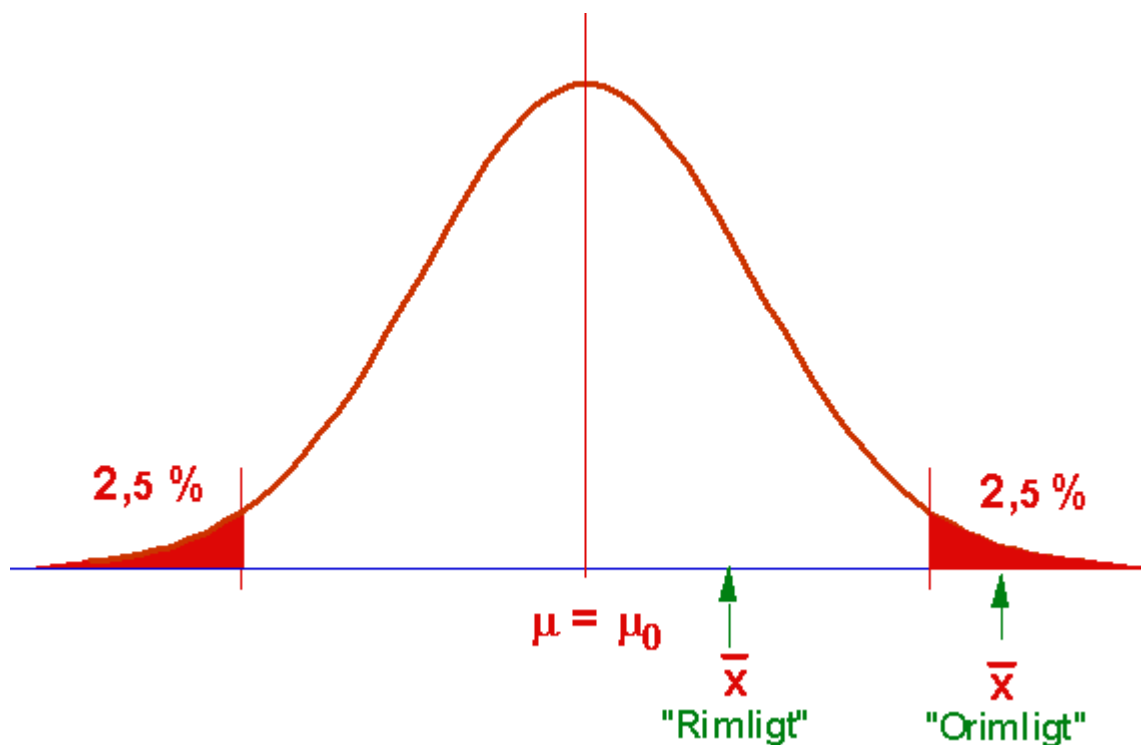
Exempel:  $H_0: \mu = \mu_0 = \text{konst}$

$H_M: \mu \neq \text{konst}$

Nästa steg är att finna en lämplig estimator för populationsstorheten (här lämpligen  $\bar{x}$  eftersom hypoteserna avser  $\mu$ ); dra ett stickprov och bedöma, huruvida det erhållna estimatet (dvs värdet på  $\bar{x}$ ) verkar "rimligt" under antagande att  $H_0$  gäller.

Vi måste alltså bestämma, vad som är "rimligt" Säg, att vi med "rimligt" menar ett  $\bar{x}$ -värde som tillhör 95 % av de  $\bar{x}$ -värden vi skulle kunna få om  $\mu$  verkligen är det värde

(=konst) som  $H_0$  säger. Om vi får ett  $\bar{x}$  som inte tillhör dessa 95 % blir vår slutsats, att det inte är troligt att  $H_0$  gäller. I så fall måste  $H_M$  gälla - den täcker ju alla övriga fall.



Figur B2.5-1 "Rimliga" och "orimliga" stickprovutfall under nollhypotesens giltighet

Sättet att ta reda på huruvida det erhållna är "rimligt" är att avgöra, hur många standardavvikelseenheter från det nollhypotetiska  $\mu$ -värdet som vårt  $\bar{x}$  hamnat. Vi bildar den så kallade *testvariabeln* (här betecknad  $z$ ) enligt principen:

$z = (\text{estimat} - \text{estimatorns väntevärde enligt } H_0) / \text{estimatorns standardavvikelse}$

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

$\sigma_{\bar{x}}$  är i regel okänd, men kan som vi tidigare sett skattas med  $\frac{s_x}{\sqrt{n}}$  så att

$$z = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$$

Testvariabeln  $z$  är "den normerade motsvarigheten" till det  $\bar{x}$ -värde vi fått, dvs den utgör ett värde "i samma skala" som den normerade normalfördelning vi har tabeller för. Genom transformationen har vi åstadkommit, att även testvariabeln  $z$  har medelvärde 0 och standardavvikelse 1.

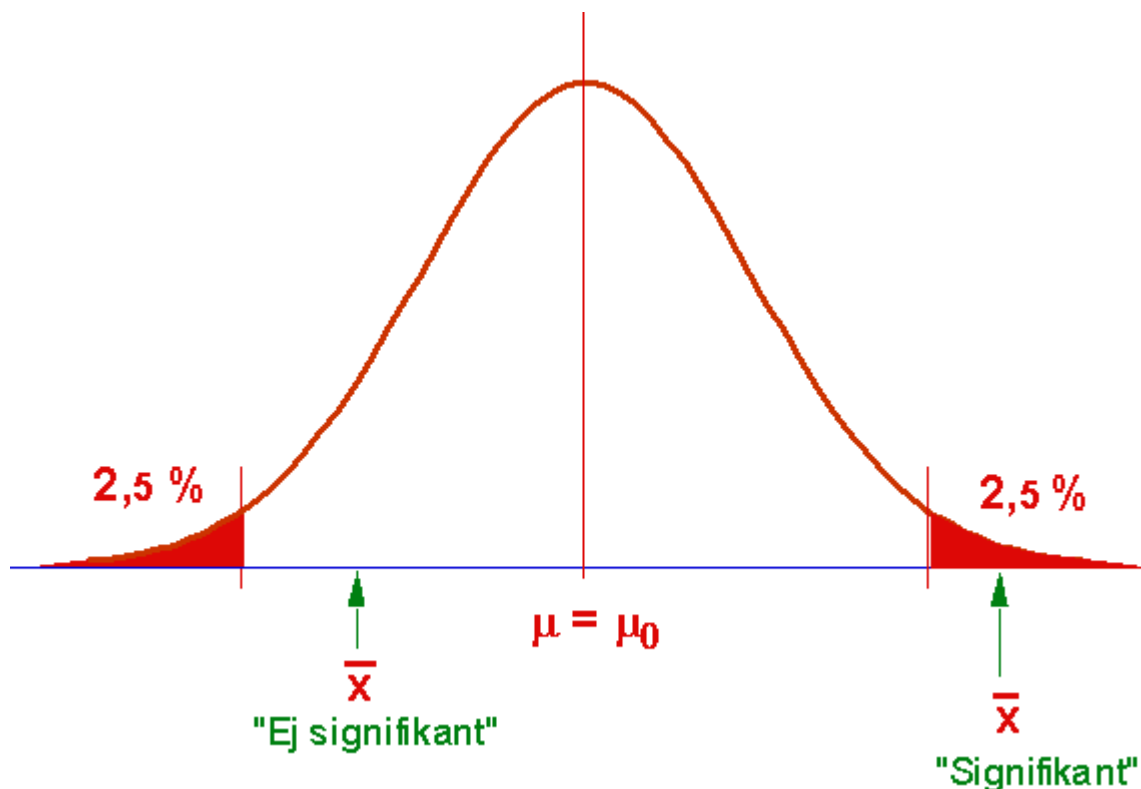
Om beloppet av vår (som vi antar normalfördelade) testvariabel  $z$  överstiger 1,96 blir slutsatsen, att  $H_0$  förkastas till förmån för  $H_M$ . Om  $z$  till beloppet är lägre än 1,96 kan vi inte förkasta  $H_0$  - vilket dock inte innebär, att vi "bevisat att  $H_0$  gäller". Vi kan aldrig "styrka" en nollhypotes - bara förkasta den eller inte förkasta den.

### B2.5.2 Exempel, dubbelsidigt test

Låt säga, att vi har gjort 49 mätningar av terrassytenivåns avvikelse från den projekterade nivån. Riktvärdet för avvikelserna är 0. Vi har fått  $\bar{x} = -3$  (mm) och en standardavvikelse i stickprovet  $s = 14$  (mm). Vi testar  $H_0 : \mu = R = 0$  mot hypotesen  $H_M: \mu = R = 0$ . Vi bildar testvariabeln:

$$z = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} = \frac{-3 - 0}{\frac{14}{\sqrt{49}}} = -1,5$$

- 1,5 är inte till beloppet större än 1,96 och är därför inte vad vi i förväg bestämt vara "orimligt", dvs det är inte "orimligt" att få ett  $\bar{x}$ -värde -3 om den faktiska medeldifferensen ( $\mu$ ) är 0, dvs = riktvärdet R. Därför kan vi inte förkasta nollhypotesen.



Figur B2.5-2 Signifikant respektive ej signifikant testresultat

Om vi i stället hade fått t ex  $\bar{x} = 5,0$  (och samma  $s$ ) i stickprovet hade testvariabeln  $z$  fått värdet 2,5. Då skulle vi ha förkastat nollhypotesen på den valda  $s$  k

*signifikansnivån* (5 %). Vi skulle säga, att vårt resultat var "signifikant" (på den valda nivån).

Ordet "signifikant" kan i det här sammanhanget översättas med "alltför avvikande för att orsaken rimligen kan vara enbart slumpvariation".

### **B2.5.3 Fel av typ I och fel av typ II, power och OC-kurvor**

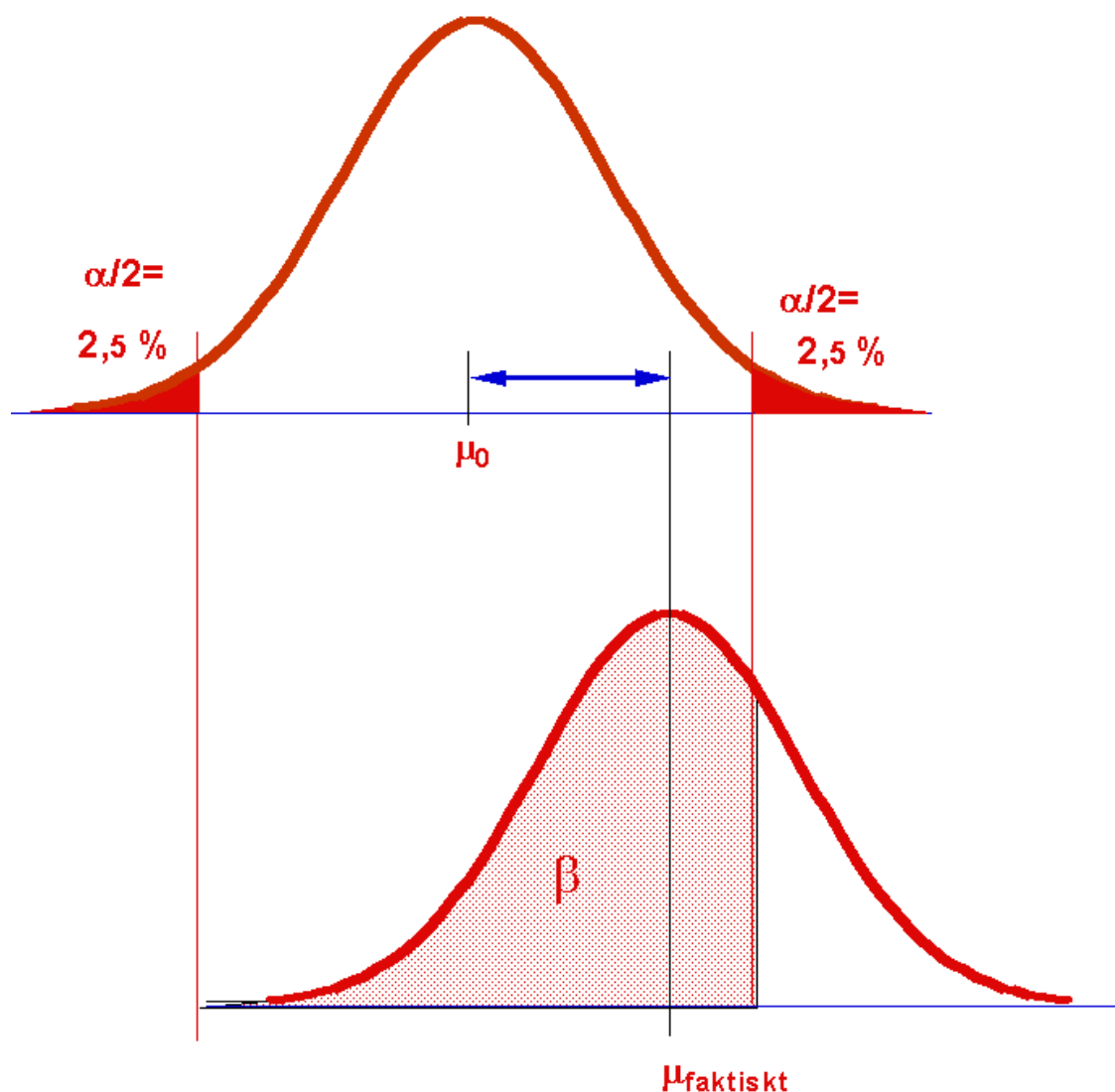
Signifikansnivån för ett test anger "den tillämpade metodens sannolikhet att felaktigt förkasta korrekta nollhypoteser": I exemplet ovan valde vi signifikansnivån 5 % - dvs vi tar risken att i 5 % av alla tänkbara fall få sådana stickprov att vi kommer att förkasta nollhypotesen *trots att* den är sann.

Detta fel brukar kallas "fel av typ I". Sannolikheten för att det inträffar anges av signifikansnivån och brukar betecknas med  $\alpha$ .

Om vi vill minska risken för att felaktigt förkasta korrekta nollhypoteser skall vi sänka signifikansnivån - t ex till 1 %. På den signifikansnivån är de kritiska gränserna för  $z$  inte längre  $\pm 1,96$  utan  $\pm 2,58$  - det krävs alltså en större avvikelse mellan det funna  $\bar{x}$  och nollhypotesens  $\mu_0$  (för samma standardavvikelse) för att vi skall komma att förkasta nollhypotesen.

Att förkasta korrekta nollhypoteser är dock inte det enda fel man kan råka ut för. Att *inte* förkasta en *falsk* nollhypotes är ju också fel. Detta kallas "fel av typ II" och dess sannolikhet brukar betecknas  $\beta$ .





Figur B2.5-3 Fel av typ I och fel av typ II

$\alpha$ , sannolikheten för fel av typ I, är lika med den signifikansnivå som vi väljer, medan storleken av  $\beta$  beror av vilket värde testvariabelns väntevärde faktiskt har. Utan att gå in på problemet alltför teoretiskt kan man kanske förstå, att sannolikheten att inte förkasta en felaktig nollhypotes blir mindre och mindre ju mer felaktig nollhypotesen är.

Med utgångspunkt i de två felriskerna  $\alpha$  och  $\beta$  försöker man konstruera test som är så effektiva som möjligt. Man vill naturligtvis helst inte förkasta *korrekta* nollhypoteser, samtidigt som man vill att sannolikheten att *inte* förkasta *felaktiga* nollhypoteser skall vara så *liten* som möjligt och åtminstone minska snabbt ju mer felaktig nollhypotesen är, dvs ju mer det faktiska värdet på kriterievariabelns väntevärde avviker från det som nollhypotesen anger.

I det här sammanhanget talar man om ett tests "makt" eller "styrka" ("*power*" på engelska, vanligt även i svenskt fackspråk). Power anger testets förmåga att avslöja falska nollhypoteser:

$$\text{Power} = 1 - \beta$$

Eftersom  $\beta$  avtar, ju mer felaktig nollhypotesen är, ökar powern på motsvarande sätt. Man eftersträvar givetvis, att powerfunktionen skall vara så brant som möjligt - dvs redan "små fel" hos nollhypotesen skall ge hög sannolikhet för avslöjande.

Ibland uttrycker man testets styrka med dess s k operationskaraktäristika ("operation characteristic" på engelska, vars grafiska bild ofta kallas "OC-kurva"). OC-kurvan visar samma sak som power-funktionen, fastän "tvärtom", dvs den visar, hur sannolikheten att begå fel minskar, när det faktiska förhållandet alltmer avviker från vad nollhypotesen utsäger.

Vilken signifikansnivå är "bäst"? Den frågan har inget entydigt svar. Ett lågt  $\alpha$  (t ex 1%) innebär samtidigt högre  $\beta$  än vad ett högre  $\alpha$  skulle medföra. Frågan är snarast, vilket av de två möjliga felen som är mest fatalt i den situation där testet utnyttjas. Det finns omfattande litteratur på området att ta del av för den intresserade.

#### B2.5.4 Enkelsidigt test

Vi har redan konstaterat, att en nollhypotes inte kan "bekräftas" - endast förkastas eller inte förkastas. Däremot innebär testlogiken, att om  $H_0$  förkastas så måste  $H_M$  gälla. Detta pekar mot att det vore önskvärt att formulera testet så att  $H_M$  uttrycker "den slutsats vi helst vill dra".

Innebörden i detta är lättast att uppfatta i samband med s k enkelsidiga test; dvs test där vi inte bara bryr oss om hur stor den absoluta avvikelserna mellan det estimerade och det hypotetiska värdet är utan också tar hänsyn till åt vilket håll avvikelserna går.

Låt oss ta ett exempel: En viss egenskap hos ett kontrollobjekt skall enligt kraven "överstiga 110". Kravet innebär alltså, att värdet inte får vara 110 eller lägre - däremot får det vara hur högt som helst.

Vi undersöker kontrollobjektet med hjälp av ett stickprov  $n = 49$  och får  $\bar{x} = 112$  och  $s = 14$ .

Vi vill dra en slutsats av typ "(det sanna) värdet överstiger 110" (dvs kontrollobjektet accepteras). Vi formulerar våra noll- och mothypoteser så att den önskvärda utsagan hamnar i mothypotesen - och så att hypoteserna tillsammans täcker alla tänkbara fall:

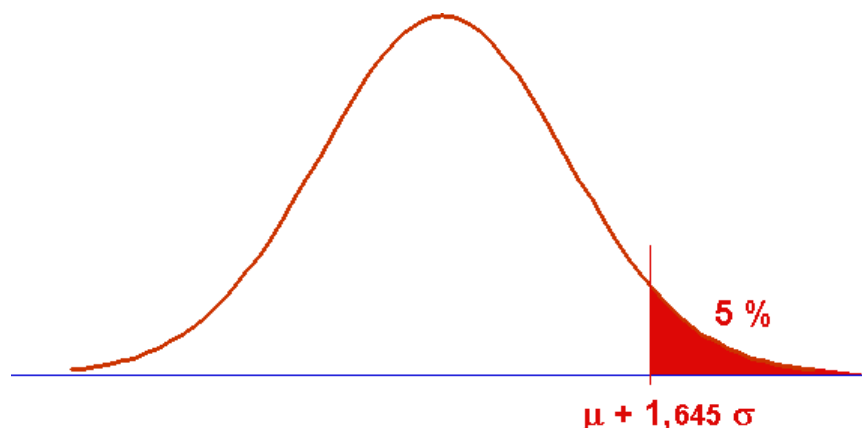
$$H_0: \mu \leq 110$$

$$H_M: \mu > 110$$

Här är nollhypotesen som synes inte entydig - den omfattar många giltighetsfall. Vi skall testa den under dess "ogynnsammaste" giltighetsfall, dvs när den är så nära  $H_M$ :s giltighet som möjligt, dvs för fallet  $\mu = 110$ . Om vi i sådant fall får ett signifikant resultat och alltså kan förkasta giltighetsfallet  $\mu = 110$  så kan vi därmed också förkasta alla giltighetsfall  $\mu < 110$ .

Vi väljer signifikansnivån 5 %; dvs tar 5 % risk att förkasta en korrekt nollhypotes ( $\mu \leq 110$ , vars ogynnsammaste giltighetsfall är  $\mu = 110$ ). Detta innebär, att vi "begår

detta fel" om nollhypotesen är sann och vi råkar få ett  $\bar{x}$ -värde från den del av samplingsfördelningens högra svans där sannolikheten att hamna är 5 %.



Figur B2.5-4 Signifikansområde

Tabellen visar, att den aktuella gränsen ligger 1,645 standardavvikelseenheter över samplingsfördelningens medelvärde.

Ligger vårt erhållna  $\bar{x}$ -värde över denna gräns? För att undersöka detta bildar vi testvariabeln:

$$z = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} = \frac{112 - 110}{\frac{14}{\sqrt{49}}} = 1$$

som alltså anger, att vårt erhållna  $\bar{x}$  inte ligger inom det område där vi ansett det vara "osannolikt" att hamna om nollhypotesen är sann.

Resultatet är inte signifikant. Vi kan inte förkasta nollhypotesen.

Vi kan alltså *inte* dra slutsatsen (enligt mothypotesen) att "värdet överstiger 110". Vi kan därmed *inte godkänna* kontrollobjektet.

Om vi däremot hade fått  $\bar{x} = 114$  i vårt stickprov (och samma standardavvikelse) så hade testvariabelvärdet varit:

$$z = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} = \frac{114 - 110}{\frac{14}{\sqrt{49}}} = 2$$

vilket skulle ha inneburit, att vi förkastade nollhypotesen (till förmån för mothypotesen som ju täcker alla de fall som nollhypotesen inte täcker). Vår slutsats hade då blivit: "μ överstiger 110, kontrollobjektet är godkänt"

## B2.6 Acceptansintervall

Acceptansintervall (i den bemärkelse termen används i [VÄG 94](#)) kan uppfattas som *generaliseringar* av de testförfaranden som beskrivits ovan. För att bespara kontrollören/utföraren arbetet med att vid varje kontroll räkna om kriterievariabler till testvariabler, slå i normalfördelnings-, t-fördelnings- eller binomialfördelningstabeller för att hitta de gränser som anger vad som skall betraktas som ett "osannolikt utfall" under antagande om nollhypotesens giltighet etc har man i förväg och för en lämplig signifikansnivå räknat ut *det intervall inom vilket kriterievariabeln (oftast  $\bar{x}$ ) får falla för att man skall kunna dra slutsatsen "kontrollobjektet är godkänt"*. Beroende på kontrollens karaktär leder detta ibland till "enkelsidiga" (asymmetriska) acceptansintervall, ibland till "dubbelsidiga" (symmetriska) acceptansintervall.



Figur B2.6-1 Enkelsidigt och dubbelsidigt acceptansintervall

Som vi sett i det föregående beror variansen/standardavvikelsen hos  $\bar{x}$  (eller annan estimator) dels av stickprovsstorleken ( $n$ ), dels av den bakomliggande populationens varians ( $\sigma^2$ ), vilken som regel skattas med  $s^2$ . Som vi också sett tidigare påverkas möjligheterna att vid ett test få signifikant resultat av variansens/standardavvikelsens storlek.

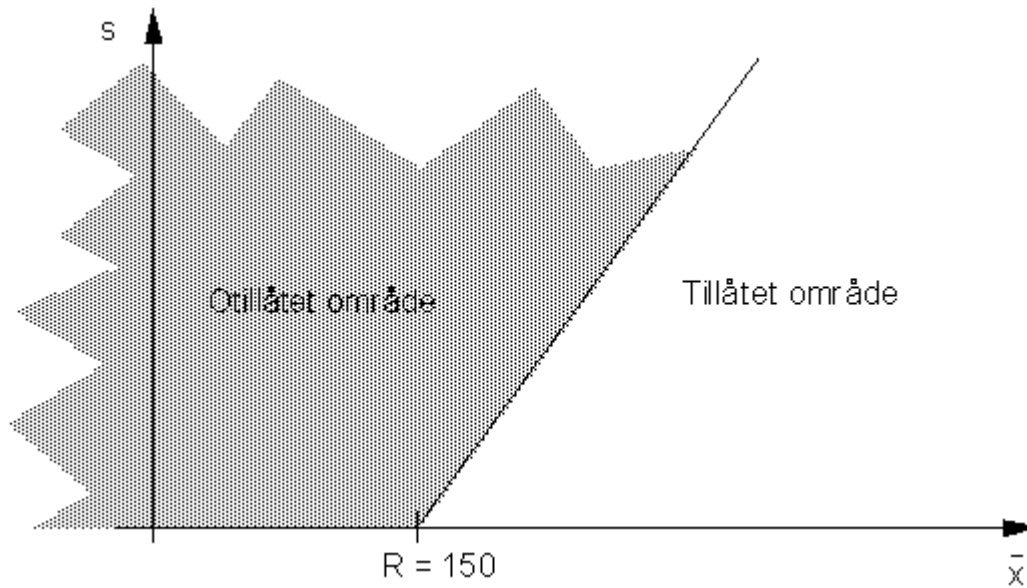
Exempel: Vi har för ett kontrollobjekt 49 mätningar av en egenskap för vilken kravet är att den skall vara "större än 150". Medelvärdet i stickprovet är  $\bar{x} = 153$  och standardavvikelsen i stickprovet  $s = 14$ . Vi testar  $H_0: \mu \leq 150$  mot  $H_M: \mu > 150$ :

$$z = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} = \frac{153 - 150}{\frac{14}{\sqrt{49}}} = 1,5 \quad (\text{ej signifikant ens på 10 \% nivå})$$

Om standardavvikelsen i stället hade varit  $s = 7$  hade vi fått

$$z = \frac{153 - 150}{\frac{7}{\sqrt{49}}} = 3 \quad (\text{signifikant t o m på 0,5 \% nivå})$$

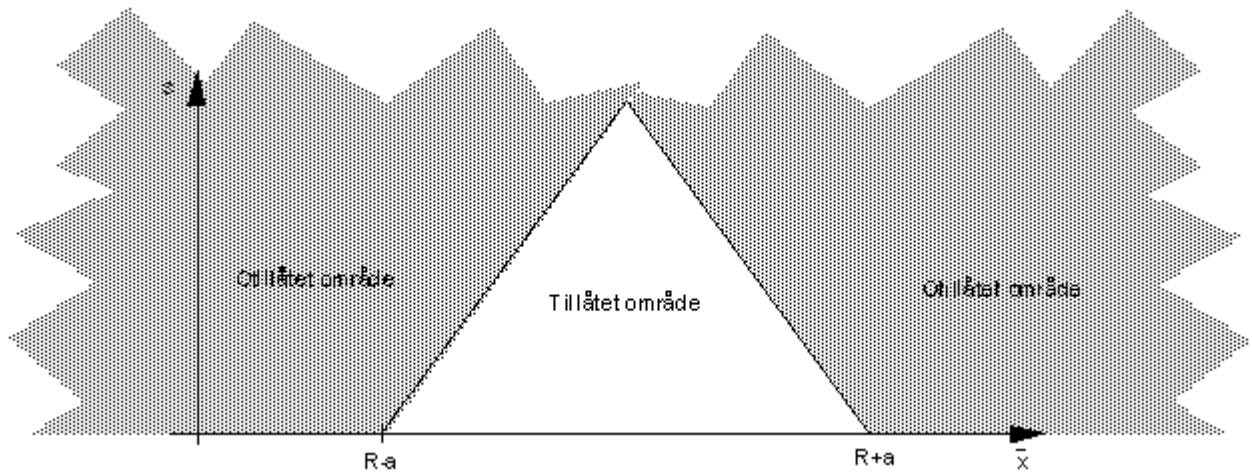
Standardavvikelsen - som ju är ett mått på spridningen i våra mätvärden ("osäkerheten" om man så vill) har - och skall ha - inverkan på hur stort acceptansintervall man kan tillåta. Ju större stickprovsstandardavvikelse, desto "svårare" skall det vara att förkasta nollhypotesen "kontrollobjektet uppfyller inte kravet" (till förmån för mothypotesen "kontrollobjektet uppfyller kravet"). Acceptansintervallens beroende av standardavvikelsen kan illustreras grafiskt av en  $s/\bar{x}$  acceptansyta i  $s/\bar{x}$ -planet. Acceptansytan är den geometriska orten för alla tänkbara acceptansintervall (för olika värden på  $s$ ).



Figur B2.6-2 Acceptansyta för enkelsidiga acceptansintervall

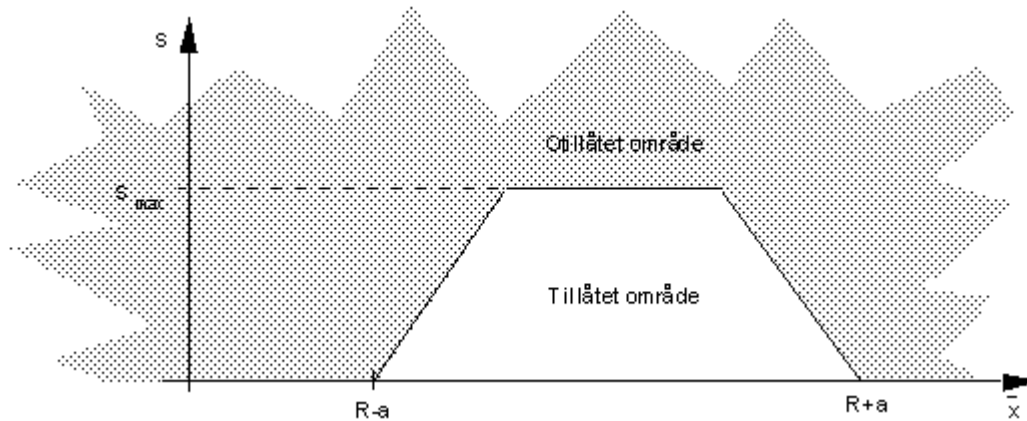
I (det orealistiska) extremfallet  $s = 0$  kan vi godkänna kontrollobjektet för alla värden på  $\bar{x}$  över 150, som är det "tekniska" gränsvärde som gäller. Så snart standardavvikelsen i stickprovet överstiger 0 (vilket den i praktiken alltid gör) måste vi höja kravet på  $\bar{x}$  (dvs kräva större täljare i testvariabeluttrycket) för att signifikansnivån skall vara den förutbestämde.

Om kontrollsituationen innebär, att dubbelsidigt test genomförs, har vi ett dubbelsidigt acceptansintervall. Det "tekniska acceptansintervallet",  $T$ , är  $R \pm a$ , som definierats för  $s = 0$ . När man tar hänsyn till slumpvariationen krymper intervallet med växande  $s$ :



Figur B2.6-3 Acceptansyta för dubbelsidiga acceptansintervall

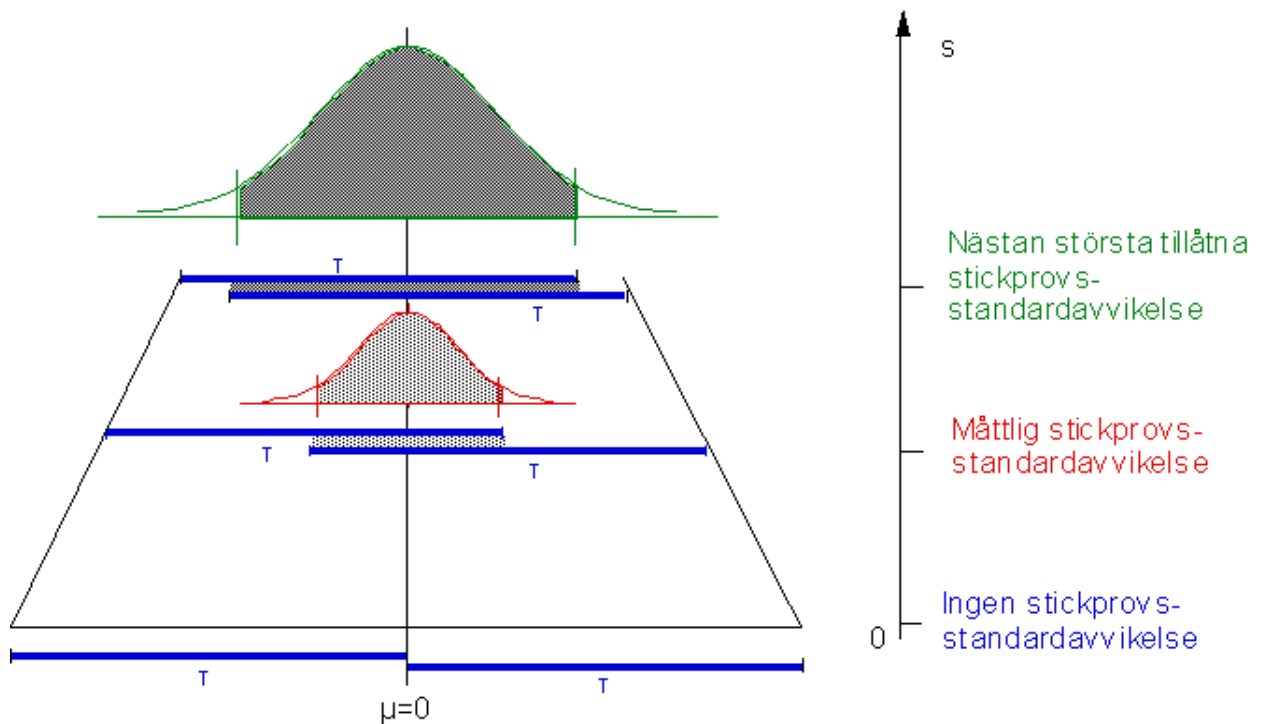
Ofta finns dessutom av olika skäl krav på att stickprovsvariationen inte får vara hur stor som helst. Det finns då en övre gräns för standardavvikelsen - stickprov med större standardavvikelse än den maximalt tillåtna får inte utnyttjas för slutsatsdragning rörande kontrollobjektet:



Figur B2.6-4 Acceptansyta med restriktion på  $s$  (för dubbelsidiga acceptansintervall)

Talrika exempel på acceptansytor av olika slag finns i bilaga 1.

Man kan också uppfatta acceptansintervall som en form av konfidensintervall för de gränser mellan vilka  $\bar{x}$  kan tillåtas falla vid ett visst värde på stickprovsstandardavvikelsen. Figur B2.6-5 illustrerar detta synsätt.



Figur B2.6-5 Acceptansintervall vid olika värden på  $s$