Stockholms universitet

Sekventiellt t-test av skillnaden i väntevärden mellan två normalfördelade stickprov

Jenny Areskogh

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2009:9
Matematisk statistik
September 2009

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematisk statistik
Stockholms universitet
Kandidatuppsats **2009:9**,
http://www.math.su.se/matstat

Sekventiellt t-test av skillnaden i väntevärden mellan två normalfördelade stickprov

Jenny Areskogh*

September 2009

Sammanfattning

I den här uppsatsen undersöker vi egenskaper hos tre olika t-test. Antag att vi har två oberoende och normalfördelade stickprov $X_1, ..., X_{n_1}$ och $Y_1, ..., Y_{n_2}$. Vidare är både väntevärdena μ_1 och μ_2 samt varianserna σ_1^2 och σ_2^2 för respektive stickprov okända. Vid test av nollhypotesen $H_0 : \mu_1 = \mu_2$ används Student's t-test om $\sigma_1^2 = \sigma_2^2$, annars används vanligen Welch's t-test. Om vi inte har tillgång till information om huruvida $\sigma_1^2 = \sigma_2^2$ eller ej, kan vi utföra ett sekventiellt t-test. Först testar vi med hjälp av ett F-test om likhet i varianserna kan antas eller ej, sedan utför vi antingen Student's t-test eller Welch's t-test. Målsättningen är att komma fram till kritiska gränser i detta sekventiella t-test som leder till ett bättre t-test än att utföra endast Student's t-test eller Welch's t-test. Genom simuleringar av storlek 10.000 märker vi att Student's t-test får fel signifikansnivå om det utförs då $\sigma_1^2 \neq \sigma_2^2$ och bör alltså inte användas under dessa förutsättningar. Samtidigt märker vi att Welch's t-test inte har mycket sämre styrka än Student's t-test, som är UMP då varianserna är lika. De kritiska gränser vi kommer fram till i det sekventiella t-testet är inte optimala då proceduren att hitta kritiska gränser i F-testet var en mycket svårare uppgift än förväntat. De sekventiella t-testen får en för låg signifikansnivå i fallet $\sigma_1^2 = \sigma_2^2$ samtidigt som Welch's t-test visar sig vara ett starkare t-test än dessa. Under de kritiska gränser som presenterats och använts i denna uppsats kommer vi fram till att Welch's t-test är att föredra framför de sekventiella t-testen.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: jenny.areskogh@hotmail.com. Handledare: Mikael Andersson.

Abstract

In this essay we are analyzing the properties of three different t-tests. Take in consideration that we have two independent and normally distributed samples $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_2}$. Furthermore, both the sample means μ_1 and μ_2 as well as the sample variances σ_1^2 and σ_2^2 of the randome samples is unknown. When testing the nullhypothesis $H_0 : \mu_1 = \mu_2$, the Student's t-test is used as $\sigma_1^2 = \sigma_2^2$, otherwise the Welch's t-test is commonly used. If the information about $\sigma_1^2 = \sigma_2^2$ is unavailable we can construct a sequential t-test. The sequential t-test starts with an F-test in order to see if we can assume that the variance is equal or not before we carry out the Student's t-test or the Welch's t-test. The ambition of this essay is to find the critical values in this specific sequential t-test, which will lead to a better t-test than the Student's t-test or the Welch's t-test. By simulating the Student's t-test with a size of 10.000 we notice that the significance level is inaccurate when $\sigma_1^2 \neq \sigma_2^2$, which concludes that it should not be used under these conditions. In addition to this, we notice that the Welch's t-test has almost as much power as the Student's t-test, which is UMP when the variance is equal. The critical values that we constructed in the sequential t-test is not optimal because the procedure to find the limits of the F-tests was much harder than expected. The sequential t-test has a significance level that is too low when $\sigma_1^2 = \sigma_2^2$ and in addition to this the Welch's t-test proves to be a more powerful t-test. We can draw the conclusion that under the critical limits that is presented and used in this essay, Welch's t-test is a better t-test compared to the sequential t-test.

Förord och tack

Denna uppsats utgör ett kandidatarbete i matematisk statistik om 15 högskolepoäng vid Stockholms Universitet. Ett stort tack riktas till min handledare Mikael Andersson, universitetslektor vid instutionen för matematisk statistik på Stockholms Universitet. Jag tackar för hans vägledning som har varit av största vikt under arbetets gång. Jag vill även rikta ett tack till nära vänner och familj som stöttat och peppat mig genom denna uppsats.

Innehåll

1 Introduktion

1.1 Inledning

Antag att vi har två oberoende och normalfördelade stickprov, $X_1, X_2, ..., X_{n_1}$ och $Y_1, Y_2, ..., Y_{n_2}$ där både väntevärdena μ_1 och μ_2 samt varianserna σ_1^2 och σ_2^2 för respektive stickprov är okända. Det är ofta av intresse att testa nollhypotesen $H_0 : \mu_1 = \mu_2$ mot den tvåsidiga alternativhypotesen $H_A : \mu_1 \neq \mu_2$ eller någon av de ensidiga alternativhypoteserna $H_A : \mu_1 > \mu_2$ och $H_A : \mu_1 < \mu_2$. Detta görs på olika sätt beroende av om vi kan göra antagandet $\sigma_1^2 = \sigma_2^2$ eller ej.

Ett exempel på när det är av intresse att testa nollhypotesen ovan är om vi vill undersöka huruvida en medicin har effekt på högt blodtryck. Vi kan anta att väntevärdena för gruppen som får medicin och för kontrollgruppen, som får placebo, är lika innan behandling. Det vi är intresserade av att testa är om väntevärdet för gruppen som får medicinen är lägre än väntevärdet för kontrollgruppen efter behandling. I detta exempel kan vi resonera på följande sätt om variansen för respektive grupp:

Vanligtvis antar vi att varianserna i de båda grupperna är lika i och med att den beskriver variationen i blodtryck och att de båda grupperna består av människor med högt blodtryck. Om behandlingen förväntas ha lika stor effekt på samtliga behandlade patienter är det rimligt att anta att variansen i gruppen är oförändrad. Är det däremot så att en del av de behandlade patienterna inte reagerar på behandlingen är det istället rimligt att anta att variansen i gruppen ökar. Detta ger oss två grupper med skilda varianser.

Under förutsättningen att varianserna är lika, d.v.s. $\sigma_1^2 = \sigma_2^2$, testas nollhypotesen $H_0 : \mu_1 = \mu_2$ med hjälp av Student's t-test. Den kritiska gränsen för testet bestäms av att teststatistikan är t-fördelad och antalet frihetsgrader beror av stickprovsstorlekarna. Har vi istället fallet med olika varianser, d.v.s. $\sigma_1^2 \neq \sigma_2^2$, finns ingen exakt lösning till hur nollhypotesen ovan ska testas. Welch's t-test är en anpassning av Student's t-test för just detta fall, då varianserna för de båda stickproven är skilda från varandra. I detta test är teststatistikan approximativt t-fördelad och antalet frihetsgrader är stokastiskt. Problemet är att det oftast inte finns information om ifall varianserna är lika eller inte.

För att hantera detta problem kan vi göra ett sekventiellt t-test där vi först utför ett F-test som visar om vi kan anta att varianserna är lika eller inte. Därefter utför vi antingen Student's t-test eller Welch's t-test beroende på utfallet av F-testet. Genom att använda denna metod tar vi hänsyn till stickprovsvarianserna när vi väljer vilket t-test som ska utföras.

1.2 Syfte och metod

Syftet med denna uppsats är att undersöka egenskaperna hos ett sekventiellt t-test med hjälp av styrkeberäkningar och simuleringar. Då vi utför det sekventiella t-testet vill vi att det ska ha den maximala styrkan bland alla tänkbara test med önskvärd signifikansnivå och självklart vill vi även dra korrekta slutsatser. Målsättningen är att hitta en procedur för att välja kritiska gränser i det sekventiella t-testet för att uppfylla dessa krav. För varje test kommer vi att göra 10.000 simuleringar för att få ett någorlunda säkert resultat. Notera att resultaten som redovisas i denna uppsats baseras på simuleringar av signifikansnivån och styrkan och därmed får vi inte exakta siffror utan kommer att ha en standardavvikelse på

$$\sqrt{\frac{p(1-p)}{10000}}. \tag{1}$$

I de fall vi använder signifikansnivån fem procent, p=0.05, får vi en standardavvikelse på 0.0022. Värt att notera är att maximal standardavvikelse är 0.005 som fås av p=0.5. Alltså kan vi dra relativt goda slutsatser om huruvida våra resultat ligger kring en viss nivå.

Vi kommer att generera normalfördelade slumptal som kommer att representera stickproven $X_1, .., X_{n_1}$ och $Y_1, .., Y_{n_2}$. Väntevärdet och variansen för de båda stickproven kommer att varieras för att kunna undersöka signifikansnivå och styrka för ett t-test av nollhypotesen $H_0 : \mu_1 = \mu_2$ under olika förutsättningar. Genom att först utföra F-testet och sedan välja något av de två t-testen, Student's t-test eller Welch's t-test, får vi ett sekventiellt t-test. De kritiska gränserna i de tre ingående testen väljer vi genom genererandet av dessa slumptal och numeriska beräkningar av styrkefunktioner. Vi kommer att få olika starka sekventiella t-test beroende av de kritiska gränserna under förutsättning att signifikansnivån för det sekventiella t-testet inte överstiger fem procent. Vi kommer att konstruera en tabell med numeriska värden på de kritiska gränserna vi använder i det sekventiella t-testet. Avslutningsvis kommer vi att jämföra det sekventiella t-testet med Student's t-test och Welch's t-test. Denna jämförelse kommer att illustreras med hjälp av bland annat styrkefunktioner.

2 Begrepp som används i denna uppsats

2.1 Signifikansnivå

Signifikansnivån för ett test är sannolikheten att förkasta en sann nollhypotes, d.v.s. P(förkasta $H_0 \mid H_0$ är sann). Om vi bestämmer oss för att vi inte vill att signifikansnivån ska överstiga fem procent innebär detta att inte mer än fem procent av gångerna vi utför testet ska en sann nollhypotes förkastas.

2.2 Kritisk gräns

Den kritiska gränsen för ett test bestämmer för vilka värden på teststatistikan som en hypotes ska förkastas. Utför vi ett tvåsidigt t-test, som i denna uppsats, förkastas nollhypotesen om $|T| > t$, där T är teststatistikan och t är den kritiska gränsen. Den kritiska gränsen, t, beror av signifikansnivån och antalet frihetsgrader.

2.3 Styrka och styrkefunktion

Styrkan i ett test är sannolikheten att förkasta nollhypotesen när denna är falsk och därmed alternativhypotesen är sann, d.v.s. P(förkasta $H_0 \mid H_0$ är falsk). Då alla andra omständigheter är lika gäller att ju större styrka ett test har desto bättre är det. Styrkefunktionen för ett test ger en bild av sannolikheten att en falsk nollhypotes förkastas. Den visar vilken styrka ett test har för olika värden på parametrarna under alternativhypotesen, d.v.s. för olika värden på skillnaden i väntevärdena.

2.4 Uniformly Most Powerful (UMP)

Att ett test är Uniformly Most Powerful, UMP, innebär att det har den största styrkan för varje parametervärde i alternativhypotesen. Antag att vi ska testa nollhypotesen $H_0 : \mu_1 = \mu_2$ mot alternativhypotesen $H_A : \mu_1 \neq \mu_2$. Då gäller att ett test är likformigt starkast om det för alla olika skillnader i väntevärdena, $|\mu_1 - \mu_2|$, har den största styrkan.

Mer detaljer kan hämtas i böckerna Lindgren (1993) samt Vejde och Leander (2000).

3 Test som används i denna uppsats

3.1 Student's t-test

Student's t-test utvecklades för ungefär 100 år sedan av William Gosset under pseudonymen Student, se *The probable error of mean* (1908), och har idag blivit ett av de främst använda testen. Antag att vi har fallet $\sigma_1^2 = \sigma_2^2$. Denna förutsättning gör det lämpligt att använda Student's t-test. I detta test har teststatistikan utseendet

$$T = \frac{\overline{X} - \overline{Y}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

och är t-fördelad med antalet frihetsgrader $\nu = n_1 + n_2 - 2$. I formeln för teststatistikan betecknar

$$s_p^2 = \frac{(n_1 - 1)s_1 + (n_2 - 1)s_2}{n_1 + n_2 - 2}$$

den sammanvägda variansen där s_1 och s_2 står för standardavvikelsen i respektive stickprov. Fördelen med Student's t-test är att det kan användas för små normalfördelade stickprov med okänd varians. Dessutom är detta test ett exakt test som är UMP (Uniformly Most Powerful), det vill säga är det test som har den största styrkan mot varje enkel hypotes i alternativhypotesen H_A. Om vi istället har fallet att varianserna är olika är T inte längre t-fördelad. Använder vi oss av Student's t-test under dessa förutsättningar ökar risken för att fel slutsatser dras, d.v.s. att vi förkastar en sann nollhypotes eller accepterar en falsk nollhypotes.

3.2 Welch's t-test

Welch's t-test utvecklades för ungefär 60 år sedan av B.L. Welch, se *The generalization of student's problem when several different population variances are involved* (1947) samt Sawilowsky (2002), och är idag det test som vanligen används när vi inte kan anta att $\sigma_1^2 = \sigma_2^2$. När vi har fallet $\sigma_1^2 \neq \sigma_2^2$ är teststatistikan T, ovan, inte längre t-fördelad. Det finns ingen exakt lösning till hur nollhypotesen $H_0 : \mu_1 = \mu_2$ ska testas, det så kallade Behrens-Fishers problem. Welch's t-test är en anpassning av Student's t-test för just detta fall, det vill säga då varianserna för de båda stickproven inte kan antas vara lika. I detta test har teststatistikan utseendet

$$T_w = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

och är approximativt t-fördelad med

$$\nu = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}$$

antal frihetsgrader. I dessa uttryck står s_1 och s_2 för standardavvikelsen i respektive stickprov. Vi ser i formeln för antalet frihetsgrader att de beror på standardavvikelsen i respektive stickprov och därmed kommer att vara stokastisk. I och med att Student's t-test är UMP kommer Welch's t-test att ha lägre styrka om vi har fallet att varianserna verkligen är lika.

3.3 Icke-centrala fördelningar

Antag att vi har ett normalfördelat stickprov och vill testa nollhypotesen $H_0 : \mu = \mu_0$. Teststatistikan har följande utseende

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

och är i fallet $\mu = \delta \neq \mu_0$ icke-centralt t-fördelad med icke-centrala parametern δ, vanligtvis noterat t_ν. Som vanligt står ν för antalet frihetsgrader. Detta innebär att dess täthetsfunktion är centrerad över δ istället för att vara centrerad över noll. Översätter vi detta till två oberoende normalfördelade stickprov där vi vill testa nollhypotesen $H_0 : \mu_1 = \mu_2$ får vi att teststatistikan

$$T = \frac{\overline{X} - \overline{Y}}{s_p\sqrt{\frac{1}{n1} + \frac{1}{n2}}}$$

är fördelad med t_ν om H_0 är sann och fördelad med $t_\nu(\delta)$ om H_0 är falsk och $\mu_1 - \mu_2 = \delta$. För mer detaljer se Lindgren (1993).

3.4 F-test

I denna uppsats används F-testet för att testa om vi kan anta att $\sigma_1^2 = \sigma_2^2$ eller ej. Teststatistikan har följande utseende

$$F = \frac{s_1^2}{s_2^2}$$

och är F-fördelad med $n_1 - 1$ och $n_2 - 1$ antal frihetsgrader. Om vi får utfallet

$$F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) < F < F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$$

accepterar vi hypotesen om lika varianser, annars förkastas den. I uttrycken ovan står α för den valda signifikansnivån, n_1 och n_2 för respektive stickprovsstorlek samt s_1 och s_2 för standardavvikelsen i respektive stickprov. För mer detaljer se Vejde och Leander (2000).

4 Genomförande och resultat

När vi genomför ett test av $H_0 : \mu_1 = \mu_2$ vill vi, i denna uppsats, att testets signifikansnivå inte ska överstiga fem procent. Denna nivå har vi utgått ifrån för Student's t-test, Welch's t-test och det sekventiella t-testet. Vi vill dessutom ha ett så starkt t-test som möjligt. De simulerade stickproven är av storlek fem om inte annat nämns.

4.1 Jämförelse av Student's t-test och Welch's t-test

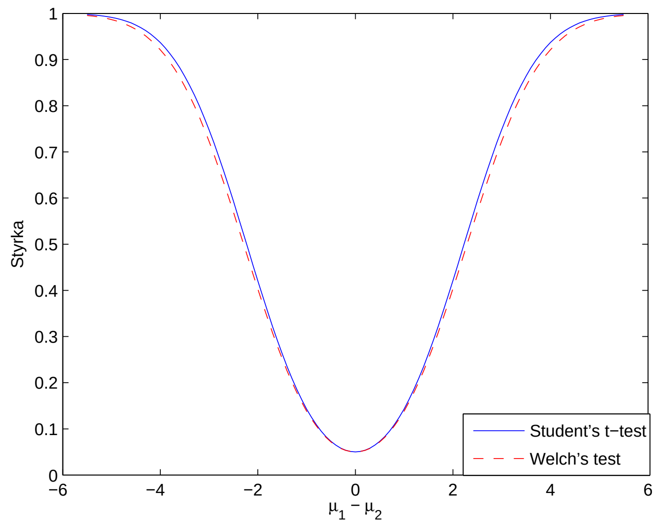
Det är känt att vid test av $H_0 : \mu_1 = \mu_2$ för två normalfördelade stickprov används Student's t-test om varianserna för de båda stickproven är lika, annars används oftast Welch's t-test. Eftersom vi sällan vet om varianserna är lika eller inte görs ibland fel test. I detta avsnitt kommer vi både att simulera och använda inbyggda fördelningsfunktioner i matlab för att undersöka hur styrka och signifikansnivå beter sig under olika förutsättningar. Den valda signifikansnivån i alla beräkningar i detta avsnitt är fem procent.

4.1.1 Fallet $\sigma_1 = \sigma_2$

Vi vet att Welch's t-test har lägre styrka än Student's t-test om varianserna är lika eftersom Student's t-test är UMP. Dock är det intressant att notera att styrkan inte är mycket lägre vilket illustreras i Figur 1-4. I Figur 1 och Figur 2 har vi använt oss av ett stickprov av storlek fem och i Figur 3 och Figur 4 har vi stickprovsstorleken tio. I samtliga figurer har vi skillnaden i väntevärdena, vi kallar den $\delta = \mu_1 - \mu_2$, på x-axeln. I Figur 1 och Figur 3 illustreras styrkefunktionen för både Student's t-test och Welch's t-test då varianserna för de båda stickproven är lika. I båda dessa styrkefunktioner har vi använt $\sigma_1 = \sigma_2 = 1$. Vi har använt oss av en inbyggd icke-central t-fördelningsfunktion i matlab för att plotta styrkefunktionen för Student's t-test. Styrkefunktionen för Welch's t-test är även den beräknad på liknande sätt med skillnaden att antalet frihetsgrader är simulerade. Att simulera frihetsgraderna för Welch's t-test är nödvändigt eftersom de beror på utfallet av stickproven och därmed är stokastiska.

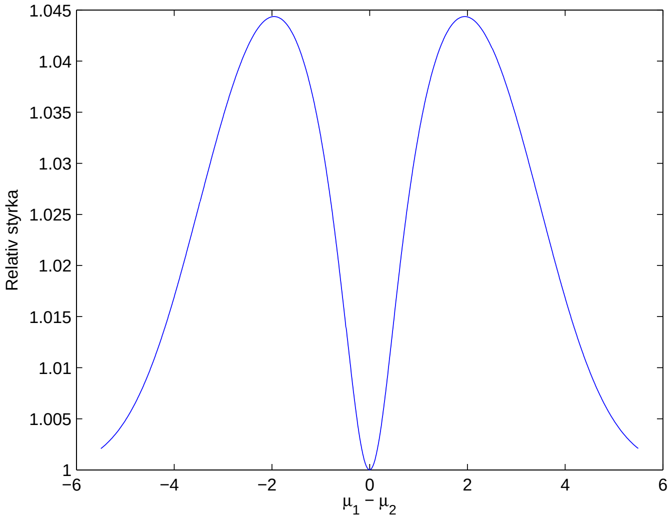
I Figur 2 och Figur 4 illustreras den relativa styrkan för Student's t-test gentemot Welch's t-test, d.v.s. styrkan för Student's t-test delat med styrkan för Welch's t-test. Vi tittar först på Figur 1 och Figur 2. Vi kan se i båda figurerna att skillnaden i styrka inte är speciellt stor och då $\delta = 0$ har båda testen samma styrka. Detta är inte speciellt anmärkningsvärt eftersom i de fall vi har verklig skillnad mellan väntevärdena lika med noll kommer båda testen förkasta nollhypotesen $H_0 : \mu_1 = \mu_2$ med sannolikhet lika med signifikansnivån. Då skillnaden i väntevärdena är stor går styrkan för båda testen mot ett, se Figur 1. Även detta resultat är inte speciellt anmärkningsvärt eftersom ju längre vi ligger från nollhypotesen ju lättare är det att förkasta den för båda testen.

Student's t-test har den större styrkan vilket vi nämnt tidigare och styrkan för Welch's t-test ligger precis under. Skillnaden i styrka ses tydligast i figurerna som visar den relativa styrkan. Utifrån Figur 2 kan vi notera att den relativa styrkan är mindre än 1.045, d.v.s. att Welch's t-test är ca 4.5 procent sämre än Student's t-test när skillnaden i styrka är som störst. Vi kan även räkna ut att maximum för den relativa styrkan, när vi har en stickprovsstorlek lika med fem, är där $|\delta| = 1.95$. Vi har alltså två maximum punkter där symmetrin kommer av att vi gör tvåsidiga test.

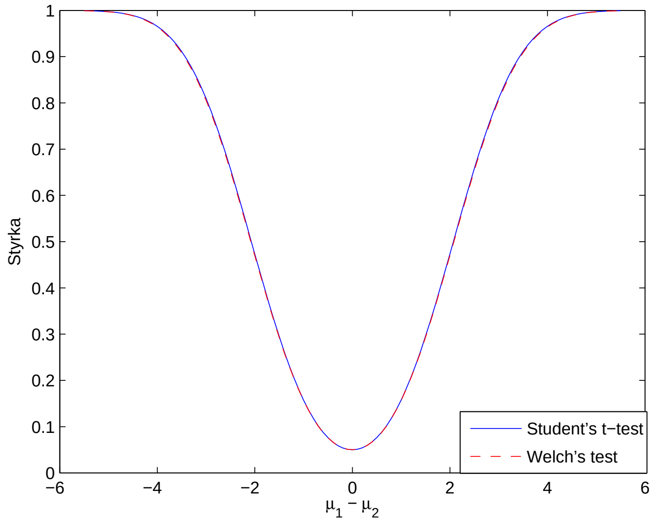


Figur 1: *Styrkefunktioner för Student's t-test och Welch's t-test vid stickprovsstorleken fem.*

Figur 3 och Figur 4 visar styrkefunktionen för Student's t-test och Welch's t-test respektive den relativa styrkan då vi har ett stickprov av storlek tio. Dessa har tagits fram på samma sätt som är beskrivet ovan med skillnaden att vi nu har använt oss av stickprovsstorleken tio. Skillnaden mot fallet med stickprovsstorleken fem är till synes att skillnaden mellan styrkefunktionerna har minskat. I Figur 3 ser det ut som om det endast finns en styrkefunktion och tittar vi på den relativa styrkan varierar den mycket mindre än tidigare. Utifrån Figur 4 kan vi notera att den relativa styrkan är mindre än 1.01, d.v.s. att under dessa förutsättningar är Welch's t-test ca 1 procent sämre än Student's t-test när skillnaden är som störst. Den relativa styrkans maximum i detta fall hittar vi där $|\delta| = 1.69$.

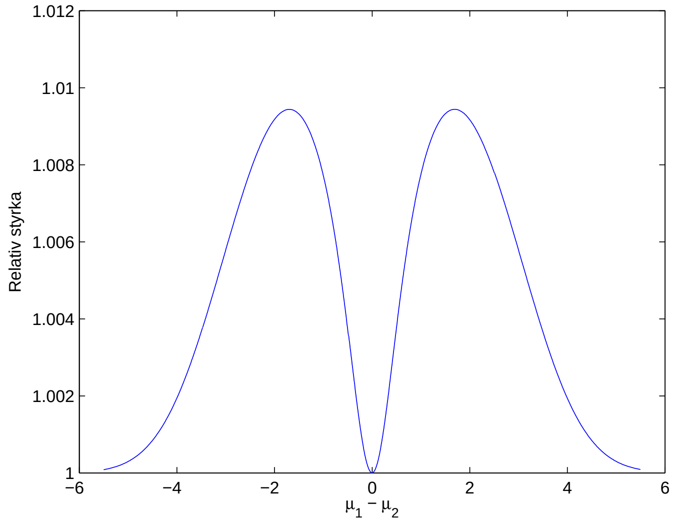


Figur 2: *Relativa styrkan för Student's t-test vid stickprovsstorleken fem.*



Figur 3: *Styrkefunktioner för Student's t-test och Welch's t-test vid stickprovsstorleken 10.*

I Figur 1-4 ser vi att vi inte förlorar mycket i styrka om vi utför Welch's t-test även om varianserna för de båda stickproven är lika. Vi kan även notera att om vi har större stickprov, i detta fall av storlek tio, förlorar vi mindre



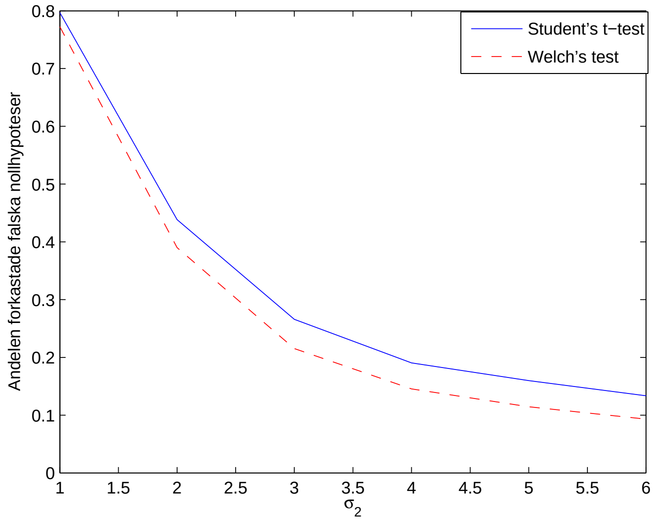
Figur 4: *Relativa styrkan för Student's t-test vid stickprovsstorleken 10.*

styrka än vid mindre stickprov, i detta fall av storlek fem.

4.1.2 Fallet $\sigma_1 \neq \sigma_2$

Det är intressant att undersöka hur Student's t-test påverkas av att vi har skillnad i varianserna eftersom teststatistikan inte längre är exakt t-fördelad i detta fall. I Figur 5 och Figur 6 visas styrka och signifikansnivå för Student's t-test och Welch's t-test mot standardavvikelsen i stickprov nummer två. Illustrationerna är gjorda med hjälp av simuleringar där Student's t-test och Welch's t-test har utförts 10.000 gånger var. Standardavvikelsen i stickprov nummer ett hålls konstant, $\sigma_1 = 1$, medan standardavvikelsen för stickprov nummer två varieras, $\sigma_2 = 1, 2, 3, 4, 5, 6$. I Figur 5 ser vi den uppskattade styrkan för Student's t-test och Welch's t-test. I denna illustration har vi använt oss av värdena $\mu_1 = 0$ och $\mu_2 = 2$. Vi valde att undersöka styrkan då skillnaden i väntevärdena är två eftersom vid nära två ligger den största skillnaden mellan styrkefunktionerna, då varianserna är lika. Vi ser att Student's t-test har större styrka även då skillnad i standardavvikelserna föreligger.

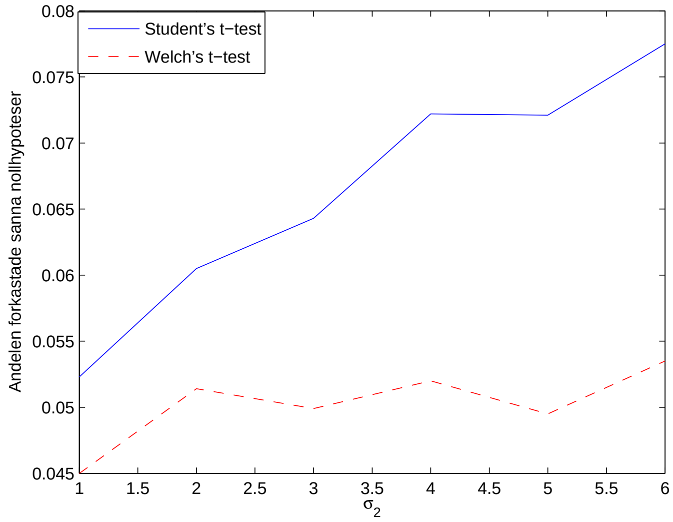
För att få en bättre bild av hur Student's t-test påverkas av skillnader i standardavvikelserna plottar vi även antalet förkastade sanna nollhypoteser, d.v.s. signifikansnivån. Denna illustration utförs på liknande sätt som ovan med skillnaden att vi nu plottar signifikansnivån och därför har använt oss av värdena $\mu_1 = \mu_2 = 0$. Detta visas i Figur 6. Vi ser tydligt att trots att signifikansnivån är vald till fem procent för båda testen stiger antalet



Figur 5: *Styrkan för Student's t-test och Welch's t-test då* $|\mu_1 - \mu_2| = 2$.

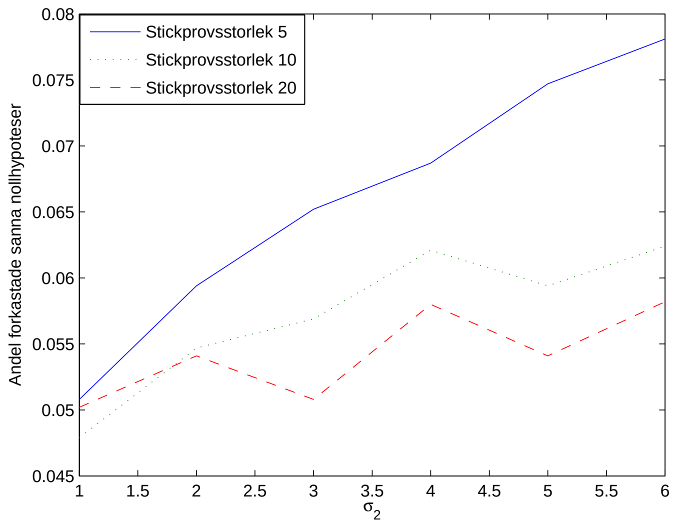
förkastade sanna nollhypoteser, signifikansnivån, mot åtta procent för Student's t-test medan signifikansnivån för Welch's t-test ligger kvar vid fem procent. Signifikansnivån för Welch's t-test ligger givetvis kvar vid fem procent i och med att det testet är utformat för att användas i de fall då det föreligger skillnad i standardavvikelserna. Däremot är det intressant att signifikansnivån för Student's t-test ökar med standardavvikelsen för stickprov nummer två. Vi kan konstatera genom både Figur 5 och Figur 6 att Student's t-test oftare förkastar nollhypotesen $H_0 : \mu_1 = \mu_2$ än vad Welch's t-test gör, oavsett om nollhypotesen är sann eller falsk, då det föreligger skillnad i varianserna.

Student's t-test är som sagt mycket bra att använda vid små stickprovsstorlekar, då vi har normalfördelade stickprov med samma, okända, varians. I Figur 6 såg vi att Student's t-test förkastar en sann nollhypotes oftare än den valda signifikansnivån då stickprovsstorleken är lika med fem. I Figur 7 ser vi hur signifikansnivån ändras för Student's t-test för stickprovstorlekarna fem, tio och tjugo. Som förut kan vi se att antalet förkastade sanna nollhypoteser, signifikansnivån, för stickprovsstorleken fem inte håller sig på den förvalda signifikansnivån fem procent. Stickprovsstorlekarna tio och tjugo ser ut att öka från signifikansnivån fem procent till straxt över femprocentsnivån. I och med att resultaten baseras på simuleringar är det svårt att utläsa om och i så fall hur mycket signifkansnivån för de större stickprovsstorlekarna ligger över nivån fem procent. Denna illustration är utförd på samma sätt som ovan med skillnaden att vi plottar för endast Student's t-test men för



Figur 6: *Signifikansnivån för Student's t-test och Welch's t-test.*

olika stickprovsstorlekar.



Figur 7: *Signifikansnivån för Student's t-test för stickprovsstorlekarna fem, tio och tjugo.*

I detta avsnitt har vi sett att Student's t-test får fel signifikansnivå då

$\sigma_1 \neq \sigma_2$. Alltså bör det inte användas när skillnad i standardavvikelserna föreligger även om det har en större styrka. Vi har även sett att Welch's t-test inte är mycket sämre än Student's t-test då $\sigma_1 = \sigma_2$.

4.2 Utformning av det sekventiella t-testet

Det sekventiella t-testet är utformat på så sätt att av de 10.000 simuleringar som görs bestämmer F-testet när Student's t-test utförs och när Welch's t-test utförs. Därefter kan vi, efter att ha uppskattat signifikansnivån och styrkan för de Student's t-test och de Welch's t-test som utförs, beräkna signifikansnivån och styrkan för det sekventiella t-testet. Enligt lagen om total sannolikhet kommer signifikansnivån för det sekventiella t-testet att ges av

$$\alpha_S = P(T)\alpha_T + P(W)\alpha_W \tag{2}$$

där α står för signifikansnivån för respektive test och P för sannolikheten att det blir just det testet. Beteckningarna S, T och W står för sekventiellt t-test, Student's t-test respektive Welch's t-test. På samma sätt ges styrkan för det sekventiella t-testet av

$$\pi_S = P(T)\pi_T + P(W)\pi_W \tag{3}$$

där π står för styrkan för respektive test.

Vi strävar efter att det sekventiella t-testet ska ha en signifikansnivå som ligger så nära fem procent som möjligt samtidigt som den inte överstiger fem procent. Detta ska gälla oavsett om det föreligger skillnad mellan varianserna eller inte. Dessutom vill vi ha ett så starkt t-test som möjligt. I detta avsnitt kommer vi att simulera Student's t-test, Welch's t-test och olika sekventiella t-test för att undersöka egenskaper, så som styka och sigifikansnivå, under olika förutsättningar. Den valda signifikansnivån för Student's t-test och Welch's t-test kommer att varieras.

4.2.1 F-test av $H_0 : \sigma_1 = \sigma_2$

För att avgöra om vi kan anta att varianserna för de båda stickproven är lika eller inte utför vi ett F-test av nollhypotesen $H_0 : \sigma_1 = \sigma_2$.

Teststatistikan F är F-fördelad och vi använder de kritiska gränser som finns för F-fördelningen. Genom att använda signifikansnivån 50 procent får vi ett smalt område där vi kan acceptera hypotesen om lika varianser för de båda stickproven. Om vi får utfallet

$$F_{0,25}(n_1 - 1, n_2 - 1) < F < F_{0,75}(n_1 - 1, n_2 - 1)$$

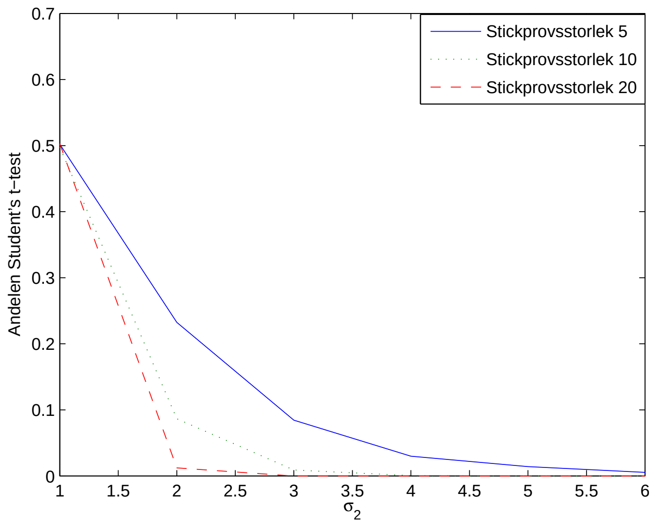
accepterar vi hypotesen om lika varianser, annars förkastas den. Valet av signifikansnivån 50 procent är godtyckligt och resulterar i att vi kommer att få ett test med kritiska gränser som förkastar den sanna nollhypotesen om lika varianser 50 procent av gångerna. Detta resulterar i att vi kommer att utföra Student's t-test hälften av gångerna och Welch's t-test den andra hälften av gångerna. I detta fall då varianserna är lika är teststatistikan i Student's t-test t-fördelad och sannolikheten att förkasta nollhypotesen $H_0 : \mu_1 = \mu_2$, när den är sann, är lika med den signifikansnivå som väljs. Dessutom blir styrkan för det sekventiella t-testet

$$\pi_S = 0.5\pi_T + 0.5\pi_W$$

och därmed högre än styrkan för Welch's t-test eftersom $\pi_T > \pi_W$.

I det andra fallet, med skilda varianser, leder ett test med dessa kritiska gränser till att vi förkastar hypotesen om lika varianser oftare då skillnaden mellan varianserna är stor. Om vi har en stor skillnad i varianserna kommer vi alltså att utföra Welch's t-test mycket oftare än Student's t-test. Detta illustreras i Figur 8. Vi utför F-testet, med de ovan nämnda kritiska gränserna, för de olika stickprovsstorlekarna fem, tio och tjugo under förutsättningen $\sigma_1 = 1$ och $\sigma_2 = 1, 2, 3, 4, 5, 6$. Om det föreligger skillnad eller ej mellan väntevärdena spelar ingen roll för andelen Student's t-test som utförs. I denna illustration använde vi oss av $\mu_1 = \mu_2 = 0$ men det hade inte gjort någon skillnad om vi hade använt oss av t.ex. $\mu_1 = 0$ och $\mu_2 = 2$.

I Figur 8 ser vi andelen Student's t-test som utförs vid stickprovsstorlekarna fem, tio och tjugo. Vi ser att andelen Student's t-test som utförs minskar drastiskt då vi ökar standardavvikelsen för stickprov nummer två. När standardavvikelserna är lika, i denna simulering $\sigma_1 = \sigma_2 = 1$, ligger andelen utförda Student's t-test på nivån 50 procent för de tre stickprovsstorlekarna. Vi kan notera att detta stämmer överens med de kritiska gränser som vi använder för F-testet. Då standardavvikelsen för stickprov nummer två ökar, minskar andelen utförda Student's t-test för samtliga stickprovsstorlekar och vi kan se att för de större stickprovstorlekarna minskar andelen utförda Student's t-test fortare. För stickprov av större storlek än fem utförs nästan aldrig Student's t-test när skillnaden i standardavvikelserna har blivit tillräckligt stor.



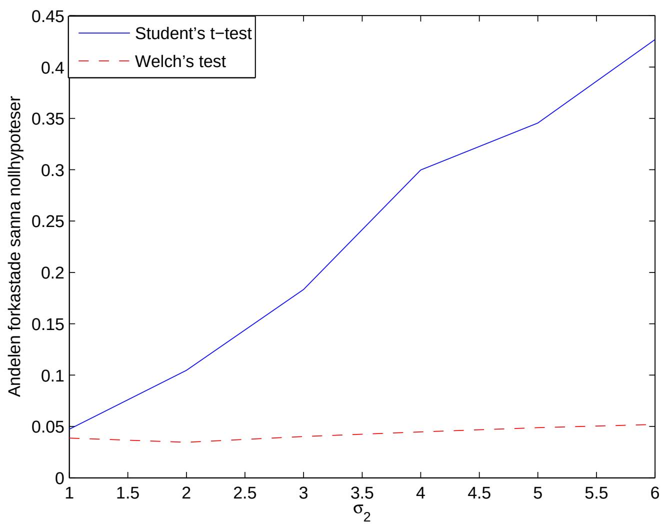
Figur 8: *Andelen Student's t-test som utförs efter vi har testat om varianserna kan anses lika eller ej.*

Sannolikheten att välja rätt test blir alltså större ju större stickprov vi har. Dock används vanligen Student's t-test vid små stickprovsstorlekar och som vi ser, i Figur 8, leder det till att det utförs en del Student's t-test under fel förutsättningar.

Nu har vi bestämt de kritiska gränserna för F-testet och låter dessa bestämma om vi ska utföra Student's t-test eller Welch's t-test. I Figur 9 ser vi hur andelen förkastade sanna nollhypoteser, d.v.s. signifikansnivån, för Student's t-test påverkas av att vi först utför ett F-test av varianserna. Illustrationen är gjord på så sätt att av de 10.000 simuleringar som görs bestämmer F-testet hur stor del av dem där Student's t-test utförs och hur stor del av dem där Welch's t-test utförs. Sedan räknar vi ut signifikansnivån, som andel t-test där nollhypotesen förkastas, för respektive t-test för de olika värdena på σ_2. Eftersom vi räknar ut signifikansnivån gäller att väntevärdena är lika och vi använder värdena $\mu_1 = \mu_2 = 0$. Resultatet blir att signifikansnivån ökar kraftigt med standardavvikelsen i stickprov nummer två. Eftersom vi först utför F-testet är andelen Student's t-test som utförs mycket liten då standardavvikelsen för stickprov nummer två är fem och sex, se Figur 8. Den andel Student's t-test som utförs, när skillnaden i varianserna är så pass stor, förkastar sanna nollhypoteser extremt ofta. Detta kan vi tydligt observera i Figur 9 och jämföra med signifikansnivån för Welch's t-test som väntat håller sig på nivån fem procent. Det är viktigt att inte blanda ihop

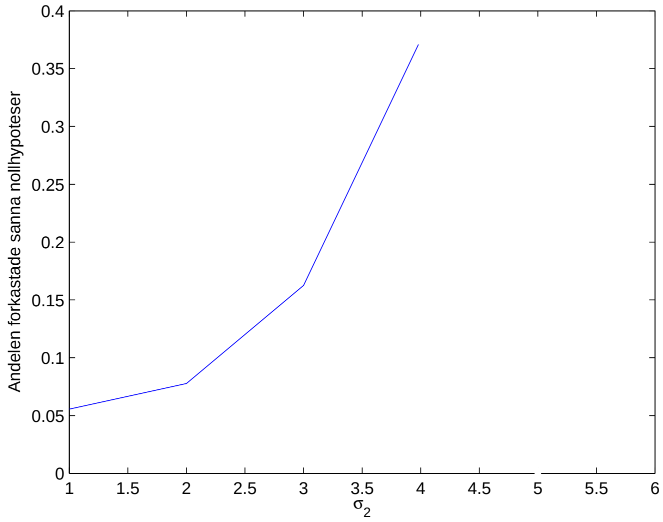
Figur 9 med Figur 6. I Figur 9 är antalet Student's t-test som utförs beroende av utfallet av F-testet, d.v.s. ett betingat test, vilket inte är fallet i Figur 6.

Det är intressant att undersöka om vi får samma extrema signifikansnivå för Student's t-test om vi istället använder oss av stickprovsstorleken tio. Simuleringen utförs på samma sätt som beskrivs ovan med skillnaden att vi ökar stickprovsstorleken från fem till tio. Resultatet ser vi i Figur 10. Då $\sigma_2 = 1, 2, 3, 4$ beter sig signifikansnivån på samma sätt som för stickprovsstorleken fem, d.v.s. startar på nivån fem procent för att sedan öka kraftigt. I denna simulering utfördes det tre Student's t-test då $\sigma_2 = 5$ med utfallet H_0 förkastas ej samt utfördes inget Student's t-test då $\sigma_2 = 6$. Detta syns tydligt i Figur 10. Eftersom det är en väldigt liten del Student's t-test som utförs för de större skillnaderna i standardavvikelserna, i denna simulering 0.03 procent då $\sigma_2 = 5$ och 0 procent då $\sigma_2 = 6$, är det svårt att säga hur signifikansnivån beter sig där. Med tanke på hur signifikansnivån växer fram till $\sigma_2 = 4$ kan vi anta att den även ökar för $\sigma_2 = 5$ och $\sigma_2 = 6$.

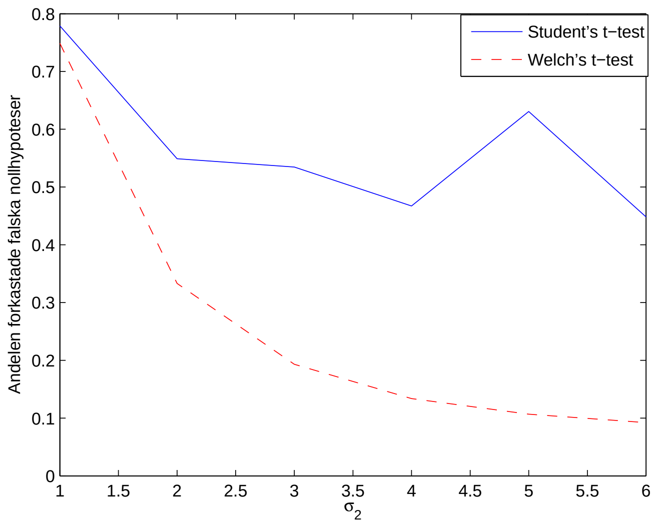


Figur 9: *Signifikansnivån för Student's t-test och Welch's t-test då vi först utför ett F-test.*

Det vi såg i Figur 9 kan vi även notera i styrkan för Student's t-test. I Figur 11 ser vi styrkan för Student's t-test och Welch's t-test då skillnaden i väntevärdena är två. Styrkefunktionen för Welch's t-test ser ut som den gör i Figur 5, den påverkas inte av F-testet, medan Student's t-test får en mycket förhöjd styrka till följd av den extremt förhöjda signifikansnivån.



Figur 10: *Signifikansnivån för Student's t-test, vid stickprovsstorleken tio, då vi först utför ett F-test.*



Figur 11: *Styrkan för Student's t-test och Welch's t-test då vi först utför ett F-test.*

Alltså kommer Welch's t-test att förkasta nollhypotesen $H_0 : \mu_1 = \mu_2$ lika ofta oavsett om vi gör ett sekventiellt t-test eller inte medan Student's t-test

Tabell 1: *Signifikansnivån för det sekventiella t-testet för stickprovsstorlekarna fem och tio samt för olika värden på standardavvikelsen i stickprov nummer två.*

	$n_1 = n_2 = 5$	$n_1 = n_2 = 10$
$\sigma_2 = 1$	0.0430	0.0489
$\sigma_2 = 2$	0.0509	0.0481
$\sigma_2 = 3$	0.0524	0.0512
$\sigma_2 = 4$	0.0530	0.0515
$\sigma_2 = 5$	0.0537	0.0498
$\sigma_2 = 6$	0.0550	0.0484

kommer att förkasta nollhypotesen mycket oftare i ett sekventiellt t-test.

4.2.2 Val av kritiska gränser i det sekventiella t-testet

I föregående avsnitt bestämde vi kritiska gränser för F-testet och såg hur andelen förkastade nollhypoteser för Student's t-test påverkas av att vi testar likhet i varianserna innan vi utför ett t-test av $H_0 : \mu_1 = \mu_2$. Den andel Student's t-test som utförs när skillnaden i varianserna är stor har en extremt hög signifikansnivå samtidigt som den andel Student's t-test som utförs är väldigt liten, se Figur 8 och Figur 9. I Tabell 1 visas hur signifikansnivån för det sekventiella t-testet påverkas av detta och förändras med standardavvikelsen för stickprov nummer två.

Värdena i Tabell 1 har räknats ut genom att göra ett sekventiellt t-test för de två stickprovsstorlekarna fem och tio. Först utför vi F-test och sedan utför vi de Student's t-test och Welch's t-test som bestämts av F-testet. Därefter räknas signifikansnivån ut enligt lagen om total sannolikhet som nämts tidigare, formel (2). Vi använder signifikansnivån fem procent för både Student's t-test och Welch's t-test i detta sekventiella t-test vilket resulterar i att vi har en standardavvikelse på 0.0022 i resultaten, se formel (1). Som tidigare använder vi även här värdet ett när vi pratar om att standardavvikelserna i de båda stickproven är lika, d.v.s. $\sigma_1 = \sigma_2 = 1$. I Tabell 1 ser vi att signifikansnivån för det sekventiella t-testet, för både stickprovsstorleken fem och tio, ökar med standardavvikelsen för stickprov nummer två. Ökningen är tydligast för stickprovsstorleken fem men vi kan misstänka att det även gäller för stickprovsstorleken tio, även om det inte syns lika tydligt.

Som vi illustrerade i Figur 8 är det Welch's t-test som utförs oftast när det föreligger stor skillnad mellan varianserna. Eftersom signifikansnivån för

Welch's t-test är fem procent och signifikansnivån för Student's t-test blir extremt stor får vi resultatet att signifikansnivån för det sekventiella t-testet kommer att ligga över fem procent, även om det endast är en mycket liten del Student's t-test som utförs. Detta inses genom att använda formel (2). Den för höga signifikansnivån som dessa kritiska gränser resulterar i gäller för båda stickprovsstorlekarna, fem och tio. Skillnaden är att för stickprovsstorleken tio utförs det knappt några Student's t-test när skillnaden i varianserna är riktigt stor. I simuleringen av värdena i Tabell 1 utförs det inte några Student's t-test för stickprovsstorleken tio när $\sigma_2 = 6$. Observera däremot att sannolikheten att det utförs Student's t-test alltid är positiv.

Det är vid $\sigma_2 = 2$ som det utförs flest Student's t-test under fel förutsättningar samtidigt som signifikansnivån för Student's t-test ökat till kring tio procent, se Figur 8-10. Vi kan misstänka att det sekventiella t-testet har för hög signifikansnivå redan vid $\sigma_2 = 2$. Alltså påverkar Student's t-test det sekventiella t-testet dels på de lägre nivåerna ($\sigma_2 = 2$ och $\sigma_2 = 3$) genom att det där fortfarande genomförs en relativt stor del Student's t-test under felaktiga förutsättningar och dels på de högre nivåerna ($\sigma_2 = 4$, $\sigma_2 = 5$ och $\sigma_2 = 6$) genom att inverka med extremt hög signifikansnivå. För att hantera detta problem kan vi justera de kritiska gränserna för Student's t-test och Welch's t-test så att den sammanvägda signifikansnivån håller sig under fem procent. De kritiska gränserna bestäms av antalet frihetsgrader och den valda signifikansnivån. I denna uppsats begränsar vi oss till stickprovsstorleken fem och kommer att undersöka olika val av signifikansnivåer.

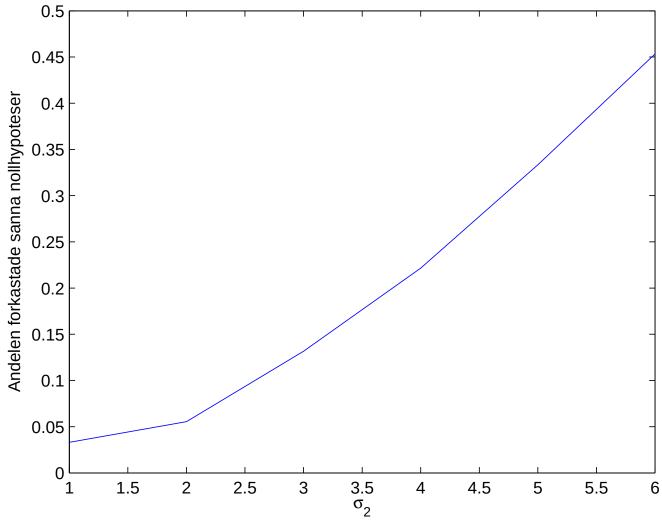
I Figur 12 ser vi hur andelen förkastade sanna nollhypoteser, signifikansnivån, för Student's t-test påverkas av att vi sänker den valda signifikansnivån. Simuleringen är utförd på samma sätt som den i Figur 9 med skillnaden att vi använder oss av en lägre signifikansnivå. Den valda signifikansnivån i denna figur är tre procent och det syns tydligt att trots att vi gjort en kraftig sänkning från fem till tre procent blir andelen förkastade sanna nollhypoteser, signifikansnivån, fortfarande extremt hög när standardavvikelsen för stickprov nummer två är stor.

Vi kan utforma det sekventiella t-testet enligt tre olika alternativ:

1. Sänka signifikansnivån för Student's t-test och låta signifikansnivån för Welch's t-test vara kvar på fem procent.

2. Sänka signifikansnivån för både Student's t-test och Welch's t-test.

3. Sänka signifikansnivån för Welch's t-test och låta signifikansnivån för



Figur 12: *Signifikansnivån för Student's t-test när den valda signifikansnivån är* $\alpha_T = 0.03$.

Student's t-test vara kvar på fem procent.

Om vi endast sänker signifikansnivån för Student's t-test kommer det sekventiella t-testet fortfarande att ha en för hög signifikansnivå när det föreligger skillnad i varianserna. Samtidigt kommer vi att ha en för låg signifikansnivå när varianserna verkligen är lika och det utförs lika många Student's t-test som Welch's t-test. Sänker vi signifikansnivån för båda testen kan vi få ett sekventiellt t-test vars signifikansnivå inte överstiger fem procent när det föreligger skillnad men då varianserna är lika kommer signifikansnivån definitivt att ligga under fem procent. Om vi endast sänker signifikansnivån för Welch's t-test kommer signifikansnivån för det sekventiella t-testet att ligga under fem procent då likhet i varianserna föreligger. I fallet med skilda varianser kommer däremot en sänkning av signifikansnivån för endast Welch's t-test resultera i att vi kan få en signifikansnivå för det sekventiella t-testet som inte överstiger fem procent. Att signifikansnivån för det sekventiella t-testet kommer att ligga under nivån fem procent, då likhet i varianserna råder, i alla tre alternativ kan vi konstatera genom formel (2). Resten av resonemanget inses genom formel (2) samt Figur 8, Figur 9 och Figur 12. Huruvida signifikansnivån för det sekventiella t-testet kommer att överstiga eller ligga nära fem procent, då skillnad i varianserna råder, för de två sista alternativen beror på hur mycket vi sänker signifikansnivån och ändrar på de kritiska gränserna.

Alla de olika kombinationer av signifikansnivåer som testats kommer inte

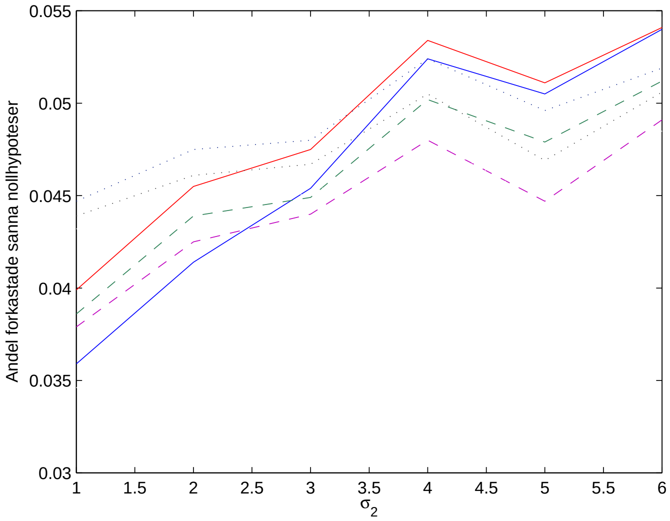
Tabell 2: *Justerade signifikansnivåer som genererar kritiska gränser vilka representerar olika sekventiella t-test.*

Sekventiellt t-test	t_T	t_W	α_T	α_W
1	2.3060	2.8308	0.05	0.046
2	2.4490	2.8099	0.04	0.047
3	2.4490	2.8521	0.04	0.045
4	2.3060	2.7896	0.05	0.048
5	2.3060	2.8740	0.05	0.044
6	2.6338	2.7504	0.03	0.05
7	2.6338	2.8099	0.03	0.047
8	2.4490	2.7504	0.04	0.05

att redovisas i denna uppsats men vi kommer att titta på exempel från de tre alternativ att utforma det sekventiella t-testet på, som redovisades ovan. I Tabell 2 redovisas de kritiska gränserna som fås av olika signifikansnivåer. Olika kombinationer av kritiska gränser representeras som olika sekventiella t-test. I den första kolumnen visas numret som det sekventiella t-testet har tilldelats och i de två följande kolumnerna visas den kritiska gränsen för Student's t-test respektive Welch's t-test. De två kolumnerna till höger visar vilken signifikansnivå som ger den kritiska gränsen.

Av de åtta olika sekventiella t-testen i Tabell 2 väljer vi att närmre undersöka signifikansnivån för sex av dem. I Figur 13 visas hur signifikansnivån för det sekventiella t-testet påverkas av olika kombinationer av valda signifikansnivåer för Student's t-test och Welch's t-test. Illustrationen är gjord på samma sätt som beskrivs ovan, då vi räknade ut signifikansnivån för stickprovsstorleken fem i Tabell 1. Skillnaden är att vi i denna simulering varierar den valda signifikansnivån för Student's t-test och Welch's t-test, d.v.s. undersöker olika kritiska gränser.

Det är svårt att urskilja vilken linje som representerar ett visst sekventiellt t-test. Vi ser dock att flera av de olika kombinationerna ger en alldeles för låg signifikansnivå när $\sigma_1 = \sigma_2 = 1$ medan några kombinationer ser ut att ge en för hög signifikansnivå när standardavvikelsen för stickprov nummer två är tillräckligt stor. De två heldragna linjerna representerar alternativ 1 och är de två kombinationerna 6 och 8 i Tabell 2. De är bland de linjer som får en för låg respektive för hög signifikansnivå då skillnaden i standardavvikelserna är liten respektive stor. De två streckade linjerna representerar alternativ 2 och är de två kombinationerna 2 och 3 i Tabell 2. Även dessa



Figur 13: *Signifikansnivån för olika sekventiella t-test.*

Tabell 3: *Signifikansnivån i de två utvalda sekventiella t-testen.*

	$\alpha_T = 0.05$ $\alpha_W = 0.046$	$\alpha_T = 0.04$ $\alpha_W = 0.047$
$\sigma_2 = 1$	0.0439	0.0386
$\sigma_2 = 2$	0.0461	0.0439
$\sigma_2 = 3$	0.0467	0.0449
$\sigma_2 = 4$	0.0505	0.0502
$\sigma_2 = 5$	0.0469	0.0479
$\sigma_2 = 6$	0.0506	0.0512

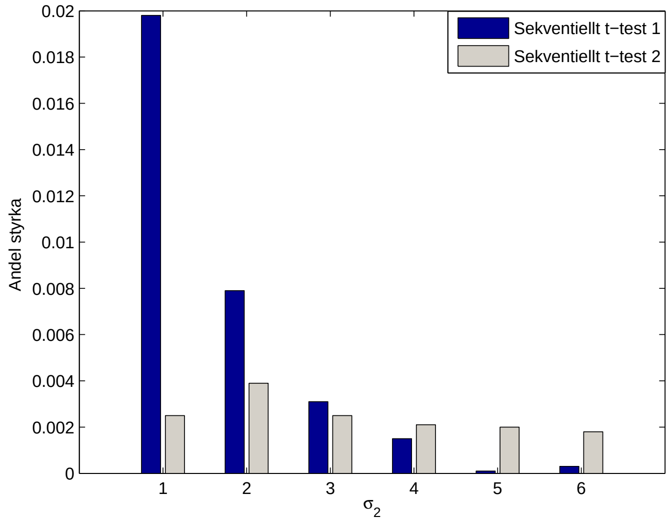
får en något låg signifikansnivå när skillnaden i standardavvikelsen är liten men ser ut att vara nära fem procent när skillnaden i standardavvikelserna är stor. Slutligen har vi de två prickade linjerna som representerar alternativ 3 och är de två kombinationerna 1 och 4 i Tabell 2. Dessa två kombinationer har en mycket bättre signifikansnivå när $\sigma_1 = \sigma_2$ och ser ut att ligga nära fem procent när skillnaden i standardavvikelserna är stor.

Vi väljer ut två sekventiella t-test ur Tabell 2. Detta gör vi utifrån den signifikansnivå som de olika sekventiella t-testen har enligt Figur 13. I Tabell 3 visas vilka kombinationer som valdes och vilken signifikansnivå de ger. Vi kommer att referera till dem som sekventiellt t-test 1 och sekventiellt t-test 2, enligt den ordning de kommer i tabellen.

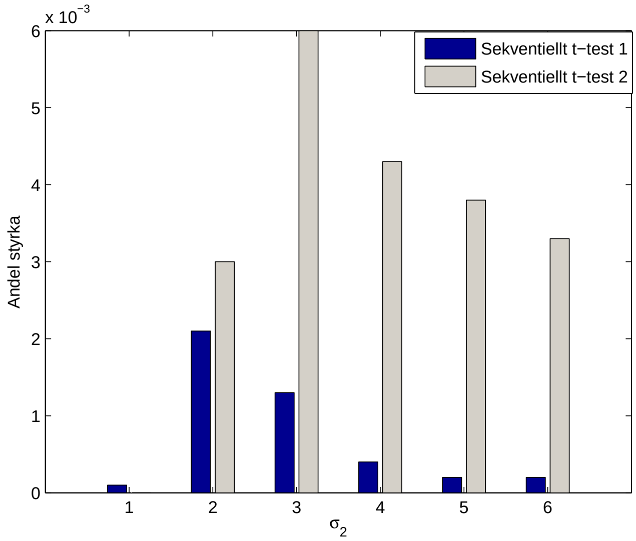
Nu kan vi fråga oss vilket av dessa två sekventiella t-test som är bäst. För att undersöka detta kan vi beräkna hur starka de olika testen är. Detta illustreras med hjälp av stapeldiagram i Figur 14 och Figur 15. För att komma fram till resultaten i Figur 14 och Figur 15 utför vi först två sekventiella t-test med ovan valda kritiska gränser, för att sedan mäta andelen gånger som de två testen förkastar respektive accepterar den falska nollhypotesen $H_0 : \mu_1 = \mu_2$. Skillnaden mellan figurerna är att i Figur 14 har vi väntevärdesskillnaden $|\mu_1 - \mu_2| = 2$ medan i Figur 15 är väntevärdesskillnaden $|\mu_1 - \mu_2| = 4$. I båda figurerna mäter y-axeln hur ofta ett visst test förkastar en falsk nollhypotes medan det andra testet accepterar den. Medan $\sigma_1 = 1$ har vi för varje värde som standardavvikelsen i stickprov nummer två antar, $\sigma_2 = 1, 2, .., 6$, alltså plottat, enligt schemat nedan, rutan som har nummer ett mot rutan som har nummer två.

	Sekventiellt t-test 1		
	Förkasta	Acceptera	
Förkasta		2	Sekventiellt t-test 2
Acceptera	1		

Det starkaste testet bör alltså förkasta en falsk nollhypotes medan det andra testet accepterar den och bör göra detta fler gånger än vad det andra testet gör det. I varje stapelpar i Figur 14 och Figur 15 representerar den vänstra stapeln sekventiellt t-test 1 och den högra stapeln sekventiellt t-test 2. I Figur 14 ser vi att vilket av de två sekventiella t-testen som är starkast varierar med standardavvikelsen i stickprov nummer två. Sekventiellt t-test 1 är starkast när skillnaden i varianserna är liten och sekventiellt t-test 2 är starkast när skillnaden i varianserna är stor. Under förutsättningarna i Figur 15, då verklig skillnad i väntevärdena är fyra, ser vi att sekventiellt t-test 2 är starkare förutom då standardavvikelserna är lika. Värt att notera i båda dessa figurer är att det handlar om en mycket liten andel där det ena t-testet är bättre än det andra och vice versa.



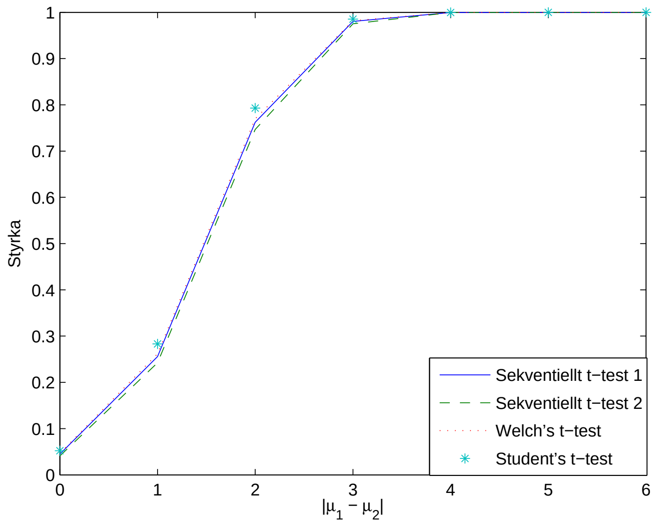
Figur 14: *Andel fall då endast det ena sekventiella t-testet förkastar en falsk nollhyppotes i fallet* $| \mu_1 - \mu_2 |= 2$



Figur 15: *Andel fall då endast det ena sekventiella t-testet förkastar en falsk nollhyppotes i fallet* $| \mu_1 - \mu_2 |= 4$

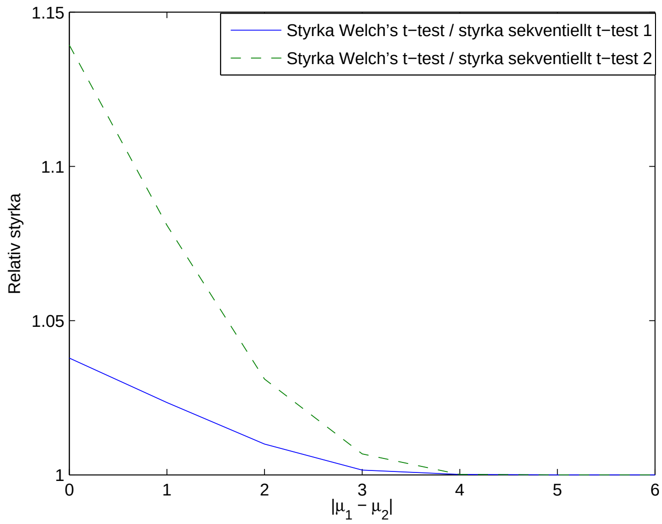
5 Jämförelse av det sekventiella t-testet, Student's t-test och Welch's t-test

Som vi såg i tidigare avsnitt har Student's t-test större styrka än Welch's t-test om standardavvikelserna i de båda stickproven är lika. Det sekventiella t-testet består av både Student's t-test och Welch's t-test vilket gör att vi kan dra slutsatsen att Student's t-test även har större styrka än det sekventiella t-testet i fallet med lika standardavvikelser. Detta kan vi även se i Figur 16 där vi plottar styrkan för de sekventiella t-testen, Student's t-test och Welch's t-test mot skillnaden i väntevärdena. Illustrationen är utförd på så sätt att styrkan för respektive test räknas ut (för de sekventiella t-testen med formel (3)) medan skillnaden i varianserna är noll, i denna simulering $\sigma_1 = \sigma_2 = 1$. Samtidigt hålls väntevärdet för stickprov nummer ett konstant, $\mu_1 = 0$, och väntevärdet för stickprov nummer två varieras, $\mu_2 = 0, 1, ..., 6$. Den valda signifikansnivån är fem procent för både Student's t-test och Welch's t-test och de sekventiella t-testen har kritiska gränser enligt sekventiellt test 1 och sekventiellt test 2, som presenterades i slutet av Avsnitt 4.2.2. I Figur 16 ser vi att Student's t-test har den största styrkan medan det är svårt att skilja Welch's t-test och de sekventiella t-testen åt.



Figur 16: *Styrkan för Student's t-test, Welch's t-test och de sekventiella t-testen i fallet då varianserna är lika.*

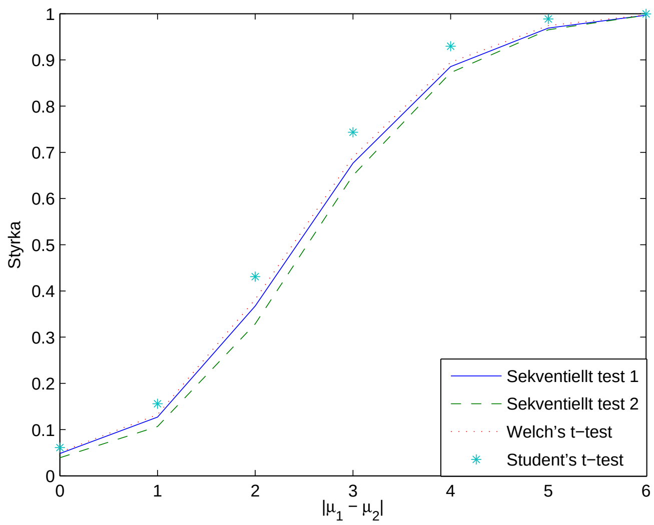
För att kontrollera vilket av Welch's t-test och de sekventiella t-testen som har störst styrka i fallet då varianserna är lika plottar vi den relativa styrkan, d.v.s. styrkan för Welch's t-test delat med styrkan för de sekventiella t-testen. Resultatet ser vi i Figur 17. Det syns tydligt att Welch's t-test är starkare än de båda sekventiella t-testen då $|\mu_1 - \mu_2| = 0, 1, 2, 3$. Detta resultat kan förklaras med att då skillnaden i väntevärdena är noll är sannolikheten att förkasta $H_0 : \mu_1 = \mu_2$ lika med signifikansnivån och Welch's t-test har en signifikansnivå på fem procent medan de sekventiella t-testen har en sänkt signifikansnivå. Den sänkta signifikansnivån för de sekventiella t-testen är troligtvis även anledningen till att Welch's t-test är starkare också då det föreligger skillnad i väntevärdena. Då skillnaden i väntevärdena blir större är Welch's t-test och de sekventiella t-testen lika starka. Styrkan för ett test går fort mot ett när skillnaden i väntevärdena är stor, detta är antagligen anledningen till att vi inte ser någon skillnad i styrkan för de tre t-testen under dessa förutsättningar. Vi hinner alltså inte märka någon skillnad i styrkan som de olika signifikansnivåerna ger innan styrkan går mot ett. I Figur 17 ser vi att Welch's t-test är maximalt 5 procent bättre än sekventiellt t-test 1 och maximalt 15 procent bättre än sekventiellt t-test 2.



Figur 17: *Relativa styrkan för Welch's t-test.*

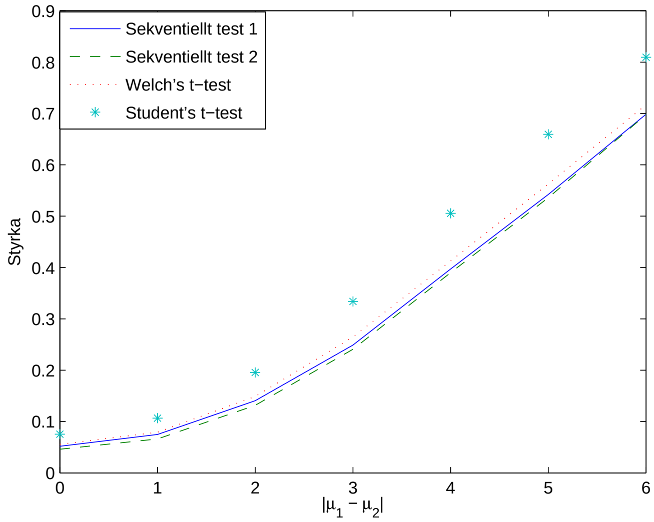
Student's t-test har, som vi vet, den största styrkan när varianserna är lika. Det har även den största styrkan när det föreligger skillnad i varianserna, vilket vi bland annat såg i Figur 5. I och med att detta är en följd av den extremt förhöjda signifikansnivån som vi såg i Figur 6, kan vi dra slutsatsen att det sekventiella t-testet är ett bättre alternativ än att endast utföra Student's t-test. I Figur 18 och Figur 19 ser vi styrkan för Student's t-test, Welch's t-test och de sekventiella t-testen. Illustrationen är utförd på liknande sätt som illustrationen för Figur 16 med skillnaden att i dessa plottar föreligger det skillnad i varianserna. Med avseende på vad vi kom fram till tidigare i detta stycke bryr vi oss inte om vad Student's t-test visar i de båda figurerna. Student's t-test finns med endast så att vi kan notera skillnaden och att det vi såg i Figur 5 inte var en tillfällighet för endast fallet $|\mu_1 - \mu_2| = 2$. I Figur 18 har vi $\sigma_2 = 2$ och i Figur 19 har vi $\sigma_2 = 4$, d.v.s i det första fallet har vi en relativt liten skillnad i varianserna och i det andra fallet är skillnaden relativt stor. Gemensamt för de båda figurerna är att de visar att Welch's t-test är starkare än de båda sekventiella t-testen. Detta ser vi tydligast då skillnaden i varianserna är stor, Figur 19.

I Avsnitt 4.1 kom vi fram till att vi maximalt förlorar 4.5 procent styrka om vi utför Welch's t-test istället för Student's t-test då varianserna är lika. I Figur 17 kan vi konstatera att Welch's t-test är bättre än de sekventiella t-testen när vi har mindre skillnader i väntevärdena och varianserna är lika. Vi kan konstatera i Figur 18 och Figur 19 att Welch's t-test är starkare än de sekventiella t-testen även då det föreligger skillnader i varianserna. I Avsnitt 4.2.2 såg vi att signifikansnivån för de sekventiella t-testen varieras, från straxt under till fem procent, med standardavvikelsen för stickprov nummer två. Dessa resultat i samband med att Welch's t-test har signifikansnivån fem procent oavsett om det föreligger skillnad eller ej i varianserna, leder till slutsatsen att Welch's t-test är att föredra framför de sekvetiella t-testen.



Figur 18: *Styrkan för Student's t-test, Welch's t-test och de sekventiella t-testen i fallet då standardavvikelsen för stickprov nummer två är två.*

I Figur 14 och Figur 15 är det svårt att utläsa vilket av de båda sekventiella t-testen som är att föredra eftersom, vilket av de två t-testen som är bäst, varierar med både skillnad i varians och skillnad i väntevärde. Vi har kommit fram till att Welch's t-test är att föredra framför de båda sekventiella t-testen. I Figur 17 ser vi att sekventiellt t-test 1 är ca 5 procent sämre än



Figur 19: *Styrkan för Student's t-test, Welch's t-test och de sekventiella t-testen i fallet då standardavvikelsen för stickprov nummer två är fyra.*

Welch's t-test medan sekventiellt t-test 2 är ca 15 procent sämre. Utifrån detta kan vi dra slutsatsen att bland de sekventiella t-testen är sekventiellt t-test 1 att föredra framför sekventiellt t-test 2.

I denna uppsats har vi även undersökt hur vissa resultat påverkas av en större stickprovsstorlek. Simuleringarna ovan är även gjorda för stickprovsstorleken tio men redovisas inte eftersom resultatet blir detsamma, Welch's t-test har större styrka. Då alla tre t-testen, Student's t-test, Welch's t-test och sekventiellt t-test, blir bättre med en större stickprovsstorlek är detta resultat naturligt. I Avsnitt 4.1 såg vi att om vi utför Welch's t-test istället för Student's t-test när stickprovsstorleken är tio, tappar vi endast en procent i styrka. Alltså är Welch's t-test att föredra även vid denna stickprovsstorlek.

Av jämförelserna kan vi dra slutsatsen att Welch's t-test är att föredra framför både Student's t-test och de sekventiella t-testen. Denna slutsats kan vi dra eftersom Welch's t-test är ett förhållandevis starkt test samtidigt som det ger rätt signifikansnivå under de förutsättningar vi har undersökt. Vill vi ändå använda oss av ett sekventiellt t-test är sekventiellt t-test 1 att föredra framför sekventiellt t-test 2. Observera att dessa slutsater har dragits för de kritiska gränserna som använts och presenterats i denna uppsats. Om vi använder andra kritiska gränser i F-testet är det mycket möjligt att utfallet blir annorlunda.

6 Diskussion

Syftet med denna uppsats var att undersöka egenskaperna hos sekventiella t-test och försöka konstruera ett sekventiellt t-test med större styrka än Welch's t-test och som kan dra rätt slutsatser, när det föreligger skillnad i varianserna, till skillnad mot Student's t-test.

I jämförelsen av de tre olika t-testen, Student's t-test, Welch's t-test och det sekventiella t-testet, kommer vi fram till att det bästa testet att utföra är Welch's t-test. Vi kan dock fråga oss hur pass bra det sekventiella t-testet som vi har utformat är. Det första vi kan anmärka på är de kritiska gränser som vi använder i F-testet, då vi testar om varianserna kan anses lika eller ej. De ger ett smalt område där vi kan acceptera hypotesen om lika varianser och 50 procent av gångerna, då varianserna verkligen är lika, kommer vi att felaktigt utföra Welch's t-test. Detta leder till att vi tappar styrka i det sekventiella t-testet i fallet med lika varianser. I fallet då varianserna är olika leder dessa kritiska gränser till att vi felaktigt utför en del Student's t-test. Efter att vi har undersökt hur Student's t-test beter sig under dessa omständigheter bör det sekventiella t-testet ha en högre styrka än Welch's t-test, speciellt då skillnaden i varianserna är stor. När vi jämför de två testen ser vi att så är inte fallet. Detta kan bero på att vi var tvungna att minska signifikansnivån för Welch's t-test för att kompensera den extremt höga signifikansnivån för Student's t-test i det sekventiella t-testet.

Att hitta optimala kritiska gränser för F-testet visade sig genom arbetets gång vara en svårare uppgift än förväntat och det behövs mer tid än vad som är avsatt för denna uppsats för att hitta dessa. Därmed bestämde vi oss godtyckligt för att använda de kritiska gränser som fås av signifikansnivån 50 procent i F-testet. Alltså kan vi förmodligen konstruera ett sekventiellt t-test, med andra kritiska gränser för F-testet, som är starkare och därmed bättre. Att hitta dessa optimala kritiska gränser kan vara en bra fortsättning på denna uppsats.

Genom denna uppsats har vi huvudsakligen använt oss av stickprovsstorleken fem samt gjort en del jämförelser med större stickprovsstorlekar. Då vi undersökte signifikansnivån för Student's t-test för de tre olika stickprovsstorlekarna fem, tio och tjugo, se Figur 7, fick vi intrycket av att signifikansnivån för de större stickprovsstorlekarna inte ökade på samma sätt som för stickprovsstorleken fem, utan höll sig kring eller straxt över nivån fem procent. När vi sedan undersökte signifikansnivån för stickprovsstorleken tio, betingat av F-testet, ökade den i samma takt som för stickprovsstorleken fem, se Figur 10. Dock utfördes det extremt få eller inga alls av Student's t-test vid de stora skillnaderna i varianserna, vilket gör det svårt att dra slutsatser om hur de större stickprovsstorlekarna beter sig. För att kunna dra en korrekt slutsats om hur de större stickprovsstorlekarna påverkas borde vi ha ökat antalet simuleringar.

När vi har undersökt hur det sekventiella t-testets egenskaper fungerar med en större stickprovsstorlek har vi t.ex. sett att andelen felaktigt utförda Student's t-test i det sekventiella t-testet minskar med större stickprov. Därmed har andelen förkastade sanna nollhypoteser knappt någon inverkan på signifikansnivån i det sekventiella t-testet då skillanden i varianserna är tillräckligt stor. Alltså får vi ett bättre sekventiellt t-test om vi har en större stickprovsstorlek. Samtidigt blir även Student's t-test och Welch's t-test bättre med större stickprovsstorlek. Det sekventiellt t-testet kommer alltså inte bli bättre än Welch's t-test vid en större stickprovsstorlek. I de fall där stickprovsstorleken är tillräckligt stor är det alltid ett bättre alternativ att normalapproximera.

7 Referenser

LINDGREN, B.W. (1993): *Statistical Theory.* Chapman Hall/CRC. Fourth edition. 28, 322-348.

SAWILOWSKY, SHLOMO S. (2002): *Fermat, Schubert, Einstein, and Behrens–Fisher: The Probable Difference Between Two Means When $\sigma_1 \neq \sigma_2$.* JMASM 1:461-472.

STUDENT [WILLIAM GOSSET]. (1908): *The probable error of mean.* Biometrika 1:1-25.

VEJDE, O. LEANDER, E. (2000): *Ordbok i statistik.* Borlänge Vejde förlag.

WELCH, B. L. (1947): *The generalization of student'sproblem when several different population variances are involved.* Biometrika 34:28-35.