

Hur skriver man statistikavsnittet i en ansökan?

Val av metod och stickprovdimensionering



Datum/föreläsare	Titel
15/9	Introduktion till logistisk regression
5/10	Datumhantering i analyser, Robert
25/10	Estimander, confounders, colliders och nyckelbegrepp inom Kausal inferens, Anna
15/11	Tips och tricks i Stata, Christel
6/12	Bootstrapping, vad är det?
18/1	Interaktion, Gabriel

Upplägg

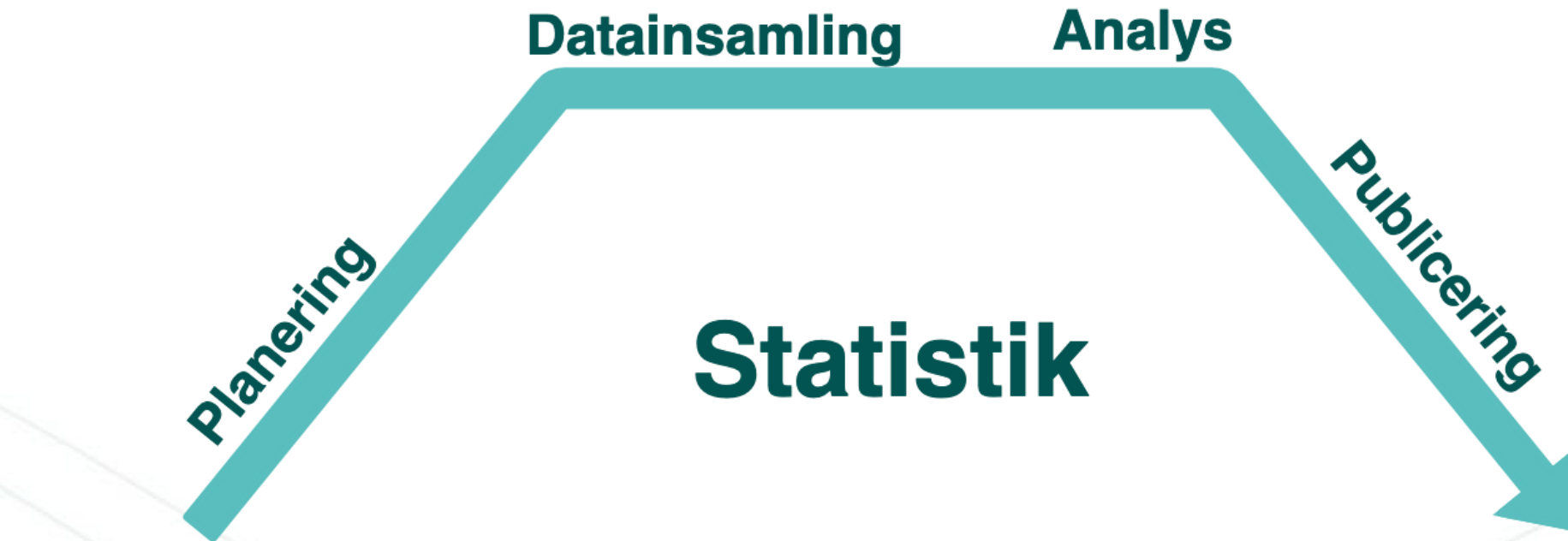
- Teori
- Frågor

Efter detta pass hoppas jag att:

- ni ska veta vad man ska tänka på
- vilka verktyg som finns
- vilket stöd som finns att få



Varför är statistik viktigt?



Frågeställning

Statistik i planeringen?

- Riskerar att dra felaktiga slutsatser
- Riskerar att inte kunna dra några slutsatser
- Ej publicerbart
- Missade anslag



Vetenskapsrådet

"Valet av statistisk analys (hypotesprövning) och/eller modellering, kvalitetssäkring eller dylikt ska redovisas tydligt. Kriterier för eventuella interimsanalyser skall anges. **Det är inte tillräckligt att ange allmänna begrepp såsom parametrisk, ickeparametrisk metod, deskriptiv statistik eller standardmetoder för analys.** Observera, att normalfördelningen kan vara en möjlig sannolikhetsfördelning för symmetriska kvantitativa data men aldrig för data från bedömningar i frågeformulär. Hur hanteras beroende data från upprepade mätningar, till exempel inom grupp eller mellan grupper? Vilka statistiska metoder planeras för komplexa data från frågeformulär och andra subjektiva bedömningar, som genererar ordnade kategoridata? Hur hanteras longitudinella data och bortfall?"

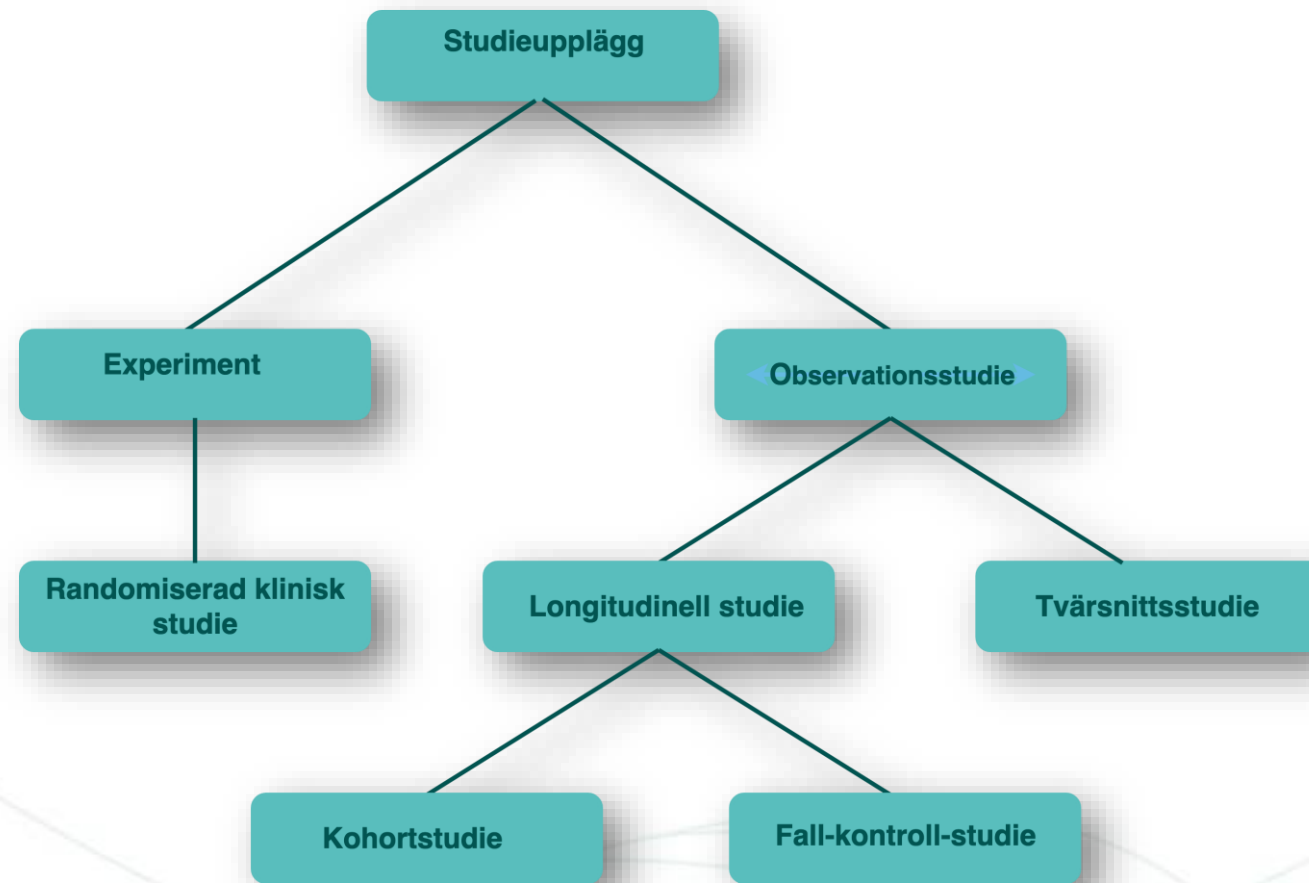


Val av statistisk metod

- Vilken studiedesign?
- Vilket är mitt utfall ?
- Hur kommer mitt data se ut?
- Val av statistisk metod



Design



Utfall

- Skillnader
- Risker
- Antalet händelser
- Tid till händelse



Datakvalitet

- Vilket bortfall bör vi räkna med?
- Mätfel
 - Slumpmässiga
 - Systematiska
- Hur många patienter kan vi få in?



Primary Outcome

- Särskilt viktigt i kliniska prövningar
- Det vi är mest intresserade av att undersöka
- Studien dimensioneras för dessa utfall
- Ska hållas nere i antal
- Kan behöva överväga justering för multipla jämförelser
- Gärna kontinuerlig (om möjligt)



Analysplanen

- Strukturera syftet i frågeställningar. Vad ska studeras och på vilka?
- Studiedesign
- Variabler och mätning
- Dimensionering
- Statistisk analysplan
- **Replikerbar**



Statistiska metoder

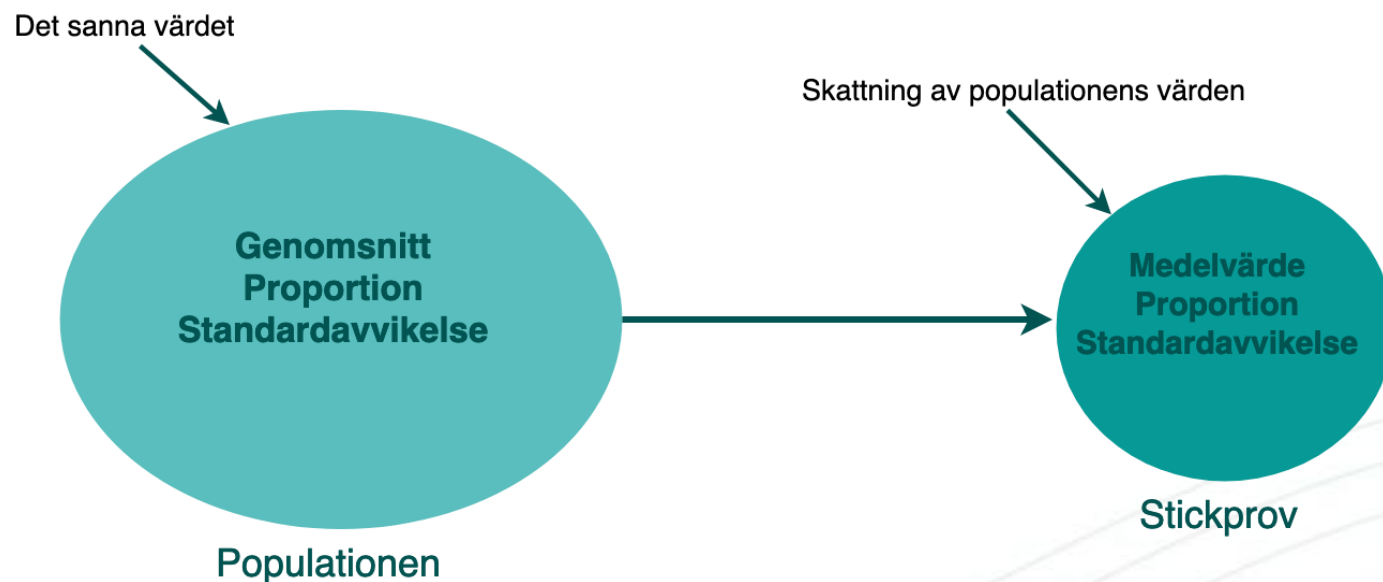
Det är lätt att köra fast i gamla hjulspår när det gäller val av statistisk metod.

Metod utifrån frågeställning och data.



Repetition och terminologi

I de flesta situationer vi ställs inför har vi inte tillgång till hela populationen vi vill undersöka. Vi tar då ett stickprov från den population vi önskar undersöka.



Lägesmått

Vi skattar exempelvis det medelvärde värdet för populationen med medelvärde i stickprovet.

Har vi en skev fördelning av mätvärden exempelvis operationstid så är det bättre att använda medianen som lägesmått.



Spridningsmått

Standardavvikelsen är ett mått på hur stor spridningen mellan observationerna i stickprovet är.

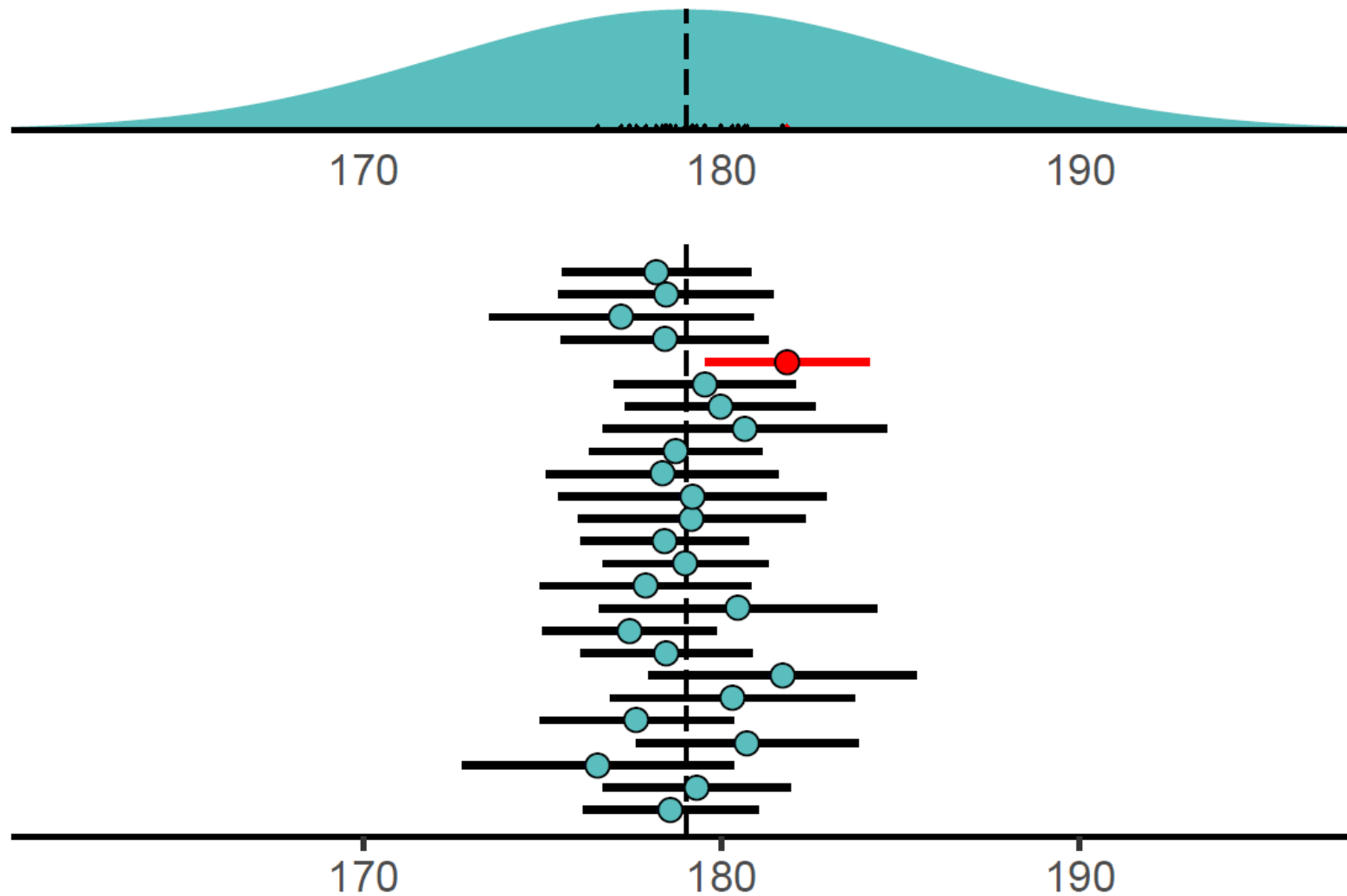
Standard error är ett mått på osäkerheten i en punktskattning. Anger precisionen i vår skattning.

Med hjälp av standard error kan vi beräkna bland annat konfidensintervall.

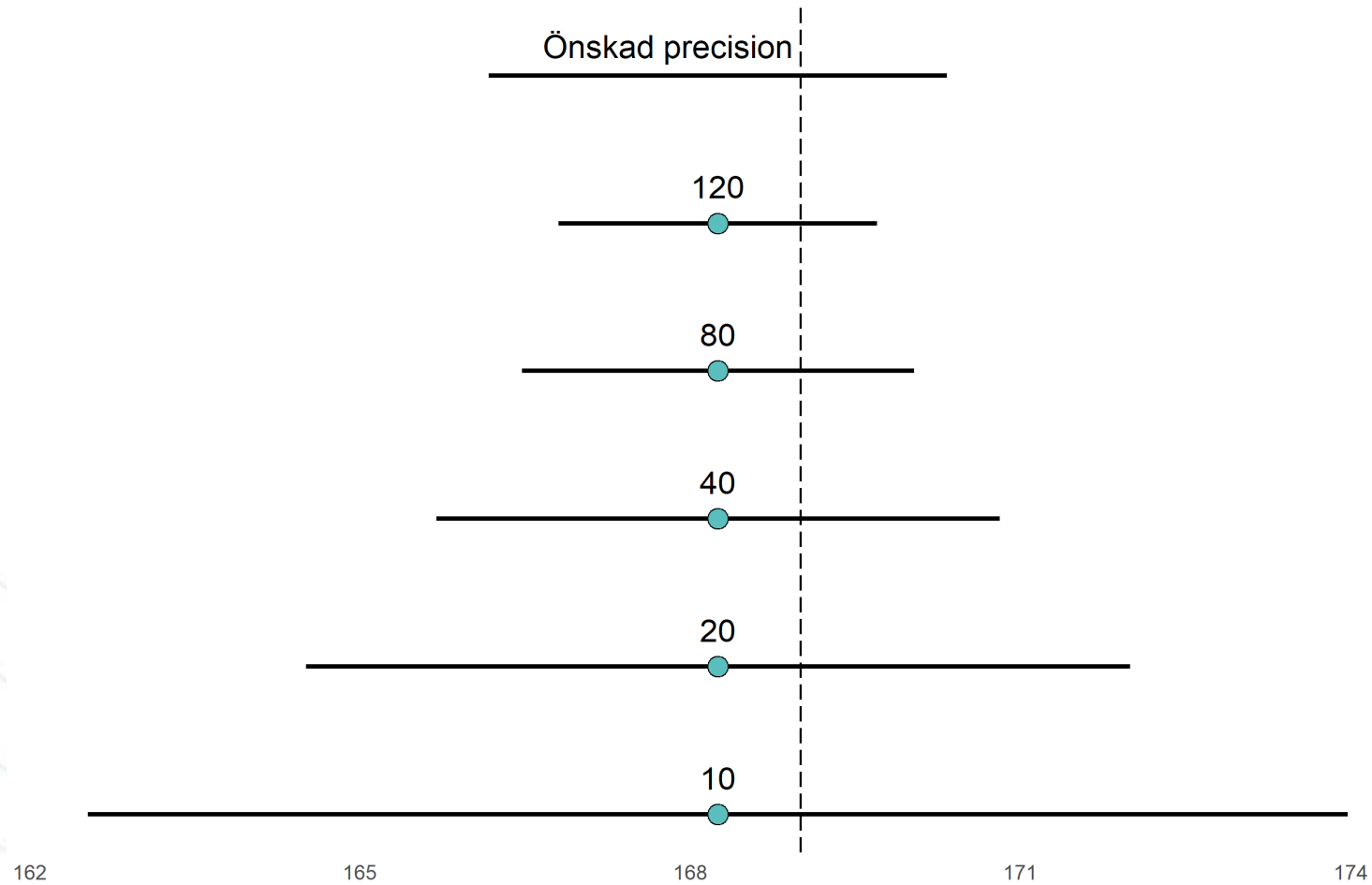
Om vi upprepar ett försök tillräckligt många gånger kommer ett 95% konfidensintervall att täcka det sanna medelvärdet 95% av gångerna.



Konfidenzintervall



Precision



Varför är power viktigt?

- För lite data
 - Missar intressanta resultat
 - Kostnad
- För mycket data
 - Kostnad
 - Ej etiskt
- **OBS!** Endast relevant före en studie är genomförd!



Vägen dit

- Vilken typ studiedesign?
- Vilket är mitt utfall ?
- Hur kommer mitt data se ut?
- Val av statistisk metod
 - **Stickprovdimensionering**



Vad är power?

Bakgrund

Konceptet med power bygger på hypotesprövning. En nollhypotes som ska vara falsifierbar och ha en mothypotes.

Nollhypotesen: Det finns ingen skillnad mellan grupperna

$$H_0: \mu_1 - \mu_2 = 0$$

Mothypotesen: Det finns en skillnad mellan grupperna

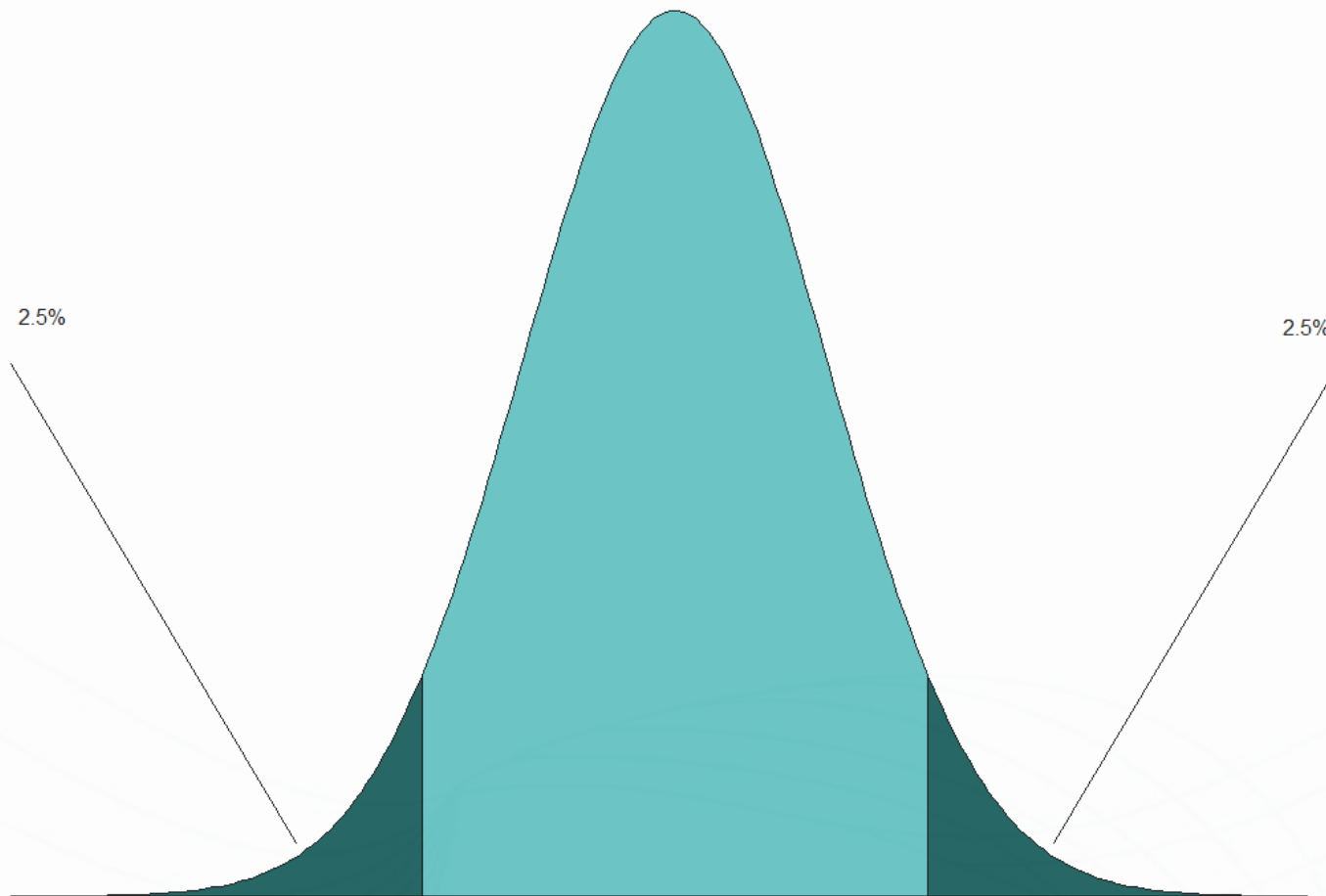
$$H_1: \mu_1 - \mu_2 \neq 0$$

Vad är power?

Sannolikheten att korrekt förkasta en falsk nollhypotes (när den alternativa hypotesen är sann)



Hypotesprövning



Power?

Verkligheten	Slutsats utifrån test	
	Ej förkasta nollhypotesen	Förkasta nollhypotesen
Nollhypotesen sann	Korrekt	Typ I-fel
Nollhypotesen falsk	Typ II-fel	Korrekt

Power = 1-sannolikheten för typ II-fel



Typ-I



Typ-II

Typ-I fel

Typ-I felets storlek är lätt att kontrollera genom signifikansnivån.

En signifikansnivå på 5% innebär att vid tillräckligt många upprepningar av ett försök kommer att upptäcka en signifikant skillnad

som i verkligheten inte existerar i 5% av fallen. Dvs vi förkastar nollhypotesen.



Exempel typ-I

Antag att vi utför ett experiment där vi undersöker om tuggummituggande är associerat med matstrupscancer. Från vårt experiment drar vi slutsatsen att tuggummituggande ger en högre risk för matstrupscancer. Om det inte stämmer med verkligheten har vi gjort oss skyldiga till ett typ-I fel.

Verkligheten	Slutsats utifrån test	
	Ej förkasta nollhypotesen	Förkasta nollhypotesen
Nollhypotesen sann	Korrekt	Typ I-fel
Nollhypotesen falsk	Typ II-fel	Korrekt



Antag att vi istället drar slutsatsen att tuggummituggande inte ger en högre risk för matstrupscancer. Om det stämmer med verkligheten har vi inte gjort oss skyldiga till något fel.

Verkligheten	Slutsats utifrån test	
	Ej förkasta nollhypotesen	Förkasta nollhypotesen
Nollhypotesen sann	Korrekt	Typ I-fel
Nollhypotesen falsk	Typ II-fel	Korrekt

Tyvärr är det betydligt svårare att kontrollera typ II-felet. Typ II-felet beror på flera variabler:

- Storlek på den skillnad vi önskar mäta
- Spridningen i materialet (standardavvikelsen)
- Signifikansnivån
- Antalet patienter

Oftast är det mer intressant att beräkna hur många patienter som behövs för att uppnå en viss power.



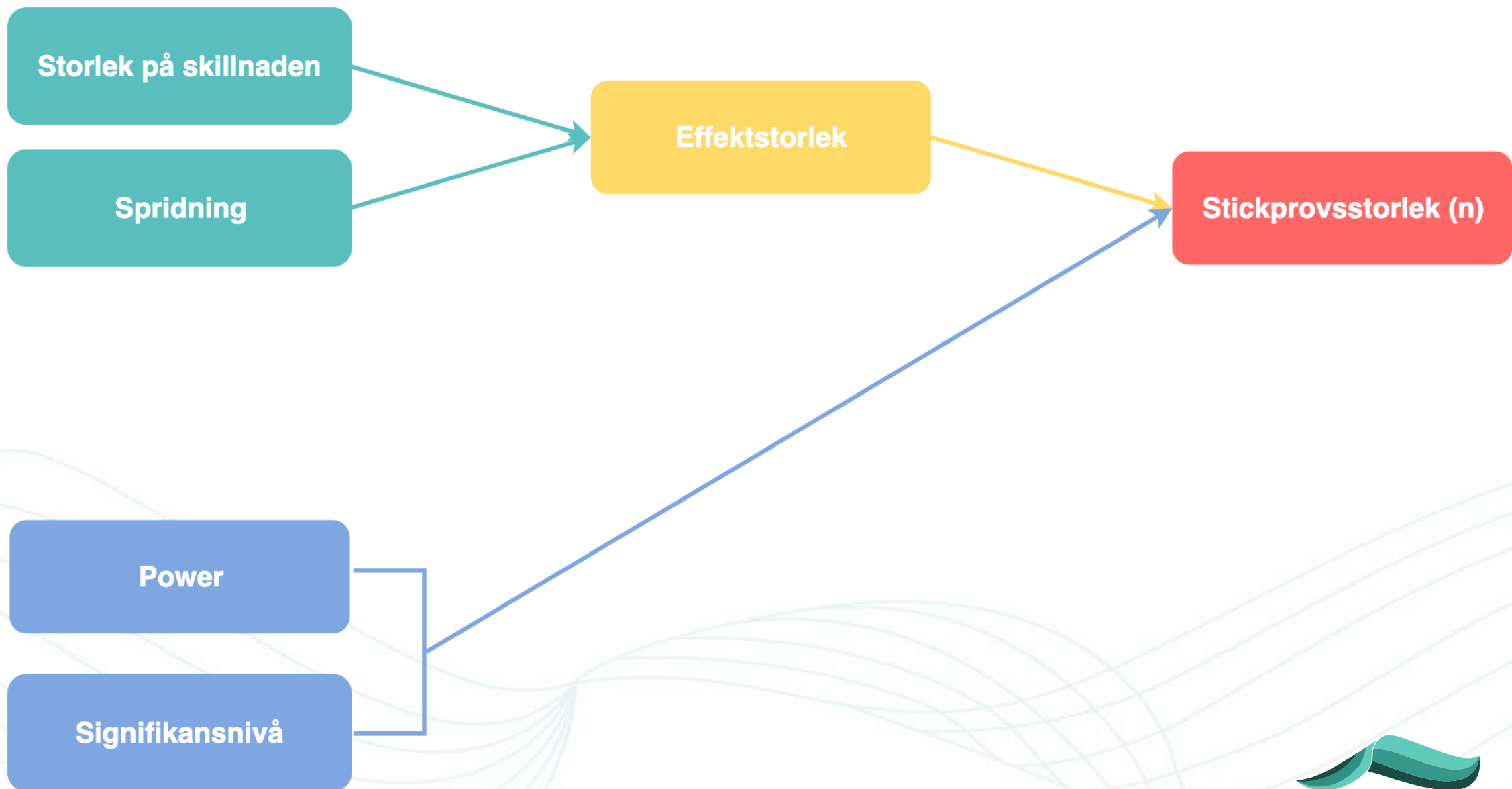
Exempel typ-II fel

Antag att vi utför ett experiment där vi undersöker om rökning är associerat med matstrupscancer. Från vårt experiment drar vi slutsatsen att rökning inte ger en högre risk för matstrupscancer. Om det inte stämmer med verkligheten har vi gjort oss skyldiga till ett typ-II fel.

Verkligheten	Slutsats utifrån test	
	Ej förkasta nollhypotesen	Förkasta nollhypotesen
	Korrekt	Typ I-fel
Nollhypotesen sann	Korrekt	Typ I-fel
Nollhypotesen falsk	Typ II-fel	Korrekt

Antag att vi istället drar slutsatsen att rökning ger en högre risk för matstrupscancer. Om det stämmer med verkligheten har vi inte gjort oss skyldiga till något fel.

Verkligheten	Slutsats utifrån test	
	Ej förkasta nollhypotesen	Förkasta nollhypotesen
Nollhypotesen sann	Korrekt	Typ I-fel
Nollhypotesen falsk	Typ II-fel	Korrekt



Storlek på skillnaden

Definiera gränsen för intressanta skillnader!

Ofta har vi en tidigare studie att luta oss mot då vi uppskattar den förväntade skillnaden.

En fråga vi måste ställa oss är -

Hur stor måste skillnaden vara för att den ska vara kliniskt relevant?

Det kan vara så att vi är intresserade av att hitta mindre skillnader än vad som uppmäts i pilotstudien.



Spridningen

Om det finns en tidigare studie kan vi hämta standardavvikelsen från denna studie.

Finns ingen sådan får vi uppskatta standardavvikelsen. Ett alternativ är att arbeta med olika scenarion och undersöka hur stickprovsstorleken förändras med standardavvikelsen.



Effektstorlek

Beroende på vilken statistisk metod man valt så beräknas effektstorleken på olika sätt olika. Ett exempel är Cohen's d men det finns väldigt många olika mått.

Gemensamt är att de är beroende av storleken på skillnaden och på spridningen. Överskattas spridningen så underskattar vi inte antalet som behövs.



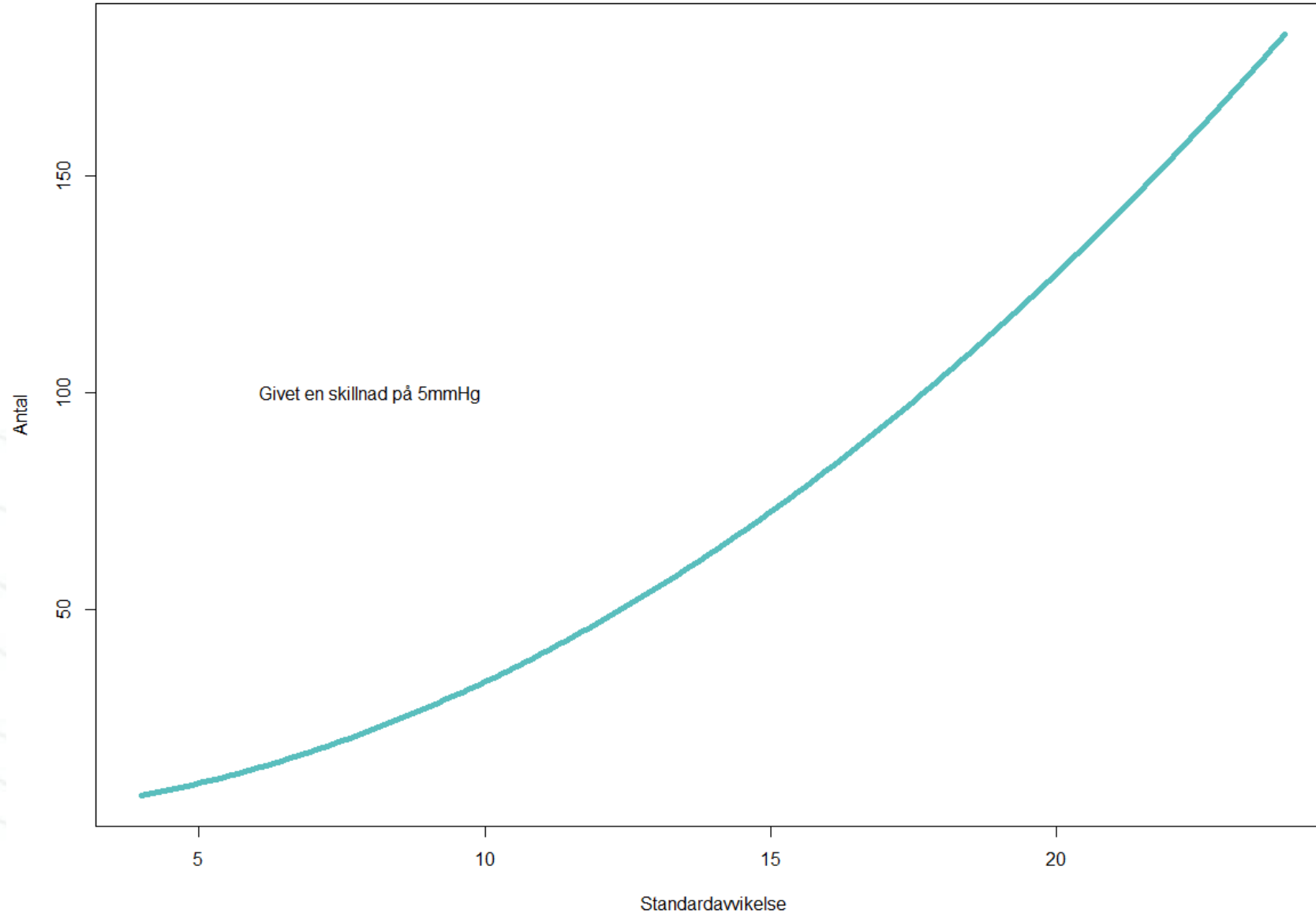
Antag att vi vill testa hur effektiv en behandling är på att sänka det diastoliskt blodtryck i en grupp patienter.

Är skillnaden mindre än 5mmHG mellan före och efter behandling bedöms skillnaden som irrelevant. Vi har dock ingen aning om hur stor spridningen är.

Vi beräknar hur många patienter som krävs för att nå upp till 80% power på 5% signifikansnivå.



Exempel



En varning!

Effect size	d
Small	0.20
Medium	0.50
Large	0.80



Övriga komponenter

Mothypotesen

Kan vara enkelsidig eller dubbelsidig. Vad vi väljer här beror på om vi

testar olikhet eller om vi redan på förhand vet vilken grupp som kommer ha högre mätvärden. Är vi osäkra ska vi **alltid** välja dubbelsidigt. För samma power kräver ett dubbelsidigt test fler patienter



Signifikansnivå

Sannolikheten för typ-I fel. Vanligtvis 0.05. Lägre signifikansnivå kräver fler patienter för att uppnå samma power.

Power

Vanligtvis dimensioneras stickprovet för att uppnå en power på 80% men 90% förekommer också.

Sammanfattning

Hur stor stickprovstorlek som behövs beror på flera saker

- Power
- Signifikansnivån
- Effect size
- Metod



- Power sätts oftast till 80% och stickprovet dimensioneras utifrån det.
- Stickprovet dimensioneras för ett fixt värde på signifikansnivå, en lägre signifikansnivå kräver en större stickprovsstorleken.
- Desto mindre skillnader vi vill kunna upptäcka, desto större behöver stickprovet vara.
- Vissa metoder ger högre power än andra.



Antag att vi vill testa hur effektiv en ny behandling är på att sänka det diastoliskt blodtryck. Vi planerar ge en grupp placebobehandling och en grupp aktiv behandling. Innan behandlingen är blodtrycket är 100mmHg i båda grupperna.

Vi vet från en tidigare studie att vi kan förvänta oss en standardavvikelse på 25mmHG. Nu vill vi veta hur många patienter vi behöver inkludera för att upptäcka en skillnad på 5mmHg.

Med andra ord om blodtrycket i gruppen som får behandling är 95mmHg eller mindre vill vi upptäcka det med en power på 80 %.



Vi vill testa:

Nollhypotesen: Ingen skillnad

Mothypotesen: Blodtrycket är lägre efter behandling



Vi är bara intresserad av fallet mindre än 95mmHg. Således är det ett ensidigt test.

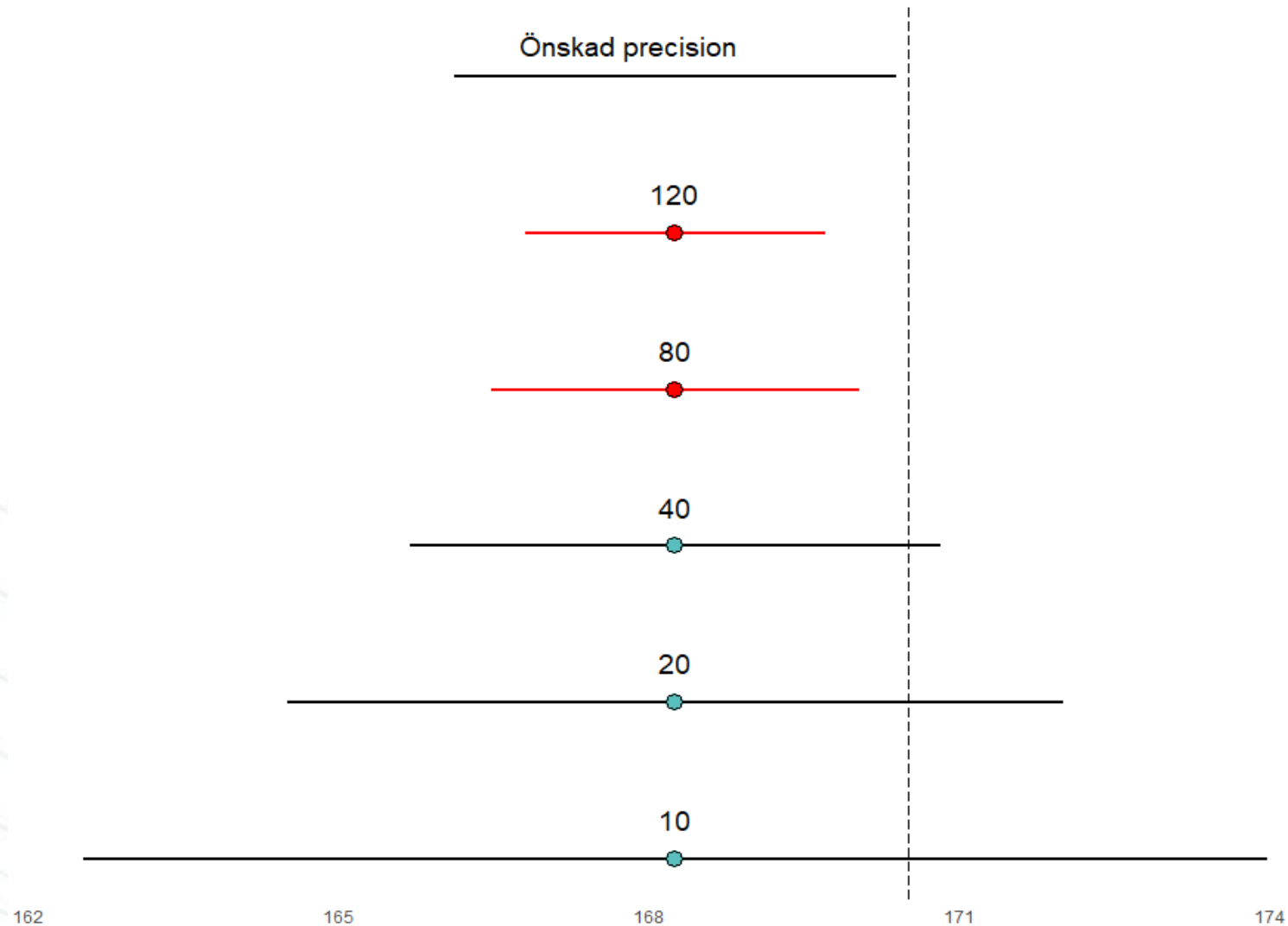
Om vi använder signifikansnivån 5% behövs 310 patienter i vardera grupp för uppnå en power på 80 %.



Om blodtrycket efter behandling är 90mmHG? I detta fall behöver vi endast 78 patienter i vardera grupp för uppnå en power på 80 %.



Ett annat sätt att beräkna stickprovsstorleken



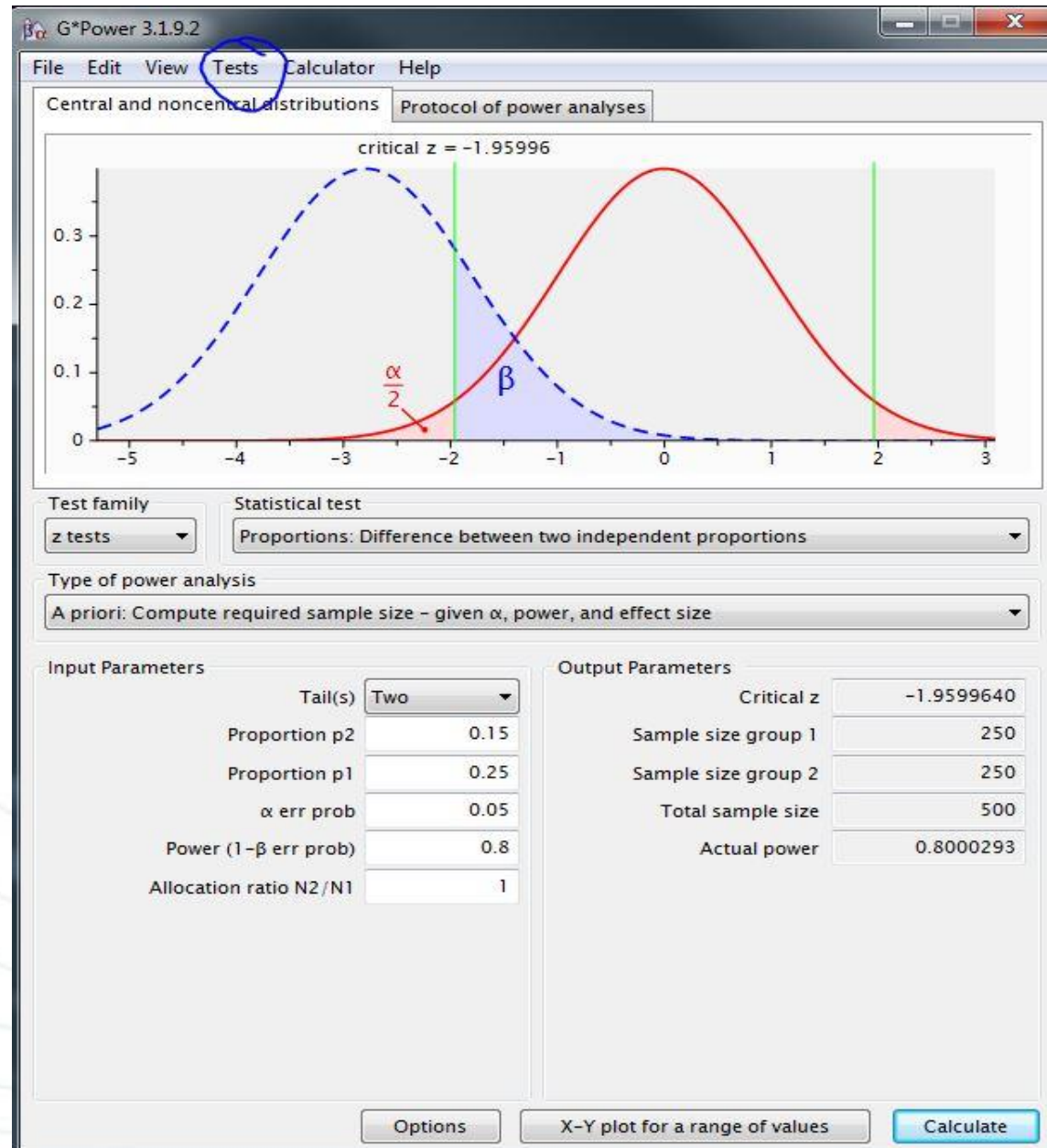
För enklare fall:

- G*Power
- <http://powerandsamplesize.com/>
- <https://www.sealedenvelope.com>

För svårare fall...



G*power



Calculate:

Sample Size

Sample Size, n_B

247

Power, $1 - \beta$

0,80

Type I error rate, α

5%

0,15

Group 'A' Proportion, p_A

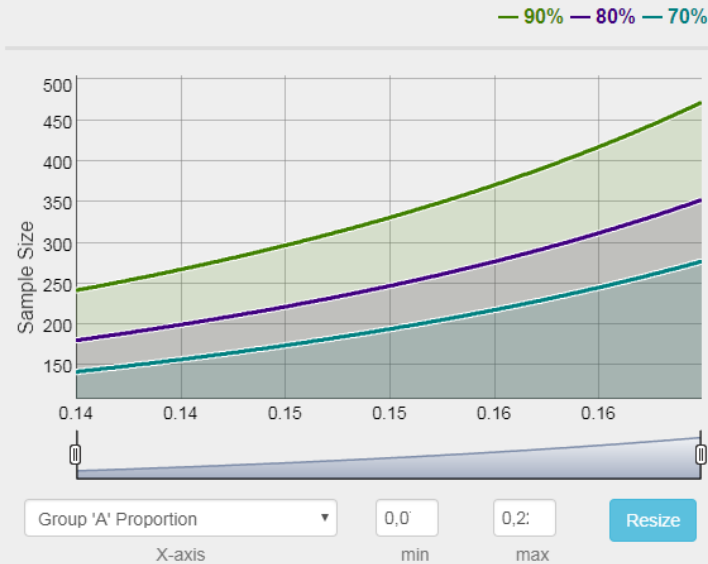
0,25

Group 'B' Proportion, p_B

1

Sampling Ratio, $\kappa = n_A/n_B$

Calculate



Documentation

Introduction

Usage

References

Validations

Test 1 Mean

Calculate Sample Size Needed to Compare 2 Proportions: 2-Sample, 2-Sided Equality

This calculator is useful for tests concerning whether the proportions in two groups are different. Suppose the two groups are 'A' and 'B', and we collect a sample from both groups -- i.e. we have two samples. We perform a two-sample test to determine whether the proportion in group A, p_A , is different from the proportion in group B, p_B . The hypotheses are

$$\begin{aligned} H_0 : p_A - p_B &= 0 \\ H_1 : p_A - p_B &\neq 0 \end{aligned}$$

where the ratio between the sample sizes of the two groups is

Sealedenvelope

Significance level (alpha)	5%
Power (1-beta)	80%
Percentage 'success' in control group	15 %
Percentage 'success' in experimental group	25 %
Calculate sample size	
Sample size required per group	248
Total sample size required	496
Adjustment for non-compliance/cross-over	
Percentage cross-over expected in control group	%
Percentage cross-over expected in experimental group	%

- Beskriv vilken metod ni valt och varför den passar.
- Baserat på vilken metod ni valt beskriv vilken information som ligger till grund för era beräkningar och vad ni kommit fram till.



Resultatet från behandlingen av det diastoliska blodtrycket kommer analyseras med ett parat t-test. Detta är lämpligt då vi mäter samma

individer före och efter behandling (matchade par) och variabeln vi mäter är kontinuerlig.



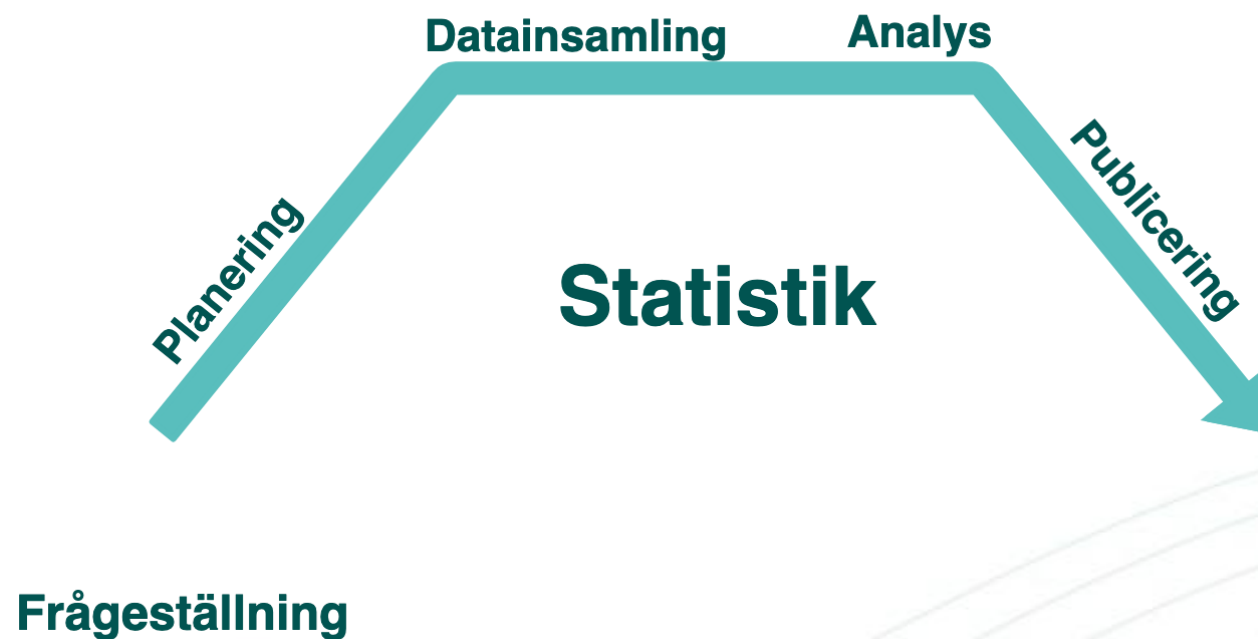
Stickprovsstorleken beräknades med ett dubbelsidigt parat t-test på en 5% signifikansnivå.

I en tidigare studie har vi hittat en genomsnittlig skillnad på 10mmHG mellan för och efter mätningar och en standardavvikelse på 15mmHG. Eftersom att vi är intresserade av att hitta mindre skillnader än 10mmHG så kommer vi räkna med en skillnad på 5mmHG (vilket vi bedömer vara den minsta skillnaden av kliniskt intresse). Denna skillnad tillsammans med den observerade standardavvikelsen på 15mmHG användes för att beräkna stickprovsstorlek . En stickprovsstorlek på 144 patienter behövs för att uppnå en power på 80%.



Sammanfattning

- Detta måste ingå i en ansökan:
- Frågeställning
- Primary outcome
- Studiedesign
- Statistisk analysplan
- Stickprovdimensionering



Genom ALF-finansiering har vi möjlighet att hjälpa er.

Alla kliniska forskare som planerar en klinisk studie, och avser att söka

forskningsanslag från forskningsråd och andra större nationella finansiärer omfattas.

[http://www.registercentrumnorr.vll.se/
statistik.rcnorr@regionvasterbotten.se](http://www.registercentrumnorr.vll.se/statistik.rcnorr@regionvasterbotten.se)