



Reproducerbarhet av p-värden: en simuleringsstudie

*Examensarbete för kandidatexamen i matematik vid Göteborgs universitet
Kandidatarbete inom civilingenjörsutbildningen vid Chalmers tekniska högskola*

Milton Bark
Lina Hammargren
Carl Larsson
William Nilsén

Reproducerbarhet av p-värden: en simuleringsstudie

Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet

Lina Hammargren William Nilsén

Kandidatarbete i matematik inom civilingenjörsprogrammet Maskinteknik vid Chalmers tekniska högskola

Milton Bark Carl Larsson

Handledare: Serik Sagitov

Examinator: Ulla Dinger och Maria Roginskaya

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2019

Notationslista

H_0 Händelsen att stickprovet kommer från nollfördelningen $N(\mu_0, \sigma^2)$

H_1 Händelsen att stickprovet kommer från den alternativa fördelningen $N(\mu_1, \sigma^2)$

S Händelsen att en studie är signifikant och publiceras

B Händelsen att en publicerad studie har omedveten bias

π Andelen av studier där stickprovet kommer från alternativa fördelningen, $\mathbb{P}(H_1)$

$1 - \pi$ Andelen av studier där stickprovet kommer från nollfördelningen, $\mathbb{P}(H_0)$

b Andelen av studier som publicerats och har omedveten bias, $\mathbb{P}(B)$

σ Standardavvikelse

n Stickprovsstorlek

Kan vi lita på forskning?

2005 publicerade John Ioannidis en artikel vars titel löd Why Most Research Findings are False och efter framförandet av artikeln publicerade många andra forskare artiklar inom samma ämnesområde. Denna explosion av nya studier startade en intensiv debatt i vetenskapsvärlden; är de statistiska metoder som används inom vetenskaplig forskning i dagsläget bristfälliga? Precis denna fråga menade en grupp av blivande ingenjörer och matematiker på Chalmers- och Göteborgs Universitet att besvara i deras examensarbete. Genom att simulera miljontals studier som efterföljde dagens statistiska metoder i verklig forskning lyckades gruppen ta fram och konkretisera vissa av dessa debatterade brister.

Studenternas resultat visade att de statistiska metoder dagens vetenskap vilar på framför allt har en brist: den mänskliga faktorn. Gruppen visade att det inte hade funnits några problem med att använda dagens vanligt förekommande statistiska metoder för validering av forskningshypoteser under ideella fall. Ett par få felaktiga procent av studier till följt av en felmarginal skulle fortfarande publiceras men dessa felaktiga studier skulle snabbt komma till ytan så länge som alla forskare delade med sig av samtliga resultat, såväl positiva som negativa. Dessvärre jobbar forskarlag som tar fram studier separat och inte i stora nätverk, det i kombination med en enorm konkurrens gör att negativa resultat inte delas med andra. Bristen på kommunikation och sammankoppling tillåter de få felaktiga antalet studier att finnas kvar i vetenskapsvärlden där fler och fler felaktiga studier ackumuleras.

Studenterna drar slutsatsen i deras uppsats att den dåliga kommunikationen mellan olika forskarlag är en följd av det mycket större problemet benämnt som *publikationsbias*. Med publikationsbias menas en föreliggande stark vilja att bli publicerad. Att forskare har en publikationsbias är inte förvånande, hur värderar samhället en enskild forskare egentligen och hur ser incitamentstrukturen ut för forskare? Forskare bedöms och belönas i stor grad efter deras publikationer, både i prestige såväl som monetär kompensation. Denna prestationspress gör troligen många forskare omedvetet eller medvetet benägna att manipulera sin data eller metod för att öka sannolikheten att bli publicerade.

Vidare skapade studenterna nya simulerade studier som beaktade olika fall av manipulation av data under de statistiska metoder forskare använder sig av i dagsläget, studenternas slutsats var oroväckande. I princip vilket resultat som helst, oavsett om det stämmer eller inte, kan valideras och konfirmeras när det finns medveten bias så länge forskaren har tillgång till tillräckligt mycket data. Det finns dock hopp, genom att samla in delar av alla publicerade resultat inom en forskningsfråga eller forskningsfält kan en uppskattning göras om hur mycket fusk som sker inom frågan eller fältet.

Mot bakgrund av publikationsbias är det inte förvånande att det inom vissa forskningsfält har det visat sig otroligt svårt att reproducera studier och få samma utfall. Ett exempel på bristen av reproducerbara studier har kartlagts i analyser av flera publicerade studier inom bland annat prekliniska cancerstudier. Bioteknikföretaget Amgens visade i ett projekt att de inte kunde replikera 47 av 53 prekliniska cancerstudier.

Så hur kan problemet lösas? Studenterna argumenterade för att vi måste gå till roten av problemet vilket är publikationsbias. Att sätta artificiella krav på tidsskrifter för att kvotera in replikations-test i sina tidningar skulle kunna vara en lösning men samtidigt är problematiken antagligen mer djupgående än sådan. Tidsskrifter har ett incitament att publicera studier som kommer att attrahera läsare som betalar för tjänsten. Om tidsskrifter inte publicerar replikationsstudier är det antagligen en följd utav att de inte genererar lika många läsare och därmed sämre intäkter. Med den utgångspunkten blir det extra viktigt att informera läsarna av vikten för forskning och vetenskap att studier faktiskt replikeras, då det är läsarnas preferenser som är basen till hur vi utvärderar publikationer.

Bidrag till allmänheten

Utöver den en populärvetenskapliga artikel som är specifik för arbetet har även ett bidrag till den icke-vinstdrivande encyklopedin Wikipedia gjorts. Bidraget har skapats genom att producera och publicera en artikel inom det fält kandidatarbetet behandlar. Syftet med artikeln är att nå ut till en bredare audiens och förklara delar av problematiken för allmänheten i verkligheten. Artikeln är i enlighet med gruppens och handledarens mål om att stärka den entreprenöriella andan inom det matematiska vetenskapliga ämnet på Chalmers tekniska högskola.

Wikipedia artikeln kan finnas på <https://sv.wikipedia.org/wiki/Replikationskrisen>

Sammanfattning

Det här kandidatarbetet studerar reproducerbarhet av vetenskapliga studier. Ämnet är viktigt eftersom det har framkommit att många studier inte är reproducerbara och således att forskningsresultat kan vara felaktiga. En orsak till icke-reproducerbara studier är bias. Bias kan vara medveten då forskare manipulerar sin forskning i syfte att bli publicerad eller kan vara omedveten som följd av att forskaren har en förutfattad mening av vad resultatet ska bli i en studie.

För att få en översikt av problemet simulerades ett flertal studier för att representera ett forskningsfält. En studie i det här arbetet representeras av ett p-värde som är antingen signifikant eller inte signifikant. I det här arbetet används en signifikansnivå $\alpha = 0.05$ för att bestämma om ett p-värde är signifikant. I en studie kan stickprovet komma från fördelningen för nollhypotesen eller den alternativa hypotesen. Det antas genomgående i texten vara känt vilken fördelning stickprovet kommer ifrån. Detta på grund av att en viktig del av arbetet är att undersöka om det går att härleda om H_0 eller H_1 föreligger i en studie. De hypoteser som använts i arbetet är

$$H_0 := \text{stickprovet kommer från } N(0, 5^2)$$

$$H_1 := \text{stickprovet kommer från } N(1, 5^2)$$

Genom grafiska framställningar undersöks studierna, det vill säga p-värdena. Först undersöks fördelningarna av p-värdena under H_0 respektive H_1 . Därefter implementerades en metod för medveten bias i syfte att undersöka påverkan av medveten bias på ett forskningsfält. Slutligen kombineras p-värdena under H_0 och H_1 både med och utan medveten bias för att kunna göra en mer realistisk bild av verklighetstillståndet.

Den andra formen av bias, omedveten bias, har undersökts med hjälp av det så kallade positiva prediktiva värdet (PPV). PPV är sannolikheten att H_1 föreligger givet att ett signifikant p-värde erhållits. Detta arbete betraktade proportionen av studier med omedveten bias som känd.

Användningen av bland annat dessa metoder resulterade i grafiska framställningar och tabeller som visar på en tydlig skillnad mellan fallen då bias förekommer och då det inte förekommer. Resultaten i simuleringen har även jämförts med en verklig studie och bekräftar resultaten av arbetet.

Slutsatsen av de grafiska representationerna av forskningsfält är att det till exempel går att utröna att medveten bias förekommit inom ett forskningsfält. En vidare slutsats är att PPV påverkas negativt av båda formerna av bias och antalet studier med signifikanta p-värden kommer att öka då studier är biased inom ett forskningsfält.

Abstract

The purpose of this bachelor's thesis is to investigate the reproducibility of scientific studies. The subject is important due to the debate about the reproducibility issue of studies which indicates that published research findings may be false in several cases. One of the reasons for non-reproducible studies is bias. There are two cases of bias that will be discussed in this thesis; unconscious and conscious bias. Conscious bias is the act of active manipulation of studies in order to get published, and unconscious bias can arise when scientists are preconceived of a certain outcome before the conduct of a study.

To better understand the problem, several studies were simulated to represent a field of research. A study in this thesis will be represented by a p-value that either can be significant or non-significant. This thesis used a significance level of $\alpha = 0.05$ to decide whether a p-value is significant.

The sample in a study can either come from the distribution for the null hypothesis or the alternative hypothesis. It is assumed to be known throughout the thesis from which hypothesis the distribution of the sample comes from. The reason why this assumption is necessary is in order to investigate the possibility if one could derive the presence of H_0 or H_1 in a study. The hypotheses that this thesis uses are the following

$$H_0 := \text{The sample comes from } N(0, 5^2)$$
$$H_1 := \text{The sample comes from } N(1, 5^2)$$

The studies or in other words, p-values, were analyzed by graphical representations. At first, the distributions of p-values from H_0 and H_1 were analyzed. After analyzing the p-values, a method for conscious bias was implemented with the purpose of analyzing the p-values further, with respect to conscious bias in a field of research. Finally, the p-values from H_1 and H_0 were combined with and without conscious bias to produce a more realistic picture of a scientific field.

The second form of bias, unconscious bias, has been analyzed with a method called positive predictive value (PPV). PPV is the probability that the sample comes from H_1 given that a significant p-value is present. This thesis considered the proportion of studies with unconscious bias to be known.

The use of these methods among others resulted in graphical representations and tables, that demonstrates an evident difference between the cases when bias is present and when it is not. The simulations have been compared to results that are based on real studies which confirms the results of this thesis.

One of the conclusions which was drawn is that the presence of conscious bias can be deduced from the graphical representation of studies in a research field. Furthermore, the PPV is negatively affected by both forms of bias. Moreover, the amount of studies with significant p-values increases when bias is present within a research field.

Innehåll

1	Inledning	1
1.1	Syfte	2
1.2	Problemformulering	2
1.3	Avgränsningar	2
1.4	Etik	2
2	Bayesiansk referensram för en studie i simuleringen	3
2.1	Beräkning av p-värde	3
2.2	P-värden som slumpvariabel	3
2.2.1	Fördelningen av p-värden under H_1	4
2.2.2	Fördelning av p-värden under H_0	4
2.3	Utfallsrummen och omedveten bias	4
2.3.1	Utfallsträd för en studie	5
2.4	Positivt prediktivt värde (PPV)	6
2.4.1	PPV för en studie	6
2.4.2	PPV för ett forskningsfält	6
2.5	Omedveten bias (OB) och dess påverkan på PPV	6
2.6	Medveten bias med hjälp av p-hacking	7
3	Redovisning och analys av simuleringar	8
3.1	Hypotestest och beräkning av p-värden utan medveten bias	8
3.1.1	Fördelning av p-värden utan medveten bias	8
3.1.2	Visualisering av utfallsrummet och beräkning av PPV med och utan omedveten bias	8
3.2	Simulering av medveten bias	9
3.3	Kombination av p-värdens fördelningar	9
3.4	Jämförelse av simulering med verkliga studier	10
4	Resultat	10
4.1	Simuleringar av p-värdenas fördelningar	10
4.2	Simuleringar av utfallsrummen med och utan manipulering	13
4.3	PPV för forskningsfält	14
5	Diskussion	15
5.1	De realiserade utfallsrummen	15
5.2	Fördelningen av p-värden	15
5.3	PPV för forskningsfält	17
5.4	Evaluering av metoderna för att studera ett forskningsfält	18
6	Slutsats och rekommendation för vidare forskning	18
6.1	Potentiella åtgärder för att öka integriteten inom vetenskaplig forskning	18
6.1.1	Registrera studier på förhand	18
6.1.2	Ställ krav på komplett rådata	19
6.1.3	Avlasta forskare genom att studenter replikerar studier	19
6.2	Publikationsbias gör att vetenskap inte är självkorrigerande	19
6.3	Rekommendation för vidare forskning	19

Förord

Det här arbetet är ett kandidatprojekt utfört på Matematiska vetenskaper vid Chalmers tekniska högskola under våren 2019. Bakom arbetet står en grupp av fyra studenter varav två personer kommer från civilingenjörsprogrammet Maskinteknik vid Chalmers och två personer från kandidatprogrammet i matematisk statistik vid Göteborgs universitet.

Varje medlem i gruppen har bidragit till arbetet i dess slutliga form men projektet har delats upp så att det finns huvudansvariga för varje del av arbetet. En loggbok har förts över var medlems insatser under arbetets gång.

Vi vill särskilt tacka vår handledare professor Serik Sagitov för vägledning i arbetet och viktiga synpunkter angående utformning av arbetet.

Arbetsfördelningen i form av huvudsakligt ansvar presenteras nedan.

Populärvetenskaplig presentation: Milton Bark

Wikipedia artikel: Milton Bark och Carl Larsson

Sammanfattning: Carl Larsson och Lina Hammargren

Förord: Lina Hammargren

1: Inledning: Carl Larsson och Lina Hammargren

1.1: Hela gruppen tillsammans

1.4: Hela gruppen tillsammans

2: Bayesiansk referensram för en studie i simuleringen: Carl Larsson och Lina Hammargren

2.6: Milton bark

3: Redovisning och analys av simuleringar: Milton Bark med bidrag från William Nilsén

4: Resultat: Milton Bark och William Nilsén

4.3: Carl Larsson och Lina Hammargren med bidrag från William Nilsén

5: Diskussion

5.1: Lina Hammargren

5.2: Milton Bark

5.3: Carl Larsson

5.4: Lina Hammargren

6: Slutsats

6.1: Lina Hammargren

6.2: Milton Bark

6.3: Milton Bark

Viktiga idéer och övriga bidrag:

Metod för simulering av fusk: Milton Bark

Framtagning av optimal α : Milton Bark

Härledning av PPV för fallet med och utan omedveten bias: Lina Hammargren

Skapandet av koden: William Nilsén med bidrag från Milton Bark

1 Inledning

Det pågår i dagsläget en diskussion huruvida p-värdet ska användas inom vetenskaplig forskning och på vilket sätt det i så fall ska användas. Exempelvis har tidsskriften *Basic and Applied Social Psychology* (BASP) förbjudit användandet av p-värden i de artiklar som de publicerat. Emellertid har debatten nyanserats något av exempelvis Olle Häggström (2017) som argumenterar för att p-värdet kan användas i kombination med en diskussion kring resultaten. Ytterligare ett bidrag till debatten gav Christie Aschwanden (2015) som argumenterade för att även effektstorleken som anger hur stor skillnad det är mellan de två hypoteserna ska diskuteras utöver p-värdet då p-värdet inte visar hur stark en effekt är.

Beräkningar av p-värden är en metod för att ta beslut om hypoteser i forskningsfrågor och användningen av p-värden är utbredd inom vetenskaplig forskning. Det är emellertid en slumpvariabel och risken finns alltid att fel beslut tas om att förkasta en hypotes baserat på p-värdet.

Det kan vara olämpligt att helt förbjuda p-värden då detta kan vara en bra analysmetod. Däremot kräver det att studien i övrigt är välplanerad i termer av stickprovsstorlek och effektstorlek. Dessutom kan p-värdet kompletteras med andra analysmetoder för att få en tydligare bild av resultatet.

Det är dock inte endast problemet att p-värdet är en slumpvariabel och att signifikanta såväl som icke-signifikanta värden kan erhållas med metoden oavsett om verklighetstillståndet är att den hypotes som studeras är sann eller inte som har identifierats inom forskningsvärlden. Inom vissa forskningsområden förekommer det även brister i metoden samt utelämnad rådata (Begley och Ellis, 2012). Dessa brister gör att studierna inte kan reproduceras och resultaten av beräkningen av p-värde blir än mindre tillförlitliga.

Reproducerbarhet är ett viktigt koncept inom vetenskaplig forskning och innebär att en studie skall kunna återupprepas och återfå samma resultat eller liknande resultat ifall ny data samlas. I dagsläget finns det ämnesområden och forskningsfält där det föreligger brist på reproducerbara studier. Ett exempel på bristen av reproducerbara studier kartlades i analyser av flera publicerade studier inom bland annat prekliniska cancerstudier där bioteknikföretaget Amgens i ett projekt inte kunde replikera 47 av 53 cancerstudier (ibid). Ämnesområden där replikationsbristen råder är dock ej begränsad till medicinsk forskning. Ziliak och McCloskey (2008) granskade 369 artiklar inom ekonomisk forskning, deras resultat visade att endast cirka en av fyra resultat stämde med det som presenterades i artiklarna de studerade.

Det finns flera olika anledningar till varför problemet har uppkommit. En betydande anledning är publiceringsbias, vilket är en stark vilja för forskare att publicera signifikanta resultat. Flera studier har visat på en tendens att endast statistiskt signifikanta resultat publiceras och att studier som inte är signifikanta bortses från (Häggström, 2016; Rosenthal, 1979). Detta kan tänkas skapa incitament för forskare, som ofta bedöms utifrån antalet publicerade resultat och dess citeringar av andra forskare, att omedvetet eller medvetet manipulera sin data eller metod för att öka sannolikheten att bli publicerade.

Ovannämnda problem har visat på behovet av att göra samlade bedömningar av forskningsfält och inte betrakta resultatet på en enskild studie som fakta. I det här projektet kommer tre metoder för evaluering av forskningsfält att presenteras baserat på erhållna p-värden. Projektet kan skapa vidare frågeställningar kring hur samhället kvantifierar och värderar forskning. Hur kan forskares neutralitet till sin egen forskning kontrolleras och hur stor påverkan har incitamentstrukturen på den forskning som produceras? Törs läsaren lita på alla forskningsresultat och hur viktigt är det att vara kritisk till det mottagaren läser och använder?

1.1 Syfte

Med utgångspunkt i det beskrivna problemet beträffande bristen av reproducerbarhet är syftet med rapporten att förklara för studenter som kommer komma i kontakt med forskningsresultat i ett tidigt stadiet om vad som kan orsaka felaktiga publicerade resultat. Uppsatsen kan tänkas resultera i en mer kritisk student som kan använda publicerad forskning mer effektivt i framtida egen forskning och tillämpning.

1.2 Problemformulering

Målet är att genom en simuleringsstudie av ett fiktivt forskningsfält kartlägga vilka studier som hade publicerats då p-värdet används som beslutsunderlag. Simuleringsstudien består av att upprepa ett frekventistiskt hypotestest för att beräkna ett p-värde, således står det en simulerad studie bakom varje p-värde.

Nedan formuleras de evalueringsmetoder som simuleringsstudien kommer undersöka.

- Simulering av p-värde kurvorna inom ett fiktivt forskningsfält, som är sammanslagningen av studier under H_0 och H_1 , för jämförelse mellan forskningsfält där det förekommit medveten bias med fallet där ingen medveten bias förekommit.
 - P-värdenas fördelningar kommer att framställas grafiskt för att undersöka förekomsten av vissa värden på p-värdet under olika omständigheter.
- En realisation av utfallsrummen i fallen då det förekommer omedveten bias eller medveten bias.
 - Utfallsrummet består av de olika händelserna som beskriver vad som leder till att en studie publiceras eller inte.
- Post-studie sannolikheten för att en studie har dragit sitt stickprov från den alternativa fördelningen tillhörande alternativhypotesen givet ett signifikant p-värde
 - Denna metod är tätt sammankopplad med realiseringen av utfallsrummen och går direkt att härleda av de definierade händelserna.

1.3 Avgränsningar

- Kandidatarbetet kommer att presentera en förenklad simulering av ett forskningsfält för att besvara frågeställningen.
- Kandidatarbetet kommer inte att ta fram p-värdens fördelningar från verkliga studier. Där emot kommer en jämförelse göras med befintliga artiklar som studerar verkliga p-värde fördelningar.
- Kandidatarbetet kommer endast att behandla en typ av medveten bias.
- Kandidatarbetet kommer inte ta upp problemet med att studier som borde publiceras inte publiceras.

1.4 Etik

Det finns en risk att resultatet av denna rapport kan medverka till att läsaren får svårare att tro på studier, forskningsresultat och den vetenskapliga metoden. Därför är det av yttersta vikt att rapporten inte blir anti-vetenskaplig eller uppmuntrar människor till att misstro vetenskap. Detta problem kan tacklas genom att föra objektiva samt nyanserade resonemang och analyser.

Nyttan med det förväntade utfallet av rapporten är att läsaren blir upplyst och bättre införstådd kring de problem som idag finns med biased publikationer av forskningsresultat.

Rapporten kommer att undersöka ett strukturellt problem inom forskningsvärlden, och ska inte användas för att ogiltigförklara enskilda forskningsresultat och forskares integritet.

2 Bayesiansk referensram för en studie i simuleringen

För att kunna bygga upp flera studier i en simulering krävs en specificering om vad som utgör en studie. I det här kapitlet presenteras den matematiska teorin om vad som ligger bakom en studie och dess utfall i simuleringen som redogörs för i kapitel (3).

Till grund för beräkning av p-värdet ligger ett slumpmässigt stickprov som dras från den valda populationen. Vi kommer att anta att variansen på populationen är känd. Uttrycket *studie* kommer i texten att användas för att beskriva detta förfarande. I simuleringen kommer flera studier att tas fram för att visualisera p-värdenas fördelningar.

2.1 Beräkning av p-värde

Innan en studie genomförs finns det en föreliggande hypotes som ska prövas. För att pröva hypotesen beräknas p-värdet för att undersöka fenomenet inom populationen.

Vi kommer att anta att populationen är normalfördelad i studierna enligt $N(\mu, \sigma^2)$ där $\mu = \mu_0$ eller $\mu = \mu_1$ och utföra ett ensidigt hypotestest för att testa om $\mu_0 < \mu_1$. Det finns således två konkurrerande hypoteser, utfallet av en studie definieras därav som

$$H_0 := \text{stickprovet är från nollfördelningen } N(\mu_0, \sigma^2)$$

$$H_1 := \text{stickprovet är från den alternativa fördelningen } N(\mu_1, \sigma^2)$$

Vi kommer att dra ett slumpmässigt stickprov av storlek n från populationen för att beräkna p-värdet. Ett slumpmässigt stickprov (x_1, \dots, x_n) är en mängd av element som är en realisation av (X_1, \dots, X_n) där X_i $i = 1, \dots, n$ är slumpvariabler. Realisationen används för att beräkna det observerade stickprovsmedelvärdet $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$. Det observerade stickprovsmedelvärdet används för att beräkna teststatistikan z_{obs} som beskrivs i uttryck (1).

$$z_{obs} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad (1)$$

Teststatistikan används för att beräkna p-värdet, vilket är sannolikheten att få ett lika extremt eller mer extremt värde som den observerade teststatistikan givet att stickprovet dragits ur fördelningen för H_0 . Komponenterna som används för att beräkna p-värdet är Φ , Z och z_{obs} där Φ och Z definieras nedan. Vi får att p-värdet blir som i uttryck (2) och betecknas PV .

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ och } \Phi(t) = \int_{-\infty}^t \phi(x) dx$$
$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}, \text{ där } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$PV = \mathbb{P}(Z > z_{obs} | H_0) = 1 - \Phi(z_{obs}) \quad (2)$$

Vi kallar händelsen att vi får ett signifikant p-värde $S := PV \leq \alpha$. Vidare antar vi i vår modell att alla studier som fått ett signifikant p-värde publiceras. Således är S händelsen att en studie publiceras.

2.2 P-värden som slumpvariabel

Som tidigare nämnt finns det två fall för en studie: verklighetstillståndet är antingen att stickprovet kommer från H_0 eller att stickprovet kommer från H_1 . Vi kommer därför att analytiskt härleda fördelningsfunktionerna för p-värdena i de två fallen. P-värdet PV beskrivs i uttryck (2).

2.2.1 Fördelningen av p-värden under H_1

För att säga något om p-värdenas fördelning härleder vi fördelningsfunktionen, där dess derivata är täthetsfunktionen som beskriver kurvans utseende och därmed p-värdenas fördelning. Z (se uttryck 1) väntevärde blir

$$\mathbb{E}(Z) = \frac{\mathbb{E}[\bar{X}] - \mu_0}{\sigma/\sqrt{n}} = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \text{ vilket medför att } Z \sim N\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}, 1\right),$$

och därför blir fördelningsfunktionen $F_{PV}(p)$

$$\begin{aligned} F_{PV}(p) &= \mathbb{P}(PV \leq p) = \mathbb{P}(1 - \Phi(Z) \leq p) = \mathbb{P}(\Phi(Z) > 1 - p) = \mathbb{P}(Z > \Phi^{-1}(1 - p)) = \\ &= 1 - \Phi\left(\Phi^{-1}(1 - p) - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right), \end{aligned}$$

Derivatan av $F_{PV}(p)$ löses med hjälp av kedjeregeln och derivata för en invers funktion

$$\left\{ \frac{d}{dx}(f(g(x))) = f'(g(x))g'(x), \frac{d}{dx}(f^{-1}(x)) = \frac{1}{f'(f^{-1}(x))} \right\}$$
$$f_{PV}(p) = \frac{d}{dp} \left(1 - \Phi\left(\Phi^{-1}(1 - p) - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) \right) = \phi\left(\Phi^{-1}(1 - p) - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) \frac{1}{\phi(\Phi^{-1}(1 - p))}$$

Vilket leder till en högerskev fördelning av p-värden.

2.2.2 Fördelning av p-värden under H_0

Med samma resonemang som i 2.2.1 får vi då vi antar att X_i $i = 1, \dots, n$ tillhör H_0 och att $\mathbb{E}(\bar{X}) = \mu_0$, vilket medför att $\mathbb{E}(Z) = 0$ och $Z \sim N(0, 1)$.

$$\begin{aligned} F_{PV}(p) &= \mathbb{P}(PV \leq p) = \mathbb{P}(1 - \Phi(Z) \leq p) = \mathbb{P}(\Phi(Z) > 1 - p) = \mathbb{P}(Z > \Phi^{-1}(1 - p)) = \\ &= 1 - \Phi(\Phi^{-1}(1 - p)) = 1 - (1 - p) = p \end{aligned}$$

Vilket leder till en likformigfördelning av p-värden under H_0 enligt $PV \sim U[0, 1]$.

2.3 Utfallsrummen och omedveten bias

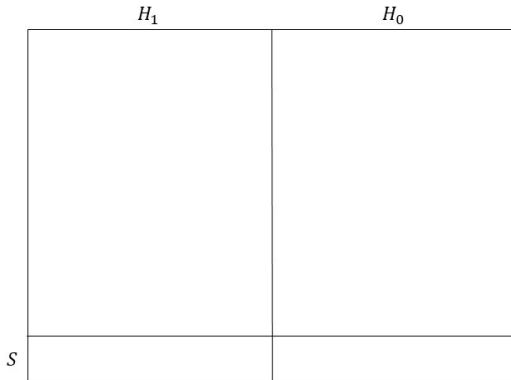
Mot bakgrund av tidigare definierade händelser ska vi nu klargöra vad utfallsrummen blir där elementen är p-värden för utförda studier.

Händelserna som bygger upp utfallsrummet utan omedveten bias är som följer och ett Venndiagram över dessa händelser finns i figur (1).

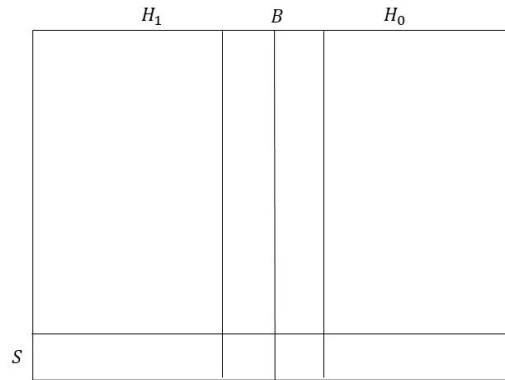
$$\begin{aligned} H_0 &:= \text{stickprovet är från nollfördelningen } N(\mu_0, \sigma^2) \\ H_1 &:= \text{stickprovet är från den alternativa fördelningen } N(\mu_1, \sigma^2) \\ S &:= \text{händelsen att p-värdet är signifikant } PV \leq \alpha \\ S^c &:= \text{händelsen att p-värdet inte är signifikant } PV > \alpha \end{aligned}$$

Ett problem är att detta utfallsrum inte tar hänsyn till verklighetstillståndet. Därav får vi introducera en ny händelse B som innebär att en publicerad studie är påverkad av omedveten bias. Händelsen B kan därför innehålla p-värden som inte hade varit signifikanta utan omedveten bias. Vi har därför att $S \cup B$ är den nya händelsen att en studie publiceras i utfallsrummet där vi tar hänsyn till omedveten bias. Ett Venndiagram över händelserna finns i figur (2).

B := händelsen att en studie som fått ett signifikant p-värde är påverkad av omedveten bias



Figur 1: Utfallsrummet i fallet då omedveten bias inte förekommer i ett forskningsfält.



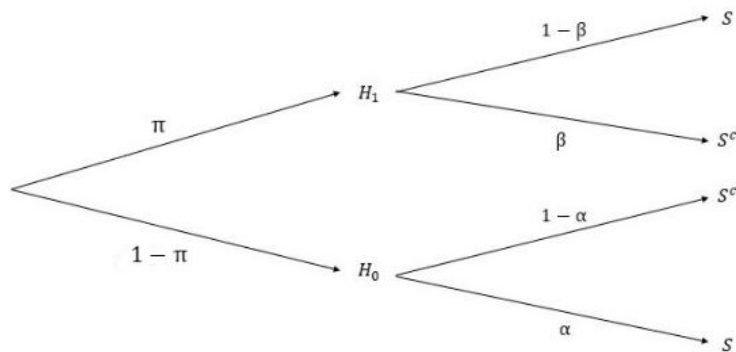
Figur 2: Utfallsrummet då omedveten bias förekommer i ett forskningsfält.

Vi antar att händelsen B är oberoende av händelserna H_1 , H_0 , S och S^c . Anledningen till antagandet att B är oberoende av H_1 och H_0 är på grund av att det är omöjligt att på förhand veta om en hypotes stämmer. Vidare antar vi att B är oberoende av S eftersom p-värdena blir signifikanta oberoende av om de hade varit signifikanta utan bias eller inte.

2.3.1 Utfallsträd för en studie

Händelseförloppet i en simulerad studie visualiseras i ett träd-diagram i figur (3) som redogör för de två utfallen i en studie, S och S^c . Som tidigare nämnts finns det två händelser, H_0 och H_1 , och om vi vidare antar att proportionen av studier där verklighetstillståndet är att H_1 är sant $\mathbb{P}(H_1) = \pi$ blir således proportionen av studier där verklighetstillståndet är att stickprovet kommer från H_0 $1 - \pi$.

$\mathbb{P}(S|H_1)$ är sannolikheten att en studie har fått ett signifikant värde givet att vi är i den alternativa fördelningen vilket är styrkan $(1 - \beta)$, $\mathbb{P}(S|H_0)$ är sannolikheten att en studie har fått ett signifikant värde givet att vi är i nollfördelningen vilket är α .



Figur 3: Träd-diagram över potentiella utfall i en studie där händelsen S innebär att en studie publiceras. Händelserna att verklighetstillståndet är H_1 respektive H_0 antas med sannolikhet π respektive $1 - \pi$. En studie publiceras om H_1 inträffat med sannolikhet $1 - \beta$ och om H_0 inträffat med sannolikhet α .

2.4 Positivt prediktivt värde (PPV)

I denna sektion kommer sannolikheten att ett stickprov kommer från H_1 givet att ett signifikant resultat erhållits att härledas. Detta är post-studie sannolikheten att ett stickprov kommer från H_1 och kallas för det positiva prediktiva värdet (PPV). Det här värdet ger en indikation på hur trovärdig en studie är och kan användas för att evaluera om ett p-värde är representativt för den genomförda studien eller resultatet av slump.

2.4.1 PPV för en studie

Vi studerar nu utfallsrummet utan omedveten bias som definierades i 2.3 och syns i figur (2). Händelsen S är att ett p-värde är signifikant, det vill säga $PV \leq \alpha$. A priori sannolikheten $\mathbb{P}(H_1) = \pi$ betraktas som en parameter i vår modell. Varje gång vi utför en ny studie blir PPV för den studien som i uttryck 3.

$$PPV = \mathbb{P}(H_1|S) = \frac{\mathbb{P}(H_1 \cap S)}{\mathbb{P}(S)} = \frac{\mathbb{P}(H_1)\mathbb{P}(S|H_1)}{\mathbb{P}(H_1)\mathbb{P}(S|H_1) + \mathbb{P}(H_0)\mathbb{P}(S|H_0)} = \frac{\pi(1 - \beta)}{\pi(1 - \beta - \alpha) + \alpha} \quad (3)$$

Uttryck 3 får vi genom definitionerna för α och $(1 - \beta)$. Vi har att $1 - \beta = \mathbb{P}(\text{förkasta } H_0 | \text{stickprovet kommer från } H_1)$ vilket blir $\mathbb{P}(S|H_1)$ och att $\alpha = \mathbb{P}(\text{förkasta } H_0 | \text{stickprovet kommer från } H_0)$ vilket blir $\mathbb{P}(S|H_0)$.

Ur uttryck (3) följer att ju större $(1 - \beta)$ är, desto högre blir PPV för en studie. Det är därför av vikt att planera styrkan $(1 - \beta)$ inför utförandet av studien. $(1 - \beta) = \mathbb{P}(\text{Förkasta } H_0 | \text{stickprovet kommer från } H_1)$.

Styrkan på en enskild studie för olika stickprovsstorlek härleds genom beräkningen nedan.

$$1 - \beta = \mathbb{P}(Z > 1.645 | H_1) = \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > 1.645 | H_1\right) = 1 - \Phi\left(1.645 - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) \quad (4)$$

I tabell 1 har vi räknat ut styrkan för några olika stickprovsstorlekar för att visa hur stickprovsstorleken påverkar styrkan.

n	30	35	40	50	100	1000
$1 - \beta$	0.29	0.32	0.35	0.41	0.64	>0.9998

Tabell 1: Styrkan enligt ekvation (4), för några olika stickprovsstorlekar. Värdena har tagits från en tabell för standard normalfördelningen.

2.4.2 PPV för ett forskningsfält

Simuleringen kommer att beräkna PPV för ett forskningsfält. Ett forskningsfält är en sammanslagning av alla studier som utförts, det vill säga både H_0 och H_1 . Detta kommer att göras genom att räkna antalet studier som hamnar i mängden $H_1 \cap S$ och dela på antalet studier som hamnar i mängden S .

2.5 Omedveten bias (OB) och dess påverkan på PPV

Figur 2 beskriver utfallsrummet i forskningsfrågor då det inte förekommer någon form av felaktighet i metod och genomförande av studier. Utfallsrummet i figur 3 tar hänsyn till *omedveten bias* genom mängden B . Det positiva prediktiva värdet för en studie med hänsyn till det nya utfallsrummet blir nu

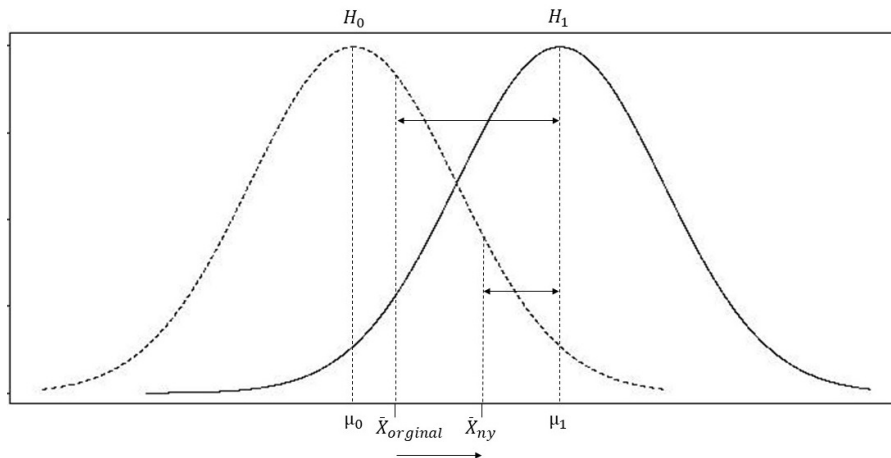
$$\begin{aligned}
PPV_{OB} &= \mathbb{P}(H_1 | (S \cup B)) = \\
&= \frac{\mathbb{P}((H_1 \cap S) \cup (H_1 \cap B))}{\mathbb{P}(S \cup B)} = \\
&= \frac{\mathbb{P}(H_1 \cap S) + \mathbb{P}(H_1 \cap B) - \mathbb{P}(H_1 \cap S \cap B)}{\mathbb{P}(S \cup B)} = \\
&= \frac{\mathbb{P}(H_1 \cap S) + \mathbb{P}(H_1 \cap S \cap B) + \mathbb{P}(H_1 \cap S^c \cap B) - \mathbb{P}(H_1 \cap S \cap B)}{\mathbb{P}(S \cup B)} = \\
&= \frac{\mathbb{P}(H_1 \cap S) + \mathbb{P}(H_1 \cap S^c \cap B)}{\mathbb{P}(S \cup B)} = \\
&= \frac{\mathbb{P}(H_1)\mathbb{P}(S|H_1) + \mathbb{P}(H_1)\mathbb{P}(S^c|H_1)\mathbb{P}(B)}{\mathbb{P}(S \cup B)} = \\
&= \frac{(1 - \beta)\mathbb{P}(H_1) + \beta\mathbb{P}(H_1)\mathbb{P}(B)}{(1 - \beta)\mathbb{P}(H_1) + \alpha\mathbb{P}(H_0) + (1 - \alpha)\mathbb{P}(H_0)\mathbb{P}(B) + \beta\mathbb{P}(H_1)\mathbb{P}(B)} = \\
&= \frac{\pi(1 - \beta + \beta b)}{\alpha + b(1 - \alpha) + \pi(1 - \beta - \alpha)(1 - b)} \\
PPV_{OB} &= \frac{\pi(1 - \beta + \beta b)}{\alpha + b(1 - \alpha) + \pi(1 - \beta - \alpha)(1 - b)} \tag{5}
\end{aligned}$$

Med samma metod som i 2.4.2 blir PPV för ett forskningsfält antalet studier som hamnar i $(H_1 \cap S) \cup (H_1 \cap B)$ delat på antalet studier som hamnar i $S \cup B$.

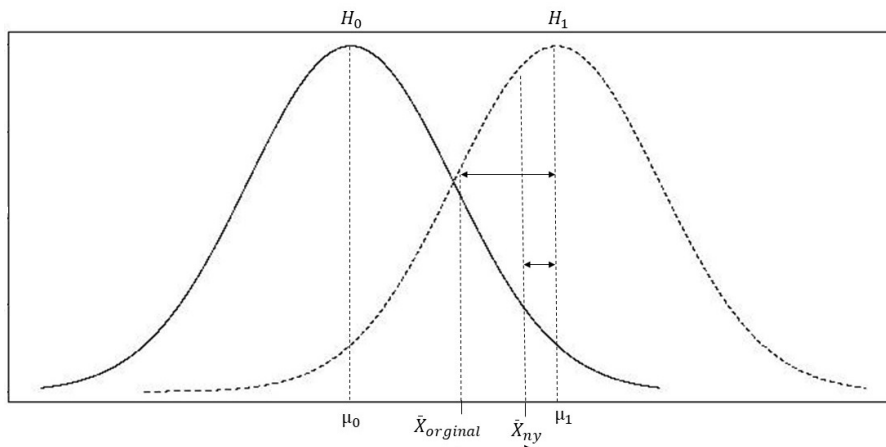
2.6 Medveten bias med hjälp av p-hacking

P-hacking är en metod för att manipulera ett stickprov i syfte att erhålla ett signifikant p-värde. Genom att selektivt ta bort det minsta elementet ur stickprovet genereras ett större värde på $\bar{x} - \mu_0$ vilket ger upphov till en större sannolikhet att framkalla ett signifikant p-värde. P-hacking är inte att förväxlas med att ta bort extrema värden efter noggrann bedömning som görs för att få en mer rättvis bild av data. Det scenario av p-hacking som kommer att användas i simuleringsstudien är fallet då medelvärdet på stickprovet selektivt ökas för att öka värdet på $\bar{x} - \mu_0$. Förhållandet mellan medelvärdet på stickprovet redogörs i figur (4) under H_0 och figur (5) under H_1 .

Figur (4) och (5) visar hur medelvärdet \bar{x} förflyttas närmare μ_1 vid p-hacking. Vi kan se från figur (5) att det sannolikt krävs mindre manipulation av data då stickprovet dras från fördelningen för H_1 för en större skattad effektstorlek.



Figur 4: Illustration av medveten bias då stickproven dras från H_0 , $\bar{x} - \mu_0$ ökar och således framkallas ett signifikant p-värde



Figur 5: Illustration av medveten bias då stickproven dras från H_1 , $\bar{x} - \mu_0$ ökar och således framkallas ett signifikant p-värde

3 Redovisning och analys av simuleringar

I det här kapitlet kommer simuleringen av flera studier att presenteras med utgångspunkt i den beskrivna teorin i kapitel (2). Simuleringen av studier är skapad i R och tar fram p-värdedistributioner, en grafisk framställning av studiernas utfallsrum samt beräknas de olika fallen för PPV. För den intresserade läsaren finns koden för simuleringen i Appendix 2.

3.1 Hypotestest och beräkning av p-värden utan medveten bias

Koden bygger på att simulera studier med tillhörande p-värden, både för den alternativa hypotesen och nollhypotesen, se sektion (2.1) för betydelse av hypoteserna.

I beräkningen av p-värdena antas parametrarna $\sigma = 5$, $\mu_0 = 0$, $\mu_1 = 1$, $\alpha = 0.05$ vara fixa och stickprovsstorleken n kommer att variera. Funktionen `rnorm()` genererade stickprov av olika storlek n ur fördelningarna $N(0, 5^2)$ respektive $N(1, 5^2)$. Steget därefter var att beräkna z_{obs} enligt uttryck (1). För att beräkna p-värdena från de ensidiga hypotestesterna användes funktionen `1-pnorm()` där `pnorm()` är fördelningsfunktionen för en standard normalfördelning.

3.1.1 Fördelning av p-värden utan medveten bias

Från sektion (3.1) genereras 25 000 000 p-värden från respektive fördelning som sammanställs i två olika diagram. Den grafiska framställningen av p-värdenas fördelning jämförs sedan med de teoretiska täthetsfunktionerna framtagna i (2.2.1) och (2.2.2) i respektive figur.

3.1.2 Visualisering av utfallsrummet och beräkning av PPV med och utan omedveten bias

Med hjälp av teorin i (2.3) utfördes en visualisering av utfallsrummen i Venndiagram. Båda utfallsrummen är skapade enligt figur (2), men i det ena fallet föreligger endast omedveten bias medan i det andra föreligger även medveten bias. Totalt genererades 400 stickprov där 200 stickprov kom ifrån respektive fördelning $N(0, 5^2)$ och $N(1, 5^2)$.

För att generera utfallsrummet i fallet då studier påverkas av omedveten bias bestämdes parametern $b = 10\%$ och används för att bygga upp mängden B , som i sin tur utgör 10% slumpade p-värden ur utfallsrummet.

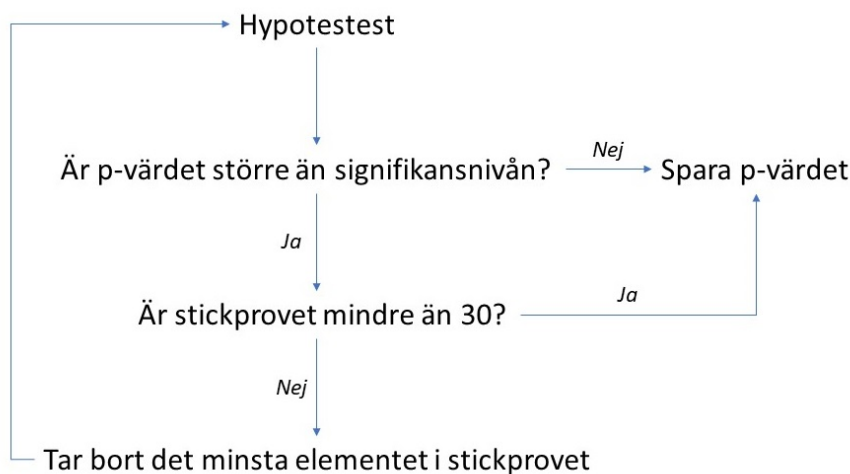
Utöver att endast studera de genererade p-värdena från sektion (3.1) beräknades PPV för ett forskningsfält via metoderna som beskrivs i 2.4.2 och 2.5.

3.2 Simulering av medveten bias

PPV med omedveten bias beräknades genom att slumpmässigt ta p-värden från utfallsrummet för att utgöra mängden B . Det kan tänkas att ett slumpmässigt antal studier har någon form av omedveten bias, men i denna sektion vill vi utforska scenariot av ren manipulation av data. Skillnaden när en aktiv manipulation eller med andra ord medveten bias föreligger medför att stickprovet justeras kontinuerligt för att generera fler signifikanta p-värden enligt figurerna (4) och (5); metoden för att uppnå detta redogörs i loopen (6).

Likt tidigare fall börjar manipulationen med ett hypotestest som i (3.1), vilket genererar ett p-värde baserat på ett stickprov. När medveten bias föreligger kontrolleras p-värdet mot signifikansnivån, om det är signifikant sparas p-värdet annars går stickprovet vidare i loopen. Nästa steg i loopen är att kontrollera att stickprovstorleken är större eller lika med 30. Ett stickprov som är större än eller lika med 30 behövs och kan ej underskridas då stickprovet är precis tillräckligt stort för att kunna använda ett Z-test tillförlitligt (Sprinthall, R. C, 2011). Om storleksgränsen på stickprovet ej uppnås sparas p-värdet, annars går stickprovet vidare i loopen. Nästa steg är det första och enda steget rådata faktiskt manipulerades.

Genom att successivt ta bort det element med det minsta värdet i stickprovet ökas medelvärdet av stickprovet. Ett ökat medelvärde resulterar i större chans att få ett signifikant p-värde som följd av den ökade skattade effektstorleken enligt sektion (2.6). Avslutningsvis itereras loopen tills antingen ett signifikant p-värde har uppkommit eller tills stickprovet inte längre är större eller lika med 30 då loopen avbryts. Genom att justera det initiala stickprovstorleken till ett stort antal ges mer utrymme för medveten bias medan ett litet stickprov ger mindre utrymme för manipulation.



Figur 6: Iterativ manipulation av stickprovet för att generera signifikanta p-värden

3.3 Kombination av p-värdens fördelningar

Efter 25 000 000 p-värden genererats från respektive fördelning då både medveten och ingen form av bias förekommer har all simulerad data samlats in. För att återskapa ett forskningsfält där ingen samt medveten bias råder sammanställs p-värdena i två olika diagram, respektive fält får då 50 000 000 p-värden. I det ena diagrammet råder ingen manipulation av stickproven medan i det andra råder manipulation av stickproven. Vidare sammanställdes grafiskt de kombinerade simulerade p-värdenas fördelning då ingen medveten bias föreligger i diagrammet som visar på hur manipulation påverkar forskningsfältet som referensmått.

3.4 Jämförelse av simulering med verkliga studier

Steget därefter då p-värden har genererats både med och utan medveten bias är att jämföra resultatet med verkliga existerande p-värden. De verkliga p-värdena inhämtas från existerande litteratur som redogör p-värden där manipulation av data starkt anas samt då antagandet görs att ingen medveten bias föreligger.

Då data samlats in sammanställdes grafiskt de verkliga p-värde fördelningarna med antagandet att de befinner sig under H_0 med medveten bias eller H_1 utan bias. Därefter jämförs stickprovstorlekar av de simulerade p-värdena tills det stickprovstorlek som bäst överensstämmer med respektive fördelning framgick. Under H_0 med medveten bias var stickprovstorleken 34 och under H_1 utan bias var stickprovstorleken 100. Då stickprovstorlekarna fastställdes slumpades 10 stycken fördelningar med 50 signifikanta p-värden ur respektive fördelning och sammanställdes grafiskt i samma diagram som de verkliga p-värdena. Även de simulerade fördelningarna med 25 miljoner p-värden utan någon form av bias sammanställdes grafiskt i respektive diagram för att få ett referensmått.

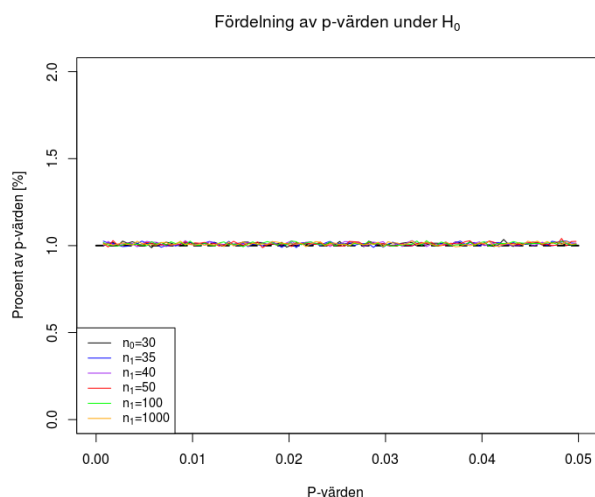
4 Resultat

I det här kapitlet kommer resultaten av kapitel (3) att redovisas. Vad som därmed följer är redogörelse av p-värdenas fördelningar utan och med medveten bias, kombinerade p-värde fördelningar utan och med medveten bias, jämförelse av simulering med verkliga p-värden, utfallsrummet med och utan medveten bias samt olika fall av PPV för ett simulerat forskningsfält.

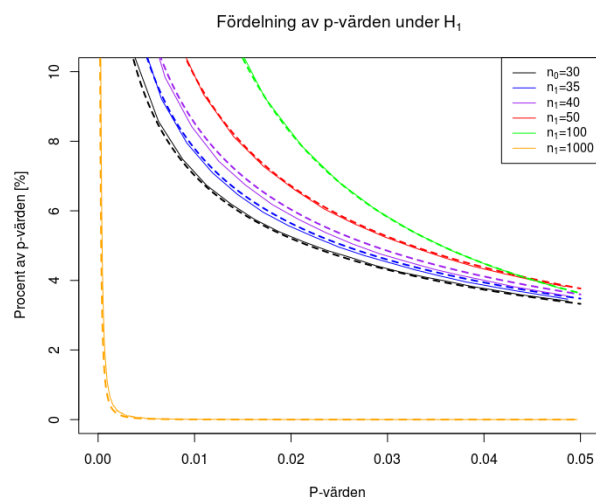
Stickprovstorleken som tidigare benämnts n kommer nu att benämnas n_0 och n_1 . Storleken $n_0 = 30$ använder vi som standardstorlek på stickprovet för jämförelse med de andra fallen. I texten används n_1 för de fall där stickprovstorleken är större än 30.

4.1 Simuleringar av p-värdenas fördelningar

I den här sektionen kommer de olika p-värde fördelningarna att redovisas grafiskt.



Figur 7: Simulerade fördelningen av 25 000 000 p-värden under H_0 , jämfört med den analytiska täthetsfunktionen under H_0 från sektion (2.2.2). Större stickprov medför en mer likformig fördelning

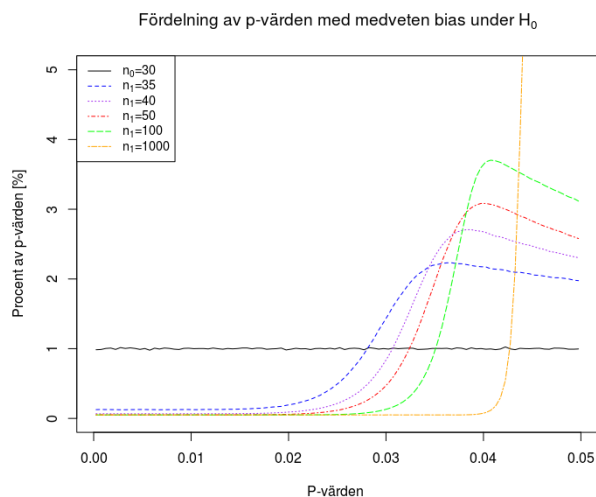


Figur 8: Simulerade fördelningen av 25 000 000 p-värden under H_1 , jämfört med den analytiska täthetsfunktionen under H_1 från sektion (2.2.1). Större stickprov medför brantare kurvor

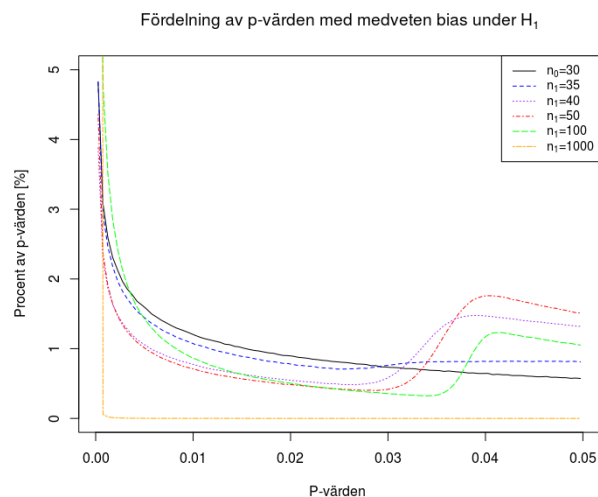
Utifrån figur (7) syns p-värdenas likformighet, och att de är likformigt fördelade under H_0 för samtliga stickprovstorlekar. Den heldragna raka linjen är den teoretiska täthetsfunktionen enligt sektion

(2.2.2). I och med att kurvan är simulerad kommer störningar framstå, i detta fall fluktuerar kurvorna mellan 0.9 och 1.1 procent.

Utifrån figur (8) syns det att då ett stickprov dras från en fördelning där det finns en skillnad mellan μ_0 och μ_1 uppkommer en höger skev fördelning av p-värdena i enlighet med beviset av täthetsfunktionen i sektion (2.2.1). Den analytiska täthetsfunktionen är representerad av en streckad linje och den simulerade p-värdes fördelningen av en heldragen linje. För figur (8) visas fördelningarna av samtliga stickprovsstorlekarna.



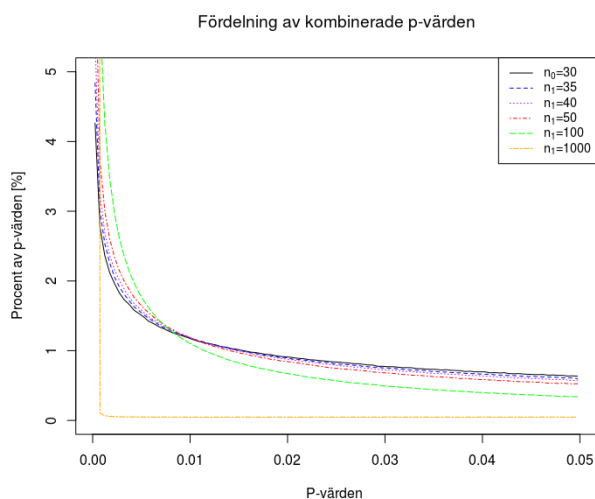
Figur 9: Fördelningen av 25 000 000 p-värden under H_0 när medveten bias förekommer, för jämförelse är den heldragna kurvan n_0 utan medveten bias



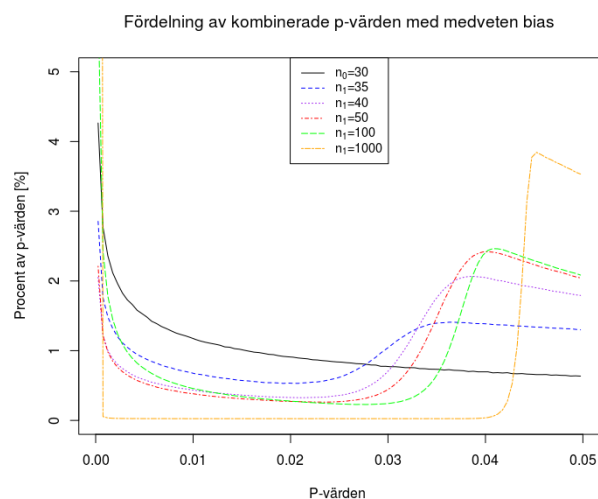
Figur 10: Fördelningen av 25 000 000 p-värden under H_1 när medveten bias förekommer, för jämförelse är den heldragna kurvan n_0 utan medveten bias

Utifrån figur (9) syns det att p-värdes fördelningarna är vänsterförskjutna. Hur vänsterförskjutna fördelningarna är beror på hur stort stickprovet är, ett större stickprov innebär mera medveten bias då stickprovet tillåts att gå igenom looperna i figur (6) flera gånger. Notera dock att inom intervallet $[0, 0.02]$ finns det mycket få p-värden både för stora och små stickprov.

Till skillnad från figur (9) kan det observeras i figur (10) att p-värdena är högerförskjutna inom intervallet $[0, 0.03]$ men blir vänsterförskjutna i intervallet $[0.03, 0.05]$. Notera här att en större stickprovsstorlek inte nödvändigtvis innebär ett större påslag för vänsterförskjutningen under H_1 . Exempel på det fenomenet är att $n_1 = 100$ ger ett mindre påslag än $n_1 = 50$ eller $n_1 = 40$.



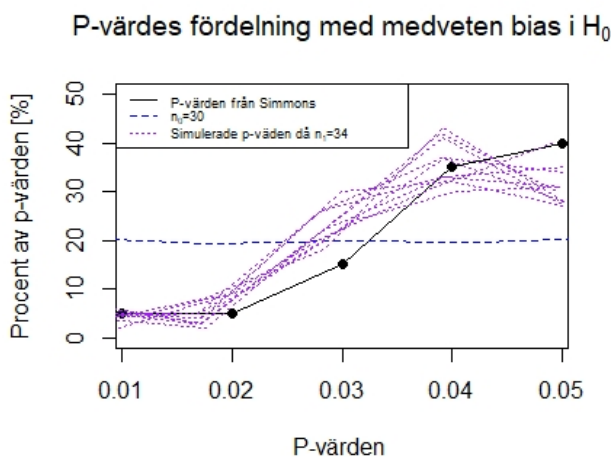
Figur 11: Fördelningen av 50 000 000 p -värden för de kombinerade händelserna under H_0 och H_1 utan medveten bias



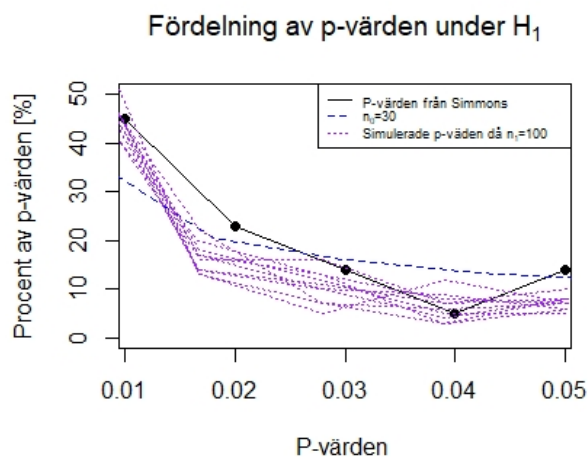
Figur 12: Fördelningen av 50 000 000 p -värden för de kombinerade händelserna under H_0 och H_1 med medveten bias

Figur (11) beskriver de sammanslagna p -värdefördelningarna med olika stickprov för de båda händelserna H_0 och H_1 då ingen medveten bias föreligger. Notera här att för samtliga stickprovstorlekar är fördelningarna tämligen lika om man bortser för det mer extrema stickprovet $n_1 = 1000$.

Figur (12) beskriver de sammanslagna p -värdefördelningarna med olika stickprov för de båda händelserna H_0 och H_1 då medveten bias föreligger. Notera att det går tydligt se skillnad mellan figur (11) och (12).



Figur 13: Fördelningen av 100 signifikanta p -värden under H_0 med medveten bias för olika stickprov. Simmons (2013), illustration av p -hacking redogörs med den punktade linjen.



Figur 14: Fördelningen av 100 signifikanta p -värden under H_1 för olika stickprov. Simmons (2013), illustration av p -hacking redogörs med den punktade linjen.

Figur (13) jämför verkliga p -värden under H_0 där det starkt anas att någon form manipulation av data eller metod föreligger. De verkliga p -värdena visas med den punktade heldragna linjen medan de simulerade fördelningarna av 100 signifikanta p -värden är illustrerade med de sträckande linjerna. Den likformiga streckade linjen, för n_0 , visar fördelningen av 25 miljoner p -värden om ingen

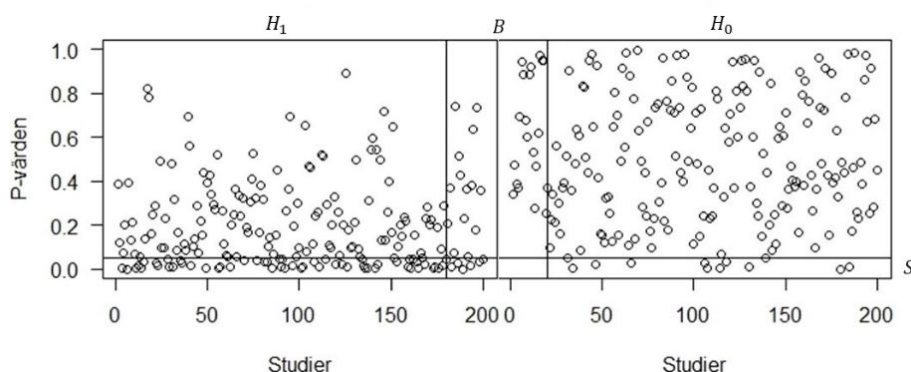
medveten bias råder. Notera här att de simulerade fördelningarna inte återspeglar den verkliga fördelningen exakt.

Figur (14) jämför verkliga p-värden där det starkt anas att ingen form manipulation av data eller metod föreligger under H_1 . Likt figur (13) visas de verkliga p-värdena med den punktade heldragna linjen medan de simulerade fördelningarna av 100 signifikanta p-värden illustreras med de sträckande linjerna. Den streckade linjen då n_0 råder visar fördelningen av 25 miljoner p-värden med syfte att få ett referensmått. Notera här att de simulerade fördelningarna inte återspeglar den verkliga fördelningen exakt.

4.2 Simuleringar av utfallsrummen med och utan manipulering

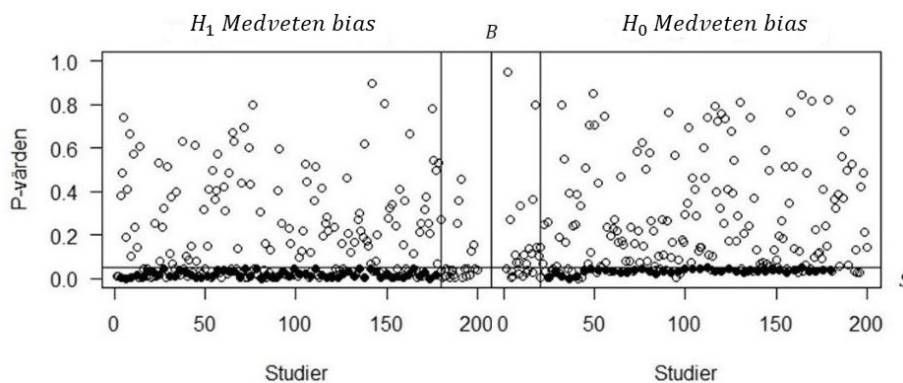
Figurerna över utfallsrummen visas i figur (15) och (16). Komponenterna i figuren beskrivs nedan

- Signifikanta p-värden markeras med symboler under den vågräta linjen och tillhör händelsen S
- P-värden som påverkas av omedveten bias och därmed publiceras markeras med punkter efter de lodräta linjerna och tillhör händelsen B . Notera att p-värdena är signifikanta, men hade inte varit det om de inte påverkats av bias. P-värdena som syns i figurerna är således de p-värdena som skulle erhållits om inte omedveten bias hade påverkat resultatet.
- Signifikanta p-värden som har uppstått till följd utav medveten bias markeras med ifyllda cirklar
- Figur (15) och (16) kan likställas med figur (2), då det är samma händelser som avses.



Figur 15: Utfallsrummet där varje punkt representerar en studie, utfallsrummet innehåller totalt 400 studier

I figur (15) syns fördelningen av p-värden under händelserna H_0 och H_1 . Notera att p-värdena är glesare för större p-värden tillhörande H_1 jämfört med H_0 . Dessutom syns det att fler p-värden blir signifikanta i H_1 , medan p-värdena är mer jämnt utspridda i H_0 .



Figur 16: Utfallsrummet där varje punkt representerar en studie där medveten bias förekommer. Utfallsrummet innehåller totalt 400 studier

I figur (16) blir p-värdena i hela utfallsrummet glesare när de går mot större p-värden. När stickprovet har manipulerats under H_0 och H_1 kan fler signifikanta p-värden observeras.

4.3 PPV för forskningsfält

Nedanstående tabell beskriver PPV för ett forskningsfält i standardfallet, då det förekommer omedveten bias, medveten bias samt då det förekommer omedveten och medveten bias. I simuleringen har en a priori sannolikhet på $\pi = 0.5$ använts. Ur tabell (2) syns respektive värde för stickprovsstorleken för alla studier och vilket fall av PPV som beräknats. PPV_{MB} beräknades för utfallet när medveten bias föreligger.

Notera att under n_0 , förekommer det inte någon medveten bias i simuleringen och därav ändras inte PPV då medveten bias föreligger. Däremot blir $PPV = 0.85$ och $PPV_{OB} = 0.57$.

n_1	PPV	PPV_{OB}	PPV_{MB}	$PPV_{MB,OB}$
35	0.87	0.60	0.58	0.51
40	0.88	0.63	0.52	0.48
50	0.89	0.67	0.50	0.46
100	0.93	0.78	0.50	0.45
1000	0.95	0.87	0.50	0.45

Tabell 2: Värden under kolumn ett visar standardfallet för PPV, kolumn två visar PPV med omedveten bias, kolumn tre visar PPV med medveten bias, kolumn fyra visar PPV med omedveten och medveten bias. De olika raderna visar olika stickprovsstorlek n_1

Ett antagande som gjorts är att utfallsrummet innehåller lika många studier under H_1 som under H_0 . Ifall antagandet i stället skulle göras att alla studier låg under H_1 skulle PPV för alla stickprov fått värdet 1 och om alla studier låg under H_0 skulle PPV för alla stickprov fått värdet 0. För visualisering av hur relationen mellan antal studier där stickprovet kommer från H_0 och där stickprovet kommer från H_1 påverkar PPV hänvisas den intresserade läsaren till Appendix I.

5 Diskussion

Metod och resultat som framförts genererar tre metoder för utvärdering av ett forskningsfält eller en forskningsfråga som kommer att diskuteras i sektionerna nedan. Mer specifikt är de tre evalueringsmetoderna: realisering av utfallssummen, fördelningen av p-värden samt PPV för forskningsfält.

5.1 De realiserade utfallsrummen

Figurerna i sektion (4.2) visar de realiserade utfallsrummen för den utförda simuleringen. Figur (15) visar resultatet i fallet då det inte förekommit någon medveten bias i studierna och figur (16) visar resultatet i fallet då det förekommer medveten bias.

Resultatet som visas i figur (15) åskådliggjorde att om 200 studier genomförts där stickprovet dras ur fördelningen för H_0 publiceras 12 artiklar som drar slutsatsen att stickprovet kommer från H_1 . Om de 188 studier som inte får ett signifikant resultat inte publiceras ser det då ut som att det finns god evidens för H_1 i forskningsfrågan vilket är ett problem.

En möjlig lösning till problemet med att studier publicerar resultat som argumenterar för att H_1 är sant när stickprovet kommer från H_0 är att sänka signifikansnivån α vilket kommer att minska antalet signifikanta resultat då stickprovet kommer från H_0 . En sänkning av α kommer dock även att minska antalet studier där stickprovet kommer från H_1 och där ett signifikant p-värde erhålls. Det finns dock en tydlig skillnad i fördelning av p-värden i figur (15) och fler studier publiceras när stickprovet kommer från H_1 . Ett lägre α kommer att medföra att mängden studier där stickprovet kommer från H_1 som publiceras minskar.

Mängden B innehåller som tidigare nämnts signifikanta p-värden som påverkats av omedveten bias. I den grafiska framställningen i figur (15) kan det tolkas som att p-värdena som publiceras inte alltid är signifikanta, men i verkligheten hade de omedvetna biased studierna inte gått att urskilja från de som inte är omedvetet biased. Mängden B påverkar i synnerhet studier där stickprovet kommer från H_0 eftersom p-värden är mer utspridda i detta fall. Det gäller dock för både H_0 och H_1 att fler resultat publiceras då det förekommer omedveten bias i ett forskningsfält. Detta kan ge en illusion av att en forskningsfråga verkar mer sann än falsk. Däremot går det att urskilja H_0 från H_1 genom att se att studierna ändå är mer utspridda under H_0 och det finns en mer jämn spridning av p-värden under H_0 .

Skillnaden i spridning av p-värden som syns i figur (15) blir mindre då data manipulerats, vilket syns i figur (16). Det blir då svårare att se en skillnad mellan H_1 och H_0 . I detta fall är metoden med realiserade utfallsrum bristfällig och det blir svårt att evaluera forskningsfältet. Mot bakgrund av detta gör medveten bias metoden svårare att använda för att analysera ett forskningsfält.

En nackdel med metoden är att den kräver icke-signifikanta p-värden som ofta inte publiceras på grund av publiceringsbias. I dagsläget finns det initiativ till att även negativa eller oväntade resultat ska publiceras i en egen journal eller databas, exempel på detta är International Journal of Negative Results av Open Access Pub som ger fri tillgång till genomförda studier med negativa resultat. Initiativ som detta är värdefulla, men ett ytterligare problem är att det finns en kultur i vetenskapsvärlden som medför att forskare bedöms på antalet positiva resultat och citeringar. I ett tidigt stadie av karriären är det i synnerhet viktigt att forskare publicerar intressanta artiklar för att forskaren ska få citeringar och vidare finansiering för sitt arbete. Vilket kan medföra att negativa resultat aldrig publiceras och därför inte kommer allmänheten tillhanda.

5.2 Fördelningen av p-värden

En metod som inte kräver att även icke-signifikanta värden publiceras för att kunna analysera ett forskningsfält är att studera fördelningen av p-värden. För fallet då p-värdenas fördelning simule-

rades under H_0 och H_1 utan medveten bias i figur (7) och (8) kan skillnader åskådliggöras. Den noterbara skillnaden här är de olika skevheterna i fördelningarna vilket förklaras matematiskt i sektion (2.2.1) och (2.2.2). Det som ligger bakom de felaktigt publicerade studierna i figur (7) är signifikansnivån. Notera här att arean som ligger under grafen är representativt för de ackumulerade antalet signifikanta p-värden, ser vi till arean får vi att 5% av alla p-värden blir signifikanta vilket motsvarar signifikansnivån. På motsvarande fördelning under H_1 i figur (8) kan det observeras att majoriteten av signifikanta p-värden är koncentrerade närmre noll. En intressant slutsats som därmed kan dras här är att en lägre signifikansnivå minskar antalet signifikanta p-värden under H_0 oproportionellt mycket mer än under H_1 . Men vad är den optimala signifikansnivån för fallet vi studerar?

Vid en signifikansnivå runt 1% kommer en ytterligare minskning av signifikansnivån sänka antalet p-värden under H_1 i större utsträckning än H_0 . Om proportionen av falskt positiva och sant positiva resultat vill minimeras bör signifikansnivån ändras från den vedertagna 5% till den signifikansnivån då antalet procent av p-värden är densamma i H_0 och H_1 , i vårt fall uppgår denna signifikansnivå till 1%. En signifikansnivå på 5% är konvention inom många fält, men anledningen till just denna signifikansnivå är främst historisk (Häggström, 2017), slutsatsen vi drar är att en signifikansnivå på 5% helt enkelt inte alltid är optimalt i alla forskningsfält.

Däremot har vi simulerat p-värden under specifika parametrar då det inte föreligger någon form av bias för kunna dra slutsatser kring en optimal signifikansnivå; vad händer när inte endast de statistiska kända felen som α råder vid hypotesprövningen? För att besvara frågan är det värdefullt att studera fallet vid manipulation av data som visas i figurerna (9) och (10). Intuitivt kan läsaren tänka sig utifrån figurerna att man bör sänka signifikansnivån från 5% till 2% snarare än 1% som vid fallet utan manipulation. Varför den sänkningen är intuitiv är då det finns mycket få presenterade positiva falska p-värden under en signifikansnivå på 2%. Intuition i det här fallet är dock missvisande, notera här hur manipulationen av data uppstår i enlighet med figur (6). Då manipulation föreligger justeras stickprovet av den bestämda signifikansnivån, en sänkning eller ökning av signifikansnivån kommer således endast förskuta början av p-värdenas procentuella ökning bakåt eller framåt. Det blir därmed mycket svårare att bestämma en optimal signifikansnivå då medveten bias föreligger, vi kan dock dra slutsatsen att större stickprov kommer att behövas för att använda sig av medveten bias vid en lägre signifikansnivå som följd utav den brantare lutningen observerad av större stickprov i figur (9).

Något som är värt att notera är att alla framtagna p-värden bygger på simuleringar, det är därmed förstäligt att ifrågasätta hur väl p-värdenas fördelningarna stämmer överens med verkligheten. I figurerna (10) och (9) har 25 miljoner p-värden genererats, att studera 25 miljoner studier där man anar starkt att medveten bias eller ingen bias föreligger är antagligen näst intill omöjligt. Med det i åtanke blir det intressant att jämföra resultaten från simuleringen med existerande litteratur som studerar verkliga studier där medveten bias och ingen medveten bias föreligger. I publikationen P-Curve: A Key to the File-Drawer av Joseph P. Simmons (2014) studerades det här fenomenet. I artikeln studerades 20 stycken publicerade och därmed signifikanta p-värden där det fanns en befogad misstanke om medveten bias, samt 22 stycken där ingen indikation av medveten bias kunde observeras. Respektive fördelning av Joseph P. Simmons resultat kan se i de heldragna punktade linjerna i figur (13) och (14).

Simmons p-värden kan i sin tur jämföras med de 100 stycken signifikanta p-värdena som drogs från vår simulering under efterliknade förhållanden i respektive figur. Notera i figurerna (13) och (14) att de simulerade p-värdenas fördelning följer ungefär samma fördelningar, men inte exakt. Flera orsaker kan tänkas föreligga beträffande varför det resultatet uppkommer. Först och främst är det ett mycket begränsat antal simulerade p-värden fördelningar vilket tillåter slumpen att bli mer framträdande. För det andra studerar Simmons verkliga studier som inte nödvändigtvis använder samma metoder som simuleringen för generering av p-värden, exempelvis kan det tänkas att ett två sidigt Z - test råder vilket invaliderar vår simulering av medveten bias etc.

Med de flera brister som kan tänkas föreligga i åtanke kan slutsatsen dras att simuleringen är nå-

gorlunda verklighetstrogen, vilket i sin tur gör slutsatserna dragna från simuleringen mer legitima. Ytterligare kritik av simuleringen är dock att flera olika former av p-hacking har studerats, bland annat i en artikel av Bruns och Ioannidis (2016) där resultaten visade att p-kurvorna var högerfördelade då vissa illegitima metoder användes vid framställning av studier. Bruns och Ioannidis resultat visar på motsatt effekt från Simmons och vårt presenterade resultat.

En nackdel med att studera p-värdenas fördelningar är kostnaden att systematiskt ta reda på olika studiers p-värden inom olika forskningsfrågor. Vidare behövs många p-värden för att få en visualisering av fördelningen som kan användas för kategorisering. Att samla in många studier inom samma fråga kan tänkas vara svårt, framförallt inom nya fält där forskningen är begränsad. Metoden blir därför inte lämplig vid små forskningsfrågor.

För att vara mer verklighetstrogen kan det tänkas att p-värden för ett helt forskningsfält samlas in, om sådant är fallet kommer det sannolikt finnas fler studier att analysera p-värdena på. Om alla simulerade signifikanta p-värden motsvarar ett forskningsfält kan vi studera utfallet av p-värde fördelningarna genom att kombinera de p-värdena under H_0 och H_1 . De kombinerade p-värdena visas i figur (11) och (12) då manipulation av stickproven råder. Något som kan ses direkt är att det är mycket svårt att veta ifall ett forskningsfält endast studerar p-värden under H_1 eller om det råder en kombination av H_1 och H_0 då ingen medveten bias föreligger. Teoretiskt borde endast p-värdets kurva under H_1 förskjutas uppåt med samma nivå som p-värden under H_0 har.

Då medveten bias föreligger enligt figur (12) uppkommer tydligare skillnader. Genom att granska p-värdets fördelning inom ett forskningsfält där manipulation av data föreligger kan en intressant slutsatsen att det med stor sannolikhet råder en tveksam metodik för produktion av studier.

5.3 PPV för forskningsfält

För att kunna ge en specifik sannolikhet för H_1 givet att ett signifikant p-värde erhållits är beräkningen av PPV en lämplig metod vars uttryck härleds i sektion (2.4.2) och (2.5). I tabell (2) visas hur de olika scenarion som presenterats i arbetet påverkar PPV. Något som kan noteras är att ett större stickprov per studie ger högre värden på ett samlat forskningsfältets PPV, detta sker indirekt som följd utav den högre styrka ett större stickprov medför. Tabellen visar även att PPV med omedveten bias har lägre värden än för standardfallet samt att PPV med medveten bias har lägre värden än både standardfallet och för fallet med omedveten bias. Dock blir värdet på PPV lägre till en viss grad när medveten bias förekommer, värdet går inte under 0.5 för $n_1 \geq 50$ och värdet ökar för standardfallet och fallet med bias.

Eftersom stickprovsstorleken spelar stor roll i skattningen av en forskningsfältets trovärdighet är det viktigt att den är tillräckligt stor för varje studie. Det finns flera faktorer som kan tänkas spela in vid val av en studies stickprov, aspekter som kan väga in skulle kunna vara tidsbegränsningar och ekonomiska begränsningar. En ytterligare begränsning är då det råder brist på individer i en population. Ett hypotetiskt exempel då det råder en sådan brist kan vara att en sällsynt sjukdom studeras där populationen med personer som har sjukdomen är liten vilket kommer att generera ett lägre stickprov. Med detta i åtanke kommer studier med större stickprov sannolikt att producera mer signifikanta resultat där stickprovet kommer från H_1 , jämfört med likartade studier där stickprovet är litet. Ett exempel på när ett forskningsfält har låg styrka som följd av begränsad stickprovsstorlek är i studien Power failure: why small sample size undermines the reliability of neuroscience. Där det visade sig efter analys av 48 relevanta studier inom neurovetenskap att den genomsnittliga styrkan per studie var 21% (Katherine S. Button et al. 2013). Konsekvensen av en sådan låg styrka är ineffektiv forskning och därmed slöseri med resurser.

Låg styrka och stickprovsstorlek per studie är inte ensamt problematiskt, en betydande nackdel för PPV är att det är svårt att med säkerhet veta vad a priori sannolikheten π är. Det är i princip omöjligt att på förhand veta hur många studier som utförts i ett forskningsfält, eftersom studier där stickprovet kommer från H_0 i många fall inte publiceras. Ioannidis (2005) använder i sin artikel parametern R , ration mellan studier där stickprovet kommer från H_1 till studier där stickprovet

kommer från H_0 . När R beräknas eller uppskattas återspeglas inte direkt π , men samma problematik kvarstår med att veta på förhand hur många studier som är utförda inom H_0 respektive H_1 inom ett fält.

Att studera PPV kan vara en bra metod vid rätt förutsättningar, men det kräver att man vet mycket om forskningsfältet på förhand. Man bör veta det ungefärliga antalet studier som utförts och antalet studier där stickprovet kommer från antingen H_0 eller H_1 . Skall man uppskatta π bör man ha belägg för det.

5.4 Evaluering av metoderna för att studera ett forskningsfält

Metoden för realisering av utfallsrummen ger en god indikator på vad den samlade bedömningen blir inom ett forskningsfält i de fallen då medveten bias inte förekommer. En nackdel är dock att det är omöjligt att generera utfallsrummen då antalet studier som studerar en fråga där stickproven dras ur fördelningen för H_0 respektive H_1 är okänt. Den här metoden är dessutom svår att implementera med verkliga studier på grund av publiceringsbias och således är nyttan av denna metod låg. Vidare lider beräkningen av PPV för hela forskningsfält av bristen att det är svårt att i verkligheten särskilja studier som lider av medveten bias och de som inte gör det och därför kan det vara bra att även i denna metod kombinera analysen med fördelning av p-värden.

Metoden som visualiserar utfallsrummen är oanvändbar så länge det inte finns möjligheter att få tag på alla utförda studier, såväl positiva resultat som negativa resultat. Det finns dock en lösning på problemet med att negativa resultat inte kommer forskningsvärlden tillhanda och det är att studier ska registreras på förhand. Skulle det finnas ett sätt för forskare att få tag på studier med negativa resultat eller studier som inte slutförts skulle samarbetet i forskningsvärlden öka och det skulle bli enklare att evaluera forskningsfält och få en översikt på status i en viss forskningsfråga.

Visualiseringen av fördelningen av p-värden kräver däremot varken negativa resultat eller en förkunskap om antalet studier som studerar fallet H_0 eller fallet H_1 . P-värdens fördelningar är således en metod som kan vara enkel att implementera i verkligheten. Metoden är även ett bra komplement till de två andra metoderna. Det är dock viktigt att klassificera en forskningsfråga korrekt och studera forskningsfrågor som är liknande. Beroende på tillgång på data och hur detaljerad en intressent är kan p-värdets fördelningar antingen studeras på en forskningsfråga liksom ett helt forskningsfält.

6 Slutsats och rekommendation för vidare forskning

I det här kapitlet kommer slutsatser och rekommendationer för vidare forskning att redogöras för. Områden som kommer att tas upp är: återgårdar för ökning av vetenskaplig integritet, publiceringsbias påverkan på forskning samt rekommendationer för vidare forskning.

6.1 Potentiella åtgärder för att öka integriteten inom vetenskaplig forskning

Nedan anges tre potentiella åtgärder för att minska problemet med att studier inte kan reproduceras. Det ska noteras att det som tidigare nämnts är ett komplext problem som inte endast har en lösning och således är det inte självklart att åtgärderna skulle vara effektiva.

6.1.1 Registrera studier på förhand

En lösning på problemet med brister i metoder är att studier ska registreras på förhand. Det finns redan initiativ på att studier ska förregistreras som Registered Reports av Center for Open Science. Registered Reports öppnar för peer review redan innan data samlats in vilket kan öka kvaliteten på enskilda studier och möjliggöra bättre evaluering av forskningsfält.

6.1.2 Ställ krav på komplett rådata

För att öka reproducerbarheten i studier är det viktigt med tillgång till rådata. En lösning kan därför vara att rådata registreras innan den processeras. Bristen på tillgång till rådata har identifierats av bland annat Begley och Ellis (2012) som studerade reproducerbarheten i prekliniska cancerstudier.

6.1.3 Avlasta forskare genom att studenter replikerar studier

Det kan vara tidskrävande att reproducera en studie och det är därför förståeligt att forskare inte vill spendera för mycket tid på detta arbete. En möjlig lösning är att studenter på kandidat- och masternivå får denna uppgift som del av sin utbildning. Emellertid är en potentiell nackdel av denna lösning att studenter kanske inte har den kunskap som krävs för att kunna göra om en studie på ett korrekt sätt.

6.2 Publikationsbias gör att vetenskap inte är självkorrigerande

Om ett forskarlag gör en studie och publicerar ett falskt resultat kan det tänkas att resultatet finns kvar utan att det felaktiga resultatet blir åtgärdat om ingen replikationsstudie görs. Forskarlag har få incitament till att korrigera felaktiga eller falska studier och vill själva bedriva egen forskning (Franco et al, 2014). En slutsats som kan dras från bristen av replikeringsbarheten är att vetenskaplig forskning inte nödvändigtvis korrigerar felaktiga slutsatser. Det är även vanligt att flera tidskrifter inom olika fält väljer att inte publicera replikationsstudier (Yeung, Andy, 2017; Martin, Clarke, Richard, 2017) vilket ytterligare minskar incitament för forskare att bedriva replikationsstudier. Samtidigt som de nuvarande kraven på statistisk signifikans för publicering ger upphov till problem måste forskning kunna utvärderas. Forskning som bedrivs måste kunna utvärderas och faktum är att ett häpnadsväckande positivt sant resultat antagligen medför mer värde till samhället än ett icke signifikant resultat.

En åtgärd för att lätta på existerande publikationsbias är att forskare börjar arbeta i nätverk. I dagsläget arbetar forskare oberoende av varandra (Ioannidis, John P.A., 2005) och det kan vara bättre om forskare arbetar i stora nätverk och samarbetar mellan forskarlag för att dra nytta av varandras resultat och upptäckter. Resultatet av det kan tänkas vara att effektiv tid inte slösas som följd av att forskare inte behöver studera något som redan studerats. Utöver minskningen av återupprepningar av forskning medför även tillgång på alla resultat att felen som uppstått som följd av signifikansnivån kan direkt utläsas genom att studera proportionen av signifikanta resultat med icke signifikanta.

En annan mycket viktig aspekt är att kravet på replikeringsbarheten måste stiga, framförallt inom fält där den är som sämst. Med utgångspunkten att det största incitamentet för forskare är publicering är en lösning för att öka antalet replikationsstudier att sätta artificiella krav på tidskrifter för att kvotera in replikationstest i tidningar. Samtidigt är problematiken antagligen mer djupgående än sådan. Tidsskrifter har ett incitament att publicera studier som kommer att attrahera läsare som betalar för tjänsten. Om flera tidsskrifter inte vill publicera replikationsstudier är det antagligen en följd utav att de inte genererar lika många läsare och därmed sämre intäkter. med den utgångspunkten blir det extra viktigt att informera läsarna av vikten för forskning och vetenskap att studier faktiskt replikeras, då det är läsarnas preferenser som är basen till hur samhället utvärderar publikationer.

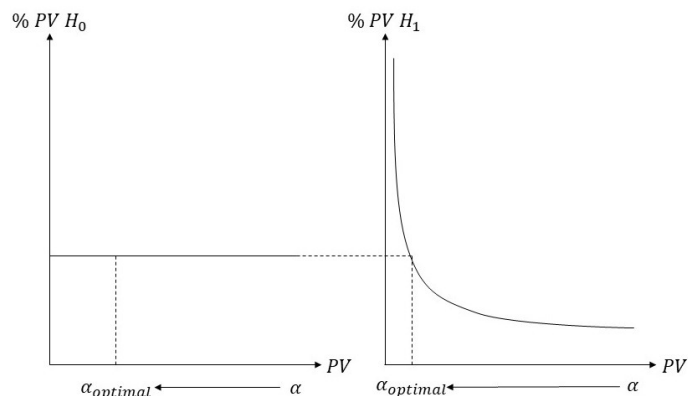
6.3 Rekommendation för vidare forskning

I den här studien har en förenklad modell eller simulering gjorts för att uppskatta ett teoretiskt forskningsfält. Även om simuleringen till viss grad stämde överens med verkliga observerade p-värden finns det flera problem som kan tänkas åtgärdas i framtida forskning. Först och främst bör modellen byggas ut med fler metoder av medveten bias, modellen tar endast hänsyn av en specifik form av manipulation när det kan tänkas finnas många tveksamma metoder för att få fram ett signifikant p-värde. Till och med en sådan sak som otillräcklig eller tvetydig redovisning av

en metod för en studie kan göra den närtill omöjlig att reproducera (Begley, C.G., 2013), dessa studiers resultat är inte nödvändigtvis felaktiga men måste tas hänsyn till.

Vidare kontrollerades resultatet av simuleringen med endast en studie och denna studie innehöll relativt få p-värden. Rekommendationen för att lösa det problemet är att samla in större mängder med verklig data.

I sektion (5.2) togs det upp att ett optimal α eller signifikansnivå kunde uppskattas från vår simulering i enlighet med figur (17) nedan.



Figur 17: Teoretisk illustration av optimal alpha då p-värden under H_0 och H_1 kan separeras

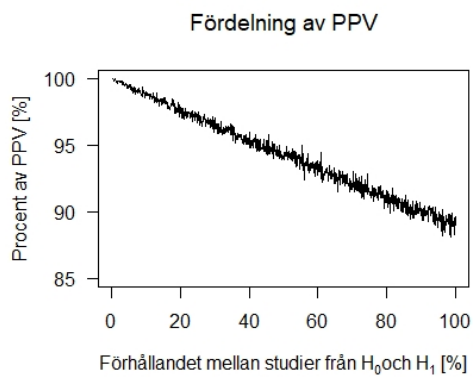
Om vidare forskning kan göras inom området kan det tänkas resultera i en matematisk kvantifierbar optimal signifikansnivå istället för dagens historiskt motiverade signifikansnivå.

För att kunna ta reda på den optimala signifikansnivån måste dock studier under H_0 och studier under H_1 separeras. Separationen av ett helt forskningsfält är därmed ett värdefullt potentiellt bidrag till forskningen. I vår simulering kan det tänkas vara tämligen enkelt att göra denna separation som följd av att endast två forskningsfrågor studeras, den ena under H_0 och den andra under H_1 . Men går det att separera ett helt forskningsfält med hundratals olika forskningsfrågor? Om sådant är fallet kan både PPV beräknas samt optimal signifikansnivå framtas.

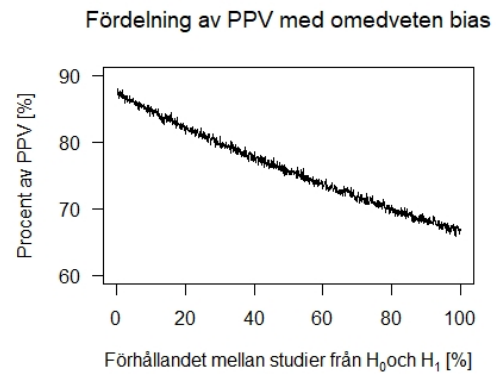
Källor

- Aschwanden, C.A., (2015). Science Isn't Broken - It's just a hell of a lot harder than we give it credit for. . FiveThirtyEight, [Blogg] 2015-08-19.
Tillgänglig: <https://hv.se.libguides.com/c.php?g=243055p=1617018>. [2019-05-12]
- Begley, C. G., and Ellis, L. M. 2012, Drug Development: Raise Standards for Preclinical Cancer Research, *Nature*, 483, 531–533.
- Begley, C.G., 2013, Reproducibility: Six red flags for suspect work, *Nature*, 497, 433–434.
- Bruns, Stephan B. and Ioannidis, John P. A. 2016, p-Curve and p-Hacking in Observational Research *PLoS ONE* 11(2): e0149144. doi:10.1371/journal.pone.0149144
- Camerer, Colin F.; Dreber, Anna; Forsell, Eskil; Ho, Teck-Hua; Huber, Jürgen; Johannesson, Magnus; Kirchler, Michael; Almenberg, Johan; Altmejd, Adam (2016-03-25). Evaluating replicability of laboratory experiments in economics. *Science*. 351 (6280): 1433–1436. Bibcode:2016Sci...351.1433C. doi:10.1126/science.aaf0918. ISSN 0036-8075. PMID 26940865.
- Grimmett, Geoffrey och Stirzaker, David (2001). *Probability and Random Processes*. New York: Oxford University Press.
- Häggström, Olle (2017) The Need for Nuance in the Null Hypothesis Significance Testing Debate, *Educational and Psychological Measurement* 77, 616-630.
- Ioannidis, John P.A. (2005) Why Most Published Research Findings Are False. *PLoS Medicine*. Volume 2. Issue 8. 696-701.
- John Bannon (2016), About 40% of economics experiments fail replication survey. *Science | AAAS*. 2016-03-03. Hämtad 2019-03-19.
- Martin, G. N.; Clarke, Richard M. (2017). Are Psychology Journals Anti-replication? A Snapshot of Editorial Practices. *Frontiers in Psychology*. 8: 523. doi:10.3389/fpsyg.2017.00523. ISSN 1664-1078. PMC 5387793. PMID 28443044
- Open Science Collaboration (2015). Estimating the reproducibility of Psychological Science (PDF). *Science*. 349 (6251): aac4716. doi:10.1126/science.aac4716. PMID 26315443.
- Rice, John A. (2006). *Mathematical Statistics and Data Analysis*. 3. Cengage Learning.
- Schooler, J. W. (2014). Metascience could rescue the 'replication crisis'. *Nature*. 515 (7525): 9. Bibcode:2014Natur.515....9S. doi:10.1038/515009a. PMID 25373639.
- Smith, Noah. Why 'Statistical Significance' Is Often Insignificant. Bloomberg.
- Simmons, Uri, Simmons, Joseph P., Nelson, Leif D., 2014, P-Curve: A Key to the File-Drawer. *Journal of Experimental Psychology: General*, Vol. 143, No.2, 534-547.
- Sprinthall, R. C. (2011). *Basic Statistical Analysis* (9th ed.). Pearson Education
- Yeung, Andy W. K. (2017). Do Neuroscience Journals Accept Replications? A Survey of Literature. *Frontiers in Human Neuroscience*. 11: 468. doi:10.3389/fnhum.2017.00468. ISSN 1662-5161. PMC 5611708. PMID 28979201.

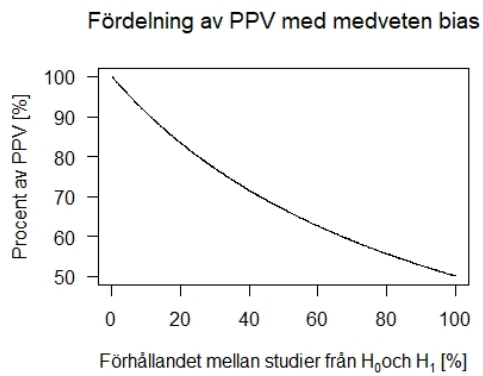
Appendix 1



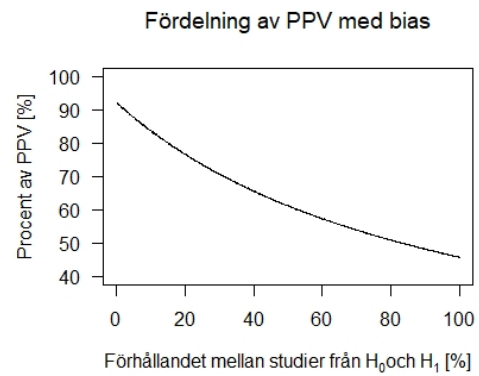
Figur 18: PPV då det finns färre studier i H_0



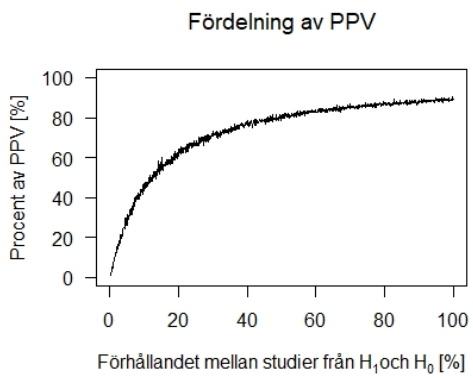
Figur 19: PPV då det finns färre studier i H_0 med omedveten bias



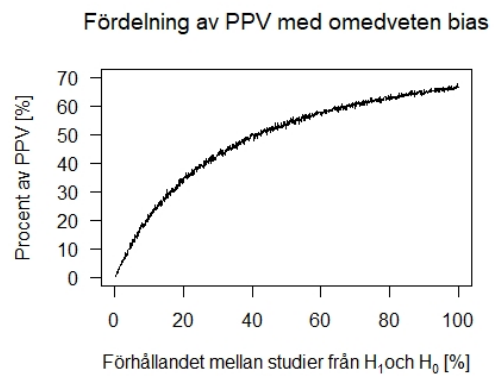
Figur 20: PPV då det finns färre studier i H_0 med medveten bias



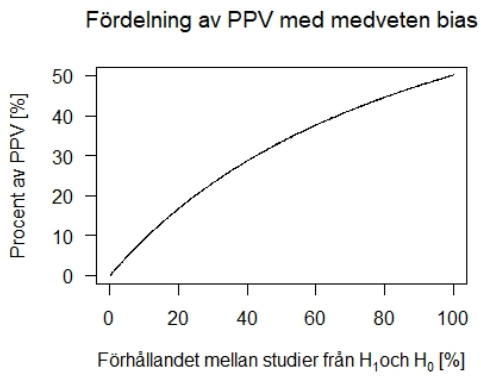
Figur 21: PPV då det finns färre studier i H_0 med bias



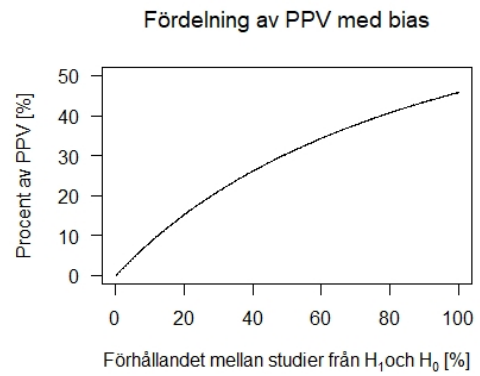
Figur 22: PPV då det finns färre studier i H_1



Figur 23: PPV då det finns färre studier i H_1 med omedveten bias



Figur 24: PPV då det finns färre studier i H_1 med medveten bias



Figur 25: PPV då det finns färre studier i H_1 med bias

Appendix 2

```
1
2 #####Beskrivning#####
3 #
4 #
5 # Syftet med följande kod är att ta fram fördelningar
6 # av p värden vid olika senarion. Utöver p värde fördelningar
7 # menar koden på att framställa utfallsrummet grafiskt för
8 # dessa olika senarion. Det positiva prediktiva värdet (PPV) beräknas
9 # samt framställs grafiskt
10 # för olika proportioner mellan antalet studier i H_0 och H_1. För att
11 # jämföra simuleringen med verkligheten
12 # tittar vi på p värdes fördelningen från artikeln "P Curve: A Key to
13 # the File Drawer av Joseph P. Simmons (2014)".
14 # Scenarion som beaktas studerar
15 # hypotestester när stickprovet kommer från H_0 eller H_1 samt när
16 # omedveten och medveten bias råder.
17 #
18 #
19 # Vid fallet då p värde fördelningar är aktuella ändra plotflag till
20 # plotflag=FALSE
21 #
22 # Vid fallet då grafisk sammställning av mängderna är aktuellt ändra
23 # plotflag till plotflag=TRUE
24 #
25 # Vid fallet då sammställning av studier för att jämföra med artikeln "
26 # p curve" ändra studieflag till studieflag=TRUE
27 #
28 # Vid fallet då grafisk sammställning av PPV är aktuellt ändra PPVflag
29 # till PPVflag=1 eller PPVflag=2.
30 # PPVflag=1 leder till att det finns färre studier som ligger i H_0 jä
31 # mfört med H_1 och PPVflag=2 leder till motsatsen.
32 #
33 # Senaste justerat datum: 16 Maj 2019
34
35 #####Simulering#####
36
37 temps < c(30, 35, 40, 50, 100, 1000) #Stickproven för varje iteration.
38 Observera att detta bara gäller för när vi har tänkt rita fö
39 rdelningarna för p värdena.
40 #30 stickprov är för ingen medveten bias, 35 till 40 för mindre
41 # medveten bias, 50 för medelstor medveten bias och 100 till 1000 för
42 # stor medveten bias.
43
44 studieflag < FALSE #Om vi vill jämföra med p kurvorna från artikeln "P
45 # Curve: A Key to the File Drawer" av Joseph P. Simmons (2014)
46
47 if (studieflag==TRUE){ #Skapar stickproven som vi är intreserade av att
48 # jämföra med artikeln
49 # temps < seq(30,40,1)
50 # }
51
52 k1 < 50000 #Totala antalet studier där stickproven kommer från H_0
53 k2 < 50000 #Totala antalet studier där stickproven kommer från H_1
```

```

41
42 PPVplot < 0 #Bestämmer om vi vill grafiskt sammanställa PPV
43
44 if (PPVplot==1){ #Då det finns fler studier i H_1 än i H_0
45   temps < rep(50,1000) #Stickprov för varje iteration
46   k1 < 10000 #Totala antalet studier där stickproven kommer från H_0
47   k2 < 10000 #Totala antalet studier där stickproven kommer från H_1
48   tempk1 < seq(10,k1,10) #Totala antalet studier där stickproven
      kommer från H_0
49   tempx < 100*seq(1/length(temps),1,1/length(temps)) #Proportionerna
      mellan studier från H_0 och H_1
50 }
51
52 if (PPVplot==2){ #Då det finns fler studier i H_0 än i H_1
53   temps < rep(50,1000) #Stickprov för varje iteration
54   k1 < 10000 #Totala antalet studier där stickproven kommer från H_0
55   k2 < 10000 #Totala antalet studier där stickproven kommer från H_1
56   tempk2 < seq(10,k2,10) #Totala antalet studier där H_0 är falsk
57   tempx < 100*seq(1/length(temps),1,1/length(temps))
58 }
59
60 tempH_0f < matrix(NA, length(temps), k1) #Sparar antalet p värdes fö
      rdelningar med medveten bias från H_0
61 tempH_1f < matrix(NA, length(temps), k2) #Sparar antalet p värdes fö
      rdelningar med medveten bias från H_1
62
63 tempH_0 < matrix(NA, length(temps), k1) #Sparar antalet p värdes fö
      rdelningar med stickprov från H_0
64 tempH_1 < matrix(NA, length(temps), k2) #Sparar antalet p värdes fö
      rdelningar med stickprov från för H_1
65
66 PPV < matrix(0, length(temps), 4) #Sparar PPV med och utan bias för
      varje stickprov
67
68 alpha < 0.05 #signifikansnivån
69 mu_0 < 0 #Medelvärdet för fördelningen under H_0
70 mu_1 < 1 #Medelvärdet för fördelningen under H_1
71 sigma < 5 #Standardavvikelsen för den fördelning vi drar stickprov frå
      n
72 b < 0.1 #Andelen studier som påverkats av omedveten bias
73
74 plotflag < FALSE #Bestämmer om utfallsrummet för mängderna ska framstä
      llas grafiskt.
75
76 f_PV < function(p, mu_1, mu_0, sigma, stickprov) { #Teoretiska tä
      thetsfunktion för p värden
77   y < dnorm((qnorm(1 - p)) * (mu_1 - mu_0) / (sigma / sqrt(stickprov))
      ) / dnorm((qnorm(1 - p)))
78   return(y)
79 }
80
81 for (q in 1:length(temps)) {
82   stickprov < temps[q] #Storleken på stickproven alltså datan som en
      specifik studie använder sig av för att göra ett hypotestest.
      Samma stickprov används för båda fördelningarna.
83

```

```

84  if (plotflag == TRUE) { #Vi väljer optimala värden för mängderna
85    k1 < 200 #Totala antalet studier från H_0
86    k2 < 200 #Totala antalet studier från H_1
87    stickprov < 33 #Storleken på stickproven alltså datan som en
      specifik studie använder sig av för att göra ett hypotestest.
      Samma stickprov används för båda fördelningarna.
88  }
89
90  if (PPVplot == 1) { #Vi väljer optimala värden för mängderna
91    k1 < tempk1[q] #Totala antalet studier från H_0
92  }
93
94  if (PPVplot == 2) { #Vi väljer optimala värden för mängderna
95    k2 < tempk2[q] #Totala antalet studier från H_1
96  }
97
98  #####Genererar p värden#####
99
100  pval1 < rep(0, k1) #Vektor av p värden med stickprov från H_0
101  pval2 < rep(0, k2) #Vektor av p värden med stickprov från H_1
102  pvaltemp1 < rep(NA, k1) #Vektor av p värden från stickprov med
      medveten bias från H_0
103  pvaltemp2 < rep(NA, k2) #Vektor av p värden från stickprov med
      medveten bias från H_1
104
105  for (i in 1:k1) { #k1 stycken hypotestest där studierna antas vara fr
      ån H_0
106    experimental_0 < rnorm(stickprov, mean = mu_0, sd = sigma) #Drar
      stickprov från fördelningen av H_0
107    Z < (mean(experimental_0) - mu_0) / (sigma / sqrt(stickprov)) #
      Test statistiska från normal fördelningen
108    pval1[i] < 1 - pnorm(Z, 0, 1) #Beräknar p värdet från en standard
      normal fördelning
109    if (pval1[i] > alpha & stickprov > 30) { #Om vi inte får ett
      signifikant p värde då minskar vi stickprovet
110      experimental_0 < sort(experimental_0) #Sorterar alla stickprov
111      for (j in 1:(stickprov - 30)) { #Vi försöker få ett potentiellt
      signifikant p värde som kommer till följd av detta lida av
      medveten bias
112        experimental_0 < experimental_0[2:length(experimental_0)] #Vi
      tar bort det minsta stickprovet från den totala mängden
113        Z < (mean(experimental_0) - mu_0) / (sigma / sqrt(length(
      experimental_0))) #Vi minskar stickproves storleken för att
      komma ifrån mu_0 för att få signifikanta p värden
114        pvaltemp1[i] < 1 - pnorm(Z, 0, 1) #Beräknar p värdet från en
      standard normal fördelning
115        if (pvaltemp1[i] <= alpha) {#Om vi lyckas få ett signifikant p
      värde med medveten bias avslutar vi loopen tidigt
116          break
117        }
118      }
119    }
120  }
121
122  for (i in 1:k2) { #k2 stycken hypotestest där studierna antas vara fr
      ån H_1

```

```

123   experimental_H_1 <- rnorm(stickprov, mean = mu_1, sd = sigma) #Vi
      drar stickprov från H_1
124   #Vid stort stickprov antas medelvärdet av stickrproven bli lika med
      den alternativa fördelningens medelvärde
125   Z = (mean(experimental_H_1) - mu_0) / (sigma / sqrt(stickprov)) #
      Test statistiska från normal fördelningen
126   pval2[i] < 1 - pnorm(Z, 0, 1) #Beräknar p värdet från en standard
      normal fördelning
127   if (pval2[i] > alpha & stickprov > 30) { #Om vi inte får ett
      signifikant p värde då minskar vi stickprovet
128     experimental_H_1 <- sort(experimental_H_1) #Sorterar alla
      stickproven
129     for (j in 1:(stickprov - 30)) { #Vi försöker få ett potentiellt
      signifikant p värde som kommer till följd av detta lida av
      medveten bias
130     experimental_H_1 <- experimental_H_1[2:length(experimental_H_1)
      ] #Vi tar bort det minsta stickprovet från den totala mäs-
      tingen
131     Z <- (mean(experimental_H_1) - mu_1) / (sigma / sqrt(length(
      experimental_H_1))) #Vi minskar stickproves storleken för
      att komma närmare mu_1 för att få signifikanta p värden
132     pvaltemp2[i] < 1 - pnorm(Z, 0, 1) #Beräknar p värdet från en
      standard normal fördelning
133     if (pvaltemp2[i] <= alpha) { #Om vi lyckas få ett signifikant p
      värde avslutar vi loopen tidigt
134       break
135     }
136   }
137 }
138 }
139
140 if(stickprov > 30){ #Vi kan bara vara påverkade av medveten bias om vi
      har ett stickprov större än 30
141
142   pval1f <- c(pval1[pval1 <= alpha], pvaltemp1[!is.na(pvaltemp1)]) #
      Slår ihop de signifikanta p värdena samt de nya p värdena som ä-
      r påverkade av medveten bias från H_0
143
144   pval11f <- sample(pval1f, k1, replace = FALSE) #Vi slumpar
      ordningen av p värdena
145
146   pval2f <- c(pval2[pval2 <= alpha], pvaltemp2[!is.na(pvaltemp2)]) #
      Slår ihop de signifikanta p värdena samt de nya p värdena som ä-
      r påverkade av medveten bias från H_1
147
148   pval22f <- sample(c(pval2[pval2 <= alpha], pval2f), k2, replace =
      FALSE) #Vi slumpar ordningen av p värdena
149
150 }
151
152 #####Mängder utan medveten bias#####
153
154 H_1 <- pval2 #Mängden signifikanta p värden från H_1
155
156 H_0 <- pval1 #Mängden signifikanta p värden från H_0
157

```

```

158 S1 < rep(NA, k1) #Mängden signifikanta p värden från H_0
159
160 for (i in 1:k1) { #Väljer p värdena i mängden H_0 som är mindre än
    eller lika med signifikans nivån
161   if (H_0[i] <= alpha) {
162     S1[i] < H_0[i]
163     H_0[i] < NA
164   }
165 }
166
167 S2 < rep(NA, k2) #Mängden signifikanta p värden från H_1
168
169 for (i in 1:k2) { #Väljer p värdena i mängden H_1 som är mindre än
    eller lika med signifikans nivån
170   if (H_1[i] <= alpha) {
171     S2[i] < H_1[i]
172     H_1[i] < NA
173   }
174 }
175
176 temp < round(length(S1) * b) #Antalet p värden som är påverkade av
    omedveten bias.
177
178 B1 < matrix(NA, k1, 2) #Mängden studier som har påverkat av
    omedveten bias slumpas fram från mängden H_1
179
180 for (i in 1:temp) { #Ser till att de p värden som är påverkade av
    omedveten bias hamnar på rätt plats i B2 vektorn
181   B1[i, 1] < H_0[i]
182   B1[i, 2] < S1[i]
183 }
184
185 H_0[1:temp] < NA #Vi sätter NA för att inte få samma p värde i mer än
    en mängd.
186
187 S1[1:temp] < NA #Vi sätter NA för att inte få samma p värde i mer än
    en mängd.
188
189 temp < round(length(S2) * b) #Antalet p värden som är påverkade av
    omedveten bias.
190
191 index < length(S2) + temp #Start index för vår mängd B2
192
193 B2 < matrix(NA, k2, 2) #Mängden studier som har påverkat av
    omedveten bias slumpas fram från mängden H_1
194
195 for (i in index:k2) { #Ser till att de p värden som är påverkade av
    omedveten bias hamnar på rätt plats i B2 vektorn
196   B2[i, 1] < H_1[i]
197   B2[i, 2] < S2[i]
198 }
199
200 H_1[index:length(H_1)] < NA #Vi sätter NA för att inte få samma p vä
    rde i mer än en mängd
201
202 S2[index:length(S2)] < NA #Vi sätter NA för att inte få samma p vä

```

```

rde i mer än en mängd
203
204 #####Mängder med medveten bias#####
205
206 if (stickprov>30){
207
208   H_1f < pval22f #Mängden studier från H_1 när vi har fått fram
      signifikanta p värden på grund av medveten bias
209
210   H_0f < pval11f #Mängden studier från H_0 när vi har fått fram
      signifikanta p värden på grund av medveten bias
211
212   S1f < rep(NA, k1) #Signifikanta p värden med medveten bias från mä
      ngden H_0
213
214   for (i in 1:k1) { #Väljer p värdena med medveten bias från mängden
      H_0 som är mindre än eller lika med signifikans nivån
215     if (H_0f[i] <= alpha) {
216       S1f[i] < H_0f[i]
217       H_0f[i] < NA
218     }
219   }
220
221   S2f < rep(NA, k2) #Mängden signifikanta p värden med medveten bias
      från mängden H_1
222
223   for (i in 1:k2) { #Väljer p värdena med medveten bias i mängden H_1
      som är mindre än eller lika med signifikans nivån
224     if (H_1f[i] <= alpha) {
225       S2f[i] < H_1f[i]
226       H_1f[i] < NA
227     }
228   }
229
230   temp < pval1f[pval1f <= alpha] #Mängden signifikanta p värden med
      medveten bias
231
232   index < round(k1 / length(temp)) #Steg längden
233
234   f1 < rep(NA, k1) #Mängden p värden som har påverkat av bias
235
236   for (i in 1:length(temp)) { #Ser till att p värdena blir placerade
      med jämnt intervall i en vektor
237     f1[i * index] < temp[i]
238   }
239
240   temp < round(length(S1f) * b) #Antalet p värden som är påverkade
      av bias (både omedveten och medveten bias)
241
242   B1f < matrix(NA, k1, 3) #Mängden p värden som har påverkats av
      bias slumpas fram från mängden H_1
243
244   for (i in 1:temp) { #Ser till att de p värdena som är påverkade av
      bias hamnar på rätt plats i B1 vektorn
245     B1f[i, 1] < H_0f[i]
246     B1f[i, 2] < S1f[i]

```



```

247     B1f[i, 3] < f1[i]
248   }
249
250   H_0f[1:temp] < NA #Vi sätter NA för att inte få samma p värde i
      mer än en mängd
251
252   S1f[1:temp] < NA #Vi sätter NA för att inte få samma p värde i mer
      än en mängd
253
254   f1[1:temp] < NA #Vi sätter NA för att inte få samma p värde i mer
      än en mängd
255
256   temp < pval2f[pval2f <= alpha] #Mängden signifikanta p värden på
      verkade av medveten bias
257
258   index < round(k2 / length(temp)) #Steg längden
259
260   f2 < rep(NA, k2) #Mängden p värden som har påverkat av bias (både
      omedveten och medveten bias)
261
262   for (i in 1:length(temp)) { #Ser till att p värdena blir placerade
      med jämnt intervall i en vektor
263     f2[i * index] < temp[i]
264   }
265
266   temp < round(length(S2f) * b) #Antalet p värden som är påverkade
      av bias.
267
268   index < length(S2f) temp #Start index för vår mängd S2f
269
270   B2f < matrix(NA, k2, 3) #Mängden studier som har påverkat av bias
      slumpas fram från mängden H_1
271
272   for (i in index:k2) { #Ser till att p värdena som är påverkade av
      bias hamnar på rätt plats i B2 vektorn
273     B2f[i, 1] < H_1f[i]
274     B2f[i, 2] < S2f[i]
275     B2f[i, 3] < f2[i]
276   }
277
278   H_1f[index:length(H_1f)] < NA #Vi sätter NA för att inte få samma
      p värde i mer än en mängd
279
280   S2f[index:length(S2f)] < NA #Vi sätter NA för att inte få samma p
      värde i mer än en mängd
281
282   f2[index:length(f2)] < NA #Vi sätter NA för att inte få samma p vä
      rde i mer än en mängd
283 }
284
285 if (plotflag == TRUE) { #För att grafiskt sammanställa utfallsrummet
      avslutar vi loopen här eftersom det krävs bara en iteration för
286 #att få värdena som bygger upp utfallsrummet
287   break
288 }
289

```

```

290 #####PPV#####
291
292 PPV[q,1] < length(pval2[pval2 <= alpha]) / (length(pval1[pval1 <=
alpha]) + length(pval2[pval2 <= alpha])) #PPV. Eftersom S är på
verkad av om vi beräknar omedveten bias i vår model kan vi inte
använda S och använder istället length(pvali[pvali <= alpha]) (i
=1 eller 2 beroende på vilken mängde vi använder). Samma sak gä
ller för H_1 och vi använder istället pval2 då.
293 PPV[q,2] < length(pval2[pval2 <= alpha]) / (length(S1[!is.na(S1)] +
length(S2[!is.na(S2)]) + length(B1[!is.na(B1)]) + length(B2[!is.
na(B2)])) #PPV med omdeveten bias. För H_1 gäller samma sak som
ovan.
294
295 tempH_0[q,1:k1] < pval1 #Sparar våra p värden från H_0 för ett
specifikt stickprov
296 tempH_1[q,1:k2] < pval2 #Sparar våra p värden från H_1 för ett
specifikt stickprov
297
298 if (stickprov>30 ){ #Om vi har ett stickprov på mer än 30 lider minst
några av våra p värden av medveten bias
299 PPV[q,3] < length(pval22f[pval22f <= alpha]) / (length(pval11f[
pval11f <= alpha]) + length(pval22f[pval22f <= alpha])) #PPV
med medveten bias Eftersom S1f är påverkad av om vi beräknar
bias i vår model kan vi inte använda S1f och använder istället
length(pvaliif[pvaliif <= alpha]) (i=1 eller 2 beroende på
vilken mängde vi använder). Samma sak gäller för H_1f och vi
använder istället pval22f då.
300 PPV[q,4] < length(pval22f[pval22f <= alpha]) / (length(S1f[!is.na(
S1f)]) + length(S2f[!is.na(S2f)]) + length(B1f[!is.na(B1f)]) +
length(B2f[!is.na(B2f)])) #PPV med bias. För H_1f gäller samma
sak som ovan.
301
302 tempH_0f[q,1:k1] < pval11f #Sparar våra p värden med medveten bias
från H_0 för ett specifikt stickprov
303 tempH_1f[q,1:k2] < pval22f #Sparar våra p värden med medveten bias
från H_1 för ett specifikt stickprov
304 } else {
305 tempH_0f[q,1:k1] < pval1 #Sparar våra p värden från H_0 för ett
specifikt stickprov
306 tempH_1f[q,1:k2] < pval2 #Sparar våra p värden från H_1 för ett
specifikt stickprov
307 }
308
309 }
310
311 #####Plot av p värden#####
312
313 if (plotflag == FALSE & studiefalg==FALSE & (PPVplot!=1 & PPVplot!=2))
{ #Vi skapar figurer för p värdes fördelningarna
314
315 color < c("black", "blue", "purple", "red", "green", "orange") # Fä
rger för figurer
316
317 #####Utan medveten bias#####
318 tempbreaks < c(20,19,18,16,13,601) #Behöver skala antalet breaks
manuellt

```

```

319
320 data < tempH_1[1,][tempH_1[1,] <= alpha]
321
322 breaks < tempbreaks[1]
323
324 data < data[data >= 0.05/breaks]
325
326 p < seq(0.05/length(data), 0.05, 0.05/length(data)) #Behöver skala
    antalet p värden som vi sätter in i täthetsfunktionen med antalet
    breaks
327 #(kan tänka på det som antalet staplar i ett histogram förutom att
    vi här inte ritar staplarna utan drar en linje kring dem)
328 #för våra simulerade p värden för att få figurer med samma antalet p
    värden så att vi kan jämföra dem.
329
330 h < hist(data, breaks = seq(min(data), max(data), length.out =
    breaks), plot = FALSE)
331 plot(x = h$mids, y = 100 * h$density / sum(h$density), ylim=c(0,10),
    xlim=c(0,0.05), ylab = "Procent av p värden [%]", xlab = "P vä
    rden ", main = expression("Fördelning av p värden under H" [1]),
    type = "l", lty=1, col=color[1]) #Ritar fördelningen av p värden
    under H_1. Notera skev fördelning åt höger.
332 lines(p, f_PV(p, mu_1, mu_0, sigma, temps[1]), type = 'l', lwd=2, lty
    = 2, col=color[1])
333
334 for (q in 2:length(temps)){ #Sammanställer grafiskt resten av våra p
    värden för andra stickprov
335
336 data < tempH_1[q,][tempH_1[q,] <= alpha]
337
338 breaks < tempbreaks[q]
339
340 data < data[data >= 0.05/breaks]
341
342 p < seq(0.05/length(data), 0.05, 0.05/length(data))
343
344 h < hist(data, breaks = seq(min(data), max(data), length.out =
    breaks), plot = FALSE)
345 lines(x = h$mids, y = 100 * h$density / sum(h$density), type = "l",
    lty=1, col=color[q]) #Ritar fördelningen av p värden under H_1
    . Notera skev fördelning åt höger.
346 lines(p, f_PV(p, mu_1, mu_0, sigma, temps[q]), type = 'l', lwd=2,
    lty = 2, col=color[q])
347
348 }
349
350 legend( x = "topright", legend = c(expression("n"[0]*"=30"),
    expression("n"[1]*"=35"), expression("n"[1]*"=40"), expression("n
    "[1]*"=50"), expression("n"[1]*"=100"), expression("n"[1]*"=1000"))
    , col = c("black", "blue", "purple", "red", "green", "orange"),
    lty = 1, cex = 0.6)
351
352 data < tempH_0[1,][tempH_0[1,] <= alpha]
353
354 breaks < 100
355

```

```

356 data <- data[data >= 0.05/breaks]
357
358 p <- seq(0.05/length(data), 0.05, 0.05/length(data)) #Interval för
    teoretiska p värden
359
360 h <- hist(data, breaks = seq(min(data), max(data), length.out =
    breaks), plot = FALSE)
361 plot(x = h$mids, y = 100 * h$density / sum(h$density), ylim=c(0,2),
    xlim=c(0,0.05), ylab = "Procent av p värden [%]", xlab = "P vä
    rden ", main = expression("Fördelning av p värden under H" [0]),
    type = "l", lty=1, col=color[1]) #Ritar fördelningen av p värden
    under H_0. Notera likformig fördelning.
362 lines(p, f_PV(p, mu_0, mu_0, sigma, temps[1]), type = 'l', lwd=2, lty
    = 2, col=color[1])
363
364
365 for (q in 2:length(temps)){
366
367     data <- tempH_0[q,][tempH_0[q,] <= alpha]
368
369     data <- data[data >= 0.05/breaks]
370
371     p <- seq(0.05/length(data), 0.05, 0.05/length(data))
372
373     h <- hist(data, breaks = seq(min(data), max(data), length.out =
    breaks), plot = FALSE)
374     lines(x = h$mids, y = 100 * h$density / sum(h$density), type = "l",
    lty=1, col=color[q]) #Ritar fördelningen av p värden under H_0
    . Notera likformig fördelning.
375 }
376 legend( x = "bottomleft", legend = c(expression("n"[0]*"=30"),
    expression("n"[1]*"=35"), expression("n"[1]*"=40"), expression("n
    "[1]*"=50"), expression("n"[1]*"=100"), expression("n"[1]*"=1000"))
    , col = c("black", "blue", "purple", "red", "green", "orange"),
    lty = 1, cex = 0.6)
377
378
379 data <- cbind(tempH_1[1, ], tempH_0[1,])
380 h <- hist(data[data <= alpha], breaks = 100, plot = FALSE)
381 plot(x = h$mids, y = 100 * h$density / sum(h$density), ylim = c(0, 5)
    , ylab = "Procent av p värden [%]", xlab = "P värden ", main =
    expression("Fördelning av kombinerade p värden"), type = "l")
382
383 for (q in 2:length(temps)) {
384     data <- cbind(tempH_1[q, ], tempH_0[q,])
385     h <- hist(data[data <= alpha], breaks = 100, plot = FALSE)
386     lines(x = h$mids, y = 100 * h$density / sum(h$density), type = 'l',
    lty = q, col = color[q]) #Ritar fördelningen av kombinerade p
    värden under H_0 och H_1
387 }
388 legend( x = "topright", legend = c(expression("n"[0]*"=30"),
    expression("n"[1]*"=35"), expression("n"[1]*"=40"), expression("n
    "[1]*"=50"), expression("n"[1]*"=100"), expression("n"[1]*"=1000"))
    , col = c("black", "blue", "purple", "red", "green", "orange"),
    lty = 1:length(temps), cex = 0.7)
389

```

```

390
391 #####Medveten bias#####
392
393 tempdata < tempH_0f[length(temps), ]
394 tempH < hist(tempdata[tempdata <= alpha], breaks = 100, plot = FALSE
    ) #Vi behöver det största procent antalet från stickproven på y
    axeln som vi får från detta stickprov
395
396 data < tempH_0f[1, ]
397 h < hist(data[data <= alpha], breaks = 100, plot = FALSE)
398 plot(x = h$mids, y = 100 * h$density / sum(h$density), ylim = c(0, 5)
    , ylab = "Procent av p värden [%]", xlab = "P värden ", main =
    expression("Fördelning av p värden med medveten bias under H"
    [0]), type = "l", col = color[1], lty = 1)
399 #Ritar fördelningen av p värden under H_0 utan medveten bias
400
401 for (q in 2:length(temps)) {
402   data < tempH_0f[q, ]
403   h < hist(data[data <= alpha], breaks = 100, plot = FALSE)
404   lines(x = h$mids, y = 100 * h$density / sum(h$density), type = 'l',
    lty = q, col = color[q]) #Ritar fördelningen av p värden under
    H_0 med medveten bias. Notera en vänster skev fördelning.
405 }
406 legend(x = "topleft", legend = c(expression("n"[0]*"=30"), expression
    ("n"[1]*"=35"), expression("n"[1]*"=40"), expression("n"[1]*"=50"),
    expression("n"[1]*"=100"), expression("n"[1]*"=1000")), col = c("
    black", "blue", "purple", "red", "green", "orange"), lty = 1:
    length(temps), cex = 0.7)
407
408 data < tempH_1f[1, ]
409 h < hist(data[data <= alpha], breaks = 100, plot = FALSE)
410 plot(x = h$mids, y = 100 * h$density / sum(h$density), ylim = c(0, 5)
    , ylab = "Procent av p värden [%]", xlab = "P värden ", main =
    expression("Fördelning av p värden med medveten bias under H"
    [1]), type = "l")
411 #Ritar fördelningen av p värden under H_1 utan medveten bias
412
413 for (q in 2:length(temps)) {
414   data < tempH_1f[q, ]
415   h < hist(data[data <= alpha], breaks = 100, plot = FALSE)
416   lines(x = h$mids, y = 100 * h$density / sum(h$density), type = 'l',
    lty = q, col = color[q]) #Ritar fördelningen av p värden under
    H_1 med medveten bias. Notera en u formad fördelning när vi
    har medveten bias.
417 }
418 legend( x = "topright", legend = c(expression("n"[0]*"=30"),
    expression("n"[1]*"=35"), expression("n"[1]*"=40"), expression("n"
    "[1]*"=50"), expression("n"[1]*"=100"), expression("n"[1]*"=1000"))
    , col = c("black", "blue", "purple", "red", "green", "orange"),
    lty = 1:length(temps), cex = 0.7)
419
420 data < cbind(tempH_1f[1, ], tempH_0f[1,])
421 h < hist(data[data <= alpha], breaks = 100, plot = FALSE)
422 plot(x = h$mids, y = 100 * h$density / sum(h$density), ylim = c(0, 5)
    , ylab = "Procent av p värden [%]", xlab = "P värden ", main =
    expression("Fördelning av kombinerade p värden med medveten bias

```

```

    ), type = "l")
423
424 for (q in 2:length(temps)) {
425   data <- cbind(tempH_1f[q, ], tempH_0f[q, ])
426   h <- hist(data[data <= alpha], breaks = 100, plot = FALSE)
427   lines(x = h$mids, y = 100 * h$density / sum(h$density), type = 'l',
         lty = q, col = color[q]) #Ritar fördelningen av kombinerade p
         värden under H_0 och H_1 med medveten bias. Notera u formad fö
         rdelning.
428 }
429 legend( x = "top", legend = c(expression("n"[0]*"=30"), expression("n
         "[1]*"=35"), expression("n"[1]*"=40"), expression("n"[1]*"=50"),
         expression("n"[1]*"=100"), expression("n"[1]*"=1000")), col = c("
         black", "blue", "purple", "red", "green", "orange"), lty = 1:
         length(temps), cex = 0.7)
430
431 }
432
433 #####Jämförelse med kurvorna i artikeln "P Curve: A Key to the File
         Drawer av Joseph P. Simmons (2014)#####
434
435 if (studieflag==TRUE){
436
437   color <- c("black", "blue", "purple", "red", "green", "orange", "
         yellow", "cyan", "brown", "grey", "pink", "violet")
438
439   y <- c(45,23,14,5,14)
440
441   x <- c(0.01,0.02,0.03,0.04,0.05)
442
443   plot(x, y, xlim=c(0.011,0.049), ylim=range(0,50), xlab="P värden ",
         ylab="Procent av p värden [%]",
444         main = expression("Fördelning av p värden under H" [1]),pch=16)
445   lines(x, y, pch=16) #Ritar linjen från studien "P Curve: A Key to the
         File Drawer" av Joseph P. Simmons (2014)
446
447   data <- tempH_1[5, ] #Antar att vi har 5 olika stickprov här
448   data <- data[data<=alpha+0.006] #Vi behöver addera 0.006 eftersom
         kommandot lines vägrar att rita linjen ända fram till 0.05 utan
         att den.
449   breaks <- 6
450
451   h <- hist(data, breaks = seq(min(data), max(data), length.out =
         breaks), plot = FALSE)
452   lines(x = h$mids, y = 100 * h$density / sum(h$density), type = "l",
         lty = 2, col = color[2]) #Ritar fördelningen av p värden under
         H_0 med medveten bias
453
454   tempd <- matrix(0,10,100)
455
456   for (q in 1:10) {
457
458     data <- tempH_1[5, ]
459     data <- data[data<=alpha+0.006]
460     data <- sample(data, 100, replace = FALSE)
461     h <- hist(data, breaks = seq(min(data), max(data), length.out =

```

```

breaks), plot = FALSE)
462 lines(x = h$mids, y = 100 * h$density / sum(h$density), type = 'l',
      lty = 3, col= color[3]) #Ritar fördelningen av p värden under H_0
      med medveten bias.
463 tempd[q,] < data
464 }
465
466 legend(x = "topright", legend = c("P värden från Simmons",expression
      ("n"[0]*"=30"),expression("Simulerade p väden då n"[1]*"=100")),
      col = c("black", "blue", "purple", "red", "green", "orange",
      yellow", "cyan", "brown", "grey"), lty = 1:3, cex = 0.55)
467
468 y < c(5,5,15,35,40)
469
470 x < c(0.01,0.02,0.03,0.04,0.05)
471
472 plot(x, y, xlim=c(0.011,0.049), ylim=c(0,50), xlab="P värden ", ylab
      ="Procent av p värden [%]",
473      main = expression("P värdes fördelning med medveten bias i H"
      [0]),pch=16)
474 lines(x, y, pch=16) #Ritar linjen från studien "P Curve: A Key to the
      File Drawer" av Joseph P. Simmons (2014)
475
476 data < tempH_0[1, ]
477 data < data[data<=alpha+0.006] #Vi behöver addera 0.006 eftersom
      kommandot lines vägrar att rita linjen ända fram till 0.05 utan
      att den.
478 breaks < 6
479
480 h < hist(data, breaks = seq(min(data), max(data), length.out =
      breaks), plot = FALSE)
481 lines(x = h$mids, y = 100 * h$density / sum(h$density), type = "l",
      lty = 2, col = color[2]) #Ritar fördelningen av p värden under
      H_0
482
483 temp_p < c(34) 30 #Positionerna för de olika stickproven som vi vill
      ha
484
485 tempd < matrix(0,10,100)
486
487 for (q in 1:10) {
488
489 data < tempH_of[temp_p[1], ]
490 data < data[data<=alpha+0.006]
491 data < sample(data, 100, replace = FALSE)
492 h < hist(data, breaks = seq(min(data), max(data), length.out =
      breaks), plot = FALSE)
493 lines(x = h$mids, y = 100 * h$density / sum(h$density), type = 'l',
      lty = 3, col= color[3]) #Ritar fördelningen av p värden under H_0
      med medveten bias.
494 tempd[q,] < data
495 }
496
497 legend( x = "topleft", legend = c("P värden från Simmons",expression
      ("n"[0]*"=30"),expression("Simulerade p väden då n"[1]*"=34")),
      col = c("black", "blue", "purple"), lty = 1:3, cex = 0.55)

```

```

498 }
499 }
500
501 #####PPV plot#####
502
503 if (PPVplot==1){ #Ritar förhållandet mellan olika mängder studier i H_0
    och H_1
504
505     plot(tempx, PPV[,1]*100, type='l', ylim = c(85, 100), yaxt = "n",
        ylab = "Procent av PPV [%]", xlab = expression("Förhållandet
        mellan studier från H"[0]*"och H"[1]*" [%]"), main = expression("
        Fördelning av PPV")) #Fördelning av PPV under olika reaktioner
        mellan H"[0]*"och H"[1]"
506
507     axis(2, at=c(85,90,95,100), labels=c(85,90,95,100), las=2)
508
509     plot(tempx,PPV[,2]*100,type='l', ylim = c(60, 90), yaxt = "n", ylab =
        "Procent av PPV [%]", xlab = expression("Förhållandet mellan
        studier från H"[0]*"och H"[1]*" [%]"), main = expression("Fö
        rdelning av PPV med omedveten bias")) #Fördelning av PPV med
        omedveten bias under olika reaktioner mellan H"[0]*"och H"[1]"
510
511     axis(2, at=c(60,70,80,90), labels=c(60,70,80,90), las=2)
512
513     plot(tempx,PPV[,3]*100,type='l', ylim = c(50,100), yaxt = "n", ylab =
        "Procent av PPV [%]", xlab = expression("Förhållandet mellan
        studier från H"[0]*"och H"[1]*" [%]"), main = expression("Fö
        rdelning av PPV med medveten bias" )) #Fördelning av PPV med
        medveten bias under olika reaktioner mellan H"[0]*"och H"[1]"
514
515     axis(2, at=c(50,60,70,80,90,100), labels=c(50,60,70,80,90,100), las=2)
516
517     plot(tempx,PPV[,4]*100,type='l', ylim = c(40, 100), yaxt = "n", ylab
        = "Procent av PPV [%]", xlab = expression("Förhållandet mellan
        studier från H"[0]*"och H"[1]*" [%]"), main = expression("Fö
        rdelning av PPV med bias")) #Fördelning av PPV bias under olika
        reaktioner mellan H"[0]*"och H"[1]"
518
519     axis(2, at=c(40,50,60,70,80,90,100), labels=c(40,50,60,70,80,90,100),
        las=2)
520
521 }
522
523 if (PPVplot==2){ #Ritar förhållandet mellan olika mängder studier i H_1
    och H_0
524
525     plot(tempx,PPV[,1]*100,type='l', ylim = c(0, 100), yaxt = "n",ylab =
        "Procent av PPV [%]", xlab = expression("Förhållandet mellan
        studier från H"[1]*"och H"[0]*" [%]"), main = expression("Fö
        rdelning av PPV")) #Fördelning av PPV under olika reaktioner
        mellan H"[0]*"och H"[1]"
526
527     axis(2, at=c(0,20,40,60,80,100), labels=c(0,20,40,60,80,100), las=2)
528
529     plot(tempx,PPV[,2]*100,type='l',ylim = c(0, 70), yaxt = "n",ylab = "
        Procent av PPV [%]", xlab = expression("Förhållandet mellan

```



```

studier från H"[1]*"och H"[0]*" [%]"), main = expression("Fö
rdelning av PPV med omedveten bias")) #Fördelning av PPV med
omedveten bias under olika reaktioner mellan H"[0]*"och H"[1]"
530
531 axis(2, at=c(0,10,20,30,40,50,60,70), labels=c
(0,10,20,30,40,50,60,70), las=2)
532
533 plot(tempx,PPV[,3]*100,type='l',ylim = c(0, 50), yaxt = "n",ylab = "
Procent av PPV [%]", xlab = expression("Förhållandet mellan
studier från H"[1]*"och H"[0]*" [%]"), main = expression("Fö
rdelning av PPV med medveten bias" )) #Fördelning av PPV med
medveten bias under olika reaktioner mellan H"[0]*"och H"[1]"
534
535 axis(2, at=c(0,10,20,30,40,50), labels=c(0,10,20,30,40,50), las=2)
536
537 plot(tempx,PPV[,4]*100,type='l',ylim = c(0, 50), yaxt = "n",ylab = "
Procent av PPV [%]", xlab = expression("Förhållandet mellan
studier från H"[1]*"och H"[0]*" [%]" ), main = expression("Fö
rdelning av PPV med bias")) #Fördelning av PPV bias under olika
reaktioner mellan H"[0]*"och H"[1]"
538
539 axis(2, at=c(0,10,20,30,40,50), labels=c(0,10,20,30,40,50), las=2)
540
541 }
542
543 #####Plot av mängder#####
544 if (plotflag == TRUE) { #Vi skapar mängderna och deras figurer
545 plot(H_0[H_0 > alpha], ylim = c(0, 1), pch = 1, yaxt = "n", col = "
black", ylab = "", xlab = "Studier", main = expression("H" [0]))
#Figure för p värden frpn mängden H_0
546 points(S1, pch = 1, col = "black")
547 points(B1[1:k1, 1], pch = 1, col = "black")
548 points(B1[1:k1, 2], pch = 1, col = "black")
549 abline(h = 0.05, col = "black")
550 abline(v = b * k1, col = "black")
551
552 plot(H_1, col = "black" , ylim = c(0, 1), pch = 1, yaxt = "n", ylab =
"P värden", xlab = "Studier", main = expression("H" [1])) #
Figure för p värden från mängden H_1
553 points(S2, pch = 1, col = "black")
554 points(B2[1:k2, 1], pch = 1, col = "black")
555 points(B2[1:k2, 2], pch = 1, col = "black")
556 abline(v = (1 - b) * k2, col = "black")
557 abline(h = 0.05, col = "black")
558
559 plot( H_0f, col = "black", ylim = c(0, 1), pch = 1, yaxt = "n", ylab
= "", xlab = "Studier", main = expression("Stickprov med bias
under H" [0])) #Figure för p värden i mängden H_0 med bias
560 points(S1f, pch = 1, col = "black")
561 points(B1f[1:k1, 1], pch = 1, col = "black")
562 points(B1f[1:k1, 2], pch = 1, col = "black")
563 points(B1f[1:k1, 3], pch = 1, col = "black")
564 points(f1, pch = 16, col = "black")
565 abline(v = b * k1, col = "black")
566 abline(h = 0.05, col = "black")
567

```

```

568 plot(H_1f, col = "black" , ylim = c(0, 1), pch = 1, yaxt = "n", ylab
      = "P värden", xlab = "Studier", main = expression("Stickprov med
      bias under H" [1])) #Figure för p värden i mängden H_1 med bias
569 points(S2f, pch = 1, col = "black")
570 points(B2f[1:k2, 1], pch = 1, col = "black")
571 points(B2f[1:k2, 2], pch = 1, col = "black")
572 points(B2f[1:k2, 3], pch = 1, col = "black")
573 points(f2, pch = 16, col = "black")
574 abline(v = (1 - b) * k2, col = "black")
575 abline(h = 0.05, col = "black")
576
577 }

```