Version, 01-09-04

Inst för MARIN EKOLOGI, G.U.

STATISTISK ANALYS OCH UTFORMNING AV EXPERIMENT
1 INTRODUKTION

Sammanställt av Per Jonsson 1995, reviderat av Mats Lindegarth 2001, Carl André 2001

Den vetenskapliga metoden

Vetenskap syftar till att förklara de fenomen vi uppfattar genom studier av mönster och de processer som genererar dessa mönster. Ett exempel på ett fenomen som kräver förklaring är observationen att det finns en negativ korrelation mellan bottenfällda larver och tätheten av adulter hos hjärtmusslan, *Cerastoderma edule*. Hjärtmusslor lever på grunda sand- och lerbottnar, ofta i höga tätheter. Modeller som kan förklara observationen är t.ex.

a) predation av adulter (vuxna musslor äter mussellarver)
b) grävning av adulter (vuxna musslor bökar runt så att larver flyr eller begravs)
c) andra predatorer gömmer sig bland adulter och äter upp larverna
d) aktivt val hos larver (larverna gillar inte vuxna)

Dessa modeller kan ge upphov till följande testbara hypoteser:
Hypotes 1. Larver finns i magen hos adulter
Hypotes 2. Fixering av adulter i sanden ger högre settling (bottenfällning av larver)
Hypotes 3. Burar som utestänger respektive innestänger andra predatorer har effekt på settling
Hypotes 4. Provytor med extrakt från adulter reducerar settling

Den vetenskapliga medtoden, som den beskrivs här, kan sammanfattas i figuren nedan.

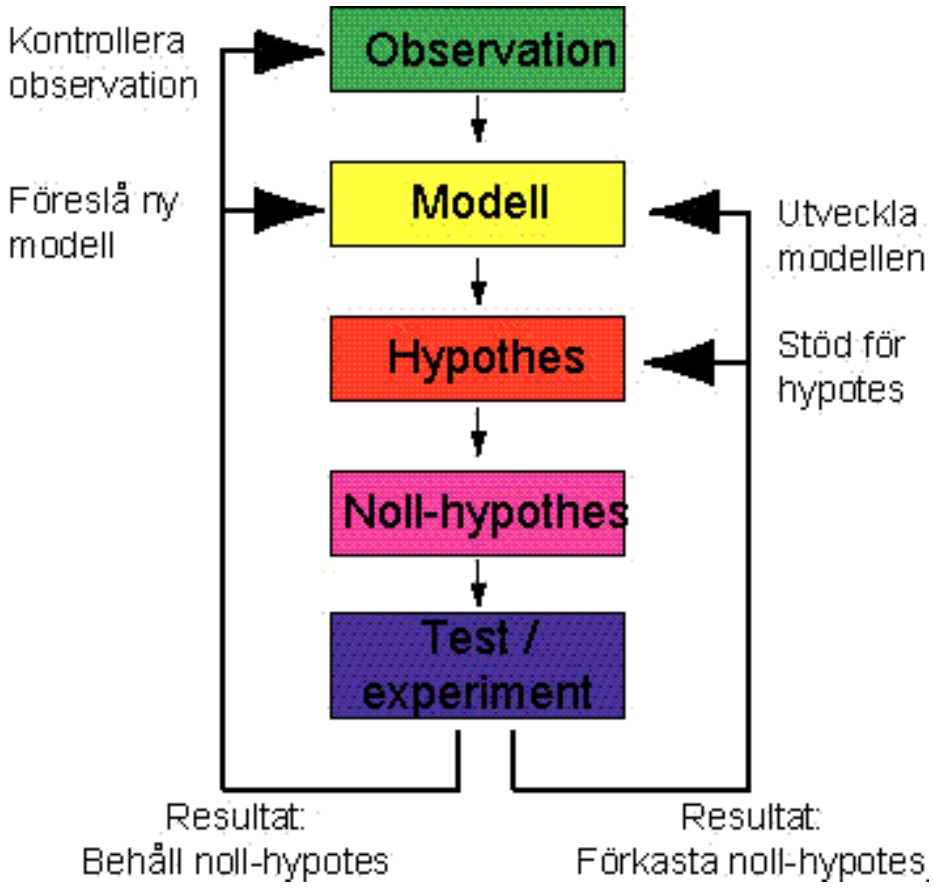

Fig. 1 Formellt logiskt schema för en vetenskaplig metod.

Hur upptäcker vi det observerade mönstret? Hur testar vi våra hypoteser?
Man skulle kunna tänka sig att man uttrycker det observerade mönstret med ord i en enkel sats. "Det finns färre settlade larver nära adulter". I en enkel värld kanske detta skulle fungera. I verkligheten är det oftast så att mönster varierar i tid och rum. Hur generellt är mönstret? Var det en tillfällighet? Hur mycket färre larver?

Dessutom observerar vi inte omvärlden på ett objektivt sätt utan påverkas av många faktorer både medvetna och omedvetna. Kanske vi hade svårare att upptäcka larverna på bottnar med många adulter där topografin är mer oregelbunden? Vi ser redan att observationen kräver en systematisk metodik. Här är ett exempel på en genomtänkt observationsstudie:

1. identifiera det område du vill uttala dig om
2. välj slumpmässigt inom området ett antal lokaler med många adulter och ett antal med få eller inga adulter
3. ta stickprov av bottenmaterial på ett representativt sätt från varje lokal och räkna larver
4. jämför antalet larver mellan de två typerna av lokaler

Vid biologiska studier, och generellt inom modern vetenskap, intar kvantifiering (mätning, frekvens) en viktig roll. Genom kvantifiering kan man precisera beskrivningen av mönster och även precisera sina hypoteser. I många fall krävs kvantitativ kunskap för att skilja mellan olika alternativa hypoteser.

Den storhet eller egenskap som vi mäter, är intresserade av och har hypoteser om kallas (beroende) variabler. I exemplet ovan räknade vi frekvensen larver i två klasser, närvaro och frånvaro av adulter. En sådan variabel kallas för nominell variabel. Andra observationer bygger på att vi mäter något. Vi skulle t.ex. kunna kartlägga antalet larver på olika avstånd från en samling adulter. En sådan mätvariabel kan vara kontinuerlig eller diskret. Slutligen har vi rankade variabler där vi kan ordna enskilda data men inte säga något om deras relativa storlek. Vissa typer av statistiska analyser behandlar flera variabler samtidigt. Det finns många exempel på sådana s.k. multivariata analyser (jämför univariata). Dessa kommer dock ej att diskuteras närmare under denna kurs men i princip kan man säga att samma logik och restriktioner gäller för dessa metoder, som för de univariata metoder som behandlas här. Slutligen är det också viktigt att skilja på multivariata och multifaktoriella analyser. Det senare är analyser som där man samtidigt undersöker hur flera saker, faktorer, påverkar en variabel.

Frekvensfördelningar

I mångt och mycket karakteriseras biologiska system av variation på många nivåer. Variationsrikedomen är ju en egenskap som gör biologi så fascinerande. Variationen innebär dock att det är mer komplicerat att identifiera mönster och deras orsaker. För fysikaliska och

kemiska mönster, som är mindre variabla, är det betydligt enklare att undersöka underliggande processer. Vi kommer att ägna mycket tid till studier av variationen i biologiska system.

Vad innebär då variationen? Låt oss ta ett enkelt exempel. Vi vill veta hur storleken av blåmusslan, *Mytilus edulis*, varierar mellan väst - och ostkusten (Fig 2).

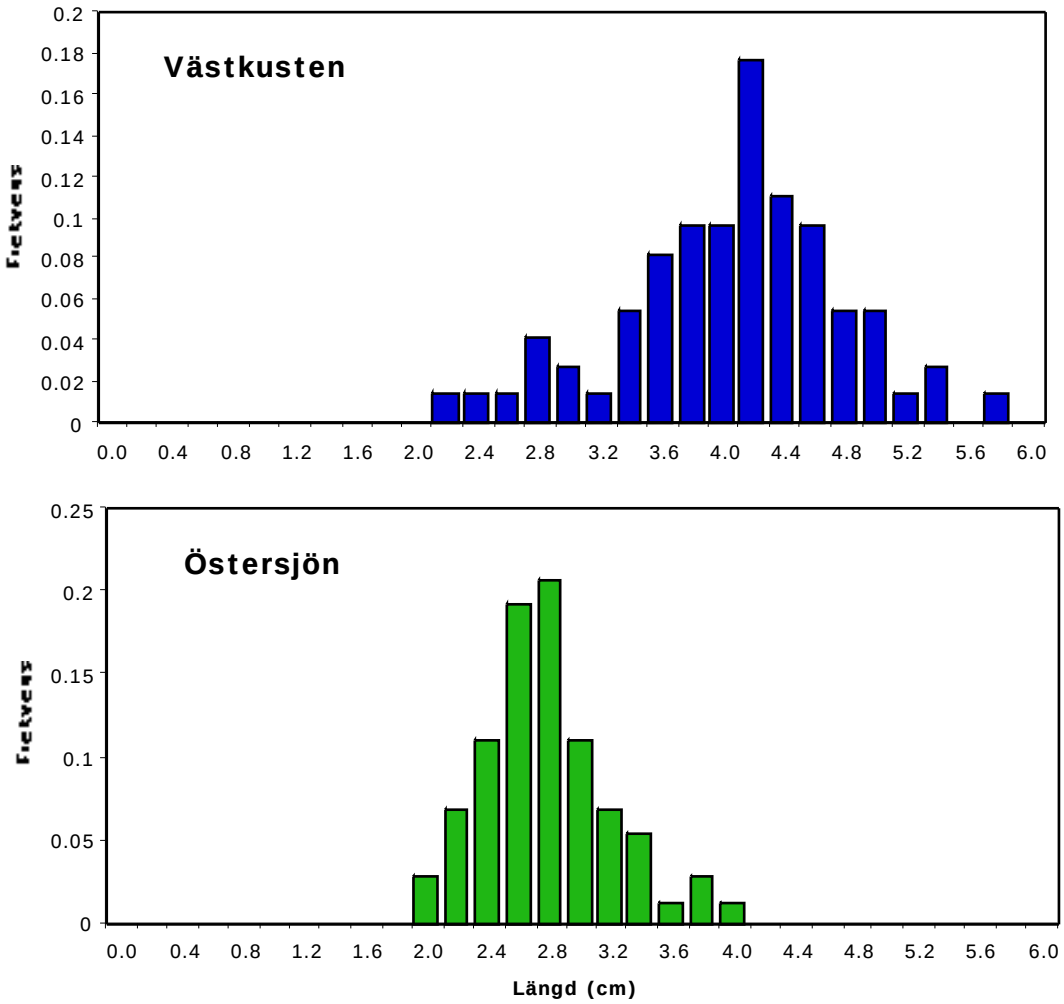

Fig. 2. Frekvensfördelningar av längder av musslor från västkusten och östersjön.

Som vi ser är inte alla individer av samma storlek. Hur ska vi kunna definiera populationerna? Det vanligaste är att vi karakteriserar populationen med två olika parametrar.

Parametrar

1. Lägesparameter. Det finns flera olika lägesparametrar men vanligast är det värde (längd) som gör att det sammanlagda medelavståndet till alla längder blir 0. Detta är det aritmetiska medelvärdet ($\frac{\Sigma x_i}{N}$).

2. Spridningsparameter

a/ intervall (min-max) bäst om populationen är fullständigt känd

b/ variansen $\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N}$ eller standardavvikelsen (SD) $\sigma = \sqrt{\sigma^2}$

Det finns även ytterligare parametrar t.ex. skevhet σ^3 och kurtosis σ^4 (toppighet). Det är viktigt att alltid titta på sina mätvärden, hur de fördelar sig. Att karakterisera en fördelning med ett fåtal parametrar är behändigt, men innebär nödvändigtvis förlust av information. Olika fördelningar kan därför ha samma parametervärden.

Stickprov

Meningen med denna studie är ju att kunna dra en slutsats om hur *Mytilus* ser ut i naturen, d.v.s. att kunna uttala sig om naturliga populationer. Vi har endast tagit ett stickprov och för att vi ska kunna dra slutsatser om populationen som stickprovet är dragit ifrån krävs att stickprovet är representativt. Det finns endast ett helt representativt stickprov där proportionen av längder är den samma som i populationen. För att öka chansen att få ett representativt stickprov eftersträvas en stickprovsstrategi där varje individ i den population som man studerar ska ha lika stor sannolikhet att bli dragen.

Hur kan vi veta om vi har tagit ett representativt stickprov. Sanningen är att det är mycket svårt att veta och kräver en hel del kompletterande information om populationen. Biologin bestämmer provtagningen. När man inte vet något om populationen brukar man ta ett s.k. slumpmässigt stickprov. Orsaken är att man då minimerar risken för subjektiva val av stickprov. Men kom ihåg att ett slumpmässigt stickprov troligen inte är representativt om t.ex. individerna är klumpade i sin fördelning.

Representativitet är ett allvarligt problem inom biologin (ett exempel är bottenprovtagning där effektiviteten varierar med olika bottentyper). Ett sätt att förbättra representativiteten är att inkludera eventuell information om populationen i provtagningsstrategin. Vet man att en population förekommer framför allt i vissa miljöer kan man göra en s.k. stratifierad provtagning där stickproven fördelas olika i de olika miljötyperna.

Centralvärdesteoremet

En mycket intressant egenskap hos skattade medelvärden (baserade på stickprov) är att de tenderar att vara normalfördelade oavsett den underliggande fördelningen av den aktuella populationen. Figur 3 visar hur en jämn och en skev fördelning ger approximativt normalfördelade medelvärden. Denna allmäna tendens kallas för centralvärdestendensen och beskrivs av det s.k. centralvärdesteoremet (CVT).

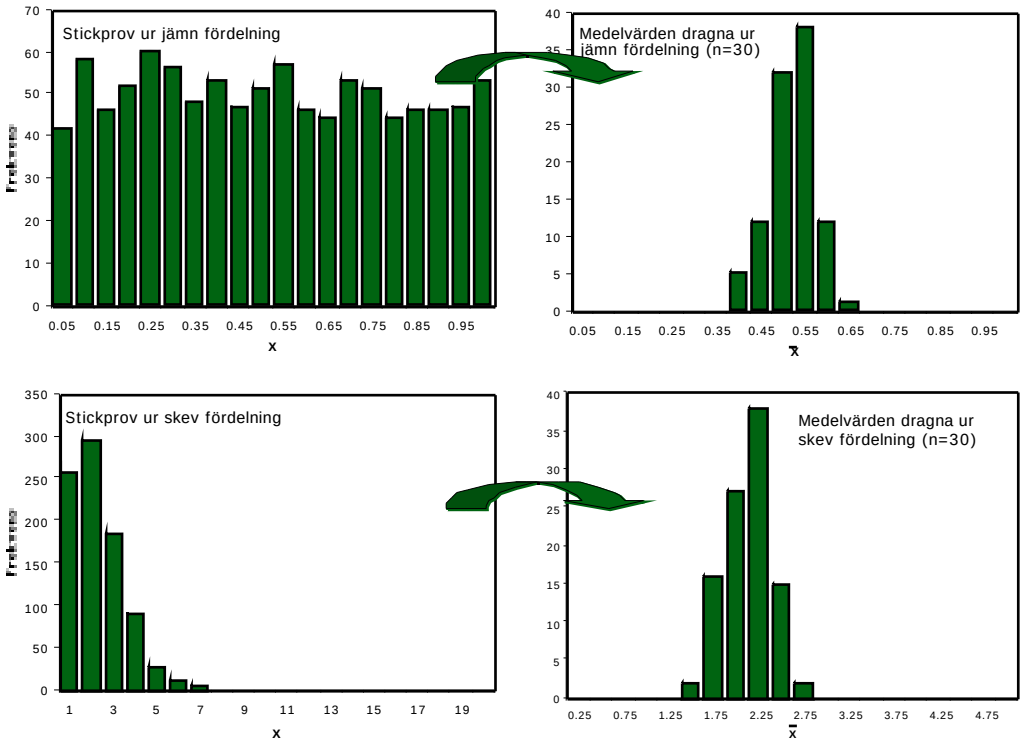

Fig. 3. Frekvensfördelningar från en jämn och en skev fördelning (n = 1000) samt fördelningar av 100 medelvärden dragna från dessa populationer.

Figur 4 visar schematiskt hur en fördelning ger approximativt normalfördelade medelvärden och hur variansen av de skattade medelvärdena är relaterad till variansen i den sanna fördelningen. Notera att man betecknar det sanna medelvärdet i den sanna populationen med μ, och det skattade medelvärdet i stickprovet med $\bar{x}$.

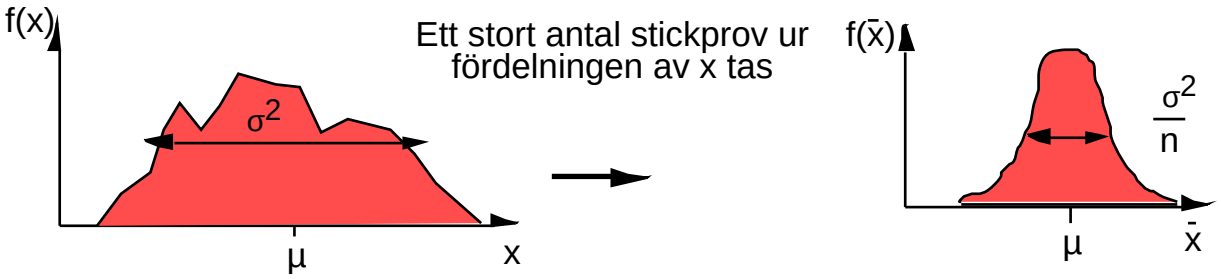

Fig. 4. Schematisk bild av sambandet mellan spridningen hos den bakomliggande populationen och fördelningen av medelvärden hos upprepade stickprov.

Det är viktigt att känna till under vilka villkor som CVT gäller. Vissa fördelningar är problematiska t.ex. extremt skeva, multimodala och binomiala. I detta sammanhang är det väsentligt att komma ihåg att många statistiska tester, s.k. parametrisk statistik, som vi kommer att bekanta oss med kräver att de skattade medelvärdena, $\bar{x}$, är normalfördelade.

Normalfördelning

Vi kan nu börja använda centralvärdesteoremet praktiskt och vi återvänder till våra blåmusslor (Fig 2). Vi beräknar ett medelvärde för det ena stickprovet. Om vi nu återupprepar detta för ett stort antal stickprov från samma population skulle vi kunna bygga upp en approximativ normalfördelning av medelvärden där vårt ursprungliga medelvärde ligger någonstans utefter denna fördelning. Utifrån detta enda stickprov kan vi räkna ut ett spridningsmått för normalfördelningen, standard error (SE, eller medelvärdets medelfel), som är standardavvikelsen (SD) för stickprovet delat med roten ur stickprovets storlek.

Standardavvikelse och standard error

Det brukar vara svårt att skilja på betydelsen av dessa två spridningsmått så vi ska titta lite närmare på dessa.

1. SD är en egenskap hos populationen liksom medelvärde och antal. SD beror därför inte på stickprovets storlek.
2. SE talar om med vilken precision vi skattar det sanna medelvärdet i populationen och fås genom:

$$SE = \frac{\sigma}{\sqrt{n}}$$

Som vi kan förvänta sjunker SE med storleken av stickprovet, precisionen ökar. Notera dock att den ökade precisionen endast ökar med $\sqrt{n}$.

Nåväl, oavsett hur fördelningen av vår population ser ut så kan vi alltså ge ett mått på den noggranhet som stickprovsmedelvärdet skattar populationens medelvärde. Att detta är meningsfullt beror på centralvärdesteoremet. Varje gång man redovisar ett medelvärde ska man ange med vilken precision det är skattat**.**

95% konfidens intervall

Enligt en godtycklig överenskommelse brukar man ange nogrannheten för ett skattat medelvärde med ett s.k. 95% konfidensintervall (CI) istället för med SE. Var någonstans på vår normalfördelning av medelvärden ska vi sätta våra gränser? Uppenbarligen ska vi hitta ett intervall som omfatter 95% av vår sannolikhetsyta. För att man inte ska behöva göra en besvärlig beräkning för varje nytt fall så gör man om den tänkta fördelningen av sina medelvärden till en s.k. standard normalfördelning. Varje speciell normalfördelning av medelvärden kan återföras till en standard normalfördelning (d) med medelvärdet = 0 och variansen = 1 genom att varje $\bar{x}$ minskas med det sanna medelvärdet (μ) och delas med SE enligt:

$$\pm d = \frac{\bar{x}-\mu}{SE}$$

Den här fördelningen (d - fördelning eller standard normalfördelning) finns i statistiska tabeller, och vi kan där finna vilka d - värden som motsvarar en viss sannolikhetsyta. Om vi slår upp d_{krit} för arean = 95% finner vi 1.96. Detta innebär att det är 95% sannolikhet att dra ett stickprov där det skattade medelvärdet, $\bar{x}$, ligger inom ± 1.96 * SE från det sanna medelvärdet μ. Eller uttryckt på annat sätt: Om vi skattar ett medelvärde så ligger det sanna medelvärdet, μ, i 95% av de stickprov som tas ur populationen inom ett konfidensintervall som begränsas av $\bar{x}$ ± 1.96 * SE.

t - fördelningen

Notera att när vi räknade ut det 95% CI ovan så användes den sanna standardavvikelsen, σ, i populationen för att räkna ut SE. Den sanna standardavvikelsen känner man ju i allmänhet lika lite som det sanna medelvärdet och man får nöja sig med att skatta den från sitt stickprov. Den skattade standardavvikelsen skrivs som s (skattad varians som s^2) och räknas ut från stickprovet som,

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

Notera att de kvadrerade avvikelserna nu delas med n - 1 vilket är antalet frihetsgrader för den skattade variansen (se stycket om frihetsgrader nedan). Det är rimligt att tänka sig att de fel i skattningen som finns i både $\bar{x}$ och s gör att 95% CI måste bli bredare än när vi beräknar d baserad på den sanna σ. Hur ska vi kunna lösa detta problem? Detta är ett centralt problem inom statistisk hypotestestning och löstes av Gosset 1907. Han gjorde på följande geniala sätt.

1. Utgående från en känd normalfördelning så drar vi ett stickprov med låt oss säga 3 individer.
2. Vi räknar ut $\bar{x}$ och $SE = s/\sqrt{n}$ från stickprovet. Enligt CVT är ju $\bar{x}$ normalfördelat med standardavvikelsen SE.
3. Vi kan nu återföra denna fördelning på en standardfördelning genom att dra ifrån det sanna medelvärdet och dela med det skattade SE, d.v.s. vi beräknar ett slags d utifrån detta men med den skattade SE och kallar den nu för t:

   $$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$
4. Vi upprepar detta ett stort antal gånger och bygger upp en fördelning av alla t - värden (Fig. 5).

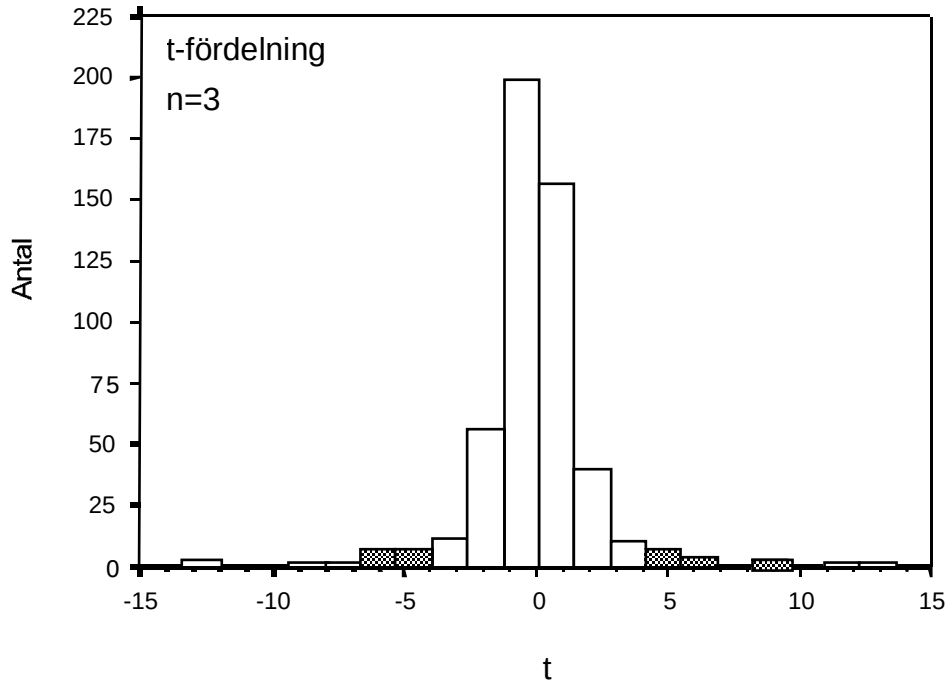

Fig. 5. Fördelning av t - värden beräknade från stickprov dragna från en normalfördelning.

5. Om s vore mätt utan fel så skulle den följa en standard normalfördelning. Men eftersom det alltid är risk för icke - representativa stickprov kommer s att varierar från prov till prov och fördelningen blir bredare än normalfördelningen; det är lättare att få extrema värden (skuggade staplar visar de 5% mest extrema värdena).
6. Nu kommer det intressanta och det är att vi nu kan använda denna t - fördelning för att konstruera ett 95% CI som är baserat på den skattade SE. Precis som för normalfördelningen gäller det att ta reda på arean med 95% sannolikhet.
7. Nu när vi har etablerat denna fördelning av t så kan vi nästa gång vi tar ett stickprov beräkna 95% CI genom att gå in i fördelningen och se vilket t - värde som motsvarar en viss sannolikhetsyta och sedan multiplicera t_{krit} med den skattade SE.
   $\mu = \bar{x} \pm SE * t_{krit, 0.05}$

En liten komplicerande faktor är att vi måste skapa en ny t - fördelning för varje storlek av stickprov. Sådana serier av fördelningar finns i tabeller. När n blir stort närmar sig t - fördelningen snabbt normalfördelningen och för stickprov över ca 30 observationer kan man använda d - värden för beräkning av CI.

Det vi har sett här är principen för hur statistiska tester fungerar. Vi beräknar en storhet, t.ex. t, som är ett exempel på en s.k. statistika. Genom att konstruera tabeller som ovan får vi fördelningen av statistikan för olika storlekar av stickprov. När vi behöver beräkna ett CI, beräknar vi statistikan och går in i en tabell för att konstruera ett intervall runt vårt skattade

medelvärde som förväntas innehålla det sanna medelvärdet med en viss specificerad sannolikhet.

Frihetsgrader

Det är nu dags att introducera nästa begrepp, nämligen frihetsgrader eller df (eng. degrees of freedom). Vi såg att fördelningen av en test - statistika berodde på stickprovets storlek. Man säger att stickprovet har ett visst antal frihetsgrader. I fallet med CI baserat på t - fördelningen har t - värdet n - 1 df. Antalet frihetsgrader beror på stickprovets storlek men är alltid färre om inte hela populationen dragits. Att riktigt förstå intuitivt vad frihetsgrader är och på vilket sätt test - statistikan beror på dessa är lite svårt att greppa.

Frihetsgrader anger hur många oberoende komponenter som ingår i teststatistikan. När vi räknar ut t för ett stickprov på låt oss säga 5 observationer måste vi först skatta medelvärdet. Där förvinner en frihetsgrad i materialet. Vet vi medelvärdet och 4 av observationerna då är även den femte bestämd. Eftersom medelvärdet används för att räkna ut s^2 har även s^2 4 df, så även t. Vet vi t - värdet och 4 observationer kan den femte inte variera.

Vi kan möjligen intuitivt förstå att vi underskattar variationen i en population när vi räknar ut variansen som avvikelserna mot det skattade medelvärdet. I själva verket blir de summerade avvikelserna som lägst när vi använder avvikelsen mot det skattade medelvärdet, vilket beror på att medelvärdet är baserat på just de data som vi ska minska med medelvärdet. En mer teknisk förklaring till frihetsgrader är att skattningen av stickprovets variation måste göras oberoende av stickprovets storlek genom att dela den underskattade summan av avvikelser med en kvantitet som kallas för frihetsgrader. Generellt får man ut hur många frihetsgrader man har för en statistika eller parameter genom att ta det antal oberoende observationer som den är baserad på och minska med antalet skattade parametrar som har använts i beräkningen. Antalet frihetsgrader är särkilt intressant då Typ II felet (se nedan) minskar med deras antal.

Hypotes testning

Test av ett stickprov mot ett sant medelvärde

Vi kan nu skatta ett medelvärde och uttrycka dess precision. Vi ska nu gå över till en intressant utbyggnad av detta. Låt oss återigen återvända till våra blåmusslor (Fig 2). Vi har deras skattade medelvärden och 95% CI. En intressant fråga kan vara om västkustpopulationen är större än 5 cm vilket kanske indikerar att det är dags att skörda musslorna. Vi ser ju inte de verkliga populationerna utan vi måste dra en slutsats baserad på de stickprov vi har. Detta kallas statistisk inferens.

Det första vi gör är att vi ställer upp en statistisk noll - hypotes (H_0). En statistisk H_0 definierar en frekvens - fördelning eller en parameter från den, t.ex. ett medelvärde. Ett exempel på en enkel H_0 är att det sanna medelvärdet i västkustpopulationen är 5 cm. H_0: $\mu = 5$. Vi testar alltså vårt skattade medelvärde som har ett visst fel, mot ett sant medelvärde utan fel. En väg är helt enkelt att undersöka om det 95% CI av det skattade medelvärdet inkluderar 5 cm eller ej. I sådant fall behåller vi H_0, alternativet är att vi förkastar H_0 och $\mu \neq 5$ (med 95% sannolikhet).

En - och tvåsidiga tester

Om H_0 förkastas, d.v.s. $\mu \neq 5$, finns två möjligheter; den är större eller mindre än 5. Detta är ett tvåsidigt test. Om H_0 sätts som $\mu \geq 5$ får man ett ensidig test. Förkastas H_0 måste $\mu < 5$. Beroende på om testet är en - eller tvåsidigt måste man hålla reda på att det kritiska värdet för test - statistikan man använder verkligen motsvarar den önskade sannolikhetsytan t.ex. 95%.

Test av ett stickprov mot ett annat stickprov

Nu gör vi det lite svårare och frågar oss om det finns någon skillnad mellan de två stickproven, Östersjön mot Västerhavet (Fig. 2). Vi ser att fördelningen av längder överlappar en del. Men även om det sanna medelvärdet är detsamma i populationerna så finns det en viss sannolikhet att man kan få två stickprov som skiljer sig åt. Vår H_0 blir nu att $\mu_1 = \mu_2$. Om vi som i Fig. 6 drar isär de två fördelningarna av medelvärden så kommer man till en gräns när de överlappar så lite att man bedömer det som osannolikt att de två stickproven kommer från populationer med samma sanna medelvärde.

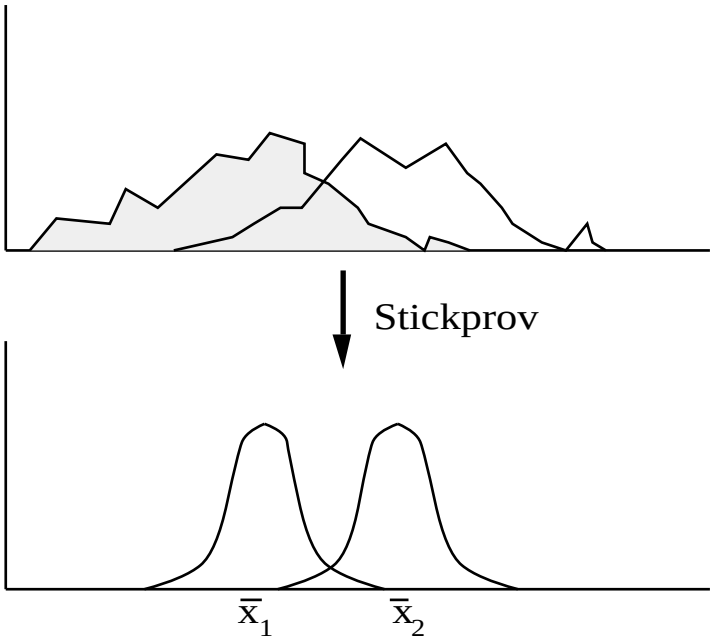

Fig. 6. Frekvensfördelningar av två schematiska populationer och samt fördelningar av medelvärden av stickprov dragna ur dessa båda populationer.

Analogt med beräkningen av 95% CI kan man nu beräkna hur stor skillnaden mellan två stickprov ($\bar{x}_2 - \bar{x}_1$) får vara för att man ska behålla H_0 med 95% sannolikhet, d.v.s. att de skattade medelvärdena är dragna ur samma population.

$\bar{x}_2 - \bar{x}_1 = SE * t_{0.05, (n-1)}$

I praktiken räknar vi ut t - värdet för en funnen skillnad mellan två stickprov enligt:

$$t_s = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{(s^2_1 + s^2_2)/n}}$$

där $\sqrt{(s^2_1 + s^2_2)/n}$ är SE för skillnaden mellan båda stickproven. Är nu t_s mindre än $t_{0.05}$ med 2 * (n - 1) antal df ligger ju uppenbarligen medelvärdena av de två stickproven så nära varandra att vi kan inte förkasta H_0. I vårt exempel med de båda musselpopulationerna är $t_s = 12.7$ vilket är ett mycket extremt t - värde med en sannolikhet under H_0 mindre än 0.05, så vi förkastar H_0 att stickproven kommer från en population med samma sanna medelvärde. Lägg märke till att antalet df är 2 * n - 2 vilket är det sammanlagda antalet observationer minus 1 för varje skattat medelvärde som använts för att beräkna SE.

Fel vid statistiska analyser

Statistisk hypotesprövning ju baserad på sannolikheter. Orsaken är naturligtvis den att vi försöker bilda oss en uppfattning om populationer i naturen baserad på ett ofta litet stickprov. Även om könskvoten är 1:1 hos t.ex. en fågelart så finns det en viss sannolikhet att vi ska dra ett icke - representativt stickprov t.ex. 2 honor och 8 hanar. Vad statistiken handlar om är ju att tala om hur sannolikt det är att ett sådant stickprov kommer från en population med jämn könskvot. När vi säger att noll - hypotesen om att väst - och ostkustmusslor har samma medellängd kan förkastas med 95% sannolikhet innebär det naturligtvis att det är 5% risk att vi faktiskt gör fel. Föjande fyra möjligheter finns vid statistiska test (Fig. 7).

	H_0 är sann	H_0 är falsk
Förkasta H_0	Typ I - fel α	Korrekt 1-β
Behålla H_0	Korrekt 1-α	Typ II - fel β

Fig. 7. Möjliga utfall och konsekvenser av statistiskt test.

I Figur 8 representerar den skuggade ytan det Typ I - fel som vi riskerar att göra. Vad påverkar då Typ-I felet? Det står under vår egen kontroll genom att vi kan välja ett godtyckligt α det vill säga vilken risk för Typ I fel vi är beredda att acceptera. Som tidigare nämnts brukar man enligt konvention sätta $\alpha = 0.05$, men i många fall där det är viktigt att inte begå ett Typ 1 fel kan α sättas lägre. Nivån på α beror naturligtvis på kostnaden att begå ett felslut. I vetenskapliga arbeten brukar man vara noga med att sätta ut vilken α nivå man bestämt sig för.

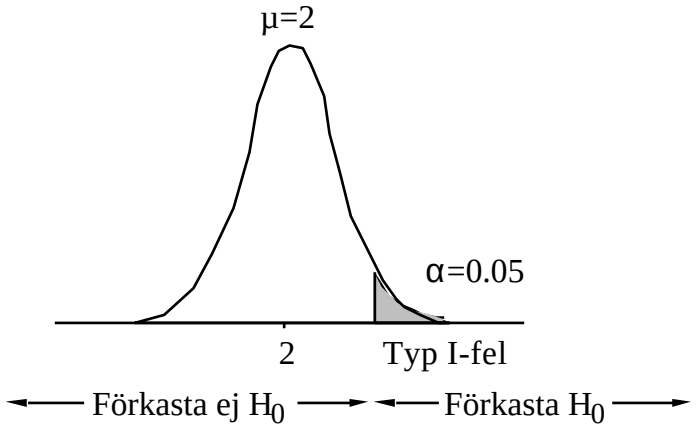

Fig. 8. Utfall av statistiskt test.

Minst lika viktigt är Typ II - felet, men här är inte forskare lika duktiga. Tyvärr finns en stor okunnighet kring Typ II - fel som ofta leder till ologiska beslut och som faktiskt kan kosta mer än ett Typ I - fel. För att göra tydligt att Typ II fel är allvarliga saker ska vi ta några exempel.

1. Vi föreslår en model för att förklara något
2. Modellen genererar en hypotes
3. Vi formulerar en logisk noll hypotes
4. Vi ställer upp en statistisk noll hypotes
5. Vi testar denna mot data

Vi ser nu på vilka fel vi kan göra och vilka konsekvenserna blir.

A. Ett Typ I fel innebär att vi förkastar H_0 fast den är sann. Att H_0 är sann innebär ju att vår modell är falsk (se Fig. 1). Det vill säga vi tror att modellen är sann när den i själva verket är falsk. Men detta behöver inte vara så allvarligt eftersom vi säkert kommer att fortsätta arbeta med modellen och testa andra förutsägelser. Troligtvis kommer vi förr eller senare på att den är falsk.

B. Gör vi ett Typ II fel däremot innebär det ju att vi behåller H_0 fast den är falsk. Det innebär att vi förkastar modellen som ju är sann och vi kanske slutar att arbeta med denna modell för att försöka finna någon annan. Det kan då dröja länge innan vi på nytt ger oss i kast med denna modell eftersom vi tror att den är felaktig.

Vad ska man då göra? Det viktigaste är naturligtvis att försöka ta reda på hur stor sannolikhet det är att man gör dessa fel. Hur vi kontrollerar Typ I felet har vi redan berört. Typ II felet är betydligt svårare att handskas med. Om H_0 faktiskt är falsk så är ju något annat sant. Problemet är bara att det kan finnas många alternativa hypoteser. Låt oss ta ett konkret exempel.

1. Vi utgår från ett exempel där man vill undersöka om halten PCB i fisk överstiger gränsvärdet 2 mg kg^{-1}. Ett stickprov om 30 fiskar tas och analyseras och ett skattat medelvärde och spridningmått beräknas. Vi vill nu testa om det funna medelvärdet är större eller mindre än det tillåtna gränsvärdet.
2. Noll - hypotesen är att det sanna värdet i målpopulationen är 2 eller mindre, d.v.s. ett ensidigt test.
3. Vi kan nu skilja på några olika utfall.
4. Titta på Figur 9 och gör antagandet att den sanna populationen av fisk innehåller 2 mg kg^{-1} i genomsnitt. Kurvan representerar sannolikheten för alla tänkbara skattade medelvärden från ett stickprov om 30 fiskar från denna population

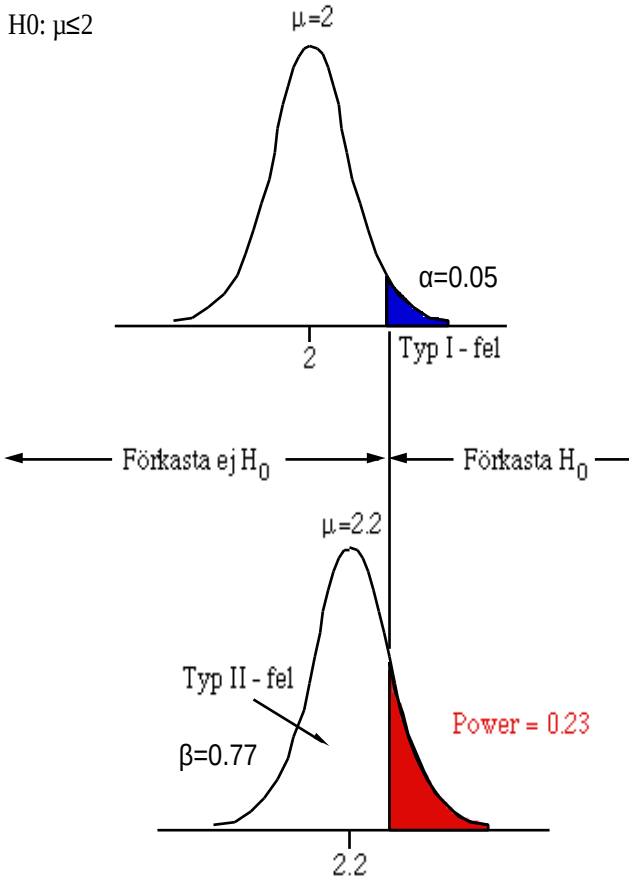

Fig. 9. Typ 1 och typ 2 fel vid test av H_0: $\mu \leq 2.0$

5. Vi bestämmer hos för att risknivån (α) för ett Typ I - fel är 0.05. Det vill säga i 5% av fallen kommer ett stickprov från denna population att ge ett medelvärde över det tillåtna gränsvärdet, och vi kommer att förkasta H_0 när den faktiskt är sann, d.v.s. vi gör ett Typ I - fel.

6. Låt oss nu istället antaga att H_0 faktiskt är fel. Populationens gifthalt är i genomsnitt 2.2 mg kg^{-1}. Vi står igen med 30 fiskar från denna population och testar ånyo vår H_0: $\mu \leq 2$ med $\alpha = 0.05$. Fortfarande finns en stor sannolikhet att vi får ett stickprov som ligger lägre än det kritiska värdet vilket gör att vi accepterar H_0 felaktigt d.v.s. vi gör ett Typ II - fel. Sannolikheten att acceptera H_0 när den faktiskt är falsk betecknas med β och kan i detta fall beräknas som 0.77.
7. Titta nu på den skuggade delen av alla tänkbara fiskstickprov. I denna sektor kommer vi att överskrida det kritiska värdet som bestäms av α och vi gör ett korrekt beslut när vi förkastar H_0. Sannolikheten (ytan) för detta beslut kallas för statistisk styrka ("power") och beräknas som $1 - \beta$. Detta är ett utomordentligt viktigt begrepp vid alla vetenskapliga analyser.

Statistisk styrka

Vad innebär statistisk styrka?

Man bör givetvis eftersträva så stor statistisk styrka som möjligt då det minskar risken för ett Typ II - fel. Av någon, kanske psykologisk, orsak bryr de flesta sig lite om statistisk styrka. Delvis beror det på att det får ett litet och opedagogiskt utrymme i läroböcker; dessutom är det få datorprogram som innehåller power - beräkningar. En psykologisk orsak kan vara följande bristande argumentering: Om jag testar för en eventuell skillnad, t.ex. födopreferens, artantal, etc. så är det bara intressant om man får en skillnad, och får man en skillnad vill man verkligen vara säker på att den är statistiskt signifikant. Finner man däremot ingen skillnad så känns det som man inte lovar något när man rapporterar att det inte var någon skillnad. Detta innebär att man inte bryr sig så mycket om den möjligheten att man inte upptäckte en skillnad fast det faktiskt fanns en skillnad. Exempel på när detta är särskilt allvarligt är t.ex. vid undersökning av utsläppseffekter och test av bieffekter hos mediciner. Kan du ge andra exempel?

Vad påverkar statistisk styrka?

1. Den statistiska styrkan står i ett omvänt förhållande till den α - nivå vi beslutar oss för att använda (Fig. 10)
2. Den statistiska styrkan ökar med minskad varians hos de undersökta populationerna eftersom minskad varians leder till säkrare skattning av medelvärdena (lägre SE) (Fig. 11). I allmänhet kan SE minskas enbart genom att stickprovets storlek ökas. Provtagnings - och analysmetodiken ger också upphov till viss variation, och ibland kan denna reduceras betydligt genom att förfina eller förändra metodiken.

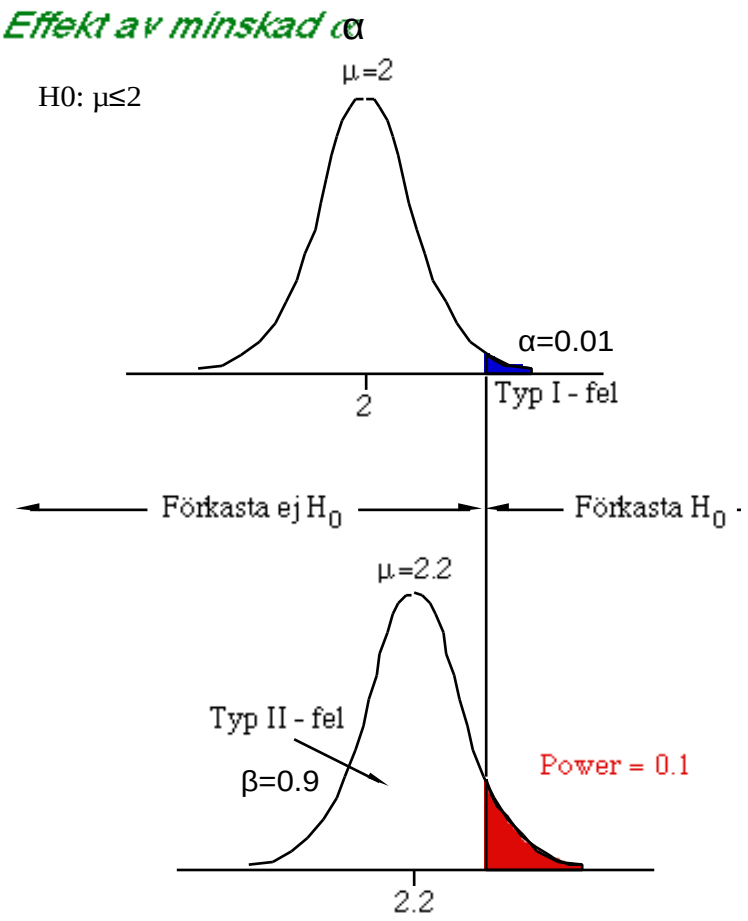

Fig. 10. Effekt av minskat typ 1 fel.

3. Den statistiska styrkan beror kraftigt på vilken skillnad mellan medelvärden (effektstorlek). man önskar upptäcka (Fig. 12). Finner man att den statistiska styrkan för att upptäcka en viss skillnad är låg får man sänka ambitionsnivån och acceptera att studien endast kan användas för att upptäcka en större skillnad. Alternativt kan man öka stickprovets storlek (n) eller eventuellt öka α (se ovan). Ibland får man dock konstatera att studien är meningslös utan ökade resurser.

Angivande av statistisk styrka

Det finns inte någon vedertagen nivå på statistisk styrka (eller Typ II - fel) på samma sätt som för α. Många menar dock att den statistiska styrkan bör vara större än 0.8 (Typ II - fel = 0.2). Det finns egentligen inget skäl varför man inte bör ha samma krav på Typ II - fel som för Typ I - felet, och enligt konvention borde då båda sättas till 0.05 (statistisk styrka = 0.95). Idealt bör nivåerna sättas så att kostnaden att göra respektive fel är lika stor.

Oavsett nivå bör man alltid rapportera statistisk styrka för en given effektstorlek när ett test inte lyckats förkasta H_0.

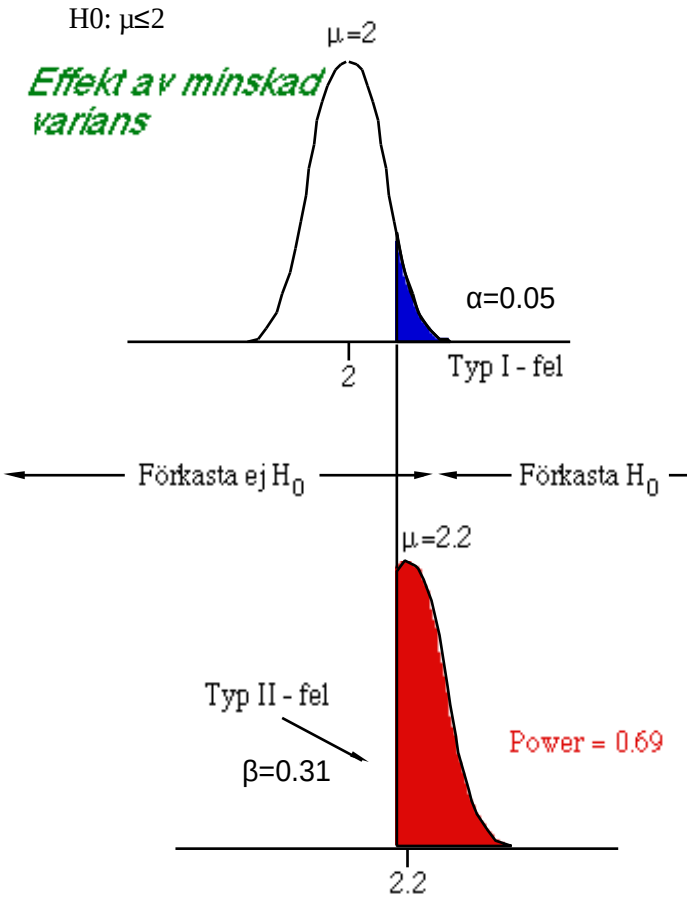

Fig. 11. Effekt av minskad varians.

Viktiga antaganden för parametrisk statistik

Vi har nu använt t - test för att pröva noll - hypoteser. Det kan nu vara bra att stanna upp lite och se vad som krävs för att vi ska kunna använda sådana statistiska metoder. De flesta antaganden som gäller t - test gäller för de flesta parametriska metoder. De huvudsakliga kraven är:

1. Normalfördelade medelvärden. Som vi redan berört är detta oftast inget problem eftersom centralvärdesteoremet visar att oavsett urspungsfördelning så tenderar stickprovens medelvärden mot en normalfördelning. Normalfördelade data (residualer) är också ett antagande. Simuleringsövningar visar dock att avvikelse från normanfördelade data oftast inte påverkar slutsatserna vid parametriska test.
2. Homogena varianser. Detta är ett viktigt antagande. Om varianserna hos de skattade medelvärdena som vi vill testa är mycket olika kommer vi att förkasta H_0 oftare än vad α anger. Vid parametriska tester bör man därför alltid testa om varianserna är tillräckligt homogena. Ett bra test är Cochrans C. Är varianserna signifikant olika kan ibland särskilda datatransformationer ge homogenitet, t.ex. genom att logaritmera sina data. Ytterligare en möjlighet är att sänka α i alla ingående test inklusive testet för homogenitet av varianser.

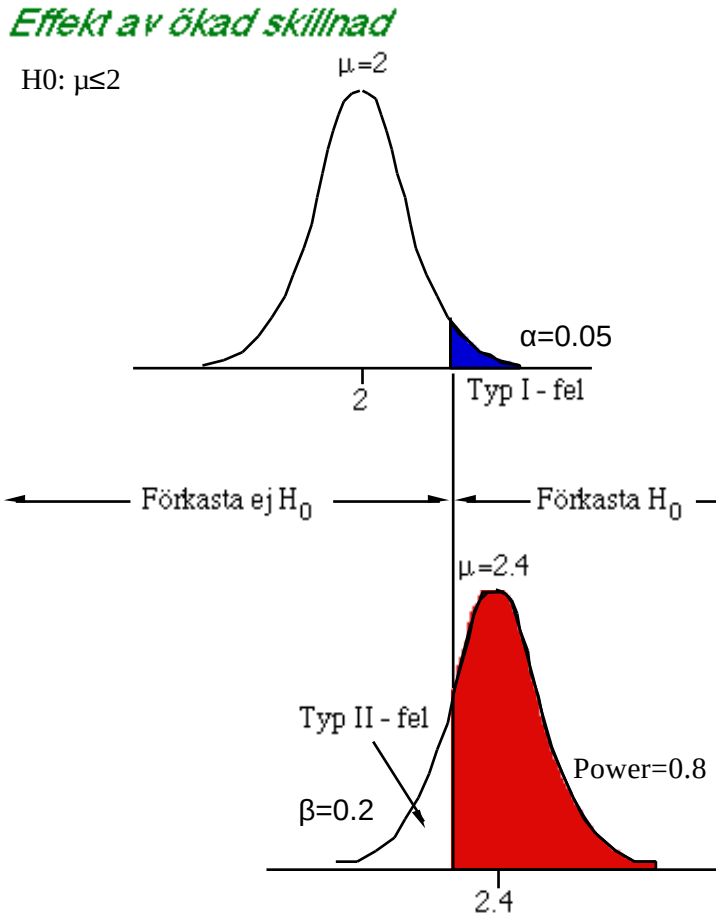

Fig. 12. Effekt av ökad skillnad mellan sanna medelvärden.

3. Oberoende stickprov. Ett absolut krav är att de skattade medelvärden är uppbyggda av oberoende observationer (mätvärden). Brott mot detta antagande innebär troligen att variansen blir felskattad, och man har färre verkliga frihetsgrader än antalet observationer anger, vilket ger felaktiga nivåer av Typ I - och Typ II - fel. Statistisk inferens, d.v.s. att dra slutsatser om den verkliga populationen från stickproven, blir här suspekt.