



Att mäta SVT-program

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

Johan Brandby
Julia Wennerblom
Stina Andersson
Therese Gardell

Att mäta SVT-program

Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet

Johan Brandby Stina Andersson

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk fysik vid Chalmers

Julia Wennerblom Therese Gardell

Handledare: Maria Roginskaya Matematiska Vetenskaper
Annika Lang Matematiska Vetenskaper
Examinator: Staffan Nilsson

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

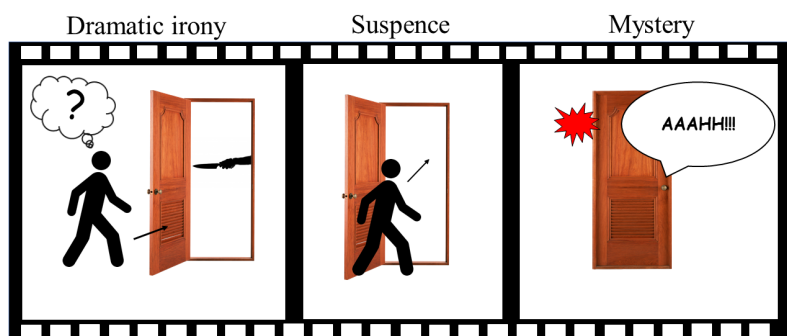
Populärvetenskaplig presentation

Att mäta SVT-program Vem vet mest och vad säljer bäst?

Vi har väl (nästan intill) alla suttit och streckkollat en serie, eller två, på diverse streamingtjänster. Kanske har vi någon gång börjat titta på en serie men gett upp efter första avsnittet då det anses "sakna potential". Det har antagligen varit för tråkigt, för hetsigt eller rent utav inte fallit oss i smaken. Men sedan så finns det de där programmen som slukar upp dig och håller dig i ett fast grepp från början till slut, och när eftertexterna rullar vill du inget hellre än att fortsätta kolla. Har du någon gång funderat över varför?

Vad det är som gör en serie bra kan bero på åtskilliga parametrar, som t.ex. budget, skådespelare, manus och regi. I kandidatarbetet *Att mäta SVT-program* har det tittats närmre på en av dessa parametrar, nämligen manus. I manus har vi det tittats på en parameter som är bland de mest kraftfulla, *de tre informationsstegen*. Den världskände dramaturgen Robert McKee skriver i sin bok, *Story*, om innehåll, struktur och stil för manusförfattning [1]. McKee pratar bland annat om *de tre informationsstegen; mystery, suspense* och *dramatic irony*. Under visning av film eller TV-serie anses tittaren besitta en viss mängd information om vad som försiggår i handlingen. De tre informationsstegen förhåller sig till huruvida tittaren vet mindre, lika mycket respektive mer än karaktärerna i den pågående scenen. I Figur 1 visualiseras ett exempel av detta i form av ett klassiskt skräckfilmsscenario som lyder:

- Det spelas dramatisk musik och en av huvudkaraktärerna rör sig mot en dörr på vid gavel. Vi som tittar får se att i dörröppningen står skurken i dramat med en kniv i högsta hugg. Huvudkaraktären har ingen aning om detta, men det har vi. Alltså har vi **dramatic irony**.
- Kameran riktas återigen mot huvudkaraktären. Vi ser hur karaktären börjar röra sig in genom dörren, men helt plötsligt står det ingen där. Vi vet nu lika mycket, eller lika lite, som huvudkaraktären. Vi har nu **suspence**.
- Huvudkaraktären går in genom dörren och stänger den efter sig. Kameran är riktad mot den slutna dörren och det hörs ett fasansfullt skrik från andra sidan. Vi vet inte vad som har hänt men det vet huvudkaraktären. Därmed har vi **mystery**.



Figur 1: Visuell beskrivning av de tre informationsnivåerna taget ur ett klassiskt skräckfilmsscenario.

I arbetet ligger fokus på hur fördelningen av dessa informationsnivåer ser ut för olika serier. Det tittas på hur ofta och hur snabbt nivåerna skiftas, hur lång tid vi spenderar i varje nivå, hur övergångarna till och från olika nivåer ser ut och slutligen hur vi kan koppla detta till TV-seriernas

popularitet¹. Marcus Purens har arbetat på Sveriges Television i mer än tjugo år. Största delen av tiden har han jobbat som klippare, men också som, projektledare, redaktör och programutvecklare. Purens är initiativtagare till projektet, tillsammans med Maria Roginskaya, biträdande professor i matematik på Chalmers Tekniska Högskola. Purens är övertygad om att disponeringen av de tre informationsstegen är en av nycklarna till att skapa en serie som tittaren "fastnar i". Undersökningen visar på att det har betydelse men det skulle behövas mer tid och resurser för att komma fram till mer säkra resultat. Likt en förstudie kan undersökningen fungera som underlag för framtida forskning inom området.

Men vad är nu upptäckterna? Den del av arbetet som tittar på hur ofta och hur snabbt nivåerna skiftas visar på ett tydligt mönster som skiljer populära och mindre populära program från varandra. Innan vi kan gå in på mönstret behöver det klarläggas att i detta arbete tittas det endast på *första* avsnittet av respektive serie. Anledningarna till detta är många. Datan, som beskriver hur de tre informationsstegen är disponerade, är insamlad av de fyra studenter som utfört kandidatarbetet och dels för att alla analyserade TV-serier ska ha tillräckligt mycket gemensamt tittas det endast på första avsnittet. Till exempel kan det första avsnittets dramaturgiska stil skilja sig en hel del från, låt säga, det sista avsnittet, eftersom det är karaktärer som ska introduceras och en handling som ska byggas upp.

Nu är vi redo att attackera resultaten! Så vad som upptäcktes var att växlingarna mellan de olika informationsnivåerna sker snabbare i program som har höga tittarsiffror. Publiken verkar alltså uppskatta när det går undan och händer mycket.

En annan del av arbetet tittar på hur stor andel av programtiden som spenderas i de olika tillstånden. Där tas det även in i beräkningarna tittarsiffror från avsnitt två. Detta då det kan visa på hur många som fortsätter följa serien efter att ha sett det första avsnittet. För att kunna hitta ett mönster testades det om det fanns ett linjärt samband mellan till exempel andel tid spenderad i *dramatic irony* och tittarsiffror. Resultaten visar på att det kan finnas ett samband och att det är som starkast för tittare på online-plattformen SVT Play i jämförelse med TV-tittare, vilket inte var fallet för undersökningen ovan. Sambandet pekar på att ju större del av tiden man vet mer än karaktärerna desto bättre. Det var också tydligare association med tittarsiffror från avsnitt två än från avsnitt ett.

Den sista delen av arbetet undersökte om skiftningar mellan de tre stadierna hade ett samband med tittarsiffrorna, och man kan se att det i vissa fall finns samband. T.ex. pekar datan på att fler skiftningar mellan *suspence* och *dramatic irony* har en positiv koppling till tittarsiffror, medan många skiftningar mellan *suspence* och *mystery* kopplar negativt till tittarsiffrorna. Detta samband syns dock endast för TV-tittare, inte för tittare på SVT play. Enkelt sagt kan man påstå att folk gillar att känna sig allsvetande, kanske för att detta är något man mycket sällan får testa på i verkligheten.

Det går alltså att hitta mönster, men i detta stadie ska de inte tas på för stort allvar. På matematikspråk kallas det att resultaten saknar statistisk signifikans vilket i korta drag betyder att det inte går att säga något om resultaten med tillräckligt hög säkerhet. Media är en komplex konstform och för att kunna dra absoluta slutsatser behöver det samlas in mer data och utföras mer heltäckande undersökningar.

¹I undersökningen baseras ett programs popularitet på tittarsiffror hämtade från MMS [2][3]. Detta inkluderar tittarsiffror från SVT Play och från TV-sändningar, de betraktas separat i undersökningen.

Sammandrag

Syftet med denna studie är att undersöka om konceptet informationsövertag i serier har en korrelation med seriens popularitet. Informationsövertag som dramaturgiskt verktyg innebär att tittaren vid varje given tidpunkt i serien antingen vet mer, mindre eller lika mycket som karaktären i serien, vilket kallas *dramatic irony*, *mystery* respektive *suspence*, och tre olika frågeställningar undersöks angående dess påverkan. Tiden som tillbringas i en informationsnivå innan det sker ett skifte undersöks med ett Anderson-Darling-test för att se om den passar någon statistisk fördelning, och om det skiljer sig mellan serier med höga och låga tittarsiffror. Slutsatsen är att det inte går att förkasta att alla väntetider är lognormalfördelade samt att växlingarna sker snabbare i program med höga tittarsiffror än i de med låga. Vidare undersöks det om totala andelen tid tillbringad i de olika informationsnivåerna har ett linjärt samband med programmets tittarsiffror. Detta utförs med linjär regression i form av minsta kvadrat-metoden. Med den insamlade datan är det svårt att säga något definitivt, men datan tyder på att det finns ett starkare linjärt samband mellan andelen tid spenderad i *dramatic irony* och tittarsiffror, än det för *mystery*. Speciellt är sambandet starkare för online-siffrorna och pekar mot att andelen tid tillbringad i *dramatic irony* korrelerar positivt med tittarsiffrorna. Sist undersöks om andelen skiftningar mellan de olika stadierna har ett linjärt samband med tittarsiffrorna, även detta med minsta kvadrat-metoden. Resultatet pekar mot att skiftningar mellan *dramatic irony* och *suspence* korrelerar positivt med tittarsiffror, medan skiftningar mellan *mystery* och *suspence* korrelerar negativt med tittarsiffror.

Abstract

The purpose of this study is to examine whether or not the concept of informational advantage in series has a correlation with the popularity of the series. Informational advantage as a dramaturgical tool means that the viewer, at any given point throughout the series, either knows more, less or the same amount as the character in the show; these stages are called *dramatic irony*, *mystery*, and *suspence* respectively. Three different questions were examined. The time spent in one stage before switching to the next is examined with an Anderson-Darling test to see if it fits any statistical distribution, and if there is a difference between series with high and low viewership numbers. The conclusion is that it cannot be rejected that the time follows a log-normal distribution and that the expected time a program will spend in one stage before switching is shorter in programs with high viewership numbers. Furthermore, whether the total percentage of time spent in the different stages has a linear relationship with the viewership numbers is examined with linear regression using the method of least squares. With the collected data it was difficult to make a definitive conclusion, however the data implies that the relationship is stronger for the time spent in *dramatic irony* and viewership numbers, than the relationship for *mystery*. The relationship is especially stronger for the online viewership. Lastly it is examined whether the percentage of switches between the different states has a linear relationship with the viewership numbers, also with the method of least squares. The result points towards switches between *dramatic irony* and *suspence* correlating positively with viewership, while switches between *mystery* and *suspence* seem to correlate negatively with viewership.

Innehåll

1	Inledning	1
1.1	Syfte och frågeställningar	1
1.1.1	Analys av väntetidens populationsfördelning	1
1.1.2	Andel tid spenderad i de tre informationstillstånden	1
1.1.3	Påverkan av skiftningar från ett stadie till ett annat	2
1.2	Resultat	2
1.3	Avgränsningar	2
2	Teori	2
2.1	De tre informationsstegen	3
2.2	Linjär regression	3
2.3	R-squared	4
2.4	Hypotesprövning	4
2.5	Q-Q-plot	4
2.6	Anderson-Darling-test	5
2.6.1	Korrigeringar för Anderson-Darling-testet vid okända parametrar	5
2.6.2	Anderson-Darling-test för exponentialfördelning	6
2.6.3	Anderson-Darling-test för gammafördelning	6
2.6.4	Anderson-Darling-test för normalfördelning	6
2.6.5	k -sample Anderson-Darling-test	7
3	Data	7
3.1	Urval av program	7
3.2	Datainsamling	7
3.3	Tittarsiffror	7
4	Analys av väntetidens populationsfördelning	8
4.0.1	Hela datamängdens väntetider	8
4.0.2	Väntetider för första och andra halvan	8
4.0.3	Väntetider för program med höga respektive låga tittarsiffror	8
4.1	Dataanalys och undersökning	8
4.2	Resultat	9
4.2.1	Hela datamängdens väntetider	9
4.2.2	Första och andra halvans väntetider	9
4.2.3	Väntetider för program med höga respektive låga tittarsiffror	11
4.3	Diskussion	13
5	Andel tid spenderad i de tre informationstillstånden	14
5.1	Dataanalys	14
5.2	Hypotes	14
5.3	Resultat	15
5.4	Diskussion	16
6	Påverkan av skiftningar från ett stadie till ett annat	17
6.1	Dataanalys	17
6.2	Resultat	17
6.3	Diskussion	18
7	Diskussion	19
7.1	Felkällor	19
7.2	Förslag för framtiden	19
7.3	Samhälleliga och etiska aspekter	20
A	Programdata	22

B	Resultat för Anderson-Darling-test av populationsfördelning	22
B.1	Resultat för enkla uppdelningar	23
B.2	Test för observationer före eller efter programmets mittpunkt, ytterligare uppdelat i TV populära eller impopulära delgrupper	24
B.3	Test för observationer före eller efter programmets mittpunkt, ytterligare uppdelat i online populära eller impopulära delgrupper	25
C	Resultat för k-sample Anderson-Darling-test	26
C.1	Separata fall för tittarsiffror och TV-programmets struktur	26
C.1.1	Homogenitet test för första halvan och andra halvan av alla program	26
C.1.2	Homogenitet test för TV populära och TV impopulära program	26
C.1.3	Homogenitet test för online populära och online impopulära program	26
C.2	Data uppdelad efter TV tittarsiffror	27
C.2.1	Test av homogenitet mellan data stratifieringen TV populära program och i första eller andra halvan av episoden	27
C.2.2	Två-stickprov test av av de fyra stratifierade dataseten	28
C.3	Data uppdelad efter online tittarsiffror	29
C.3.1	Test av homogenitet mellan data stratifieringen online populära program och i första eller andra halvan av episoden	29
C.3.2	Två-stickprov test av de fyra stratifierade dataseten	30
D	Skattade parametrar för lognormalfördelningen	31
E	Modifierade och asymptotiska förkastningsregioner	31
F	Residualsumma	32
G	Figurer	33
G.1	Alla programs väntetider	33
G.2	Första och andra halvans väntetider	35
G.2.1	Första halvan	35
G.2.2	Andra halvan	37
G.3	Höga tittarsiffror TV	39
G.4	Låga tittarsiffror TV	41
G.5	Höga onlinesiffror	43
G.6	Låga onlinesiffror	45
G.7	Första halvan av program med höga TV-siffror	47
G.8	Andra halvan av program med höga TV-siffror	49
G.9	Första halvan av program med låga TV-siffror	51
G.10	Andra halvan av program med låga TV-siffror	53
G.11	Första halvan av program med höga onlinesiffror	55
G.12	Andra halvan av program med höga onlinesiffror	57
G.13	Första halvan av program med låga onlinesiffror	59
G.14	Andra halvan av program med låga onlinesiffror	61
G.15	Plottad data för andel tid spenderad i olika stadier	62
G.16	Plottad data för skiftningar mellan stadier	65
G.17	Residualplottar för de anpassade linjerna för skiftningarna	71

Förord

En loggbok har förts över de enskilda medverkandes prestationer under arbetet.

Gemensamt arbete Avgränsning av projektet var en stor del av arbetet då projektbeskrivningen var otroligt bred. Avgränsningen samt datainsamlingen har varit gemensamt arbete.

Julias delar

Julia har enskilt ansvarat för allt angående *Påverkan av skiftningar från ett stadie till ett annat*.

- Sammandrag
- Abstract
- 1.1 Syfte och frågeställningar
- 1.1.3 Påverkan av skiftningar från ett stadie till ett annat
- 1.3 Avgränsningar
- 2.2 Linjär Regression (första delen, inte delen om de fem villkoren)
- 2.4 Hypotesprövning
- 6 Påverkan av skiftningar från ett stadie till ett annat (alla underdelar)
- 7.1 Felkällor
- 7.2 Förslag till framtiden
- F Residualsumma
- G.16 Plottad data för skiftningar mellan stadier
- G.17 Residualplottar för de anpassade linjerna för skiftningarna

Johans delar

Johan har gemensamt med Stina haft enskilt arbetat med *Analys av väntetidens populationsfördelning*. Vidare, har Johan haft större delen av informationssökningen för hypotesprövning för fördelning och homogenitetstest av stickprov. Därtill har Johan stått för implementeringen av de test för deras del *Analys av väntetidens populationsfördelning*.

- 1.1 Syfte och frågeställningar
- 1.1.1 Analys av väntetidens populationsfördelning
- 2.5 Q-Q-plot
- 2.6 Anderson-Darling-test
- 3.2 Datainsamling
- 3.3 Tittarsiffror
- 4 Analys av väntetidernas populationsfördelning
- 4.3 Diskussion
- B Resultat för Anderson-Darling-test av populationsfördelning
- C Resultat för k-sample Anderson-Darling-test
- D Skattade parametrar för lognormalfördelningen

Stinas delar

Stina har gemensamt med Johan haft enskilt arbetat med *Analys av väntetidens populationsfördelning*. Tillsammans med Johan ansvarat för teorin kring hypotesprövningen för testens fördelningar samt homogenitetstest.

- 1 Inledning
- 1.1 Syfte och frågeställningar
- 1.1.1 Analys av väntetidens populationsfördelning
- 2.2 Linjär Regression
- 2.3 R-squared
- 2.6 Anderson-Darling-test
- 3.1 Urval av program
- 3.2 Datainsamling
- 3.3 Tittarsiffror
- 4.2 Resultat
- 4.3 Diskussion
- 7.2 Förslag för framtiden (sista stycket)
- E Modifierade och asymptotiska förkastningsregioner
- G.1-G.14

Thereses delar

Therese har enskilt ansvarat för allt angående *Andel tid spenderad i de tre informationstillstånden*.

- Populärvetenskaplig rapport
- 2.1 De tre informationsstegen
- 5 Andel tid spenderad i de tre informationstillstånden
- 7.3 Samhälleliga och etiska aspekter
- G.15 Plottad data för andel tid spenderad i olika stadier
- A Programdata

Tack till

Vi vill passa på att tacka Marcus Purens för hans assistans och visdom och enorma engagemang. Den har guidat oss mycket på vägen. Vi vill också tacka Maria Roginskaya och Annika Lang för deras tid och stöd under arbetets gång.

1 Inledning

Manusförfattande ses idag nog mer som en "business" som genererar miljarder än som en konstform där en person sitter ensam och skriver på ett mästerverk, skribenternas intuition och rika kunskap om dramaturgiska regler används till att förutsäga vad som potentiellt kan bli en bra berättelse. Inom såväl litteratur som musik har det gjorts försök att, med utgångspunkt i matematiska modeller, analysera vad det är som skapar en framgångsrik produkt. Detta verkar emellertid inte gjorts förut inom dramaturgi. Det finns femåriga utbildningar inom manusskrivande men det har aldrig gjorts uppmärksammade försök till att visualisera dramaturgi och dess effekter. Det är med bakgrund i detta som Marcus Purens från SVT sökt sig till Chalmers för att se om det går att hitta ett samband mellan ett programs dramaturgiska strukturer och dess popularitet.

De dramaturgiska strukturer som analyseras i detta arbete är *de tre informationsstegen: mystery, suspense* och *dramatic irony*, förkortat MSD. Dessa är begrepp myntade av den världskända dramaturgen Robert McKee [1] och handlar om mängden information tittaren, i den aktuella scenen, har tillgång till i förhållande till seriens karaktärer. *Mystery* innebär att tittaren vet mindre än karaktärerna, *suspence* att tittare och karaktärer vet lika mycket och *dramatic irony* att tittaren sitter på viktig information som karaktärerna saknar. Populariteten hos ett program mäts i det här arbetet av dess tittarsiffror, vilka i sin tur delas in efter online- och TV-siffror. Purens teori är att TV-serier som nått stor framgång är skrivna av manusförfattare som är mycket skickliga på att växla mellan dessa tre informationsnivåer. Framförallt, förmodar han, är det skiftningar mellan *mystery* och *dramatic irony* som fångslar TV-tittare.

1.1 Syfte och frågeställningar

Syftet med detta arbete är att undersöka huruvida det går att hitta en korrelation mellan ett TV-programs tittarsiffror och dess innehåll av de tre informationsstegen. Arbetet har delats in i tre delar, som i rapporten kommer presenteras var för sig, med dataanalys, resultat och diskussion. Första delen handlar om att undersöka beteendet av det som fortsättningsvis betecknas som *väntetider*. Detta är tiden som spenderas i ett informationssteg innan det sker en växling till ett annat. I denna del del undersöks väntetidernas fördelningar. I andra delen studeras hurvida andelen tid spenderad i respektive informationssteg går att associeras till ett programs tittarsiffror i det första och andra avsnittet. Slutligen betraktas antalet hopp från *mystery* till *dramatic irony*, såväl som hopp mellan alla andra stadier, för att se om detta har någon korrelation med populariteten. Frågeställningarna associerade till respektive delfråga finns presenterade mer ingående nedan.

1.1.1 Analys av väntetidens populationsfördelning

Arbetets första del behandlar studiet av den tid som spenderas i ett informationssteg innan programmet växlar till ett nytt. Med andra ord studeras tidsintervallet från och med att en informationsnivå påbörjas tills det att en ny nivå inleds; denna tid kallas för *väntetid*. Målet är att undersöka väntetidernas fördelning, vilket görs i tre delar. Först studeras fördelningen för *alla* väntetider. Därefter delas programmets tid på hälften och det undersöks om det finns någon skillnad för hur väntetiderna är fördelade mellan *första* halvan och *andra* halvan. Det undersöks också vilka fördelningar dataseten följer. Slutligen delas programmen in efter *höga* och *låga* tittarsiffror och även här studeras om det, mellan dessa, går att hitta en skillnad i väntetidernas fördelning, samt vilka dessa fördelningar är. Hypotesen är att väntetiderna är kortare i program med höga tittarsiffror än i program med låga tittarsiffror.

1.1.2 Andel tid spenderad i de tre informationstillstånden

Har det någon betydelse hur stor andel av programmet som spenderas i de olika tillstånden? Större delen av ett program tenderar spenderas i tillståndet *suspence*; hypotesen är att det behövs ett flertal inslag av *mystery* och *dramatic irony* för att bibehålla tittarens intresse och delaktighet i programmet. Det testas därför i detta avsnitt om ett linjärt samband går att hitta mellan andel tid spenderad i *mystery* respektive *dramatic irony* och programmets uppfattade popularitet. Populariteten mäts i denna del även av tittarsiffror från avsnitt två i serien.

1.1.3 Påverkan av skiftningar från ett stadie till ett annat

Den tredje och sista frågeställningen berör huruvida antalet hopp mellan olika specifika stadier påverkar programmets popularitet. Hypotesen är att många hopp mellan *mystery* och *dramatic irony* påverkar positivt. De hopp som undersöks är *mystery*→*dramatic irony* liksom hopp mellan alla andra möjliga stadier.

1.2 Resultat

Resultatet av analysen av väntetidens populationsfördelning blev att den enda undersökta fördelningen för datan som inte kunde förkastas var lognormalfördelningen. Detta gällde för samtliga indelningar av datan. Dessutom observerades för flera av indelningarna att det fanns en skillnad i den underliggande fördelningen, vilket innebär att de har olika värden på parametrarna.

Angående andelen tid spenderad i de olika tillstånden pekade resultaten mot att andelen tid tillbringad i *dramatic irony* hade en positiv korrelation med tittarsiffror. Det observerades att denna korrelation var mer pålitlig för online-siffror än för TV-siffror.

Vid undersökning av skiftningarna mellan stadier pekade resultaten mot att skiftningar mellan *dramatic irony* och *suspence*, i båda riktningarna, hade ett positivt samband med tittarsiffror. Skiftningar mellan *mystery* och *suspence* verkade däremot ha ett negativt samband med tittarsiffrorna i båda riktningarna. Detta mönster går dock endast att se när man undersöker en persons data separat och jämför med TV-siffror.

1.3 Avgränsningar

Åtskilliga parametrar avgör hur media uppfattas, och det finns en uppsjö av media att undersöka, så för att kunna göra ett genomförbart projekt har många avgränsningar gjorts. Med sin intuition och erfarenhet inom området rekommenderade Marcus Purens att konceptet MSD undersöks, då detta har stor potential att på egen hand ha en korrelation med programmets popularitet.

Fokus i rapporten ligger på fiktiva serier då dessa antogs av Purens ha tydligare användning av MSD än t.ex. lekprogram och dokumentärer. Eftersom MSD beskriver kunskap och information analyseras endast det *första* avsnittet i serien, detta för att tittarsiffrorna i största möjliga mån ska reflektera det aktuella avsnittet och inte påverkas av hur intresseväckande tidigare avsnitt varit. I viss mån undersöks även hur många som fortsätter titta på avsnitt två. Dessutom, för att ha tillgång till pålitlig data, används bara program som har sänts på SVT. Endast program som har sänts både på TV och på SVT-play har betraktats. Anledningen är att samma program ska finnas med i analysen oavsett om datan delats in efter TV- eller onlinesiffror. Avvägningen gjordes också att begränsa datan till program mellan 40-60 min långa, detta för att få ett så stort urval som möjligt, och fortfarande betrakta program stor del av befolkningen tittar på. Såvida inget annat nämns syftar *tittarsiffror* på pilotavsnittets tittarsiffror. Dessutom delas tittarsiffrorna in efter online och broadcast, där det första syftar på antalet tittare på streamingtjänsten SVT-play och det senare avser mängden som tittat på programmet på TV.

Självklart är det inte uppenbart att avgöra huruvida programmet uppfattas som ”bra”, så i detta projektet avgörs det av tittarsiffror. Detta kan vara missvisande eftersom andra faktorer påverkar mycket, som t.ex. PR, budget, sändningstid, estetik i filmningen, medverkande av kända skådespelare, vilka program som går samtidigt på TV, ämne och manus, för att nämna några stycken. Det kan dock ses som ett rimligt antagande att mängden människor som sett ett program har ett samband med uppskattning hos publiken.

2 Teori

I avsnitten nedan presenteras den teori som arbetet bygger på. Informationsnivåerna presenteras i detalj i avsnitt 2.1. Teorin för linjär regression, som används i avsnitt 5 och 6, introduceras i avsnitten 2.2-2.4. Den teori för Anderson-Darling-test som används i avsnitt 4 presenteras i 2.6 nedan.

2.1 De tre informationsstegen

Mystery, *suspence* och *dramatic irony* är begrepp myntade av den världskända dramaturgen Robert McKee. McKee kallar dem *de tre informationsstegen* [1], vi kallar dem kort för MSD. Under visning av film eller TV-serie anses tittaren besitta en viss mängd information om vad som försigår i handlingen. De tre informationsstegen förhåller sig till om tittaren vet mindre, lika mycket respektive mer än karaktärerna i den pågående scenen. För att göra det tydligt kommer här ett exempel ur ett klassiskt skräckfilmsscenario.

- Det spelas dramatisk musik och en av huvudkaraktärerna rör sig mot en stängd dörr. Kameran visar för oss som tittar att bakom dörren står skurken i dramat med en kniv i högsta hugg. Huvudkaraktären har ingen aning om detta, men det har vi. Alltså har vi *dramatic irony*.
- Kameran riktas återigen mot huvudkaraktären. Vi ser hur dörren öppnas och helt plötsligt står det ingen där. Vi vet nu lika mycket, eller lika lite, som huvudkaraktären. Vi har nu *suspence*.
- Huvudkaraktären går in genom dörren och stänger den efter sig. Kameran är riktad mot den slutna dörren och det hörs ett fasansfullt skrik från andra sidan. Vi vet inte vad som har hänt men det vet huvudkaraktären. Därmed har vi *mystery*.

2.2 Linjär regression

Linjär regression baserar sig på antagandet att det finns ett linjärt samband mellan den beroende variabeln y och den förklarande variabeln x . Modellen som antas är $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Datan anpassas till modellen med hjälp av minsta kvadratmetoden, vilket finns förklarat i bland annat *Mathematical Statistics and Data Analysis* av John A. Rice [4]. Målet med minsta kvadratmetoden är att ta fram β_0 och β_1 som minimerar

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2, \quad (1)$$

där y_i och x_i , $i = 1, 2, \dots, n$, är de observerade datapunkterna. De skattade parametrarna som minimerar Q kallas för $\hat{\beta}_0$ och $\hat{\beta}_1$. Regressionslinjen ges då av $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

För att minsta kvadratmetoden ska vara en lämplig skattningsmetod måste datan uppfylla fem villkor. Det går att undersöka om datan uppfyller villkoren genom att plotta de så kallade *residualerna*. De definieras som $e_i = y_i - \hat{y}_i$, där y_i är observation i , $i = 1, \dots, n$, och \hat{y}_i är motsvarande punkt på den skattade regressionslinjen. Villkoren, samt hur de kontrolleras, finns beskrivna i detalj nedan.

1. Modellen är lämplig för datan

Ser en linjär modell ut att passa datan? Om en linjär modell är lämplig för datan ska det inte finnas några trender eller grupper bland punkterna då e_i plottas mot x_i [5].

2. Residualerna är symmetrisk fördelade kring 0

Minska kvadratmetoden tar fram de skattningar som minimerar summan av de *kvadrerade* felen. På grund av kvadraten kommer positiva och negativa fel behandlas lika. Skattningen tar alltså inte hänsyn till om de observerade punkterna ligger ovanför eller under den skattade linjen. För att minsta kvadratmetoden ska vara en lämplig skattningsmetod ska de positiva och negativa felen vara ungefär lika stora [5]. Alltså ska residualerna, i residualplotten, vara symmetriskt fördelade kring 0 [6].

3. Okorrelerade residualer

I minsta kvadratmetoden bidrar alla observationer lika mycket till skattningarna. Om dessa observationer på något sätt delar information kommer residualerna vara korrelerade. Detta kan hända om en observation är en upprepad mätning eller om observationerna på något sätt är grupperade (kommer från studenter i samma klass, personer i samma stad etc). Om upprepade eller grupperade observationer finns kommer minsta kvadratmetoden ge felaktiga

skattningar då den inte tar hänsyn till data som denna. Det finns emellertid inget sätt att testa villkoret om okorrelerade residualer. Istället måste datainsamling och experimentdesign studeras så att de inte ger upphov till oönskade grupperingar eller korrelerad data [5].

4. Residualerna har konstant varians

Eftersom alla observationer bidrar lika mycket till minska kvadrat-skattningarna görs antagandet, när metoden används, att bruset i alla observationer är ungefär lika stort [5]. Om detta gäller kommer residualerna vara konstanta och inte bero av x . Alltså kommer de ha lika spridning kring 0 [6].

5. Inga outliers

I linjär regression antas alla observationer komma från samma underliggande modell. Om någon observation inte uppfyller villkoret (på grund av mätfel, ett ovanligt utfall etc) kan den komma att ha mycket stor påverkan på den resulterande regressionslinjen. En observation som kraftigt avviker från resterande värden kallas *outlier*. Vissa outliers kan hittas i residualplotten. Dock kan en outlier ha stor påverkan på regressionslinjen och kan då i själva verket ha små residualer. Outliers som dessa kan inte hittas i residualplotten utan kan eventuellt upptäckas när observationerna plottas [5].

2.3 R-squared

Ett mått på hur väl regressionen passar datan är R-squared, R^2 , vilket anger andelen av den totala variansen i y som förklaras av den linjära modellen [7]. Definiera variansen som kan förklaras av modellen som $SS_{REG} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, där \hat{y}_i är regressionsmodellens skattning av observation i och \bar{y} är medelvärdet av alla observerade värden y_i , $i = 1, \dots, n$. Den totala variansen i y definieras av $SS_{TOT} = \sum_{i=1}^n (y_i - \bar{y})^2$. R^2 beskrivs då av sambandet [7]

$$R^2 = \frac{SS_{REG}}{SS_{TOT}}. \quad (2)$$

Det gäller att $R^2 \in [0, 1]$ och då $R^2 = 1$ följer modellen den observerade datan perfekt. Alltså är R^2 ett mått på styrkan av det linjära sambandet mellan x och y , där värden nära 1 indikerar ett starkt linjärt samband.

2.4 Hypotesprövning

Vid statistisk analys är det intressant att undersöka hur statistiskt signifikant ett resultat är. För att göra detta kan man testa hypotesen mot en alternativ hypotes. Nollhypotesen H_0 som undersöks är huruvida linjens lutning $\beta_1 = 0$, och testas mot den alternativa hypotesen H_1 att $\beta_1 \neq 0$ vilket är beskrivet bland annat av Rice [4]. Detta kallas att tväsidigt test. Teststatistikan fås av

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}},$$

där $\hat{\beta}_1$ är lutningen på linjen och $s_{\hat{\beta}_1}$ är roten ur den skattade variansen - standardfelet. Förkastningsregionerna ges sedan av $|t| > t_{\alpha/2}$ där α är signifikansnivån. Testets p -värde är sannolikheten att, under H_0 , få ett högre värde på teststatistikan t än det observerade. Låt t_{obs} vara teststatistikans värde för datan som testas. Testets p -värde ges då av $p = \mathbb{P}(|t| > t_{obs} | H_0 \text{ sann})$ [7].

2.5 Q-Q-plot

Q-Q-plottar är ett grafiskt analysverktyg som används för att jämföra kvantiler från två fördelningsfunktioner [4][8]. Då ett stickprovs kvantiler jämförs med de från en hypotetisk fördelning F kan skillnaden mellan dessa visualiseras och genom detta informellt klargöra i vilken grad datan passar den antagna fördelningsfunktionen.

Konstruktionen av en Q-Q-plot görs genom att ordna observationerna X_1, \dots, X_n så att de är i storleksordning, $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ [4]. Därefter plottas den i :te kvantilen från den hypotetiska fördelningen mot den i :te observationen $X_{(i)}$ i ett punktdiagram. Om dessa kommer från en liknande fördelning följer punkterna approximativt en rät linje i punktdiagrammet. Avviker de systematiskt från linjen så finns det belägg för att de inte har liknande fördelning.

Om det uppstår så kallade ties i datan, vilket betyder att flera observationer har samma värde, kommer de ha samma stickprovskvantil. Observationerna kommer alltså ligga på en vågrät linje i QQ-plotten, där varje observation har ett teoretiskt kvantilvärde. För att få en korrekt Q-Q-plot tas medelvärdet av alla dessa teoretiska kvantiler och plottas mot en av observationerna i fråga [8].

2.6 Anderson-Darling-test

För att avgöra hur väl ett stickprov följer en statistisk fördelning används så kallade *goodness-of-fit*-test, vilka är metoder som avgör hur väl ett dataset överensstämmer med en teoretisk fördelning. Noll- och alternativhypotesen för ett *goodness-of-fit*-test ges av

$$\begin{aligned} H_0 &: \text{Stickprovet } x_1, \dots, x_n \text{ kommer från } F(x; \theta), \\ H_1 &: \text{Stickprovet } x_1, \dots, x_n \text{ kommer inte från } F(x; \theta), \end{aligned} \quad (3)$$

där $F(x; \theta)$ är en fördelningsfunktion och θ är en vektor med dess parametrar [8]. Till exempel gäller för normalfördelningen att $\theta = [\mu, \sigma^2]$. Parametrarna i θ kan vara kända eller okända.

Ett exempel på *goodness-of-fit*-test är Anderson-Darling-testet som testar om ett stickprov kommer från en teoretisk fördelning med hjälp av dess empiriska fördelningsfunktion. Den empiriska fördelningsfunktionen är en trappfunktion som, då den konstrueras från stickprovet, är en approximation av populationens fördelningsfunktion [8] och definieras av

$$F_n(x) = \frac{\sum_{i=1}^n \mathbb{1}_{\{x \leq x_i\}}}{n}, \quad -\infty < x < \infty, \quad (4)$$

där $\mathbb{1}$ är indikatorfunktionen och n är stickprovsstorleken. Den empiriska fördelningsfunktionen är en konsistent skattning av populationens fördelningsfunktion [8]. Det gäller alltså att för alla $x \in \mathbb{R}$ och för alla $\epsilon > 0$ finns ett $n_0^{x, \epsilon} \in \mathbb{N}$ sådant att

$$\forall n \geq n_0^{x, \epsilon} \Rightarrow |F_n(x) - F(x)| < \epsilon. \quad (5)$$

Låt x_1, \dots, x_n vara ett slumpmässigt stickprov. Teststatistikan för ett Anderson-Darling-test med H_0 och H_1 definierat som i (3). Testvariabeln ges då av

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x, \theta)]^2}{F(x, \theta)[1 - F(x, \theta)]} dF(x, \theta), \quad (6)$$

där n är stickprovsstorleken, $F_n(x)$ är stickprovets empiriska fördelningsfunktion och $F(x, \theta)$ fördelningsfunktionen för den fördelning som testas [8]. Nollhypotesen förkastas vid signifikansnivå α om det, för det kritiska värdet R_α , gäller att $A^2 > R_\alpha$. Metoden är särskilt användbar då ett stickprov testas för en familj av fördelningar och fördelningens parametrar måste skattas från datan [8]. Emellertid får detta konsekvenser för testets riktighet. Då parametrar skattas från samma stickprov som Anderson-Darling-testet används på, beror teststatistikan även på vilken skattningsmetod som används samt stickprovets riktiga fördelning.

2.6.1 Korrigeringar för Anderson-Darling-testet vid okända parametrar

För Anderson-Darling-testet finns välutvecklad asymptotisk teori, då $n \rightarrow \infty$, för att få fram förkastningsregionerna [8]. Emellertid, som nämns i föregående avsnitt, ändras testvariabelns fördelning under nollhypotesen då en eller flera parametrar i den antagna populationfördelningen är okända. Däremot är det känt att testvariabelns fördelning under nollhypotesen, för ändliga stickprovsstorlekar, snabbt konvergerar mot de asymptotiska värdena [8]. I fallet för normal- och exponentialfördelningen kan modifieringar av testvariabeln göras för att minska effekten av de okända parametrarna och därmed få korrekta resultat även i detta fall [8][9]. Dessa korrigeringar presenteras i de följande avsnitten.

Trots att det finns modifieringar för exempelvis normal- och exponentialfördelningarna finns inget motsvarande för gammalfördelningen. Orsaken är att den innehåller en formparameter, m [8][9]. Anledningen till att närvaro av m är problematisk är att den inte kan standardiseras bort ur beräkningsformlerna. Detta medför att nollfördelningen beror på det riktiga värdet av formparametern, och således kan förkastningsregionerna inte beräknas asymptotiskt. Emellertid påpekar D'Agostino och Stephens att de asymptotiska konfidensgränserna är stabila för små ändringar i m och Anderson-Darling-testet, för gammalfördelningen, kan användas approximativt då den estimerade formparametern \hat{m} används istället för m [8].

2.6.2 Anderson-Darling-test för exponentialfördelning

En exponentialfördelning definieras av fördelningsfunktionen

$$F(x; \beta) = 1 - \exp\left(\frac{-x}{\beta}\right), \quad \beta > 0, \quad (7)$$

där β är fördelningens parameter. Anderson-Darling-testet för exponentialfördelningen prövar hypoteserna

$$\begin{aligned} H_0: & \text{ Stickprovet } x_1, \dots, x_n \text{ kommer från } \text{Exp}(\beta), \\ H_1: & \text{ Stickprovet } x_1, \dots, x_n \text{ kommer inte från } \text{Exp}(\beta). \end{aligned}$$

Då β inte känns till sedan innan måste parametern skattas från datan. Anderson-Darling-testet bygger på att denna skattning görs med Maximum Likelihood-metoden [8]. Den modifierade teststatistikan då β skattas från datan ges av

$$A_{\text{mod}}^2 = A^2 \left(1, 0 + \frac{0,6}{n}\right), \quad (8)$$

där n återigen är stickprovsstorleken och A^2 är definierat som i ekvation (6) [8]. Testets konfidensgränser presenteras i Tabell 53 i Appendix E.

2.6.3 Anderson-Darling-test för gammalfördelning

Ett Anderson-Darling-test för gammalfördelningen testar hypoteserna

$$\begin{aligned} H_0: & \text{ Stickprovet } x_1, \dots, x_n \text{ kommer från } \text{Gamma}(m, \beta), \\ H_1: & \text{ Stickprovet } x_1, \dots, x_n \text{ kommer inte från } \text{Gamma}(m, \beta), \end{aligned}$$

där $\text{Gamma}(m, \beta)$ har täthetsfunktionen $f(x) = \frac{x^{m-1}}{\Gamma(m)\beta^m} \exp(-x/\beta)$, med $\beta > 0$ och $m > 0$. Vidare definieras funktionen $\Gamma(x) = \int_0^\infty z^{x-1} e^{-z} dz$.

I fallet då både β och m är okända och måste skattas från datan, kommer skattningen av formparametern m vara beroende av parameterens faktiska värde [8]. Detta gör att Anderson-Darling-testet, även för stora datamängder, endast kan användas som ett approximativt test. Det har dock visat sig att approximationen ligger nära de faktiska värdena, även för små stickprovsstorlekar [8]. Teststatistikan ges av ekvation (6) och förkastningsregionerna hittas i Tabell 54 i Appendix E. Nollhypotesen förkastas om $A^2 > R_\alpha$.

2.6.4 Anderson-Darling-test för normalfördelning

Anderson-Darling-testet är ett av de mest kraftfulla testen för normalfördelning [8]. Noll- och alternativhypotesen ges av

$$\begin{aligned} H_0: & \text{ Stickprovet } x_1, \dots, x_n \text{ kommer från } \mathcal{N}(\mu, \sigma^2), \\ H_1: & \text{ Stickprovet } x_1, \dots, x_n \text{ kommer inte från } \mathcal{N}(\mu, \sigma^2). \end{aligned}$$

Om μ och σ^2 är okända skattas de från datan med Maximum Likelihood-metoden. I detta fall ges den modifierade teststatistikan för Anderson-Darling-testet av [8]

$$A_{\text{mod}}^2 = A^2 \left(1, 0 + \frac{0,75}{n} + \frac{2,25}{n^2}\right). \quad (9)$$

Kritiska punkter, R_α , för normalfördelningstestet då μ och σ^2 skattas från datan hittas i Tabell 52 i Appendix E.

2.6.5 k -sample Anderson-Darling-test

Scholz och Stephens har presenterat en utökning av Anderson-Darling-testet som kan pröva om k olika stickprov kommer från samma population [10]. Med andra ord kan testet pröva nollhypotesen att k stickprov är homogena mot alternativet att åtminstone ett inte är det. Notera att två homogena stickprov kommer från *samma* fördelningsfamilj med *samma* värden på parametrarna. Testet är icke-parametriskt; det görs alltså inga antaganden om datans fördelning. Teststatistikan ges av

$$A_{kN}^2 = \sum_{i=1}^k n_i \int_{B_N} \frac{[F_{in_i}(x) - H_N(x)]^2}{H_N(x)[1 - H_N(x)]} dH_N(x), \quad (10)$$

där n_i är antalet observationer för stickprov i , $F_{in_i}(x)$ är den empiriska fördelningsfunktionen för stickprov i , $N = n_1 + \dots + n_k$ är summan av alla stickprovsstorlekar, $H_N(x)$ är den empiriska fördelningsfunktionen för alla stickprov och $B_N = \{x \in \mathbb{R} : H_N(x) < 1\}$. Teststatistikan (10) kan korrigeras sådant att den tar hänsyn till fallen då kontinuerlig data diskretiseras och därmed kan ge upphov till observationer med lika värden. Dessutom standardiseras testvariabeln, vilket gör testet mindre beroende av stickprovsstorleken [10].

3 Data

I detta avsnitt beskrivs datan som används i arbetet samt hur datainsamlingen gått till.

3.1 Urval av program

Första delen i arbetet handlade om att välja ut vilka program som skulle analyseras. För att få ett så tillförlitligt resultat som möjligt var det viktigt att program med tittarsiffror i hela spannet skulle finnas med. Därför valdes tre nivåer; färre än 400 000 tittare, 400 000 - 900 000 tittare samt fler än 900 000 tittare. Dessa tittarsiffror är sammanslagningen av online- och TV-siffrorna. Därefter valdes ungefär lika många program i varje intervall bland programmen som sänts på SVT under perioden 2012-01-01 till 2018-12-31. Det blev totalt 37 stycken.

3.2 Datainsamling

Som nämns i inledningen baseras studien på dramaturgen Robert McKees tre informationsnivåer, vilka finns beskrivna i avsnitt 2.1. Första steget i arbetet var att samla in data. Detta gjordes genom att markera MSD-skiftningarna i de serier som valts ut för analysen. Varje gång det skedde en förändring i hur mycket information tittaren har i förhållande till karaktärerna gjordes en markering, 1, 2 eller 3, där siffran är kopplad till den nya informationsmängden. 1, 2 och 3 motsvarar *mystery*, *suspence* respektive *dramatic irony*. Det noterades också hur många minuter och sekunder in i programmet förändringen gjordes. Alla medlemmar i gruppen har analyserat ungefär lika många avsnitt. I ett försök att minimera skillnaden i analysen mellan olika personer har alla personer i gruppen, var för sig, tittat på första avsnittet av serien *Vår tid är nu*. Detta gjorde även Purens, då han är den som besitter mest kunskap och erfarenhet inom området. Gruppmedlemmarna diskuterade sedan, tillsammans med Purens, hans analys och satte sig därefter och tittade igenom avsnittet igen med Purens ord och analys i bakhuvudet. Målet var att alla gruppmedlemmar skulle vara överens om hur avsnitten skulle analyseras och att resultatet därmed skulle bli mindre beroende av vem som samlat in datan.

3.3 Tittarsiffror

Förutom McKees informationsnivåer behövdes tittarsiffror för varje analyserat avsnitt, vilka tillhandagavs av Mediamätning i Skandinavien AB (MMS). I det här arbetet används två typer av mätdata från MMS: antalet tittare under TV-sändningen (broadcast) och antalet som tittat via streamingtjänsten SVT-play (online). Siffrorna för broadcast är baserat på TV-tittandet hos en testpanel bestående av 3000 svenska hushåll [2]. Dessa ska vara ett representativt urval av Sveriges befolkning. Det har visat sig att variablerna *utbildningsnivå*, *tittarkonsumtion*, *familjesituation* och

ålder har större påverkan på TV-tittandet än andra. Därför läggs större vikt vid dessa vid urvalet av panelen. Varje panelhushåll har en mätare och varje person i hushållet har ett individuellt inlogg. Därefter registreras vilka program individen tittar på samt alla knapptryckningar. Det är sedan denna data som ligger till grund för skattningen av hela befolkningens TV-tittande [2].

För det webbaserade tittandet skattar MMS tittarsiffrorna genom att mäta tiden ett program spelas samt antalet starter. Tittarsiffran för avsnittet fås sedan fram genom att summera all uppspelad tid och dela denna med antalet starter. Enligt MMS är detta mått det som är mest jämförbart med det för TV-tittandet [3]. Det finns alltså, för online-siffrorna, inga panelhushåll utan all sådan aktivitet på streamingtjänsterna registreras. För båda måtten samlas data in från och med TV-avsnittets publiceringsdag och under de 7 kommande dagarna [2][3].

4 Analys av väntetidens populationsfördelning

I den här delen beskrivs tillvägagångssättet för dataanalysen, dess resultat samt en kortare diskussion för arbetets första frågeställning, den gällande väntetidens populationfördelning. Senare, i avsnitt 7, fortsätter diskussionen med en analys av felkällor samt förslag inför eventuella framtida utvecklingar av studien. Undersökningen består av tre delar. Dataseten som analyseras i varje del finns beskrivna i detalj nedan.

4.0.1 Hela datamängdens väntetider

Detta dataset består av väntetiderna för alla program. För denna data tas ingen hänsyn till eventuella grupperingar i datan, utan alla observationer slås samman. Målet är att undersöka om de här väntetiderna följer en känd fördelning.

4.0.2 Väntetider för första och andra halvan

På grund av såväl erfarenhet från datainsamlingen som konsultation med Purens på SVT ansågs det rimligt att anta att väntetiderna är olika beroende på vart i programmet skiftningarna sker. Därför studeras också fördelningar för väntetiderna för första respektive andra halvan av programmen. Det undersöks om väntetiderna i de två dataseten kommer från samma underliggande fördelning samt vilken eller vilka fördelningarna är.

4.0.3 Väntetider för program med höga respektive låga tittarsiffror

Målet är att testa Purens hypotes att tiden som spenderas i ett informationssteg innan det sker en växling skiljer sig åt mellan program som har höga och program som har låga tittarsiffror. För att undersöka detta delas programmen in i två grupper, populära och impopulära, baserade på antalet tittare. Indelningen görs efter både online- och TV-siffror. Det testas om dataseten kan antas komma från samma fördelning samt vilka de underliggande fördelningarna är. Dessutom delas de populära och impopulära programmen in i första och andra halvan. Därefter testas om väntetiderna i *första* halvan i *populära* program kan antas komma från samma underliggande fördelning som väntetiderna i *första* halvan i *impopulära* program. Motsvarande görs även för andra halvan. Slutligen undersöks vilka fördelningar dessa dataset kommer ifrån.

4.1 Dataanalys och undersökning

Analysen började med att avgränsa populära och impopulära program genom att sätta ett tröskelvärde för tittarsiffrorna för TV respektive online. Vart gränsen för höga och låga tittarsiffror skulle dras diskuterades med en av SVT:s marknadsanalytiker. Han ansåg att gränsen bör sättas till 800 000 tittare för TV och 80 000 för online. Dessa gränser ansågs rimliga utifrån vad SVT anser är höga och låga tittarsiffror; samtidigt som de ger någorlunda lika stora dataset då datan delas in efter antalet tittare. Skillnaden mellan de olika tröskelvärdena beror på att det endast är en bråkdel av allt tittande som sker via streamingtjänsten.

Därefter konstruerades ett histogram för alla programs väntetider. Detta gjordes i syfte att få en uppfattning av vilka fördelningar som skulle vara lämpliga att testa datan för. Följden blev

att fördelningarna exponential, gamma och lognormal skulle undersökas närmre. Nästa steg var en grafisk analys av hur väl de tre fördelningarna passar datan. Den grafiska analysen baserades på väntetidernas Q-Q-plottar och gav en indikation på vilka fördelningar som bäst kunde passa datan. Emellertid kan dessa undersökningarna bara ge en antydning om lämpliga fördelningar och formella test behövdes därför för att ge konkreta resultat.

För att testa datan för de givna fördelningarna användes Anderson-Darling-test eftersom testet har välutvecklad asymptotisk teori samt modifieringar av teststatistikan då okända parametrar existerar. Mer detaljerade beskrivningar av testet hittas i avsnitten 2.6.2, 2.6.3 samt 2.6.4. Emellertid, då datan testades för lognormalfördelning gjordes Anderson-Darling-testet för normalfördelningen på logaritmen av väntetiderna. Detta eftersom det gäller att om X är lognormalfördelad medförs att $Y = \log X$ normalfördelad [11].

Som nämnts ovan fanns det anledning att anta att beteendet hos väntetiderna skiljer sig åt för programmets första och andra halva samt för program med höga och låga tittarsiffror. Det var därför av intresse att undersöka fördelningarna av dessa grupperingar. Analysen gjordes enligt följande. Först undersöktes om väntetiderna i programmets första och andra halva kan antas komma från samma underliggande fördelning. Motsvarande gjordes för populära och impopulära program. Detta undersöktes med ett k -Sample Anderson-Darling-test, vilket testar homogeniteten mellan k stickprov [10]. Testet valdes eftersom det bygger på samma grundteori som det vanliga Anderson-Darling-testet och inte kräver något antagande om den underliggande populationsfördelningen.

Sista delen av dataanalysen gick ut på att undersöka vilka fördelningar väntetiderna i ovan grupperingar följer. Detta gjordes på precis samma sätt som analysen av hela datamängdens väntetider. Första steget var alltså grafiska undersökningar av den grupperade datans histogram samt Q-Q-plottar för de fördelningar som, utifrån histogrammen, ansågs lämpliga. Därefter testades datan för de olika fördelningarna med Anderson-Darling-test.

4.2 Resultat

I detta avsnitt presenteras resultatet för analysen av programmets väntetider.

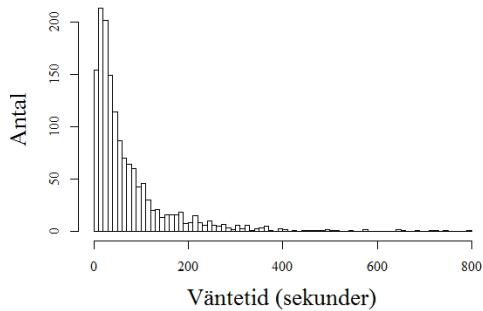
4.2.1 Hela datamängdens väntetider

Nedan presenteras grafiska undersökningar och resultatet för Anderson-Darling-testen för hela datamängden. Histogrammet för alla väntetider hittas i Figur 2. Utifrån denna bestämdes att datan skulle testas för exponentialfördelning, gammafördelning samt lognormalfördelning. Figur 3 visar kvantilerna för den logaritmerade datan mot normalfördelningens kvantiler. Övriga figurer som ligger till grund för den grafiska analysen hittas i Appendix G.1. Det är tydligt från Figur 3 att datan är relativt konsistent med linjen. Detta gäller emellertid inte i Q-Q-plottarna för exponential- och gammafördelningen, där datan för större kvantiler avviker från de teoretiska kvantilerna. Alltså antyder de grafiska undersökningarna att den logaritmerade datan följer en normalfördelning. Detta skulle innebära att datan är lognormalfördelad. Emellertid måste dessa indikationer testas med hypotestest.

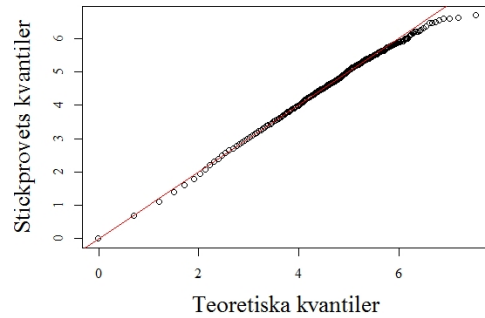
Hypotesprövningen bygger på att datan, med Anderson-Darling-test, testas för exponential- och gammafördelning. Dessutom testas den logaritmerade datan för en normalfördelning, även detta med Anderson-Darling-testet. Resultatet samt kritiska punkter presenteras i Tabell 1 respektive 2. Som nämnts i avsnitt 2.6 är nollhypotesen för ett Anderson-Darling-test att ett stickprov x_1, \dots, x_n kommer från fördelningen $F(x; \theta)$. Nollhypotesen förkastas vid signifikansnivå α då $A^2 > R_\alpha$, där A^2 är teststatistikan och R_α är det kritiska värdet. Genom att jämföra värdena för A^2 och $R_{0,05}$ för motsvarande fördelning i Tabell 1 blir det tydligt att den enda nollhypotesen som inte kan förkastas vid signifikansnivån 5% är att den logaritmerade datan följer en normalfördelning. Detta, precis som Figur 3, antyder att datan är lognormalfördelad.

4.2.2 Första och andra halvans väntetider

För att testa huruvida väntetiderna för första och andra halvan kommer från samma fördelning används ett k -sample Anderson-Darling-test med $k = 2$. Nollhypotesen är, som tidigare nämnts, att de båda stickproven har samma underliggande fördelning. Testvariablernas värden samt testets



Figur 2: Histogram över alla programs väntetider.



Figur 3: QQ-plot för normalfördelningen för logaritmen av alla väntetider.

Tabell 1: Värderna på teststatistikan då hela datasetet testas med ett Anderson-Darling-test för exponential-, gamma- och normalfördelning. Testet för normalfördelning görs på den logaritmerade datan. Gammafördelningens formparameter har värdet 1,0001.

	Exponential	Gamma	Normal
A^2	15,5853	15,5825	0,5874

Tabell 2: Kritiska värden för Anderson-Darling-testet vid signifikansnivå 5%. Dessa är hämtade från tabellerna 52, 53 samt 54 i Appendix E.

	Exponential	Gamma	Normal
$R_{0,05}$	1,321	0,768	0,752

p -värde presenteras i Tabell 3. Eftersom p -värdet är 0,0004 kan nollhypotesen att första och andra halvans väntetider kommer från samma fördelning förkastas på signifikansnivån 5%. Detta betyder att beteendet hos tiden som spenderas i ett informationssteg innan det sker en växling skiljer sig åt i programmets första och andra halva.

Tabell 3: Asymptotiskt p -värde samt värde på testvariabeln och den standardiserade testvariabeln för ett 2-sample Anderson-Darling-test av väntetiderna för programmets första och andra halva.

A^2	Standardiserad A^2	Asympt. p -värde
6,87	7,722	0.0004

Ovan konstateras att nollhypotesen, att första och andra halvans väntetider kommer från samma underliggande fördelning, kan förkastas. Nedan undersöks vilka dessa två fördelningar kan vara. I Appendix G.2.1 presenteras grafiska undersökningar för första halvans väntetider. Histogram hittas i Figur 12 och QQ-plottar i Figur 13, båda i Appendix G.2.1. Utifrån histogrammet i delfigur 12a anses exponential-, gamma- och lognormalfördelning vara tänkbara fördelningar för första halvans väntetider.

Datan testas mot ovan nämnda fördelningar genom ett Anderson-Darling-test. Resultatet hittas i Tabell 4. Notera att lognormalfördelningen återigen undersöks genom att den logaritmerade datan testas för en normalfördelning. I Tabell 2 hittas testets kritiska punkter vid signifikansnivån 5%. I Tabell 4 ses tydligt att den enda nollhypotes som inte kan förkastas är att logaritmen av datan kommer från en normalfördelning. Detta implicerar att första halvans väntetider skulle komma från en lognormalfördelning.

Tabell 4: Värdet på teststatistikan för Anderson-Darling-testen för exponential-, gamma- och normalfördelning för första halvans väntetider. Normalfördelningstestet görs på den logaritmerade datan. Gammafördelningens formparameter har värdet 1,0305.

	Exponential	Gamma	Normal
A^2	7,6408	8,0007	0,5649

I Tabell 5 presenteras resultatet av ett Anderson-Darling-test för andra halvans väntetider. De grafiska undersökningarna hittas i Appendix G.2.2. Därifrån, i histogram 14a, fås att lämpliga fördelningar att testa även här är exponential-, gamma- och lognormalfördelning. Precis som ovan är det den logaritmerade datan som testas för en normalfördelning. Även i det här fallet går det inte, vid signifikansnivå 5%, att förkasta nollhypotesen att logaritmen av datan kommer från en normalfördelning då testvariablerna jämförs med de kritiska punkterna i Tabell 2. Därför kan väntetiderna i programmets andra halva antas komma från en lognormalfördelning.

Tabell 5: Värdet på testvariabeln för Anderson-Darling-testen för exponential-, gamma- och normalfördelning för andra halvans väntetider. Normalfördelningstestet görs på den logaritmerade datan. Gammafördelningens formparameter har värdet 0,9909.

	Exponential	Gamma	Normal
A^2	7,3206	7,1571	0,3471

4.2.3 Väntetider för program med höga respektive låga tittarsiffror

I testen nedan har datan, utifrån tittarsiffror, delats in i två grupper. Detta görs först för TV och sedan för online. Målet är att se om det finns någon signifikant skillnad mellan de underliggande fördelningarna för program med höga och program med låga tittarsiffror. Notera att gränsen för höga tittarsiffror för TV går vid 800 000 och för online vid 80 000. Återigen används ett 2-sample Anderson-Darling-test. Tabell 6 visar teststatistikan och det resulterande p -värdet för testet för både online och broadcast. Nollhypotesen, att det inte finns en skillnad i väntetidernas fördelning för program med höga respektive låga TV-siffror, kan förkastas vid signifikansnivån 5%. Detta gäller emellertid inte då datan delas in efter tittarsiffror för online. Då p -värdet är 0,0587 kan nollhypotesen inte förkastas på en femprocentig signifikansnivå.

Tabell 6: Resulterande teststatistikor samt asymptotiska p -värden för 2-sample-Anderson-Darling-test för program med höga tittarsiffror mot program med låga tittarsiffror. Tabellen visar resultatet då uppdelningen sker både efter TV-siffror och efter onlinesiffror.

Tittarsiffror från/indelningar som testas	A^2	Standardiserad A^2	Asympt. p -value
TV/höga mot låga	8,270	9,560	> 0,0001
Online/höga mot låga	2,360	1,790	0,0587

För att undersöka fördelningen för populära och impopulära programs väntetider används, som tidigare, Anderson-Darling-test. Grafiska undersökningar för höga respektive låga TV-siffror hittas i Appendix G.3 samt G.4. Motsvarande för höga och låga onlinesiffror presenteras i Appendix G.5 respektive G.6. De grafiska undersökningarna indikerar att det återigen är lämpligt att undersöka exponential-, gamma- och lognormalfördelning. Detta gäller för både program med höga och program med låga tittarsiffror och när datan delas in efter såväl TV-siffror som onlinesiffror. Testvariablernas värde efter Anderson-Darling-testen hittas i Tabell 7. Tillhörande kritiska värden hittas i Tabell 2. Tabellerna gör det tydligt att den enda nollhypotes som inte kan förkastas är att den logaritmerade datan är normalfördelad. Detta gäller för alla fyra dataset, vilket skulle betyda att de är lognormalfördelade.

Tabell 7: Värden på teststatistikan då datan, indelad efter antalet tittarsiffror, testas med Anderson-Darling-test för exponential-, gamma- och normalfördelning. Normalfördelningstestet görs på den logaritmerade datan. Gammafördelningens formparameter har värdena 1,1615 och 0,9438 för höga respektive låga TV-siffror, samt 1,2145 och 0,9594 för höga respektive låga online-siffror.

Tittarsiffror från	Exponential	Gamma	Normal
TV hög	4,4970	4,0207	0,6026
TV låg	11,3395	9,5543	0,4914
Online hög	3,1906	2,4535	0,3565
Online låg	15,0000	13,4691	0,4992

Det finns, enligt resultaten i avsnitt 4.2.2 ovan, en skillnad i väntetidernas fördelning för första och andra halvan. Slutligen undersöks om detta gäller både för populära och impopulära program. Det görs för både TV och online. Alltså delas programmen in efter tittarsiffror och sedan efter första och andra halvan. Målet är att testa om det finns en skillnad mellan fördelningen i första halvan för populära program och första halvan för impopulära program samt motsvarande för andra halvan. Detta testas med ett 2-sample Anderson-Darling-test. Teststatistikor och p -värden för testen hittas i Tabell 8. Den enda nollhypotes som inte kan förkastas är att väntetiderna i *första halvan* i program med *höga* och *låga* onlinesiffror kommer från homogena fördelningar. De andra dataseten som testas kan alltså antas komma från olika fördelningar. Notera att detta innebär att de antingen kommer från olika fördelningsfamiljer eller att de kommer från samma fördelningsfamilj men att parametrarna skiljer sig åt.

Tabell 8: Resultat för 2-sample-Anderson-Darling-test då väntetiderna i *första* halvan i program med *höga* tittarsiffror jämförs med *första* halvan i program med *låga* tittarsiffror. Motsvarande testas även för programmens *andra* halva. Tittarsiffrorna delas upp efter både online och broadcast.

Tittarsiffror från/del av program	A^2	Standardiserad A^2	Asympt. p -värde
TV/första halvan	3,520	3,314	0,0150
TV/andra halvan	5,230	5,568	0,0022
Online/första halvan	0,6700	-0,4347	0,5845
Online/andra halvan	4,310	4,355	0,0062

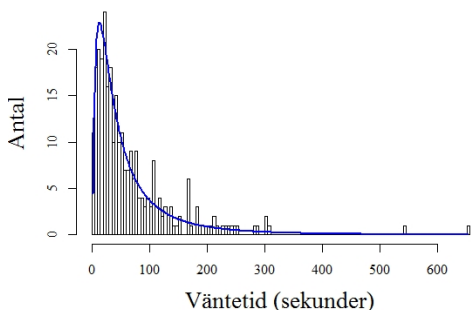
Slutligen testas, med Anderson-Darling-test, fördelningen för väntetiderna i *andra* halvan för *populära* program samt *andra* halvan för *impopulära* program. Detta görs för både TV- och onlinesiffror. Testen för den första halvan hittas i Appendix C.2. Resultatet presenteras i Tabell 9 nedan. De grafiska undersökningarna hittas i Appendix G.8 och Appendix G.12 för datan indelad efter TV-siffror respektive onlinesiffror.

Tabell 9: Värden på teststatistikan efter Anderson-Darling-test för väntetidernas fördelning i *andra* halvan för *populära* program samt *andra* halvan för *impopulära* program. Indelningen görs efter både TV- och onlinesiffror. Det är den logaritmerade datan som testas för normalfördelning. Gammafördelningens formparameter har värdena 1,1297 och 0,9518 för höga respektive låga TV-siffror, samt 1,3089 och 0,9407 för höga respektive låga onlinesiffror.

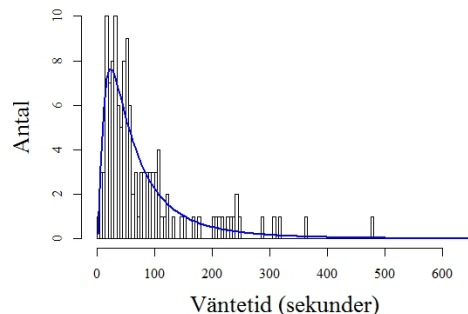
	Exponential	Gamma	Normal
TV/ hög/andra halvan	1,8528	1,8495	0,3348
TV/låg/andra halvan	5,2492	4,5128	0,3917
Online/hög/andra halvan	2,8418	2,3946	0,4100
Online/låg/andra halvan	7,1293	5,8905	0,5306

Då värdena i Tabell 9 jämförs med de kritiska värdena presenterade i Tabell 2 är den enda nollhypotesen som inte kan förkastas att den logaritmerade datan är normalfördelad. Detta gäller för alla fyra test, alltså andra halvan för både populära och impopulära program oavsett om indelningen

sker efter TV- eller onlinesiffror. Detta skulle medföra att respektive dataset är lognormalfördelat. Figuren 4 och 5 visar histogrammen med anpassad lognormalfördelning för de två dataseten.



Figur 4: Histogram över väntetiderna i andra halvan för program med höga TV-siffror med en anpassad lognormalfördelning.



Figur 5: Histogram över väntetiderna i andra halvan för program med höga onlinesiffror med en anpassad lognormalfördelning.

4.3 Diskussion

Resultaten i analysen av väntetidens populationsfördelning överensstämmer i vis grad med SVT:s hypotes att det finns en skillnad i väntetiderna mellan populära och impopulära program. Det gick, som sagt, att hitta en skillnad mellan dessa då programmen delades in efter höga och låga TV-siffror. Detta gäller emellertid inte för onlinesiffrorna vid en signifikansnivå av 5%, se Tabell 6. En möjlig anledning är att det fortfarande endast är en bråkdel av allt tittande som sker via streamingtjänsten. Det är alltså färre personer som tittat totalt vilket gör att *bra* och *dåligt* baseras på färre personers åsikter. Intressant är också att då de populära och impopulära programmen dessutom delades in efter första och andra halvan observeras en signifikant skillnad i den underliggande fördelningen för programmets andra halva, för både TV och online tittarsiffror. Vidare ses denna skillnad även för TV-siffrornas första halva, men inte för online. Detta skulle också kunna kopplas till att det är färre personer som tittar via streamingtjänsten.

I resultaten ovan fås att det inte går att förkasta att den underliggande fördelningen är en lognormalfördelning, vilket gäller för alla indelningar av datan. I Tabell 51 i Appendix D hittas parametrarna för den underliggande normalfördelningen för alla olika indelningar. Det förväntade värdet för väntetiderna i *andra halvan* i program med *höga* TV-siffror är 71 sekunder. Motsvarande för *andra halvan* i program med *låga* TV-siffror är 103 sekunder. Det gäller alltså att växlingarna, i snitt, sker snabbare i den andra halvan av populära program än vad de gör i samma del av de impopulära programmen. Samma sak gäller för den andra halvan då programmen delas in efter onlinesiffror. Då fås att väntevärdet för populära program är 85 sekunder medan det är 90 sekunder för de impopulära programmen. Precis samma gäller i första halvan, både för online och broadcast, alltså att program med höga tittarsiffror har en lägre förväntad väntetid än program med låga tittarsiffror. Dessa resultat stämmer helt överens med Purens hypotes, vilken var att bra program kommer växla snabbare mellan de olika informationsnivåerna än dåliga program. Slutsatsen är alltså att detta gäller både i episodernas första och andra halva.

Något som skulle kunna öka resultatens pålitlighet är om fler än en typ av test använts, både för fördelningstesten och homogenitetstesten. Om alla dessa hade visat på samma sak skulle det varit en tydligare indikation på att testresultaten är tillförlitliga. Nollhypotesen i Anderson-Darling-testen är att datan följer en viss fördelning. Att nollhypotesen inte kan förkastas betyder nödvändigtvis inte att datan följer fördelningen som testas, utan snarare att det inte går att utsluta att den gör det. Att flera olika test skulle visa att en fördelning inte kan förkastas är alltså en tydligare indikation på att datan följer den fördelningen.

Ett annat problem är datan som används i analysen. Väntetiderna har endast kunnat samlas in med en sekunds noggrannhet, även om de i själva verket är kontinuerliga. Konsekvensen är att data som egentligen är kontinuerlig blir diskret och alla väntetider kommer alltså att vara avrundade.

En följd av denna diskretisering är att det uppstår så kallade *ties* i datan; alltså att flera väntetider har samma värde. Ties är mycket ovanligt i det kontinuerliga fallet och är därför något som måste tas hänsyn till vid hypotesprövningen eftersom den görs för kontinuerliga fördelningar. Det k -sample Anderson-Darling-test som används är speciellt framtaget för fall då problem med ties kan uppstå [10]. Det finns emellertid ingen liknande modifiering för det vanliga Anderson-Darling-testet. D’Agostino och Stephens skriver dock följande i *Goodness-of-Fit Techniques* (s. 171): “Of course, in practice all continuous data are subject to the limits of the accurate measurement, but the inherent grouping may be so fine as to have negligible effect.” Därför, om datan är såpass finfördelad så att den inte lägger för mycket vikt vid enskilda observationer bör testet ge pålitliga resultat. Således, eftersom datan har en sekunds precision och den totala längden av ett program är 40-60 minuter ansågs att intervallen var tillräckligt små för att antagandet skulle vara uppfyllt. Vidare är det känt att normalfördelningstest baserade på den empiriska fördelingsfunktionen blir mer konservativa då ties finns i datan [8]. Med andra ord är den faktiska signifikansnivån mindre än den nominella. Således bör resultaten vara någorlunda pålitliga, men om en ny studie genomförs bör en bättre precision i mättekniken användas.

Orsaken till att Anderson-Darling-testet valdes är i huvudsak att det finns välutvecklad asymptotisk teori för testvariabeln och dessa har verifierats genom uttömmande Monte-Carlo undersökningar [12]. Därtill, anledningen till att k -sample Anderson-Darling-testet valdes är att den har visat sig vara bättre för att hitta diskrepanser i datan samtidigt som det har mer styrka än exempelvis Kolmogorov-Smirnov-test, vid jämförelse av två stickprov [13].

Ett steg i att förbättra analysen är att göra hypotestesten på andra dataset än de som parametrarna skattas från; alltså att dela in datan i två delar där den ena används för att skatta parametrarna och den andra till att göra hypotestesten. På grund av detta arbete byggde på data från endast 37 program ansågs det viktigt att, i analysen, använda all information som fanns tillgänglig. Framförallt skulle mycket information förloras om olika data användes för skattning och hypotestest då observationerna delas in efter första/andra halvan och populära/impopulära program. Detta skulle ge mycket få observationer i varje del.

5 Andel tid spenderad i de tre informationstillstånden

Denna del ämnar att undersöka huruvida totala tiden ett program spenderar i respektive informationstillstånd, MSD, associerar till programmets popularitet.

5.1 Dataanalys

Den totala tiden som programmet spenderar i var och ett tillstånd summeras ihop och andelen av den totala programtiden noteras. Andelen ettor, treor samt andelen ettor adderat med andelen treor, från det första avsnittet, plottas mot tittarsiffror för det aktuella avsnittet samt det efterkommande. Genom att titta på tittarsiffror för det andra avsnittet blir det tydligt hur många som, baserat på första avsnittet, väljer att fortsätta följa serien. Därefter tas en regressionslinje fram med hjälp av ett tvåsidigt t -test. Detta i syfte att upptäcka om hypotesen nedan går att förkasta eller inte. Under insamling av data upptäcktes att sju av de trettiosju TV-programmen som undersöktes hade en särskiljande struktur i tidsfördelningen av de tre tillstånden. De kan betraktas som outliers, eller utstickare, då majoriteten av programtiden *inte* spenderades i tillståndet *suspence* till skillnad från resterande TV-program. Det utfördes därför också en undersökning av regressionslinjen då dessa sju program utestlöts.

5.2 Hypotes

De hypoteser som ska testas lyder

H_0 : Lutningen på regressionslinjen är lika med noll,

H_1 : Lutningen på regressionslinjen är skild från noll,

där H_0 är nollhypotesen och H_1 är den alternativa hypotesen. Regressionslinjen är den som beskrivs i avsnitt 5.1, det vill säga den alternativa hypotesen säger att andelen tid spenderad i något

informations har ett linjärt samband med tittarsiffror. Nollhypotesen säger att det inte finns något linjärt samband, det vill säga att det är slumpmässigt.

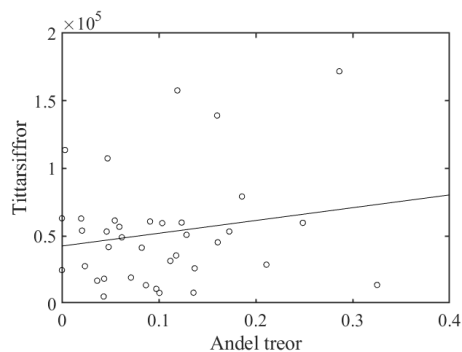
5.3 Resultat

Den linjära regressionen som utfördes på *all* insamlad data visade inte på positiva resultat. Respektive linjes lutning, $\hat{\beta}_1$, enligt

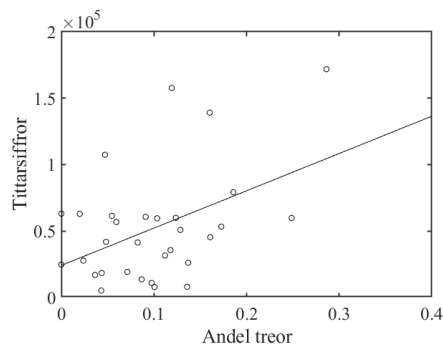
$$y = \hat{\beta}_0 + \hat{\beta}_1 x,$$

presenteras i Tabell 10. Tillhörande p -värde och värde på R-squared presenteras i Tabell 11 respektive Tabell 12. De presenterade p -värdena ger att β_1 inte är statistiskt signifikant då det lägsta p -värdet ligger på 28,2% vilket inte går under någon rimlig signifikansnivå. Se Figur 6 för visualisering av den mätning som gav upphov till ovan nämnda p -värde. Figuren representerar andelen treor i varje *första avsnitt* plottat mot online-tittarsiffror motsvarande *andra avsnittet*.

I nedre halvan av tabellerna 10, 11 och 12, som markerats med asterisk, presenteras resultaten där sju av programmen uteslutits från beräkningarna. Vad man kan se är att det då framträder p -värden under en signifikansnivå på 5%. Detta för online-tittarsiffror tillhörande avsnitt ett respektive avsnitt två. Se Figur 7 för visualisering av den mätning som motsvarar andel treor i avsnitt ett plottat mot onlinetittarsiffror för avsnitt två, för fallet då sju program uteslutits. Dessa har även ett högre värde på R-squared som presenteras i nedre halvan av Tabell 12. I Tabell 10 ser vi en signifikant skillnad på riktningskoefficienten $\hat{\beta}_1$ för datan markerad med asterisk respektive utan. För resterande plottar som värdena i tabellerna 11, 10 och 12 är baserade på se Appendix G.15.



Figur 6: Andel treor i avsnitt ett plottat mot tittarsiffror online för avsnitt två. Detta för *alla* trettiosju avsnitt. Linjen är den tillhörande regressionslinjen.



Figur 7: Andel treor i avsnitt ett plottat mot tittarsiffror online för avsnitt två. Detta då sju av TV-programmen uteslutits. Linjen är den tillhörande regressionslinjen.

Tabell 10: Tabell över riktningskoefficienten för den anpassade linjen för andel ettor, andel treor respektive andel ettor adderat med andel treor. Nedre delen av tabellen visar resultaten då sju av TV-programmen uteslutits från beräkningarna på grund av deras utstickande karaktär.

Tittarsiffror från:	$\hat{\beta}_1$ för andel ettor	$\hat{\beta}_1$ för andel treor	$\hat{\beta}_1$ för andel ettor och treor
Avsnitt 1 online	5430	91900	38100
Avsnitt 1 broadcast	-68000	$-2,73 \cdot 10^5$	$-1,70 \cdot 10^5$
Avsnitt 2 online	8180	94500	42100
Avsnitt 2 broadcast	-67600	-80400	$-1,02 \cdot 10^5$
Avsnitt 1 online*	-49100	$2,62 \cdot 10^5$	86200
Avsnitt 1 broadcast*	$-6,33 \cdot 10^5$	$2,01 \cdot 10^6$	$3,02 \cdot 10^5$
Avsnitt 2 online*	-33300	$2,79 \cdot 10^5$	$1,18 \cdot 10^5$
Avsnitt 2 broadcast*	$-4,74 \cdot 10^5$	$1,88 \cdot 10^6$	$4,52 \cdot 10^5$

Tabell 11: Tabell över p -värdet för den anpassade linjen för andel ettor, andel treor respektive andel ettor adderat med andel treor. Nedre delen av tabellen visar resultaten då sju av TV-programmen uteslutits från beräkningarna på grund av deras utstickande karaktär. De fetmarkerade p -värdena är de som accepteras av en femprocentig signifikansnivå.

Tittarsiffror från:	p -värde för andel ettor	p -värde för andel treor	p -värde för andel ettor och treor
Avsnitt 1 online	0,919	0,323	0,490
Avsnitt 1 broadcast	0,908	0,792	0,782
Avsnitt 2 online	0,871	0,282	0,420
Avsnitt 2 broadcast	0,904	0,935	0,862
Avsnitt 1 online*	0,554	0,0323	0,375
Avsnitt 1 broadcast*	0,470	0,129	0,770
Avsnitt 2 online*	0,675	0,0165	0,203
Avsnitt 2 broadcast*	0,579	0,155	0,654

Tabell 12: Tabell över R-squared för den anpassade linjen för andel ettor, andel treor respektive andel ettor adderat med andel treor. Nedre delen av tabellen visar resultaten då sju av TV-programmen uteslutits från beräkningarna på grund av deras utstickande karaktär.

Tittarsiffror från:	R-squared för andel ettor	R-squared för andel treor	R-squared för andel ettor och treor
Avsnitt 1 online	0,000303	0,0280	0,0137
Avsnitt 1 broadcast	0,000388	0,00201	0,00222
Avsnitt 2 online	0,000767	0,0330	0,0186
Avsnitt 2 broadcast	0,00042	0,000191	0,000876
Avsnitt 1 online*	0,0127	0,153	0,0282
Avsnitt 1 broadcast*	0,0189	0,0805	0,0031
Avsnitt 2 online*	0,00636	0,189	0,0571
Avsnitt 2 broadcast*	0,0111	0,0747	0,00728

5.4 Diskussion

Övre halvan av tabellerna 10, 11 och 12 bidrar inte med något statistiskt signifikant resultat men när de sju utstickarna plockats bort ur beräkningen börjar vi kunna tyda ett mönster. Det går fortfarande inte att säga något om andelen ettor och andelen ettor adderat med andelen treor, men för andelen treor sker något intressant. Vi ser ett lägre p -värde och ett högre värde på R-squared. Vad som också är intressant är att andelen treor associerar bättre till online-tittarsiffror än broadcast-tittarsiffror. Varför det är så kan diskuteras, det kan ha att göra med det aktiva valet att gå in på SVT-play och välja det en ska se till skillnad från att sitta och knappa på TV-kontrollen. En annan anledning skulle kunna vara olikheterna i målgrupperna som vänder sig till online mot de som tittar på broadcast, till exempel ålder.

Sedan kan det även diskuteras hurvida det går att säga något alls om resultaten ovan, dels på grund av de varierande resultaten men också på grund av felkällor i datainsamling som nämns närmare i en samlad diskussion i slutet av rapporten. Något annat som är problematiskt är att det finns så pass lite insamlad data, genom att bara plocka bort en mätpunkt kan resultaten förändras drastiskt. Det skulle vara av intresse att utföra samma mätningar på ett större omfång mätpunkter, det vill säga fler analyserade TV-program. Då skulle det kunna vara lättare dra slutsatser om det till exempel finns utstickare eller om det rör sig om en annorlunda programstruktur. I det senare fallet skulle det kunna göras liknande undersökningar som ovan fast på varje enskild programstruktur. Då det nu varken går att förkasta eller bekräfta hypotesen för resultaten ovan kan denna undersökning snarare användas som en utgångspunkt för framtida forskning där dessa resultat tas i beaktning.

6 Påverkan av skiftningar från ett stadie till ett annat

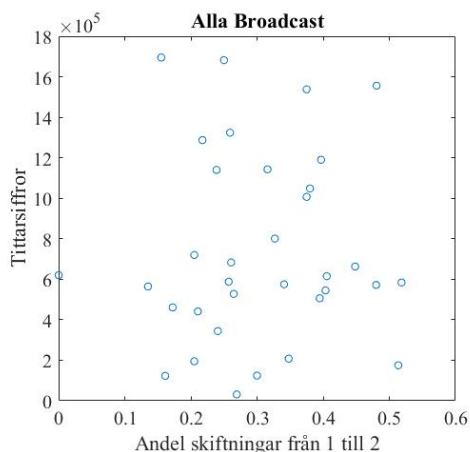
Den tredje och sista frågeställningen berör huruvida antalet hopp mellan olika specifika stadier påverkar programmets popularitet. De hopp som menas är $1 \rightarrow 2$, $1 \rightarrow 3$, $2 \rightarrow 1$, $2 \rightarrow 3$, $3 \rightarrow 1$, och $3 \rightarrow 2$.

6.1 Dataanalys

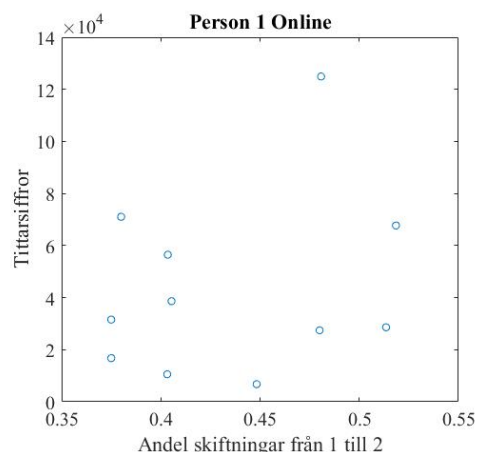
Datan analyserades i MATLAB genom att räkna antalet hopp mellan två specifika stadier och dividera med det totala antalet hopp för att få andelen hopp som sker mellan just dessa stadier. I kategorin *Alla* har alla personers data analyserats tillsammans medan i kategorierna *Person 1* och *2* har datan analyserat separerat enligt vem som samlat in den. I kategorin *Broadcast* undersöks tittarsiffrorna från TV-sändningarna, och i kategorin *Online* undersöks tittarsiffrorna från SVT-play. Datan plottades och en linje anpassades med hjälp av den linjära regressionsmetoden beskriven i avsnitt 2.2. Hur välanpassad linjen var till datan bedömdes sedan med ett tvåsidigt t -test som gav ett p -värde som mått på signifikansnivån, och R-squared som mått på hur välanpassad linjen var. Residualerna summerades och undersöktes för att se om datan passade linjen.

6.2 Resultat

I figurerna 48 och 75 återfinns datan för skiftningarna plottade mot tittarsiffror. I Figur 48 plottas andel skiftningar från stadie 1 till 2 mot broadcast-tittarsiffror, och i Figur 75 plottas data från *Person 1* för skiftningar från stadie 1 till 2 mot online-tittarsiffror. I Appendix G.17 går det att se resterande plottar.



Figur 8: Plot av skiftningarna från stadie 1 till 2 mot broadcast-tittarsiffror. Notera att formen inte indikerar annat samband än linjärt.



Figur 9: Plot av skiftningarna från stadie 1 till 2 mot online-tittarsiffror. Notera att formen indikerar annat än linjärt samband.

För alla övergångar mellan olika stadier följer nedan i Tabell 13, 14 och 15 värden för de anpassade linjernas lutning, p -värden respektive R-squared. I kategorierna *Person 1 Broadcast* $1 \rightarrow 2$, $2 \rightarrow 1$, $2 \rightarrow 3$ och $3 \rightarrow 2$ har programmet *Jordskott* tagits bort ur analysen då den ansågs vara en utstickare, detta då både p -värdet och R-squared förbättrades betydligt när den togs bort. Dessa är markerade med asterisk i tabellerna.

Tabell 13: Tabell över lutningen för de anpassade linjerna för de olika skiftningarna.

Typ av data	1 → 2	1 → 3	2 → 1	2 → 3	3 → 1	3 → 2
Alla Broadcast	146530	-1479200	-24700	633560	850940	-124030
Alla Online	-79776	-55583	-84780	131860	248670	89312
Person 1 Broadcast	-5270400*	5919800	-6264400*	9184400*	6375800	8428900*
Person 1 Online	162360	288720	206810	-336860	262080	-422200
Person 2 Broadcast	-757710	307470	-900150	1536700	1754100	1064500
Person 2 Online	-29672	-175230	-66302	295130	-11661	89312

Tabell 14: Tabell över p -värden för de olika skiftningarna.

Typ av data	1 → 2	1 → 3	2 → 1	2 → 3	3 → 1	3 → 2
Alla Broadcast	0,835	0,325	0,972	0,465	0,655	0,892
Alla Online	0,216	0,693	0,19	0,0963	0,154	0,29
Person 1 Broadcast	0,0132*	0,386	0,00337*	0,000318*	0,212	0,0121*
Person 1 Online	0,441	0,596	0,332	0,231	0,527	0,166
Person 2 Broadcast	0,481	0,83	0,385	0,228	0,425	0,534
Person 2 Online	0,865	0,44	0,696	0,143	0,974	0,54

Tabell 15: Tabell över R-squared för de olika skiftningarna.

Typ av data	1 → 2	1 → 3	2 → 1	2 → 3	3 → 1	3 → 2
Alla Broadcast	0,00138	0,0302	$3,89 \cdot 10^{-5}$	0,0168	0,00632	0,0006
Alla Online	0,0474	0,00494	0,053	0,0841	0,0624	0,0349
Person 1 Broadcast	0,557*	0,0843	0,679*	0,819*	0,167	0,566*
Person 1 Online	0,0674	0,0325	0,104	0,155	0,0458	0,202
Person 2 Broadcast	0,0461	0,00438	0,0693	0,129	0,087	0,0362
Person 2 Online	0,00274	0,0552	0,0144	0,185	0,0001	0,0351

6.3 Diskussion

Som framgår i Tabell 14 har de allra flesta anpassade linjer ett väldigt högt p -värde. För en signifikansnivå på 5% ska p -värdet vara $< 0,05$, och det syns i Tabell 14 att väldigt få linjer kommer under den gränsen. Därmed kan nollhypotesen att lutningen $\beta_1 = 0$ inte förkastas. För kategorin *Alla* är detta inte förvånande eftersom ett av villkoren för linjär regression är att datan inte får vara grupperad, och en kan argumentera för att datan är grupperad eftersom det är fyra olika personer som samlat in den. Dessutom används data från olika genrer, vilket kan ge upphov till ännu en gruppering i datan.

Betrakta istället de fall där nollhypotesen kan förkastas med en signifikansnivå på 5%: *Person 1 Broadcast* 1 → 2, 2 → 1, 2 → 3 och 3 → 2. Alla dessa kommer från data från samma person och har därmed inte problemet med korrelation som tidigare diskuterats. Det framgår dessutom i Tabell 15 att dessa linjer har relativt höga R-squared, vilket indikerar att de är relativt välanpassade. Vid undersökning av residualplottarna Figur 86, 98, 104 och 116 syns det dock att residualerna inte är jämnt utplacerade kring nollan, och det går att se mönster i samtliga plottar, vilket inte är önskvärt. Detta pekar på att minsta kvadratmetoden inte är optimal för att modellera denna data, och i framtiden rekommenderas istället att använda en annan metod. Det är också värt att nämna att datan inte är optimalt och det finns dessutom inte så mycket av den, vilket diskuteras vidare i avsnitt 7.1. Detta betyder att modellen inte beskriver verkligheten tillräckligt väl, utan det finns fler variabler att ta hänsyn till, men det går fortfarande att diskutera vad resultaten pekar mot.

Det går att konstatera att de linjer som beskriver skiftningar mellan stadie 1 och 2 (både 1 → 2 och 2 → 1) ger en negativ lutning, medan de som beskriver skiftningar mellan stadie 3 och 2 (både 3 → 2 och 2 → 3) har en positiv lutning. Detta går i linje med vad Purens hypotiserade: att folk gillar att veta mer än karaktärerna, då detta är ett stadie en befinner sig i otroligt sällan i verkliga

livet. Å andra sidan går en otroligt ofta runt och vet väldigt lite om vad som händer, och därmed är det rimligtvis inte lika uppskattat med en serie som skiftar mycket mellan stadie 1 och 2.

7 Diskussion

Nedan följer den gemensamma diskussionen för alla tre frågeställningar.

7.1 Felkällor

Den absolut största felkällan i detta arbete är insamlingen av data. Eftersom media är en komplex konstform som förlitar sig på subtila ledtrådar för att guida tittaren till slutsatser blir bedömningen av vilket stadie en befinner sig i subjektiv, och de individer som i detta fallet samlat in datan kommer oifrånkligt tolka undertexten olika. Dessutom kan en byta stadie på grund av endast en mening, och olika personer tolkar längden av det stadiet olika. Det är alltid svårt att avgöra när och hur länge information är relevant och programmet kan ge missledande ledtrådar som vidare komplicerar vilket stadie en befinner sig i. Vidare är det ofta svårt att avgöra vilken karaktär tittaren jämförs med. Ett informationsövertag implicerar att tittaren har *mer* information än någon och vem detta är tolkas lätt olika av olika personer. Alla dessa faktorer försvåras dessutom av språkbarriären som existerar i vissa av programmen, då de i vissa program talade engelska, norska eller danska.

Angående urvalet av program valdes både egenproducerade SVT-serier och inköpta serier som visats genom andra kanaler. Resultatet har blivit att serier som t.ex. *Game of Thrones* har haft förhållandevis dåliga tittarsiffror, trots att den är allmänt ansedd att vara en av världens bästa serier. Detta kommer sig troligtvis av att den finns tillgänglig dygnet runt på streamingsidor som HBO, samt att den visas på SVT långt efter den först kom ut, så att den stora massan som velat se *Game of Thrones* redan sett den på andra sidor. I och med att just denna serien är så välkänd var den lätt att utesluta från undersökningarna, men andra serier är inte lika uppenbara och kan lätt vinkla resultatet av dataanalysen.

I denna digitala värld är det också värt att nämna att online-tittandet har vuxit sig stort, och kommer troligtvis bara fortsätta växa. Detta gör att fördelningen av tittarsiffrorna förändras och kan i vissa fall vinkla resultaten av undersökningen. SKAM har t.ex. varit en enormt framgångsrik serie men broadcast-tittarsiffrorna gör den inte rättvisa. Däremot speglas det tydligt i online-tittarsiffror, vilket med största sannolikhet kommer av att serien riktar sig starkt mot den yngre generationen. Eftersom tittarsiffrorna från TV och online samlas in på så olika sätt är det också problematiskt att bara addera dem.

Dessvärre är urvalet av serier inte så stort, detta till stor del på grund av den begränsade tiden och resurser som fanns att tillgå detta projektet. Därför är det svårt att tillskriva stor statistisk signifikans till något resultat i rapporten.

7.2 Förslag för framtiden

I fortsatta undersökningar rekommenderas starkt ett större urval av datainsamlare och serier. Därtill är det viktigt att erkänna att olika genrer av film följer olika mönster och bör behandlas separat för större chans för resultat. En deckare kommer t.ex. tillbringa mycket tid i *mystery* eftersom hela premissen bygger på att man inte vet vad som hänt, och behandlas den tillsammans med andra genrer kan detta bilda grupperingar i datan och vinkla resultaten.

Det är också värt att påpeka att rena tittarsiffror inte alltid är en god indikator av popularitet. Eftersom olika program riktar sig till olika grupper bör detta tas hänsyn till i analysen. Ett enkelt exempel är att barnprogram ofta inte är populära i någon annan ålderskategori, och hade tittarsiffrorna undersökts där kan slutsatsen lätt dras att ett barnprogram är mycket impopulärt, när det i själva verket kan vara omtyckt av en stor del av åldersgruppen. Samma princip gäller för andra kategorier. En film som kretsar kring gravida kvinnor förväntar sig troligtvis att publiken består av gravida kvinnor och föräldrar. En film med stark politisk riktning har troligtvis inte förväntat sig vara omtyckt av hela befolkningen, men kan fortfarande vara omtyckt inom sin målgrupp.

Tittarsiffror kan självklart påverkas av andra faktorer också, så som sändningstid och vilka andra program som sändes samtidigt. Om en dramaserie sänds kl 14.00 en tisdag är det rimligt att anta

att den får färre tittare än en som sänds kl 21.00 en fredag, och spelar svenska herrlaget i fotboll VM-match samtidigt lär tittarsiffrorna lida ännu mer.

En intressant aspekt att undersöka är i vilket stadie filmskaparna tror tittaren befinner sig, och hur det relaterar till i vilket stadie tittaren tror sig befinna sig.

I avsnitt 4.3 diskuterades problem som uppstår till följd av diskretisering av datan. I framtida studier skulle risken för ties kunna minskas genom att tilldela varje mätvärde ett slumpmässigt brus. Detta kan motiveras med antagandet att datainsamlaren ungefär lika många gånger avrundar tiden uppåt som nedåt. Dock får inte programmets totala längd ändras.

7.3 Samhälleliga och etiska aspekter

Om just detta arbete kommer ha stor samhällelig och etisk påverkan kan ifrågasättas. Möjligen kan manusförfattare och filmskapare ta illa vid av detta försök att analysera och plocka isär deras konstform. Men om vi tar det ett steg längre, eller tio steg längre, så kan påverkan bli större. Låt säga att ett tydligt och detaljrikt mönster hittas och ett recept för den perfekta dramaturgin görs tillgänglig för allmänheten. Kommer yrken inom filmindustrin upphöra att existera då de inte längre behövs? Det kommer antagligen inte att gå så långt och modellen som tas fram kommer kanske snarare att användas som ett hjälpmedel som alltså inte kan stå på egna ben. Men i dagsläget svävar det redan ett orosmoln över den klassiska musikindustrin, där en, med flera, artificiella intelligenser tagits fram. En som står i blickfånget är AIVA som är uppbyggd med hjälp av Deep Learning [14]. AIVA har till och med skapat upphovsrättsskyddade verk och det har visat sig vara upprörande för vissa musiker.

En annan vinkel som kan analyseras är den som ifrågasätter hur en bestämmer om ett TV-program är “bra” eller inte, och vad det får för konsekvenser. I detta arbete baseras “kvalitén” av programmet på tittarsiffror som ska vara representativa för den svenska befolkningen. Det tittas inte på speciella målgrupper. Effekten av detta kan leda till att program riktade till t.ex. minoriteter i samhället får dåliga tittarsiffror i förhållande till hela den svenska befolkningen. För att vara övertydlig kan en dra det så långt som att program med dåliga tittarsiffror plockas bort från programutbudet på grund av just detta. Det skulle vara problematiskt på många nivåer och rent konkret skulle det gå emot Sveriges Televisions ledord, att de ska vara till för *alla* [15]. En framtida lösning på detta problem skulle kunna vara att faktiskt ta hänsyn till specifika målgrupper och minoriteter i bedömningen av ett programs “kvalité”.

Det är också högst troligt att även om den matematiska modell som tas fram här ger positiva resultat så kan det finnas andra modeller som även de associerar väl till antalet tittarsiffror. Att lita blint på att denna modell är den enda möjliga skulle alltså också kunna leda till att program som uppfattas bra, med andra mått mätta, sällas bort för att de inte passar denna mall.

Referenser

- [1] McKee R. *STORY: Substance, Structure, Style and the Principles of Screen Writing*. Methuen & Co Limited, York, 2014.
- [2] Så mäts TV-tittandet [Internet]. Stockholm: Mediamätning i Skandinavien AB, 2015 [citerad 2 April 2019]. Tillgänglig från: http://mms.se/?page_id=29.
- [3] Så mäts online-tittandet [Internet]. Stockholm: Mediamätning i Skandinavien AB, 2015 [citerad 2 April 2019]. Tillgänglig från: http://mms.se/?page_id=35.
- [4] Rice J. A. *Mathematical Statistics and Data Analysis*. Brooks/Cole, Belmont, third edition, 2007.
- [5] Jörnsten R. MSG500/MVE190 Linear Statistical Models - Lecture 1 [Föreläsningsanteckningar på Internet]. Göteborg: Chalmers, 2017 [citerad 15 Maj 2019]. Tillgänglig från: <http://www.math.chalmers.se/Stat/Grundutb/GU/MSG500/A18/Lecture1.pdf>.
- [6] Picchini U. MSG500/MVE190 Linear Statistical Models, 2018-19, Lecture 2 [Powerpointpresentation på Internet]. Göteborg: Chalmers, 2018 [citerad 15 Maj 2019]. Tillgänglig från: http://www.math.chalmers.se/Stat/Grundutb/GU/MSG500/A18/slides_2.pdf.
- [7] Picchini U. MSG500/MVE190 Linear Statistical Models, 2018-19, Lecture 4 [Powerpointpresentation på Internet]. Göteborg: Chalmers, 2018 [citerad 15 Maj 2019]. Tillgänglig från: http://www.math.chalmers.se/Stat/Grundutb/GU/MSG500/A18/slides_4.pdf.
- [8] D'Agostino R. B., Stephens M. A. *Goodness-of-Fit Techniques*. Marcel Dekker, New York, 1986.
- [9] van der Merwe, S. On Determining the Distribution of a Goodness-of-Fit Test Statistic. Master's thesis, University of the Free State, Box 339, Bloemfontein, 9300, South Africa, May 2014. Tillgänglig från: https://www.ufs.ac.za/docs/librariesprovider22/mathematical-statistics-and-actuarial-science-documents/technical-reports-documents/teg430-2583-eng.pdf?sfvrsn=743df921_0.
- [10] Scholz F. W., Stephens M. A. K-Sample Anderson-Darling Tests. *JASA*, Vol. 82(399):918–924, September 1987. Tillgänglig från: <https://www.jstor.org/stable/2288805>.
- [11] Log-normal distribution [Internet]. I Wikipedia, 2019 [uppdaterad 19 April 2019; citerad 5 Maj 2019]. Tillgänglig från: https://en.wikipedia.org/wiki/Log-normal_distribution.
- [12] Stephens M.A. Edf statistics for goodness of fit and some comparisons. *JASA*, 69(347):730–732, April 2012. Tillgänglig från: <https://doi.org/10.1080/01621459.1974.10480196>.
- [13] Engmann S., Cousineau D. Comparing distributions: the two-sample anderson-darling test as an alternative to the kolmogorov-smirnov test. *JAQM*, 6(3):1–17, September 2011. Tillgänglig från: https://www.researchgate.net/publication/276918573_Comparing_distributions_the_two-sample_Anderson-Darling_test_as_an_alternative_to_the_Kolmogorov-Smirnov_test.
- [14] About, AIVA - Artificial Intelligence Virtual Artist [Internet]. Luxembourg: AIVA, 2018 [citerad 17 April 2019]. Tillgänglig från: <https://www.aiva.ai/about>.
- [15] Vår roll - Vi är till för alla i Sverige [Internet]. Stockholm: Sveriges Television AB, 2019 [citerad 21 Februari 2019]. Tillgänglig från: <https://www.svt.se/omoss/var-roll/>.

A Programdata

Totalt samlades 37 avsnitt in vilka hade både TV och online tittarsiffror. I Tabell 16 presenteras alla avsnitts tittarsiffror.

Tabell 16: I denna tabell presenteras tittarsiffror för respektive analyserat avsnitt. Detta för broadcast och online för avsnitt ett respektive två.

Program	Broadcast avsnitt 1	Online avsnitt 2	Broadcast avsnitt 2	Online avsnitt 2
KRONJUVELERNA	1007000	16748	995000	16984
THE SPIRAL	663000	6745	577000	5289
30 GRADER I FEBER	1538000	31557	1441000	31664
HALVBRÖDERNA	528000	18465	528000	18465
FRÖKEN FRIMANS KRIG	1682000	55302	1718000	56847
ÄNGELBY	1048000	71045	998000	61504
BLÅ ÖGON	801000	85362	788000	59640
THE TEAM	550000	52605	525000	41414
DET BLODRÖDA FÄLTET	572000	27457	514000	27764
FRIKÄND	1058000	51909	1002000	53394
BONUSFAMILJEN	1324000	186422	1141000	157691
LIFE IN SQUARES	163000	27280	173000	19335
CUCUMBER	117000	7576	92000	7996
OLD SCHOOL	123000	24668	126000	8142
SPRINGFLODEN	31000	98768	45000	107487
DEUTSCHLAND 83	620000	55752	572000	50938
VÅR TID ÄR NU	1695000	154286	1660000	171841
LUTHER 1	441000	19585	299000	26178
SCOTT AND BAILEY	615000	38630	589000	41944
BILLIONS	344000	42127	221000	35689
MONSTER	506000	70885	216000	62975
BARRACUDA	208000	15734	199000	13729
HOTELL HALCYON	1287000	62912	1133000	60003
CALL THE MIDWIFE	1190000	10563	735000	10966
JAMESTOWN	564000	62493	492000	59822
TRUST ME	461000	48822	393000	45396
THE STATE	175000	28602	141000	24818
KIRIENKVINNA	584000	67706	529000	63147
THE SPLIT	507000	73835	427000	60788
EN ENGELSK SKANDAL	557000	153278	361000	139172
BERLIN-UNDER SAMMA HIMMEL	683000	54676	487000	53383
IN THE CLUB	262000	15781	300000	13862
TJOCKARE ÄN VATTEN	1143000	49496	1146000	54049
JORDSKOTT	1556000	124927	1279000	113667
THE HONORABLE WOMAN	181000	38137	139000	28865
PAULA	545000	56470	488000	49080
HEMMA IGEN	195000	100114	162000	79311

B Resultat för Anderson-Darling-test av populationsfördelning

De beräknade testvariablerna för alla dataset; dvs. värdena av de resulterade A^2 för exponential-, gamma- och lognormalfördelningen. Observera att testvariablerna för exponential- och lognormalfördelningen är modifierade för att vara konsekvent med teorin presenterad av D'Agostino och Stephens (1986)[8]. Därutöver notera att gammafördelningens test är approximativt på grund av

att distributionen innefattar en formparametern m .

B.1 Resultat för enkla uppdelningar

I den här sektionen presenteras alla test resultat för de tester som antingen använder all data eller delar upp den i två delar.

All data sammanslaget

Tabell 17: De resulterade A^2 testvariablerna för alla episoders data; dvs. för A^2 testvariablerna för exponential-, gamma- och lognormalfördelningen. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är lognormalfördelad inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,0001.

Exponential	Gamma	Normal
15,5853	15,5825	0,5874

Observationerna före episodens mittpunkt

Tabell 18: De resulterade testvariablerna för observationerna före episodens mittpunkt. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är lognormalfördelad inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,0305.

Exponential	Gamma	Normal
7,6408	8,0007	0,5649

Observationerna efter episodens mittpunkt

Tabell 19: De resulterade testvariablerna för observationerna efter episodens mittpunkt. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är lognormalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 0,9909.

Exponential	Gamma	Normal
7,3206	7,1571	0,3471

Episoderna med höga TV tittarsiffror

Tabell 20: De resulterade testvariablerna för episoderna med höga TV tittarsiffror. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är lognormalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,1615.

Exponential	Gamma	Normal
4,4970	4,0207	0,6026

Episoderna med låga TV tittarsiffror

Tabell 21: De resulterade testvariablerna för episoderna med låga TV tittarsiffror. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 0,9438.

Exponential	Gamma	Normal
11,3395	9,5543	0,4914

Episoderna med höga online rating

Tabell 22: De resulterade testvariablerna för episoderna med höga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,2145.

Exponential	Gamma	Normal
3,1906	2,4535	0,3565

Episoderna med låga online rating

Tabell 23: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 0,9594.

Exponential	Gamma	Normal
15,0000	13,4691	0,4992

B.2 Test för observationer före eller efter programmets mittpunkt, ytterligare uppdelat i TV populära eller impopulära delgrupper

Första halvan av episoderna med höga TV tittarsiffror

Tabell 24: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,2101.

Exponential	Gamma	Normal
2,8019	2,0568	0,4533

Andra halvan av episoderna med höga TV tittarsiffror

Tabell 25: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,1297.

Exponential	Gamma	Normal
1,8528	1,8495	0,3348

Första halvan av episoderna med låga TV tittarsiffror

Tabell 26: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 0,9592.

Exponential	Gamma	Normal
5,7424	5,1363	0,5628

Andra halvan av episoderna med låga TV tittarsiffror

Tabell 27: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 0,9518.

Exponential	Gamma	Normal
5,2492	4,5128	0,3917

B.3 Test för observationer före eller efter programmets mittpunkt, ytterligare uppdelat i online populära eller impopulära delgrupper

Första halvan av episoderna med hög online rating

Tabell 28: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,2003.

Exponential	Gamma	Normal
1,1473	0,8192	0,4429

Andra halvan av episoderna med hög online rating

Tabell 29: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 1,3089.

Exponential	Gamma	Normal
2,8418	2,3946	0,4100

Första halvan av episoderna med låg online rating

Tabell 30: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 0,9948.

Exponential	Gamma	Normal
7,6214	7,5387	0,5631

Andra halvan av episoderna med låg online rating

Tabell 31: De resulterade testvariablerna för episoderna med låga online rating. Med hjälp av förkastningsregionerna i sektionen E fås att endast den tredje nollhypotesen, att datan är log-normalfördelad, inte kan förkastas. Gammafördelningens formparametern \hat{m} har värdet 0,9407.

Exponential	Gamma	Normal
7,1293	5,8905	0,5306

C Resultat för k-sample Anderson-Darling-test

In den här sektionen presenteras alla resultat av de ickeparametriska test som görs med k-sample Anderson-Darling. Testerna grupperas in i tre kategorier: en som studerar skillnaderna mellan tittarsiffror och programmets struktur separat, en för datan som delats in efter höga och låga TV tittarsiffror, samt en för datan som delats in efter höga och låga online tittarsiffror.

C.1 Separata fall för tittarsiffror och TV-programmets struktur

C.1.1 Homogenitet test för första halvan och andra halvan av alla program

Test av homogenitet för första halven och andra halvan av alla program. Nollhypotesen är att väntetiderna för första halvan av ett program har samma population som andra halvan av ett program. Notera att det, enligt det här testet, finns en signifikant skillnad emellan deras populationsfördelningar.

Tabell 32: Test av homogenitet för första halven och andra halvan av alla program. Notera nollhypotesen, att båda stickproven har samma fördelning, kan förkastas med en signifikansnivå av 5%.

A^2	Standardiserad A^2	Asympt. p -value
6,87	7,722	0,0004

C.1.2 Homogenitet test för TV populära och TV impopulära program

Test av homogenitet för TV populära av och TV impopulära alla program. Nollhypotesen är att väntetiderna för TV populära program har samma population som TV impopulära program. Notera att det, enligt det här testet, finns en signifikant skillnad emellan deras populationsfördelningar.

Tabell 33: Test av homogenitet för TV populära av och TV impopulära alla program. Notera nollhypotesen, att båda stickproven har samma fördelning, kan förkastas med en signifikansnivå av 5%.

A^2	Standardiserad A^2	Asympt. p -value
8,270	9,560	> 0,0001

C.1.3 Homogenitet test för online populära och online impopulära program

Test av homogenitet för online populära av och online impopulära alla program. Nollhypotesen är att väntetiderna för online populära program har samma population som online impopulära program. Notera att det, enligt det här testet, finns ingen signifikant skillnad emellan deras populationsfördelningar.

Jämför det här resultatet med resultatet för TV programmen i Tabell 33. Där kan det ses att det finns en signifikant skillnad, medan för online kan ingen ses. Det här påvisar att stratifieringen av datan är viktigt och bör studeras mer för att få en bättre analys.

Tabell 34: Test av homogenitet för online populära av och online impopulära alla program. Notera nollhypotesen, att båda stickproven har samma fördelning, kan förkastas med en signifikansnivå av 5%.

A^2	Standardiserad A^2	Asympt. p -value
2,360	1,790	0,0587

C.2 Data uppdelad efter TV tittarsiffror

Testerna i den här sektionen har nollhypotesen att olika stickprover kommer från samma population och jämför det mot att åtminstone en inte gör det. Dataseten som jämförs är en stratifiering av den sammanslagna, där definitionen av uppdelningen kan ses i Tabell 35.

Tabell 35: Definitionerna av hur dataseten är stratifierade för TV tittarsiffror. Notera att ingen observationerna i de här dataseten återfinns i någon av de andra, på grund av hur de är definierade.

Definition av data grupperingar: Alla observationer som är i ...
... första halvan av TV populära program
... andra halvan av TV populära program
... första halvan av TV impopulära program
... andra halvan av TV impopulära program

Notera att genom de här definitionerna, är snittet emellan dem tomma mängden. Det vill säga att inga observationer delas emellan dataseten.

C.2.1 Test av homogenitet mellan data stratifieringen TV populära program och i första eller andra halvan av episoden

I Tabellen 36 ses att p -värdet av testet att *första halvan av TV populära program, andra halvan av TV populära program, första halvan av TV impopulära program och andrahälvan av TV impopulära program* skulle komma från samma population. Observera att nollhypotesen kan förkastas eftersom p -värdet är mindre än signifikansnivån 5% och därmed är det en signifikant skillnad i deras skattade populationer.

Tabell 36: Resultatet av ett k -sample Anderson-Darling-test för homogeniteten mellan dataseten *första halvan av TV populära program, andra halvan av TV populära program, första halvan av TV impopulära program och andrahälvan av TV impopulära program*. Testet är baserat på version 2 av testerna presenterade av Scholz and Stephens [10] och kan hantera observationer med samma värden. Notera att nollhypotesen, de olika stickproven kommer från samma population, kan förkastas med en signifikansnivå av 5% givet det asymptotiska p -värdet.

A^2	Standardiserad A^2	Asympt. p -value
15,6	9,574	> 0,0001

C.2.2 Två-stickprov test av av de fyra stratifierade dataseten

Test av homogenitet mellan dataseten *första halvan av TV populära program* och *andra halvan av TV populära program*

Tabell 37: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av TV populära program* och *andra halvan av TV populära program*. Observera att ingen signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan inte nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
2,030	1,355	0,0884

Test av homogenitet mellan dataseten *första halvan av TV populära program* och *första halvan av TV impopulära program*

Tabell 38: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av TV populära program* och *första halvan av TV impopulära program*. Observera att ingen signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan inte nollhypotesen förkastas, nämligen homogeniteten mellan stickproven. Emellertid är den här skillnaden liten eftersom den bestämda signifikansnivån är 5% och kan antyda att det kanske finns en skillnad, men mer data behövs för ett definitivt påstående.

A^2	Standardiserad A^2	Asympt. p -value
3,520	3,314	0,0150

Test av homogenitet mellan dataseten *första halvan av TV populära program* och *andrahalvan av TV impopulära program*

Tabell 39: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av TV populära program* och *andra halvan av TV impopulära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
13,80	16,85	> 0,0001

Test av homogenitet mellan dataseten *andra halvan av TV populära program* och *första halvan av TV impopulära program*

Tabell 40: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *andra halvan av TV populära program* och *första halvan av TV impopulära program*. Observera att ingen signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan inte nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
0,6600	-0,4478	0,5932

Test av homogenitet mellan dataseten *andra halvan av TV populära program* och *andrahalvan av TV impopulära program*

Tabell 41: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *andra halvan av TV populära program* och *andra halvan av TV impopulära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven. Det här betyder att fördelningarna av väntetiden, i andra hälften av TV program, är signifikant olika emellan populära och impopulära.

A^2	Standardiserad A^2	Asympt. p -value
5,230	5,568	0,0022

Test av homogenitet mellan dataseten *första halvan av TV impopulära program* och *andrahalvan av TV impopulära program*

Tabell 42: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av TV impopulära program* och *andra halvan av TV impopulära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven. Jämför det här resultatet med det i 37 vilket är motsvarigheten i populära program, vilket inte blev signifikant.

A^2	Standardiserad A^2	Asympt. p -value
5,210	5,536	0,0023

C.3 Data uppdelad efter online tittarsiffror

Testerna i den här sektionen har nollhypotesen att olika stickprover kommer från samma population och jämför det mot att åtminstone en inte gör det. Dataseten som jämförs är en stratifiering av den sammanslagna, där definitionen av uppdelningen kan ses i Tabell 43.

Tabell 43: Definitionerna av hur dataseten är stratifierade för online tittarsiffror. Notera att ingen observationerna i de här dataseten återfinns i någon av de andra, på grund av hur de är definierade.

Definition av data grupperingar: Alla observationer som är i ...
... <i>första halvan av online populära program</i>
... <i>andra halvan av online populära program</i>
... <i>första halvan av online impopulära program</i>
... <i>andra halvan av online impopulära program</i>

Notera att genom de här definitionerna, är snittet emellan dem tomma mängden. Det vill säga att inga observationer delas emellan dataseten.

C.3.1 Test av homogenitet mellan data stratifieringen online populära program och i första eller andra halvan av episoden

I Tabellen 44 ses att p -värdet av testet att *första halvan av online populära program*, *andra halvan av online populära program*, *första halvan av online impopulära program* och *andra halvan av online impopulära program* skulle komma från samma population. Observera att nollhypotesen kan förkastas eftersom p -värdet är mindre än signifikansnivån 5% och därmed är det en signifikant skillnad i deras skattade populationer.

Tabell 44: Resultatet av ett k -sample Anderson-Darling-test för homogeniteten mellan dataseten *första halvan av online populära program*, *andra halvan av online populära program*, *första halvan av online impopulära program* och *andra halvan av online impopulära program*. Testet är baserat på version 2 av testerna presenterade av Scholz and Stephens [10] och kan hantera observationer med samma värden. Notera att nollhypotesen, de olika stickproven kommer från samma population, kan förkastas med en signifikansnivå av 5% givet det asymptotiska p -värdet.

A^2	Standardiserad A^2	Asympt. p -value
11,00	6,067	0.0003

C.3.2 Två-stickprov test av de fyra stratifierade dataseten

Test av homogenitet mellan dataseten *första halvan av online populära program och andra halvan av online populära program*

Tabell 45: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av online populära program* och *andra halvan av online populära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
4,580	4,736	0,0045

Test av homogenitet mellan dataseten *första halvan av online populära program och första halvan av online impopulära program*

Tabell 46: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av online populära program* och *första halvan av online impopulära program*. Observera att ingen signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan inte nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
0,6700	-0,4347	0,5845

Test av homogenitet mellan dataseten *första halvan av online populära program och andra halvan av online impopulära program*

Tabell 47: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av online populära program* och *andra halvan av online impopulära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
2,460	1,926	0,0518

Test av homogenitet mellan dataseten *andra halvan av online populära program och första halvan av online impopulära program*

Tabell 48: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *andra halvan av online populära program* och *första halvan av online impopulära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
7,040	7,954	0,0003

Test av homogenitet mellan dataseten *andra halvan av online populära program och andra halvan av online impopulära program*

Tabell 49: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *andra halvan av online populära program* och *andra halvan av online impopulära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
3,45	3,229	0,0162

Test av homogenitet mellan dataseten *första halvan av online impopulära program* och *andra halvan av online impopulära program*

Tabell 50: Resultatet av 2-sample Anderson-Darling-test mellan stickproverna *första halvan av online impopulära program* och *andra halvan av online impopulära program*. Observera att en signifikant skillnad kan upptäckas emellan de här stickproverna och därmed kan nollhypotesen förkastas, nämligen homogeniteten mellan stickproven.

A^2	Standardiserad A^2	Asympt. p -value
4,310	4,355	0,0062

D Skattade parametrar för lognormalfördelningen

Det gick inte, i något fall, att förkasta nollhypotesen att datan kommer från en lognormalfördelning. Det betyder att den logaritmerade datan är normalfördelad. I Tabell 51 ses de Maximum likelihood- estimerade parametrarna för den underliggande normalfördelningen. Väntevärde och varians för motsvarande lognormalfördelning ges av $\mu = \exp[\text{meanLog} + \text{sdLog}^2/2]$ samt $\sigma^2 = \exp[\text{sdLog}^2 - 1] \exp[2\text{meanLog} + \text{sdLog}^2]$.

Tabell 51: Fördelningsparametrarna för de olika grupperingarna av datan, estimerade med Maximum Likelihood-metoden i R. Parametrarna är väntevärden och variansen i den underliggande normalfördelningen.

Data grupperingar	meanLog	sdLog
All data	3,7562	1,1101
Första halvan	3,6549	1,0944
Andra halvan	3,8703	1,1166
TV populär	3,6309	1,0311
TV impopulär	3,8438	1,1541
Online populär	3,8155	0,9956
Online impopulär	3,7414	1,1364
TV populär första hälften	3,5601	1,0074
TV populär andra hälften	3,7124	1,0519
TV impopulär första hälften	3,7223	1,1476
TV impopulär andra hälften	3,9784	1,1464
Online populär första hälften	3,6586	1,0191
Online populär andra hälften	4,0199	0,9250
Online impopulär första hälften	3,6539	1,1139
Online impopulär andra hälften	3,8365	1,1528

E Modifierade och asymptotiska förkastningsregioner

Tabell 52: Konfidensgränser och signifikansnivåer för Anderson-Darling-test för normalfördelning med μ och σ^2 okända [8].

α	0,15	0,1	0,05	0,025	0,01	0,005
R_α	0,561	0,631	0,752	0,873	1,035	1,159

Tabell 53: Konfidensgränser och signifikansnivåer för Anderson-Darling-test för exponentialfördelning med $\alpha = 0$ och β okänd [8].

α	0,15	0,1	0,05	0,025	0,01	0,005
R_α	0,916	1,062	1,321	1,591	1,959	2,244

Tabell 54: Konfidensgränser och signifikansnivåer för Anderson-Darling-test för gammafördelning med $\alpha = 0$ samt β och \hat{m} okända [8].

m	0,25	0,1	0,05	0,025	0,01	0,005
1	0,486	0,657	0,786	0,917	1,092	1,227
2	0,477	0,643	0,768	0,894	1,062	1,190
3	0,475	0,639	0,762	0,886	1,052	1,178
4	0,473	0,637	0,759	0,883	1,048	1,173

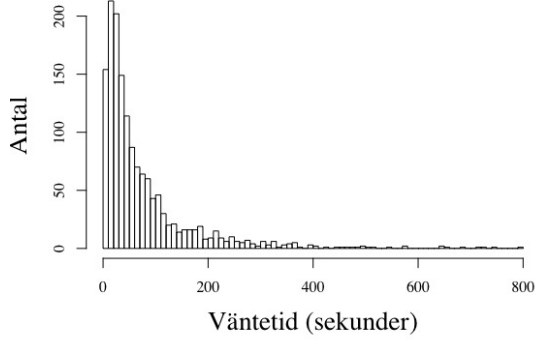
F Residualsumma

Tabell 55: Tabell över residualsumman för de olika skiftningarna.

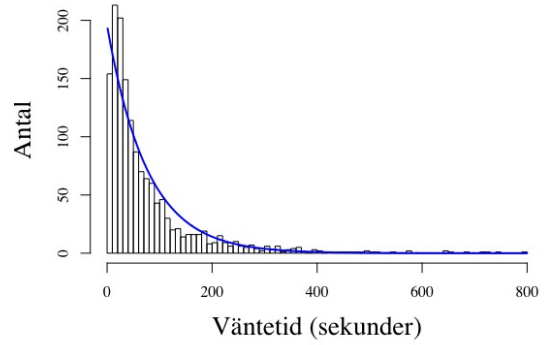
Typ av data	1 → 2	1 → 3	2 → 1	2 → 3	3 → 1	3 → 2
Alla Broadcast	$-3 \cdot 10^{-9}$	$-2 \cdot 10^{-9}$	$-9 \cdot 10^{-9}$	$-5 \cdot 10^{-9}$	$-4 \cdot 10^{-9}$	$-3 \cdot 10^{-9}$
Alla Online	$-2 \cdot 10^{-10}$	$-2 \cdot 10^{-10}$	$-4 \cdot 10^{-10}$	$-4 \cdot 10^{-10}$	$-4 \cdot 10^{-10}$	$-2 \cdot 10^{-10}$
Person 1 Broadcast	$7 \cdot 10^{-9*}$	$7 \cdot 10^{-10}$	$-3 \cdot 10^{-9*}$	$3 \cdot 10^{-10*}$	$7 \cdot 10^{-10}$	$-8 \cdot 10^{-10*}$
Person 1 Online	$-8 \cdot 10^{-11}$	$7 \cdot 10^{-12}$	$-7 \cdot 10^{-11}$	$9 \cdot 10^{-11}$	$9 \cdot 10^{-11}$	$3 \cdot 10^{-11}$
Person 2 Broadcast	$-3 \cdot 10^{-9}$	$-3 \cdot 10^{-9}$	$-3 \cdot 10^{-9}$	$-2 \cdot 10^{-9}$	$-3 \cdot 10^{-9}$	$-2 \cdot 10^{-9}$
Person 2 Online	$-4 \cdot 10^{-10}$	$-6 \cdot 10^{-10}$	$-5 \cdot 10^{-10}$	$-1 \cdot 10^{-10}$	$-5 \cdot 10^{-10}$	$-1 \cdot 10^{-10}$

G Figurer

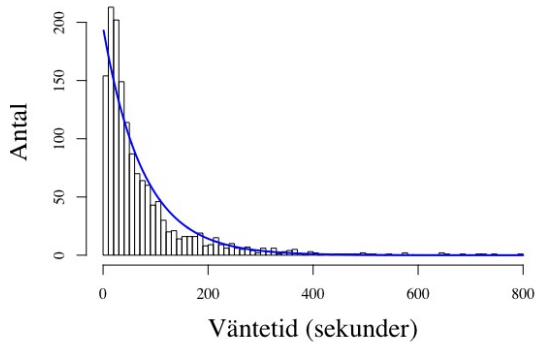
G.1 Alla programs väntetider



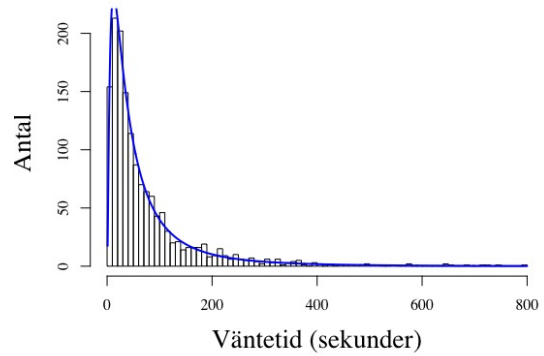
(a) Histogram över alla programs väntetider.



(b) Histogram över alla programs väntetider med en anpassad exponentialfördelning.

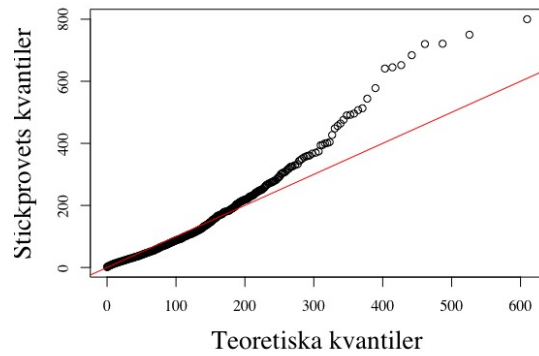
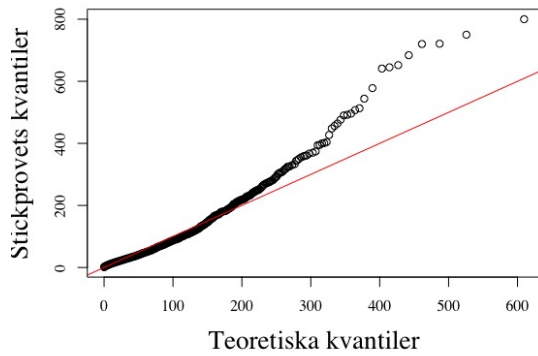


(c) Histogram över alla programs väntetider med en anpassad gammafördelning .

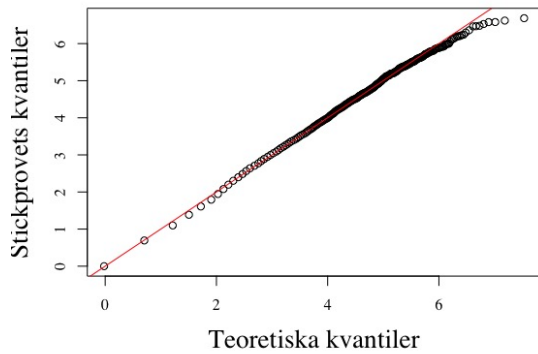


(d) Histogram över alla programs väntetider med en anpassad lognormalfördelning.

Figur 10: Histogram med och utan anpassad fördelning för *alla* väntetider i *alla* program.



(a) QQ-plot för exponentialfördelningen av alla vän- (b) QQ-plot för gammafördelningen av alla vänteti-
tetider. der.

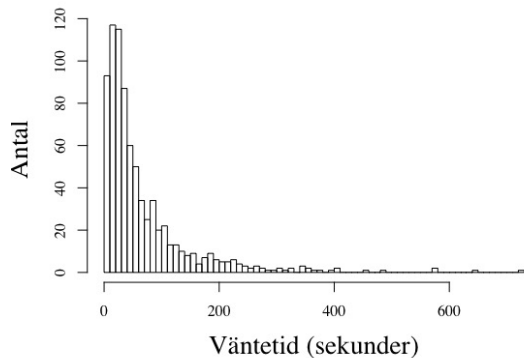


(c) QQ-plot för normalfördelningen för logaritmen av
alla väntetider.

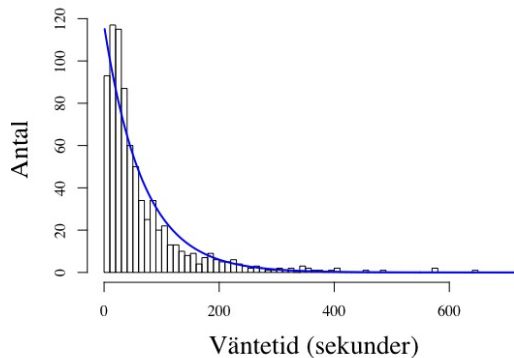
Figur 11: QQ-plots för hela datans väntetider. Figur 11a och 11b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 11c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

G.2 Första och andra halvans väntetider

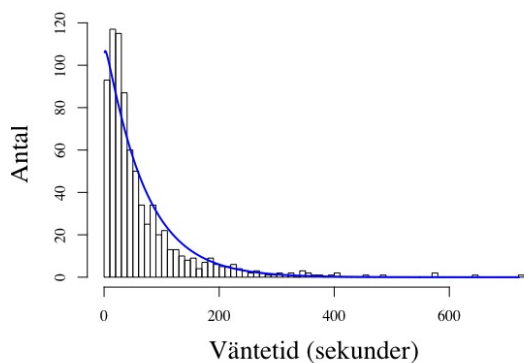
G.2.1 Första halvan



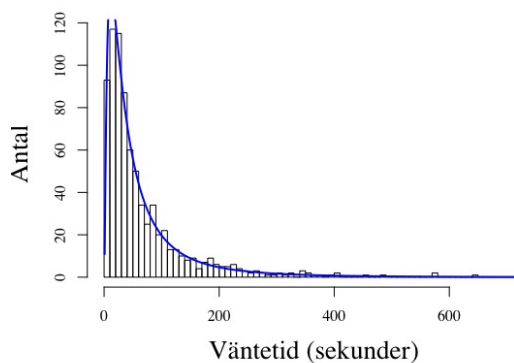
(a) Histogram över väntetiderna i programmens första halva.



(b) Histogram över väntetiderna i programmens första halva med en anpassad exponentialfördelning

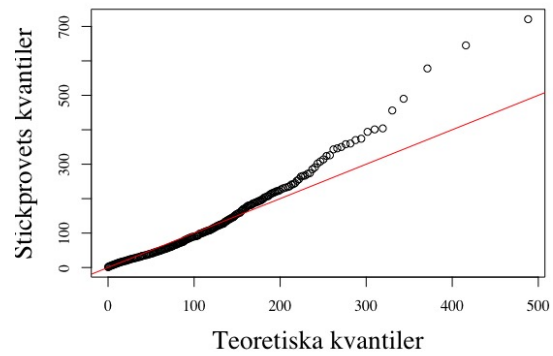
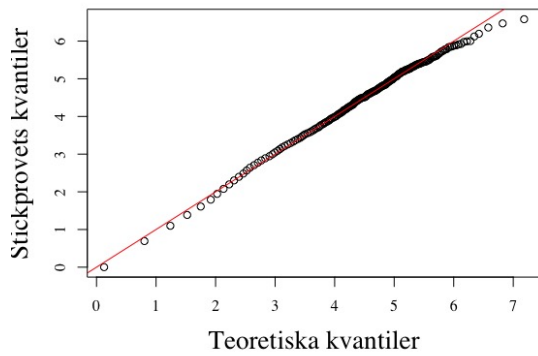


(c) Histogram över väntetiderna i programmens första halva med en anpassad gammafördelning.

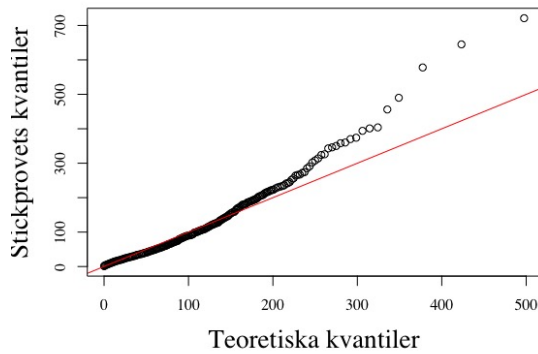


(d) Histogram över väntetiderna i programmens första halva med en anpassad lognormalfördelning.

Figur 12: Histogram med och utan anpassad fördelning för väntetiderna i programmens första halva.



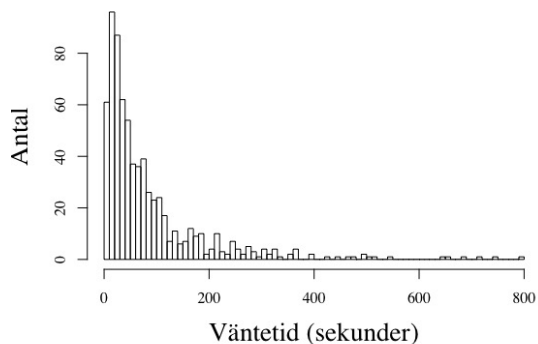
(a) QQ-plot för exponentialfördelningen av första halvans väntetider. (b) QQ-plot för gammafördelningen av första halvans väntetider.



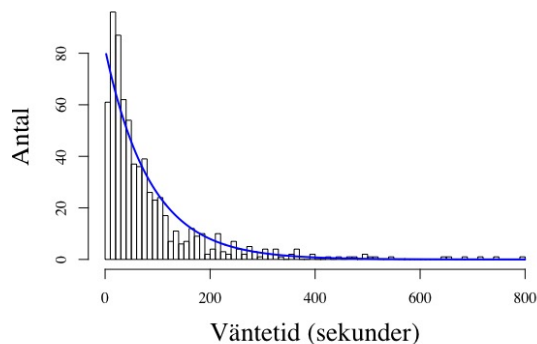
(c) QQ-plot för normalfördelningen för logaritmen av första halvans väntetider.

Figur 13: QQ-plots för väntetiderna i programmets första halva. Figur 13a och 13b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 13c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

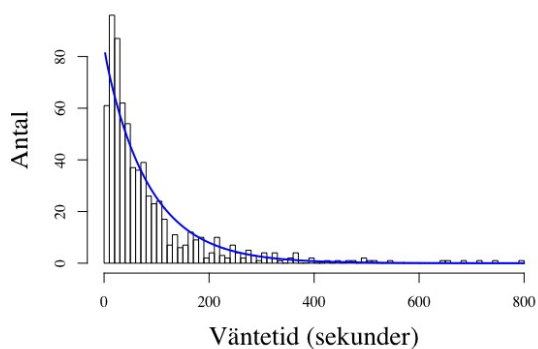
G.2.2 Andra halvan



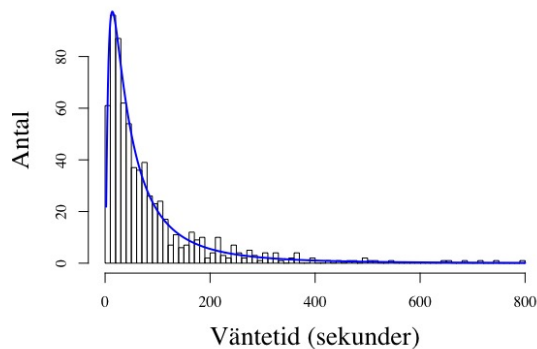
(a) Histogram över väntetiderna i programmens andra halva.



(b) Histogram över väntetiderna i programmens andra halva med en anpassad exponentialfördelning

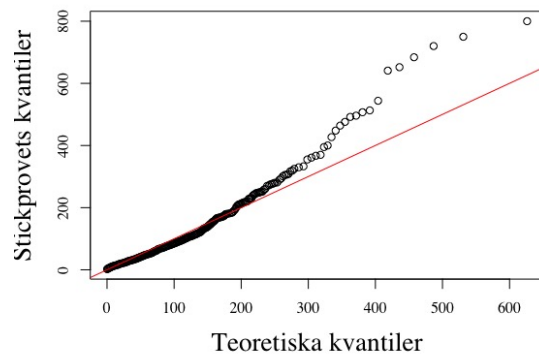
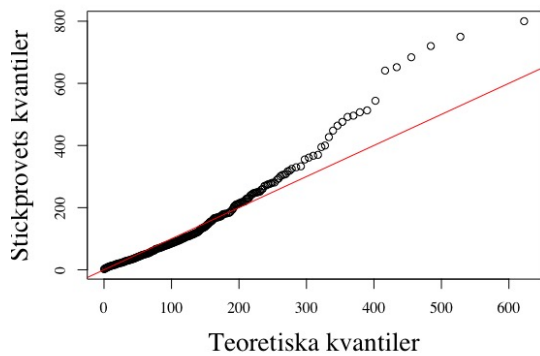


(c) Histogram över väntetiderna i programmens andra halva med en anpassad gammafördelning.

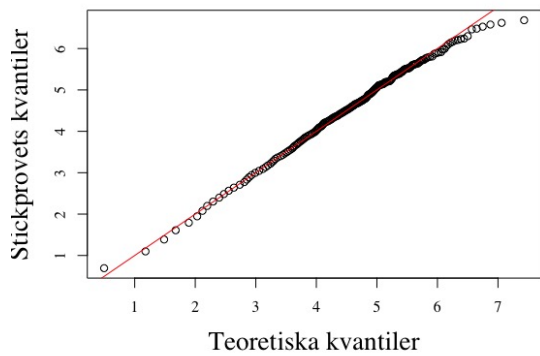


(d) Histogram över väntetiderna i programmens andra halva med en anpassad lognormalfördelning.

Figur 14: Histogram med och utan anpassad fördelning för väntetiderna i programmens andra halva.



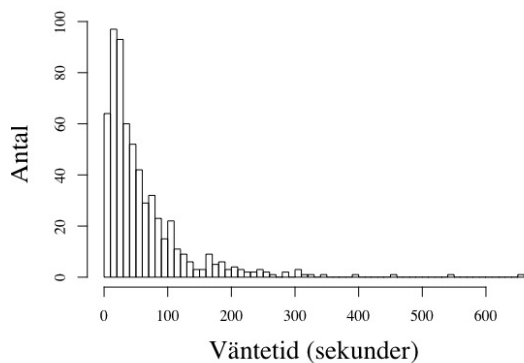
(a) QQ-plot för exponentialfördelningen av andra halvans väntetider. (b) QQ-plot för gammafördelningen av andra halvans väntetider.



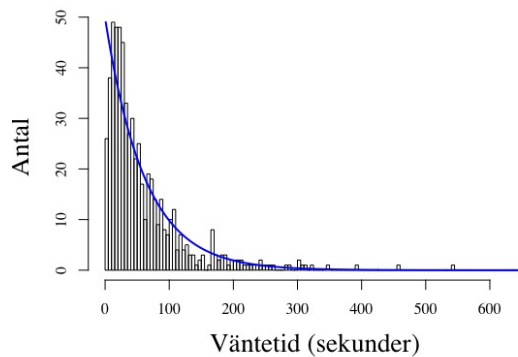
(c) QQ-plot för normalfördelningen för logaritmen av andra halvans väntetider.

Figur 15: QQ-plots för väntetiderna i programmets andra halva. Figur 15a och 15b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 15c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

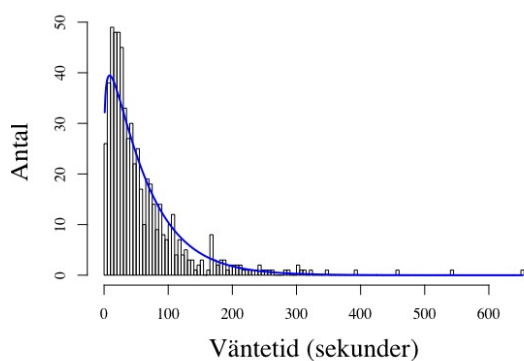
G.3 Höga tittarsiffror TV



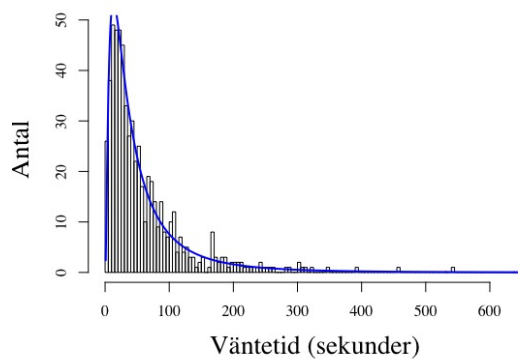
(a) Histogram över väntetiderna för program med höga TV-siffror med en anpassad exponentialfördelning.



(b) Histogram över väntetiderna för program med höga TV-siffror med en anpassad exponentialfördelning.

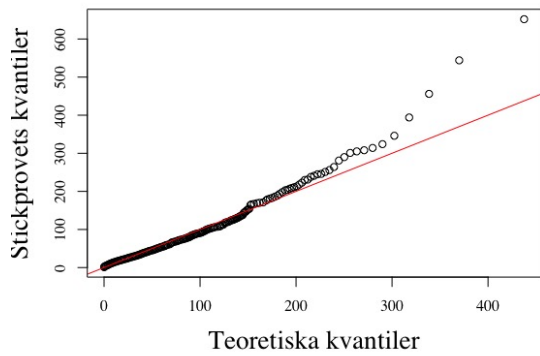


(c) Histogram över väntetiderna för program med höga TV-siffror med en anpassad lognormalfördelning.

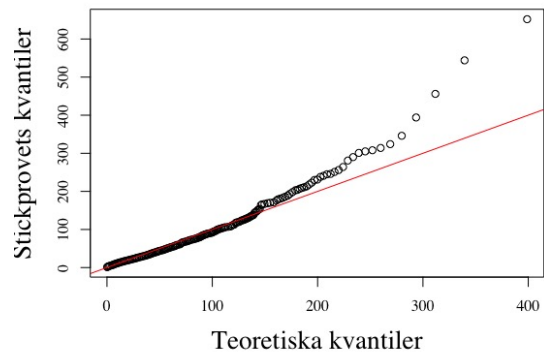


(d) Histogram över väntetiderna för program med höga TV-siffror med en anpassad gammafördelning.

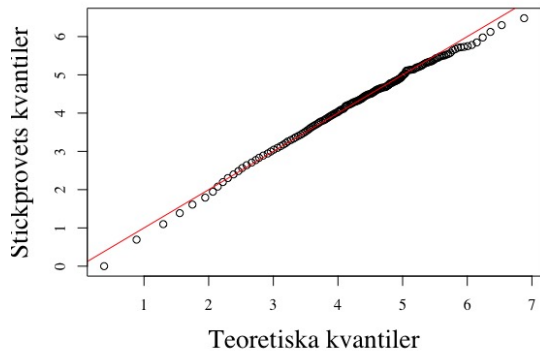
Figur 16: Histogram med och utan anpassad fördelning för väntetiderna program med höga TV-siffror ($> 800\,000$).



(a) QQ-plot för exponentialfördelningen av program med höga TV-siffror.



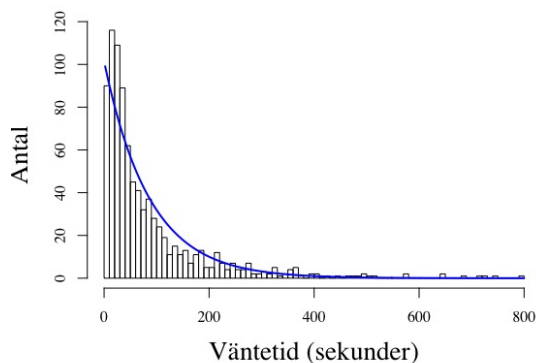
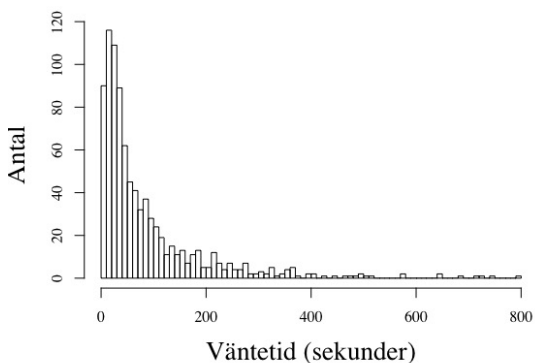
(b) QQ-plot för gammalfördelningen av program med höga TV-siffror.



(c) QQ-plot för normalfördelningen för logaritmen av program med höga TV-siffror.

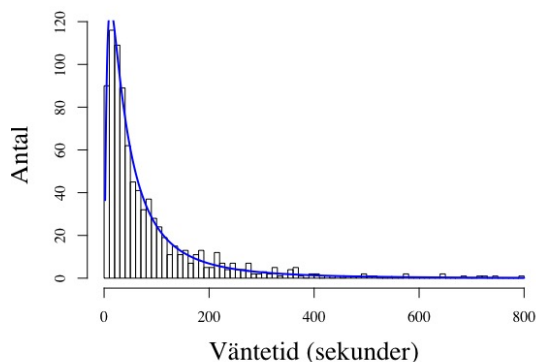
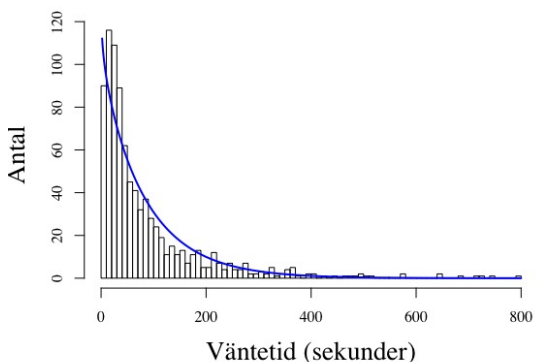
Figur 17: QQ-plots för väntetiderna i program med höga TV-siffror ($> 800\,000$). Figur 17a och 17b visar de teoretiska kvantilerna för exponential- respektive gammalfördelningen plottade mot stickprovets kvantiler. Figur 17c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

G.4 Låga tittarsiffror TV



(a) Histogram över väntetiderna för program med låga TV-siffror.

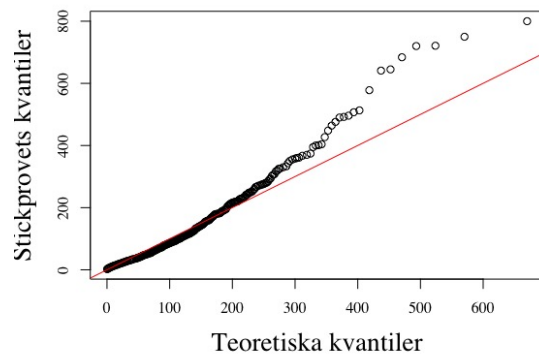
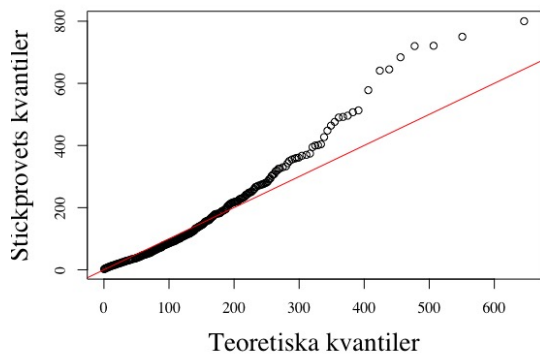
(b) Histogram över väntetiderna för program med låga TV-siffror med en anpassad exponentialfördelning.



(c) Histogram över väntetiderna för program med låga TV-siffror med en anpassad gammafördelning.

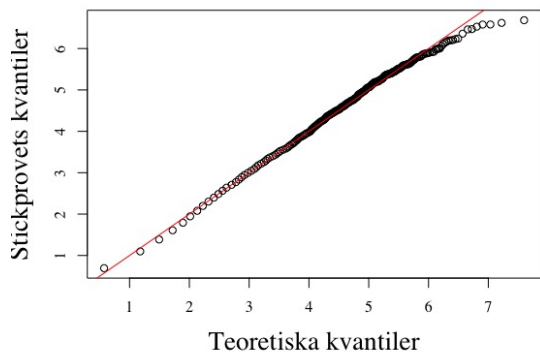
(d) Histogram över väntetiderna för program med låga TV-siffror med en anpassad lognormalfördelning.

Figur 18: Histogram med och utan anpassad fördelning för väntetiderna program med låga TV-siffror ($\leq 800\ 000$).



(a) QQ-plot för exponentialfördelningen av program med låga TV-siffror.

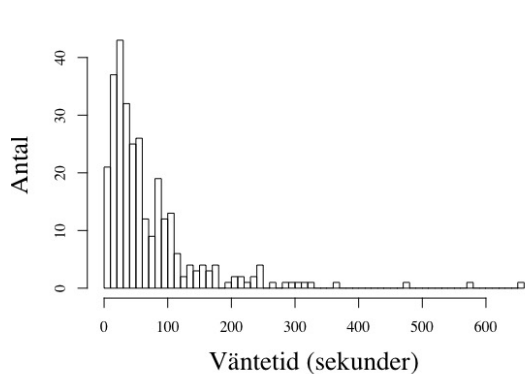
(b) QQ-plot för gammafördelningen av program med låga TV-siffror.



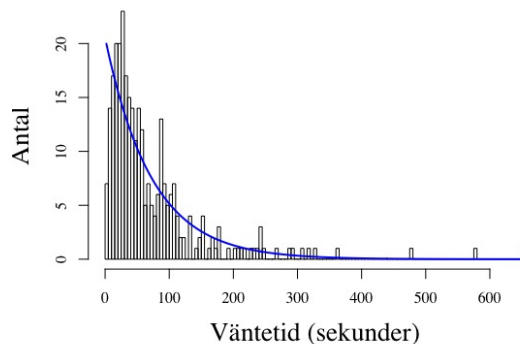
(c) QQ-plot för normalfördelningen för logaritmen av program med låga TV-siffror.

Figur 19: QQ-plots för väntetiderna i program med låga TV-siffror ($\leq 800\ 000$). Figur 19a och 19b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 19c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

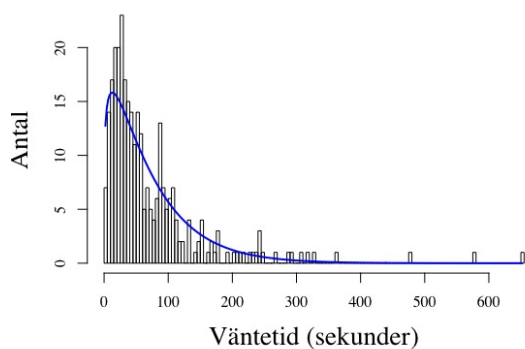
G.5 Höga onlinesiffror



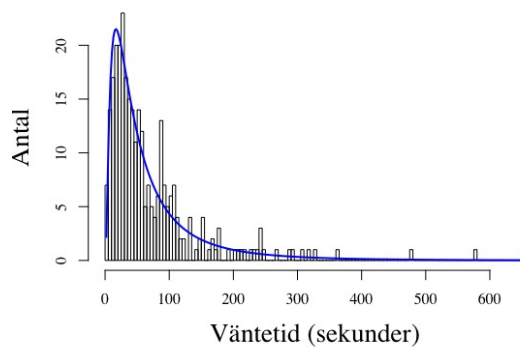
(a) Histogram över väntetiderna för program med höga onlinesiffror.



(b) Histogram över väntetiderna för program med höga onlinesiffror med en anpassad exponentialfördelning.

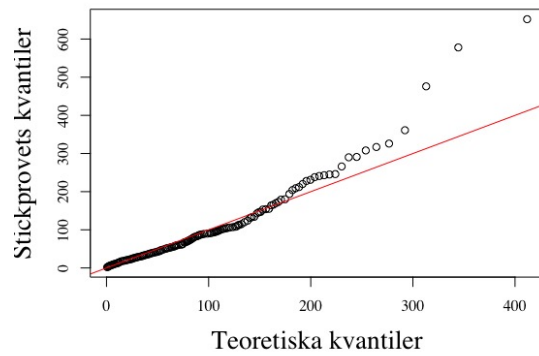
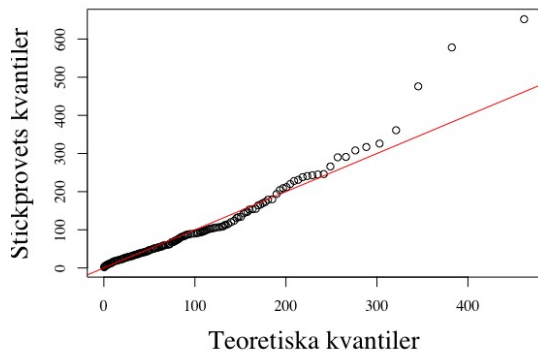


(c) Histogram över väntetiderna för program med höga onlinesiffror med en anpassad lognormalfördelning.

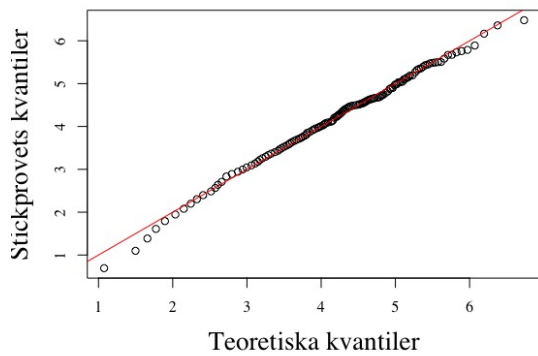


(d) Histogram över väntetiderna för program med höga onlinesiffror med en anpassad gammafördelning.

Figur 20: Histogram med och utan anpassad fördelning för väntetiderna program med höga onlinesiffror ($> 80\,000$).



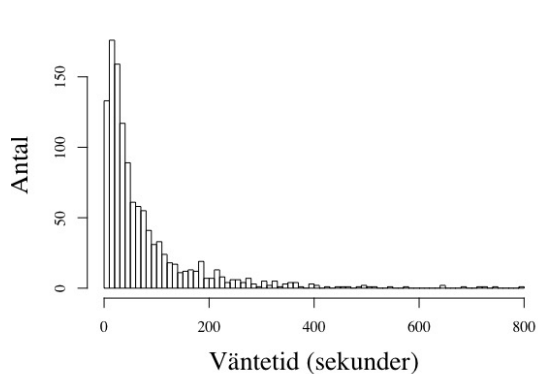
(a) QQ-plot för exponentialfördelningen av program med höga onlinesiffror. (b) QQ-plot för gammalfördelningen av program med höga onlinesiffror.



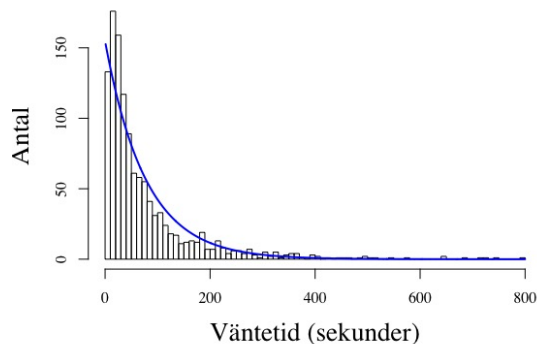
(c) QQ-plot för normalfördelningen för logaritmen av program med höga onlinesiffror.

Figur 21: QQ-plots för väntetiderna i program med höga onlinesiffror ($> 80\,000$). Figur 21a och 21b visar de teoretiska kvantilerna för exponential- respektive gammalfördelningen plottade mot stickprovets kvantiler. Figur 21c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

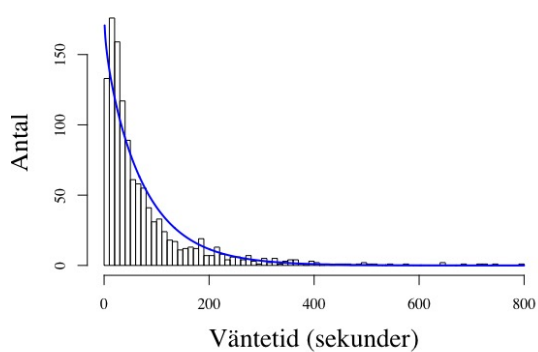
G.6 Låga onlinesiffror



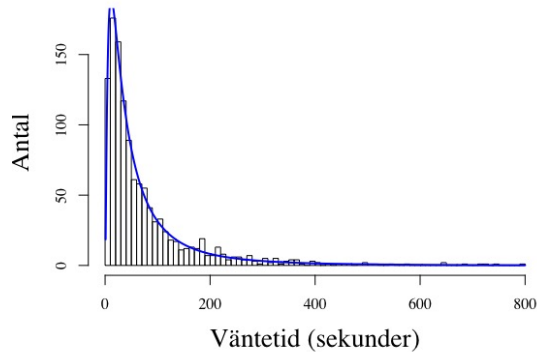
(a) Histogram över väntetiderna för program med låga onlinesiffror.



(b) Histogram över väntetiderna för program med låga onlinesiffror med en anpassad exponentialfördelning.

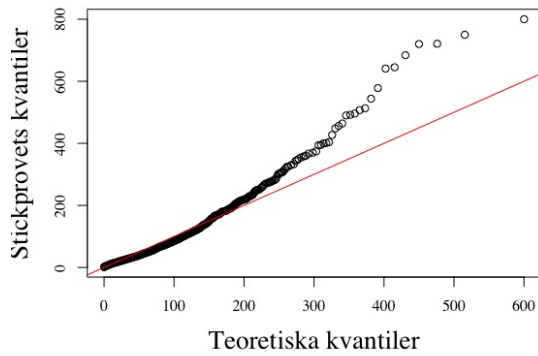


(c) Histogram över väntetiderna för program med låga onlinesiffror med en anpassad gammafördelning.

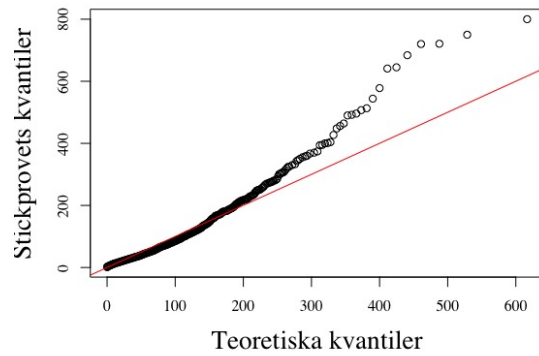


(d) Histogram över väntetiderna för program med låga onlinesiffror med en anpassad lognormalfördelning.

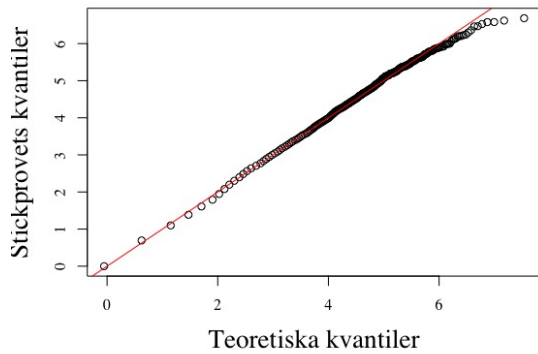
Figur 22: Histogram med och utan anpassad fördelning för väntetiderna program med låga onlinesiffror ($\leq 80\ 000$).



(a) QQ-plot för exponentialfördelningen av program med låga onlinesiffror.



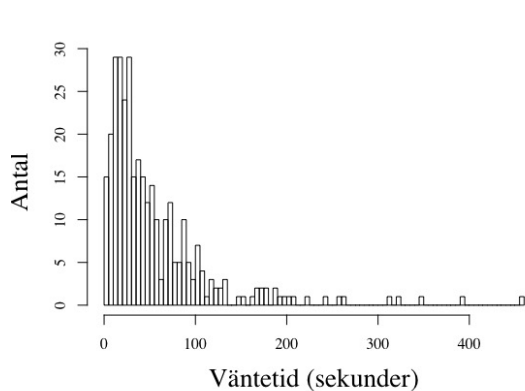
(b) QQ-plot för gammafördelningen av program med låga onlinesiffror.



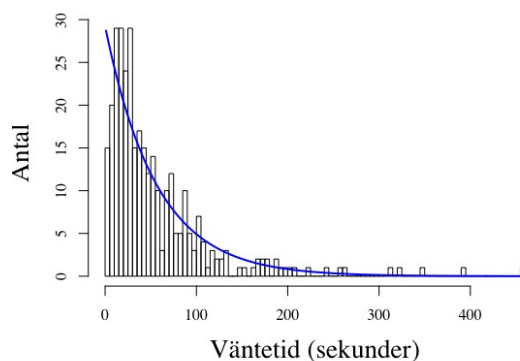
(c) QQ-plot för normalfördelningen för logaritmen av program med låga onlinesiffror.

Figur 23: QQ-plots för väntetiderna i program med höga onlinesiffror ($> 80\ 000$). Figur 23a och 23b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 23c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

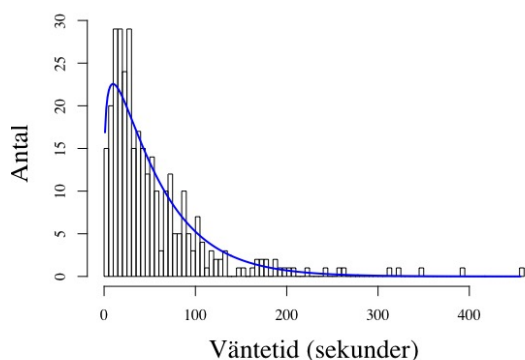
G.7 Första halvan av program med höga TV-siffror



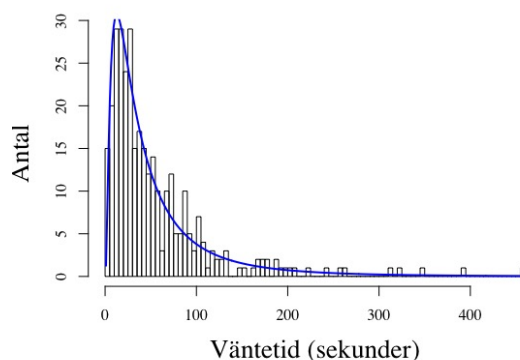
(a) Histogram över väntetiderna i första halvan för program med höga TV-siffror.



(b) Histogram över väntetiderna i första halvan för program med höga TV-siffror med en anpassad exponentialfördelning.

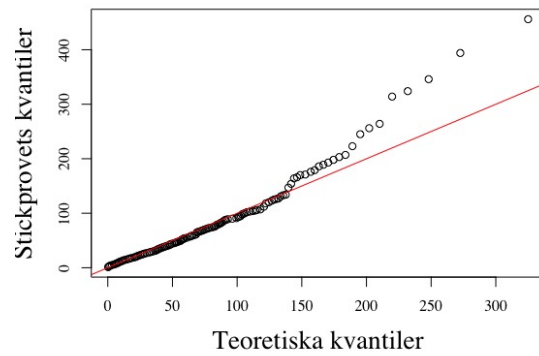
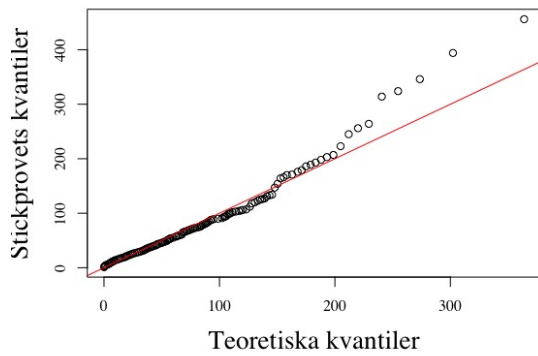


(c) Histogram över väntetiderna i första halvan för program med höga TV-siffror med en anpassad gammafördelning.



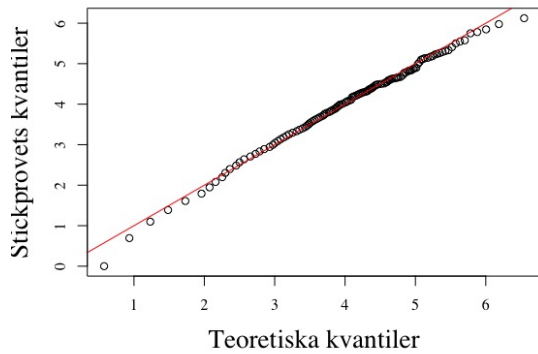
(d) Histogram över väntetiderna i första halvan för program med höga TV-siffror med en anpassad lognormalfördelning.

Figur 24: Histogram med och utan anpassad fördelning för väntetiderna i *första* halvan av program med *höga* TV-siffror ($> 800\,000$).



(a) QQ-plot för exponentialfördelningen av första halvan i program med höga TV-siffror.

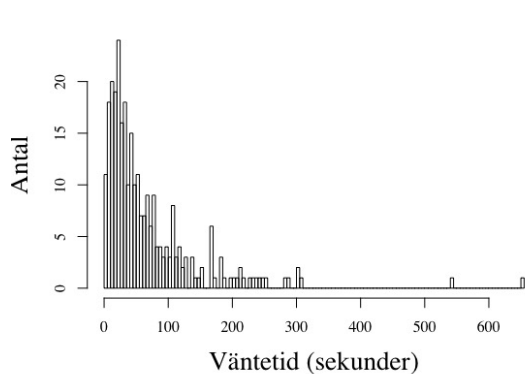
(b) QQ-plot för gammafördelningen av första halvan i program med höga TV-siffror.



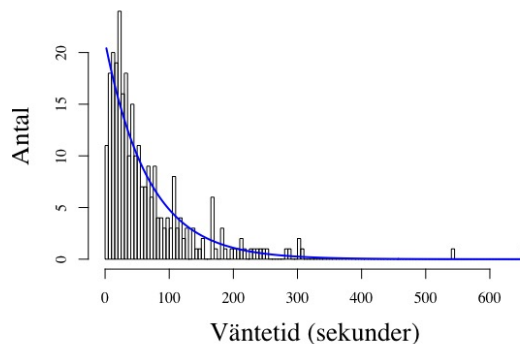
(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i första halvan av program med höga TV-siffror.

Figur 25: QQ-plots för väntetiderna i första halvan av program med höga TV-siffror ($> 800\,000$). Figur 25a och 25b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 25c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

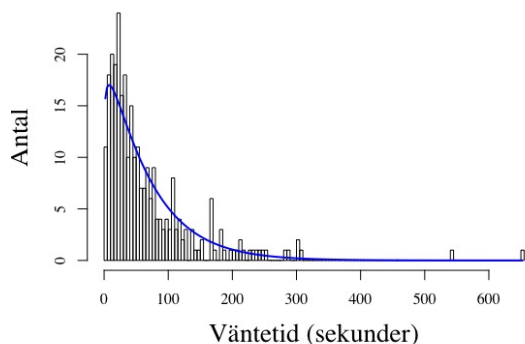
G.8 Andra halvan av program med höga TV-siffror



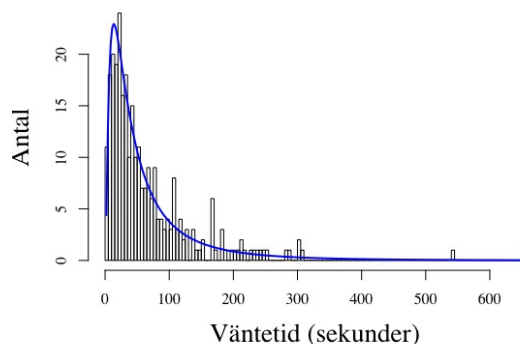
(a) Histogram över väntetiderna i andra halvan förprogram med höga TV-siffror.



(b) Histogram över väntetiderna i andra halvan förprogram med höga TV-siffror med en anpassad exponentialfördelning.

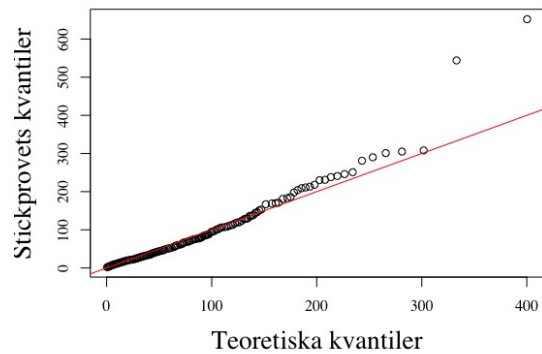
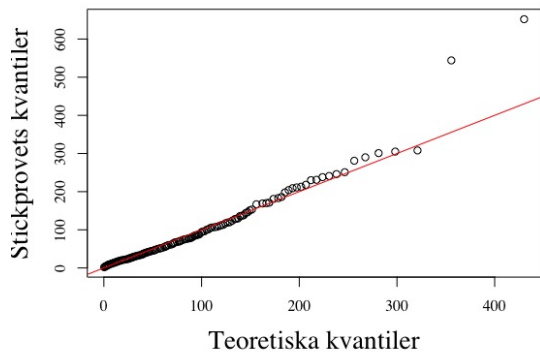


(c) Histogram över väntetiderna i andra halvan förprogram med höga TV-siffror med en anpassad gammafördelning.



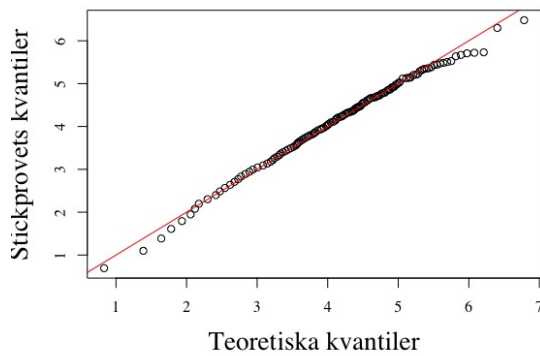
(d) Histogram över väntetiderna i andra halvan förprogram med höga TV-siffror med en anpassad lognormalfördelning.

Figur 26: Histogram med och utan anpassad fördelning för väntetiderna i *andra* halvan av program med *höga* TV-siffror ($> 800\,000$).



(a) QQ-plot för exponentialfördelningen av andra halvan i program med höga TV-siffror.

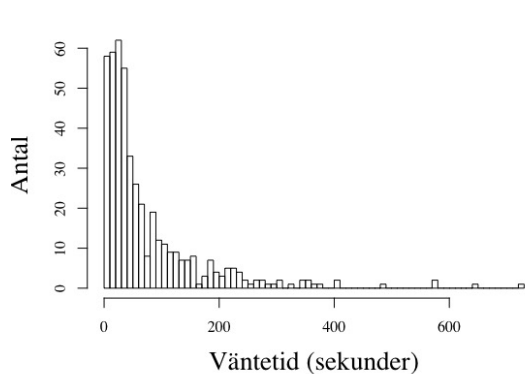
(b) QQ-plot för gammalfördelningen av andra halvan i program med höga TV-siffror.



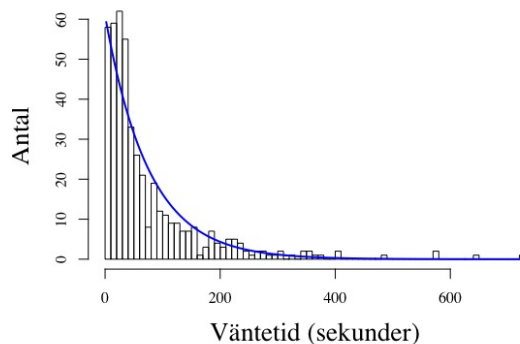
(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i andra halvan av program med höga TV-siffror.

Figur 27: QQ-plots för väntetiderna i andra halvan av program med höga TV-siffror ($> 800\,000$). Figur 27a och 27b visar de teoretiska kvantilerna för exponential- respektive gammalfördelningen plottade mot stickprovets kvantiler. Figur 27c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

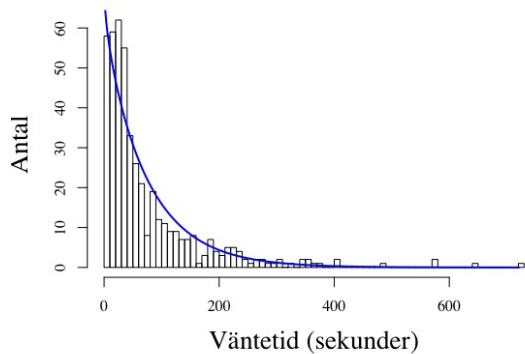
G.9 Första halvan av program med låga TV-siffror



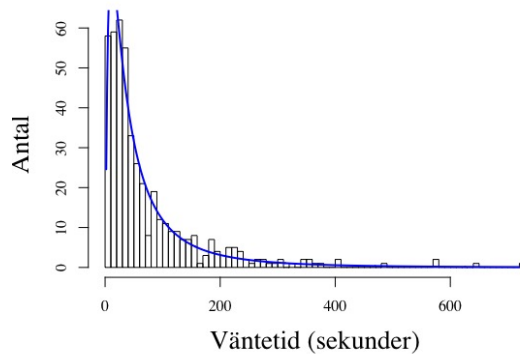
(a) Histogram över väntetiderna i första halvan förprogram med låga TV-siffror.



(b) Histogram över väntetiderna i första halvan förprogram med låga TV-siffror med en anpassad exponentialfördelning.

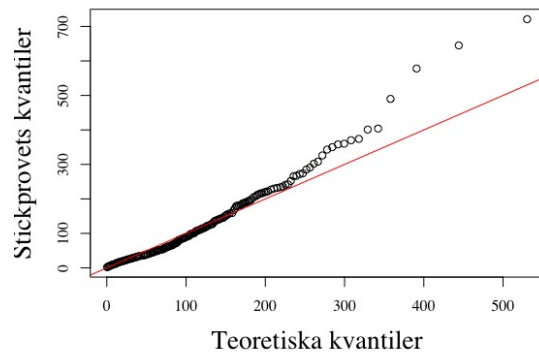
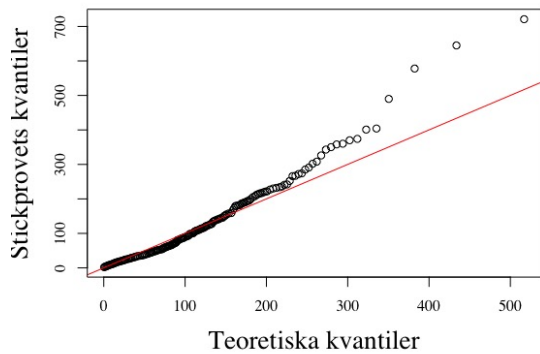


(c) Histogram över väntetiderna i första halvan förprogram med låga TV-siffror med en anpassad gammafördelning.



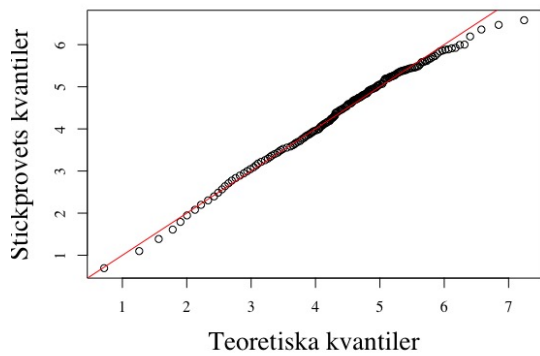
(d) Histogram över väntetiderna i första halvan förprogram med låga TV-siffror med en anpassad lognormalfördelning.

Figur 28: Histogram med och utan anpassad fördelning för väntetiderna i första halvan av program med låga TV-siffror ($\leq 800\ 000$).



(a) QQ-plot för exponentialfördelningen av första halvan i program med låga TV-siffror.

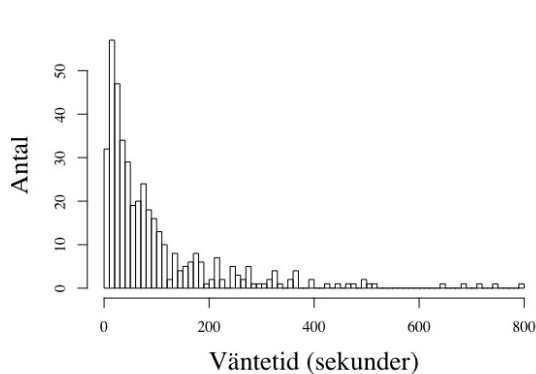
(b) QQ-plot för gammafördelningen av första halvan i program med låga TV-siffror.



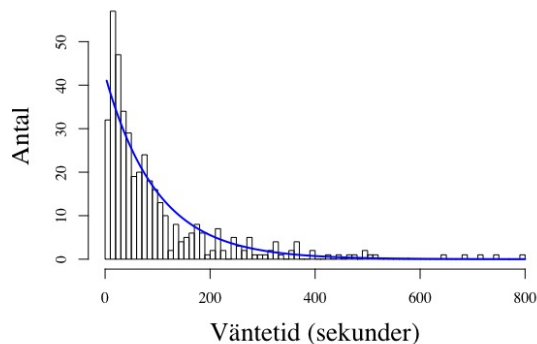
(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i första halvan av program med låga TV-siffror.

Figur 29: QQ-plots för väntetiderna i första halvan av program med låga TV-siffror ($\leq 800\ 000$). Figur 29a och 29b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 29c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

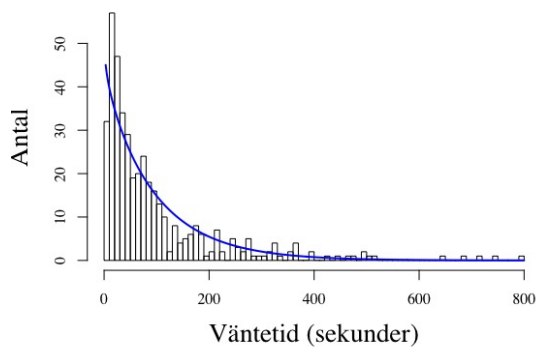
G.10 Andra halvan av program med låga TV-siffror



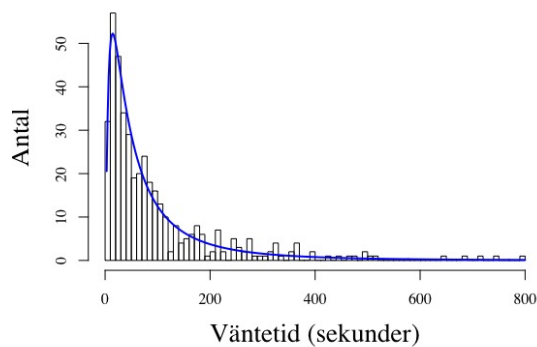
(a) Histogram över väntetiderna i andra halvan för program med låga TV-siffror.



(b) Histogram över väntetiderna i andra halvan för program med låga TV-siffror med en anpassad exponentialfördelning.

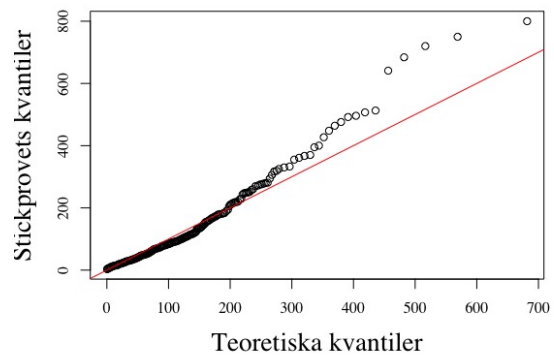
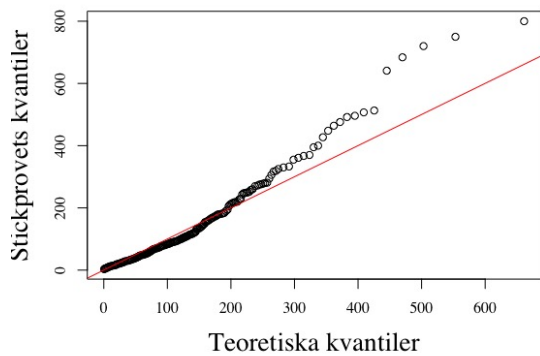


(c) Histogram över väntetiderna i andra halvan för program med låga TV-siffror med en anpassad gammafördelning.



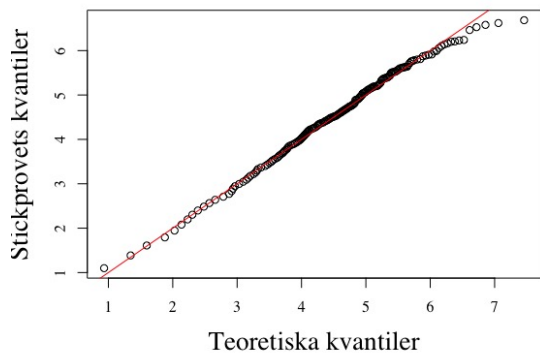
(d) Histogram över väntetiderna i andra halvan för program med låga TV-siffror med en anpassad lognormalfördelning.

Figur 30: Histogram med och utan anpassad fördelning för väntetiderna i *andra* halvan av program med låga TV-siffror ($\leq 800\ 000$).



(a) QQ-plot för exponentialfördelningen av andra halvan i program med låga TV-siffror.

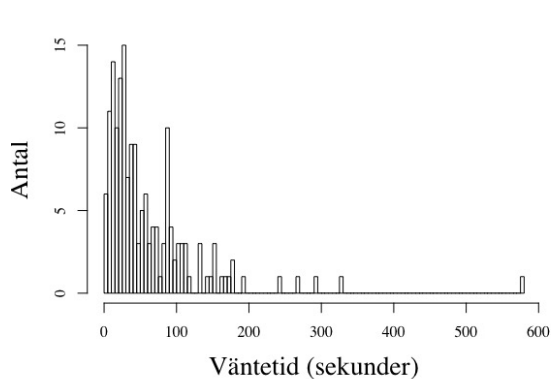
(b) QQ-plot för gammalfördelningen av andra halvan i program med låga TV-siffror.



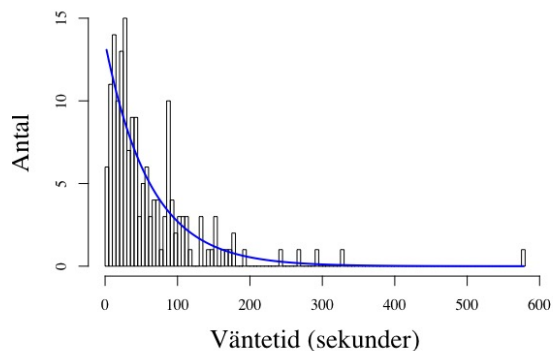
(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i andra halvan av program med låga TV-siffror.

Figur 31: QQ-plots för väntetiderna i andra halvan av program med låga TV-siffror ($\leq 800\ 000$). Figur 31a och 31b visar de teoretiska kvantilerna för exponential- respektive gammalfördelningen plottade mot stickprovets kvantiler. Figur 31c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

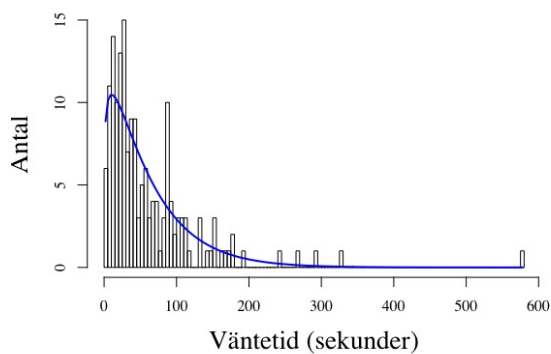
G.11 Första halvan av program med höga onlinesiffror



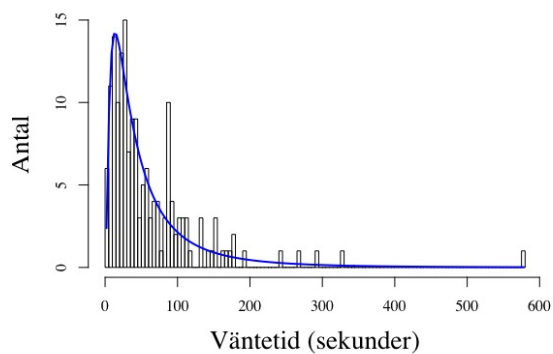
(a) Histogram över väntetiderna i första halvan för program med höga onlinesiffror.



(b) Histogram över väntetiderna i första halvan för program med höga onlinesiffror med en anpassad exponentialfördelning.

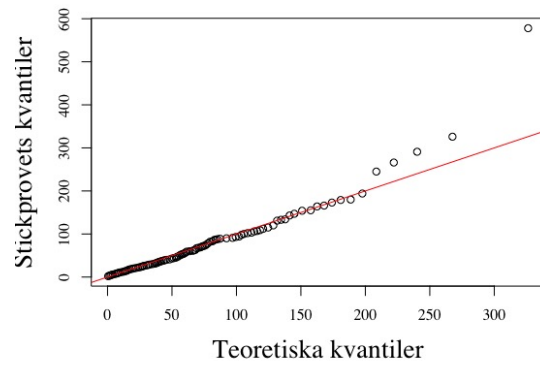
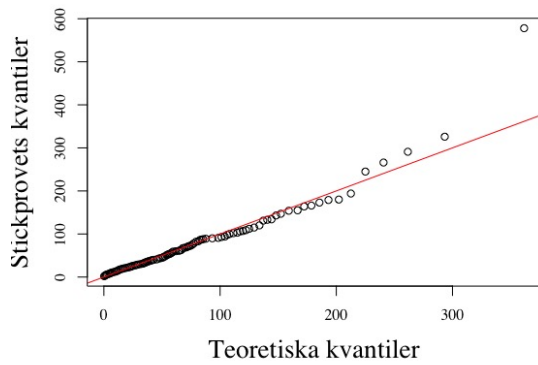


(c) Histogram över väntetiderna i första halvan för program med höga onlinesiffror med en anpassad gammafördelning.



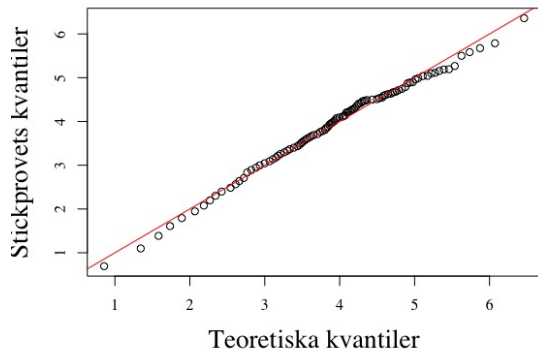
(d) Histogram över väntetiderna i första halvan för program med höga onlinesiffror med en anpassad lognormalfördelning.

Figur 32: Histogram med och utan anpassad fördelning för väntetiderna i *första* halvan av program med *höga* onlineeffror ($> 80\ 000$).



(a) QQ-plot för exponentialfördelningen av första halvan i program med höga onlinesiffror.

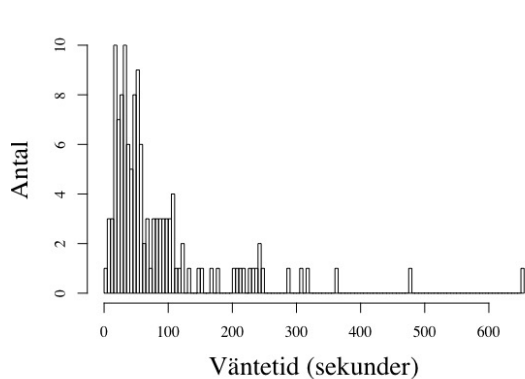
(b) QQ-plot för gammafördelningen av första halvan i program med höga onlinesiffror.



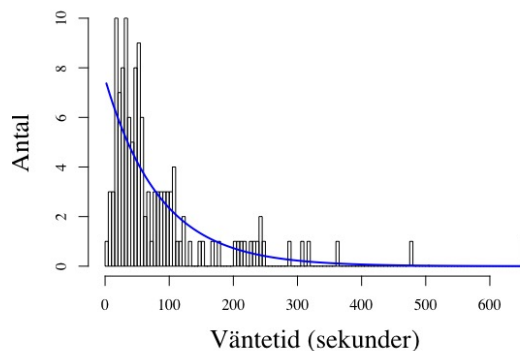
(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i första halvan av program med höga onlinesiffror.

Figur 33: QQ-plots för väntetiderna i första halvan av program med höga onlinesiffror ($> 80\,000$). Figur 33a och 33b visar de teoretiska kvantilerna för exponential- respektive gammafördelningen plottade mot stickprovets kvantiler. Figur 33c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

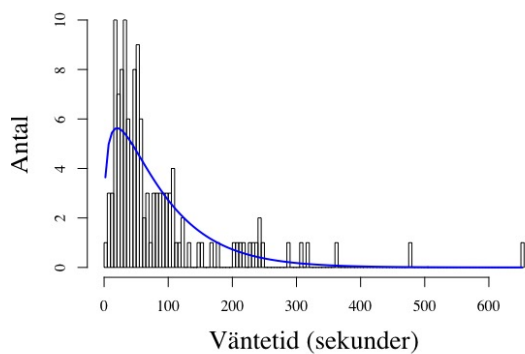
G.12 Andra halvan av program med höga onlinesiffror



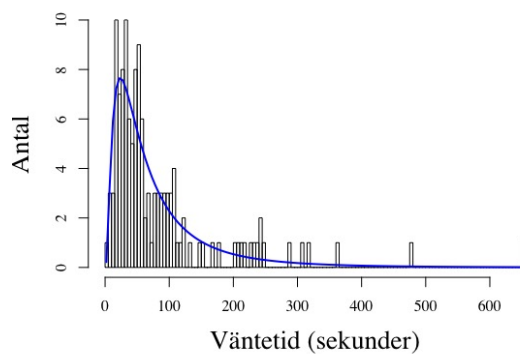
(a) Histogram över väntetiderna i andra halvan förprogram med höga onlinesiffror med en anpassad exponentialfördelning.



(b) Histogram över väntetiderna i andra halvan förprogram med höga onlinesiffror med en anpassad exponentialfördelning.

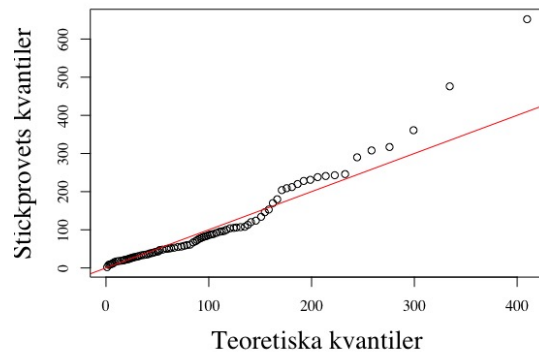
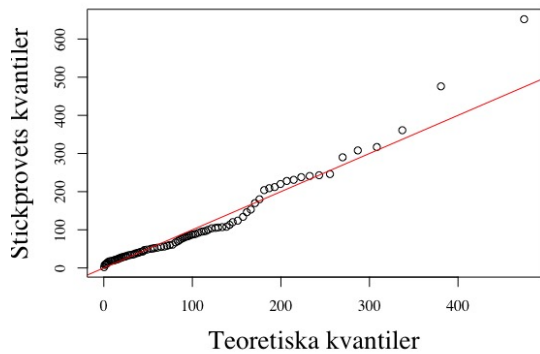


(c) Histogram över väntetiderna i andra halvan förprogram med höga onlinesiffror med en anpassad gammafördelning.



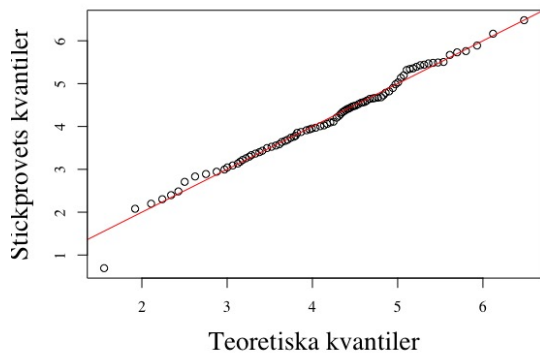
(d) Histogram över väntetiderna i andra halvan förprogram med höga onlinesiffror med en anpassad lognormalfördelning.

Figur 34: Histogram med och utan anpassad fördelning för väntetiderna i *andra* halvan av program med *höga* onlineiffror ($> 80\,000$).



(a) QQ-plot för exponentialfördelningen av andra halvan i program med höga onlinesiffror.

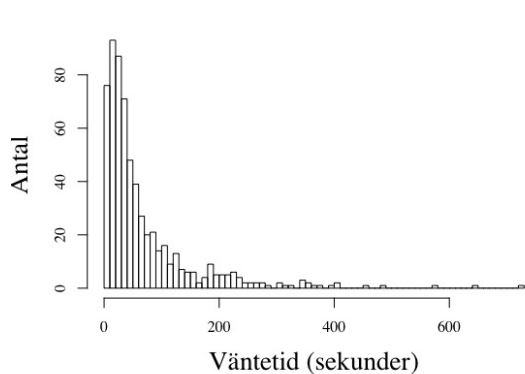
(b) QQ-plot för gammalfördelningen av andra halvan i program med höga onlinesiffror.



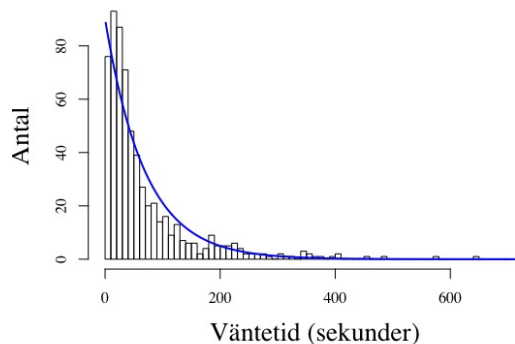
(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i andra halvan av program med höga onlinesiffror.

Figur 35: QQ-plots för väntetiderna i andra halvan av program med höga onlinesiffror ($> 80\,000$). Figur 35a och 35b visar de teoretiska kvantilerna för exponential- respektive gammalfördelningen plottade mot stickprovets kvantiler. Figur 35c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

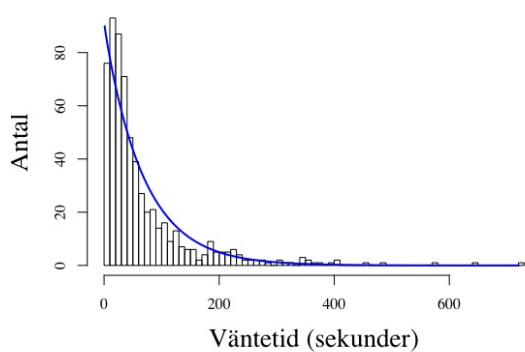
G.13 Första halvan av program med låga onlinesiffror



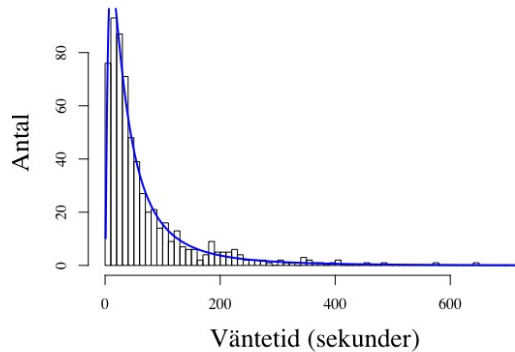
(a) Histogram över väntetiderna i första halvan för program med låga onlinesiffror.



(b) Histogram över väntetiderna i första halvan för program med låga onlinesiffror med en anpassad exponentialfördelning.

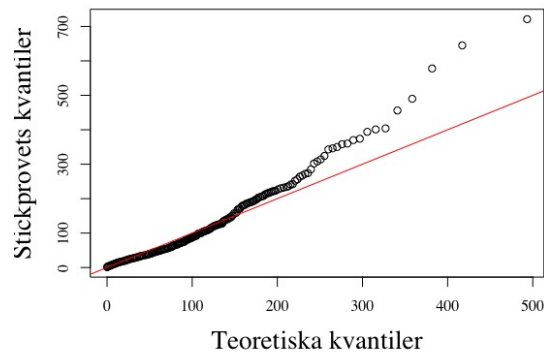
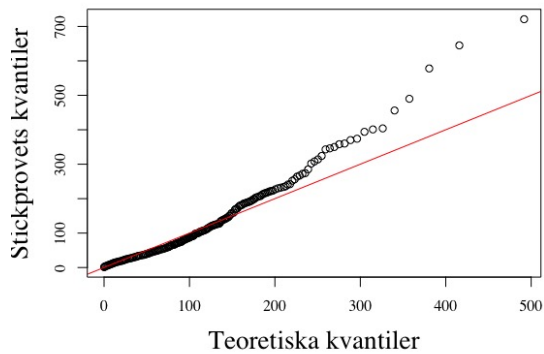


(c) Histogram över väntetiderna i första halvan för program med låga onlinesiffror med en anpassad gam-fördelning.

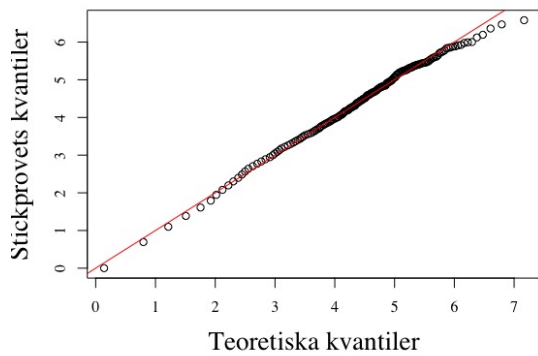


(d) Histogram över väntetiderna i första halvan för program med låga onlinesiffror med en anpassad lognormalfördelning.

Figur 36: Histogram med och utan anpassad fördelning för väntetiderna i första halvan av program med låga onlineiffror ($\leq 80\ 000$).



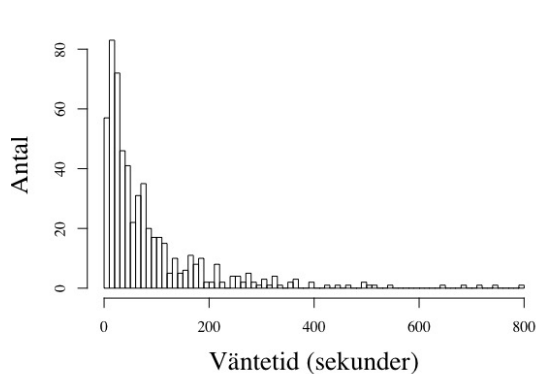
(a) QQ-plot för exponentialfördelningen av första halvan i program med låga onlinesiffror. (b) QQ-plot för gammalfördelningen av första halvan i program med låga onlinesiffror.



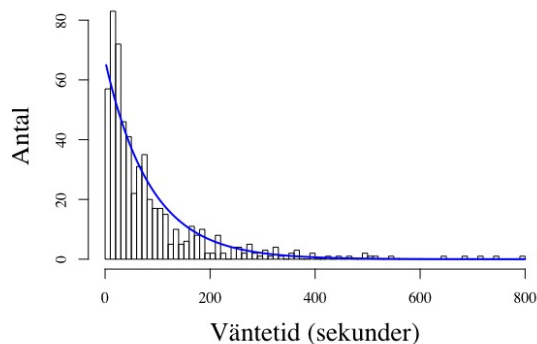
(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i första halvan av program med låga onlinesiffror.

Figur 37: QQ-plots för väntetiderna i första halvan av program med låga onlinesiffror ($\leq 80\,000$). Figur 37a och 37b visar de teoretiska kvantilerna för exponential- respektive gammalfördelningen plottade mot stickprovets kvantiler. Figur 37c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

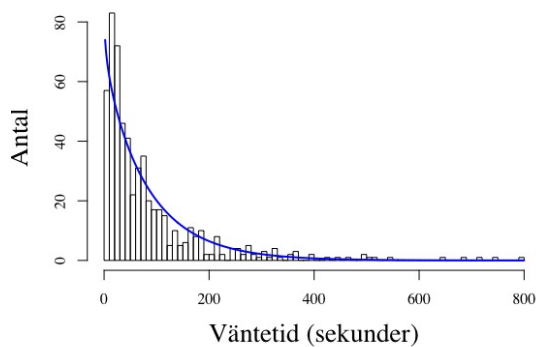
G.14 Andra halvan av program med låga onlinesiffror



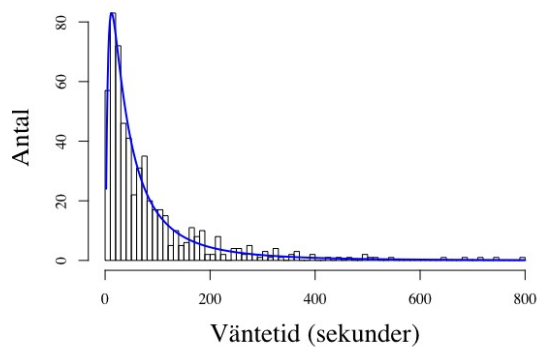
(a) Histogram över väntetiderna i andra halvan för program med låga onlinesiffror.



(b) Histogram över väntetiderna i andra halvan för program med låga onlinesiffror med en anpassad exponentialfördelning.

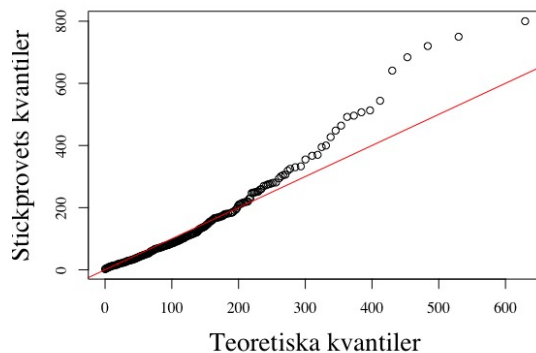
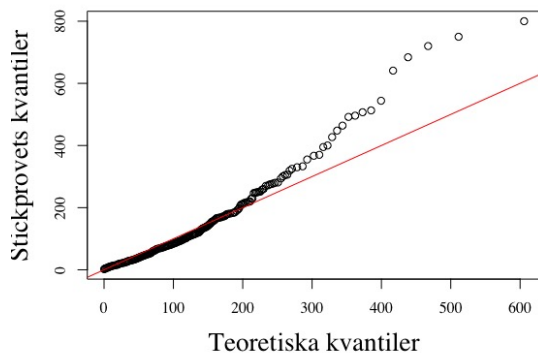


(c) Histogram över väntetiderna i andra halvan för program med låga onlinesiffror med en anpassad gam-mafördelning.

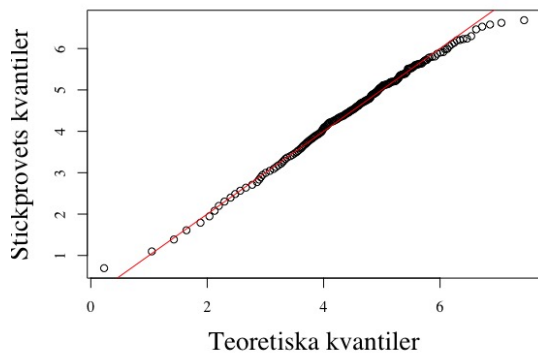


(d) Histogram över väntetiderna i andra halvan för program med låga onlinesiffror med en anpassad lognormalfördelning.

Figur 38: Histogram med och utan anpassad fördelning för väntetiderna i *andra* halvan av program med *låga* onlineiffror ($\leq 80\ 000$).



(a) QQ-plot för exponentialfördelningen av andra halvan i program med låga onlinesiffror. (b) QQ-plot för gammalfördelningen av andra halvan i program med låga onlinesiffror.

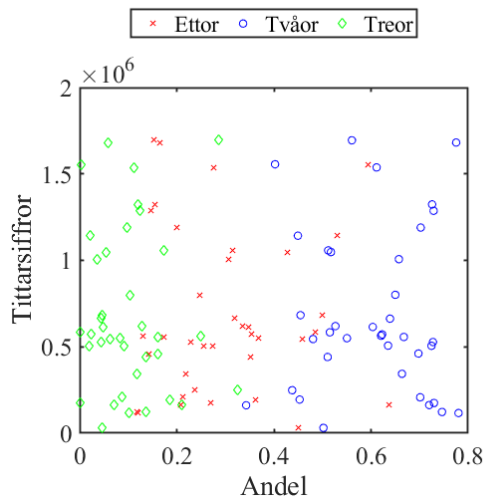


(c) QQ-plot för normalfördelningen för logaritmen av väntetiderna i andra halvan av program med låga onlinesiffror.

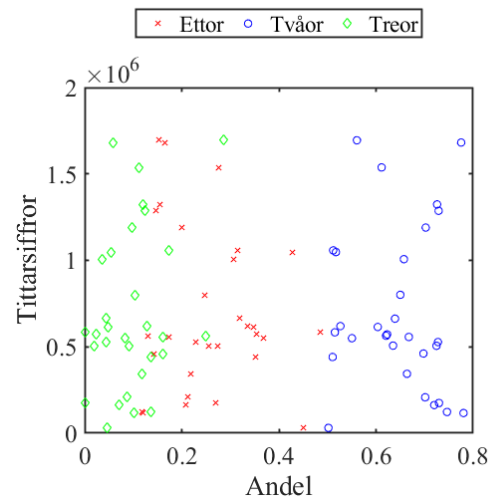
Figur 39: QQ-plots för väntetiderna i andra halvan av program med låga onlinesiffror ($\leq 80\,000$). Figur 39a och 39b visar de teoretiska kvantilerna för exponential- respektive gammalfördelningen plottade mot stickprovets kvantiler. Figur 39c illustrerar normalfördelningskvantilerna mot stickprovskvantilerna för den logaritmerade datan.

G.15 Plottad data för andel tid spenderad i olika stadier

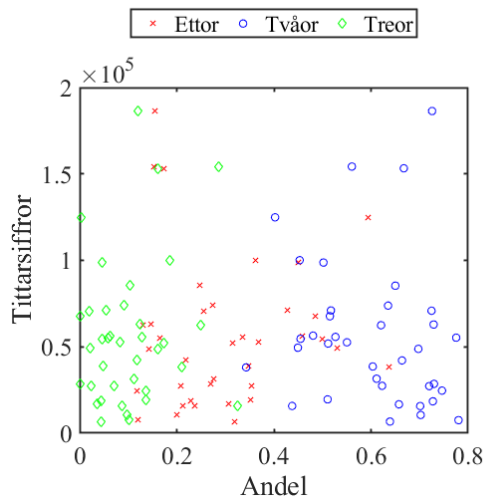
Här presenteras den samling plottar som ligger till grund för den linjära anpassningen i avsnitt 5.



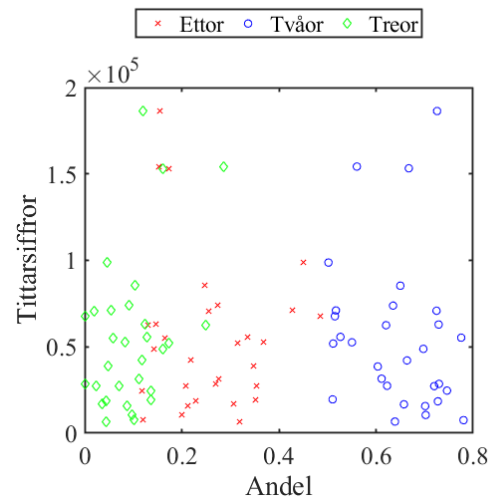
Figur 40: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt ett på broadcast**. Detta för *alla* analyserade program.



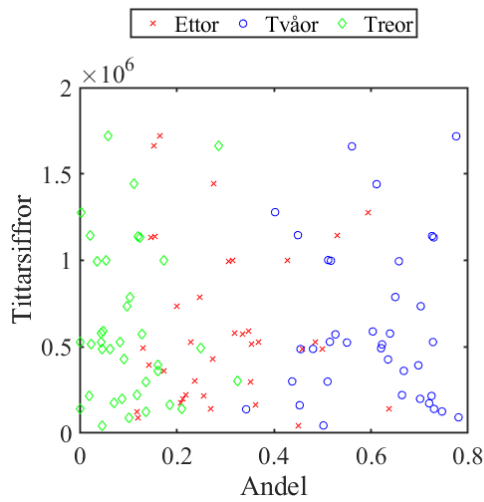
Figur 41: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt ett på broadcast**. Detta för alla analyserade program *förutom* de sju med utstickande karaktär.



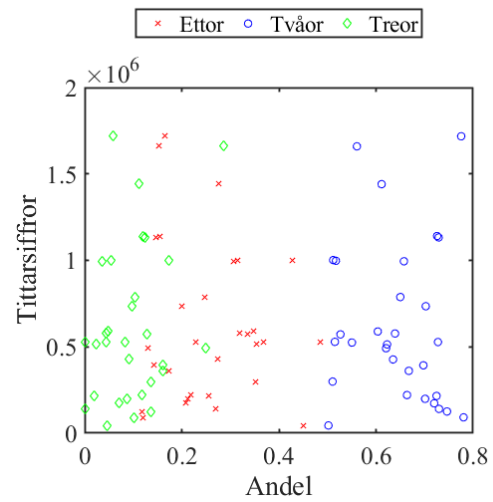
Figur 42: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt ett online**. Detta för *alla* analyserade program.



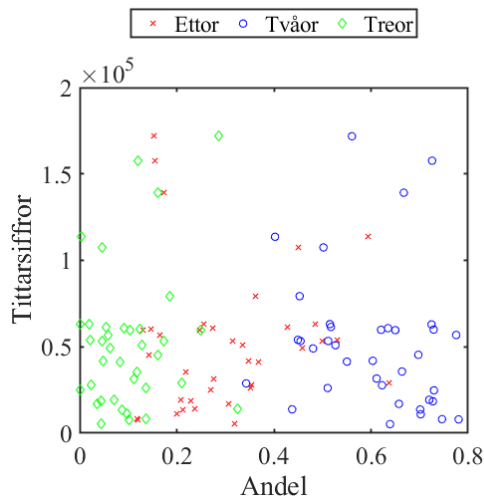
Figur 43: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt ett online**. Detta för alla analyserade program *förutom* de sju med utstickande karaktär.



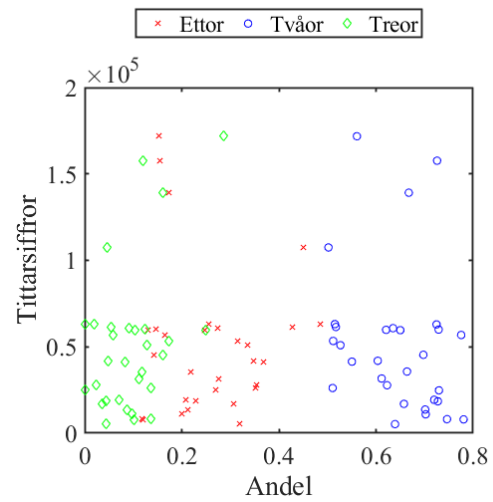
Figur 44: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt två på broadcast**. Detta för *alla* analyserade program.



Figur 45: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt två på broadcast**. Detta för alla analyserade program *förutom* de sju med utstickande karaktär.

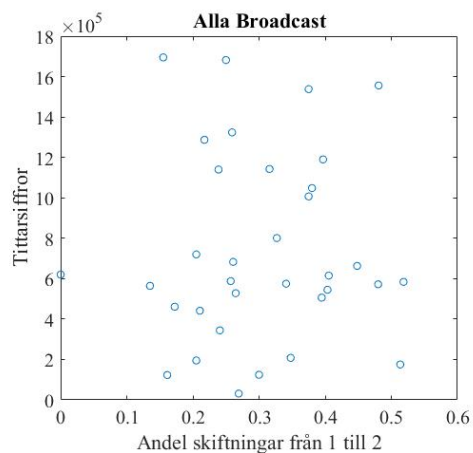


Figur 46: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt två online**. Detta för *alla* analyserade program.

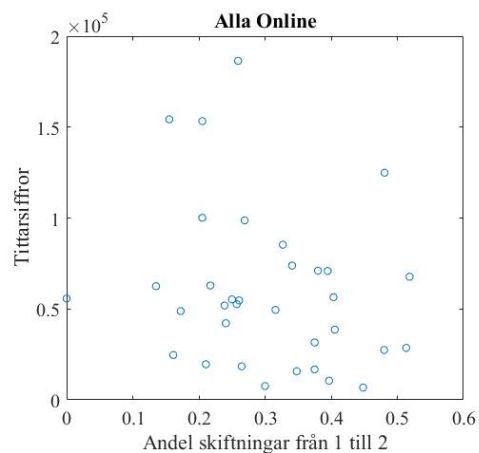


Figur 47: Plot av andel ettor, tvåor respektive treor mot tittarsiffror för **avsnitt två online**. Detta för alla analyserade program *förutom* de sju med utstickande karaktär.

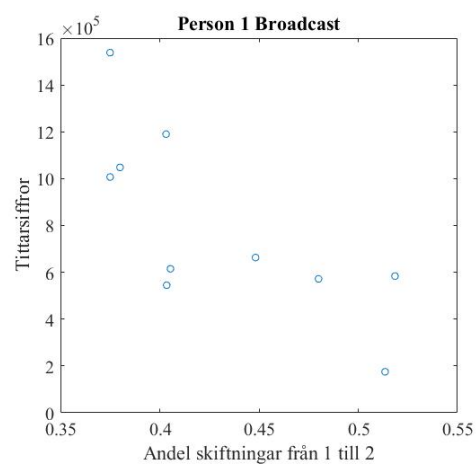
G.16 Plottad data för skiftningar mellan stadier



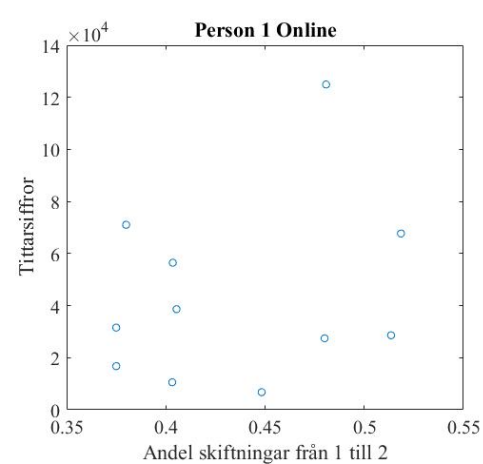
Figur 48: Plot av skiftningarna från stadie 1 till 2 mot broadcast-tittarsiffror.



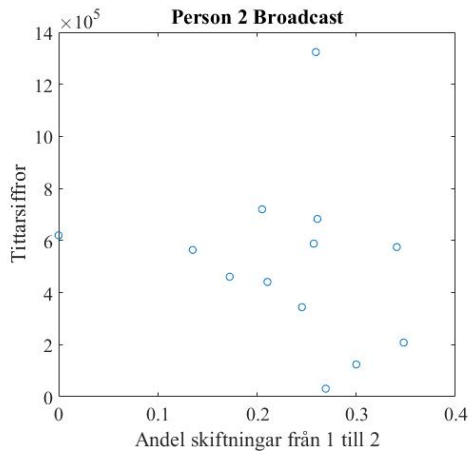
Figur 49: Plot av skiftningarna från stadie 1 till 2 mot online-tittarsiffror.



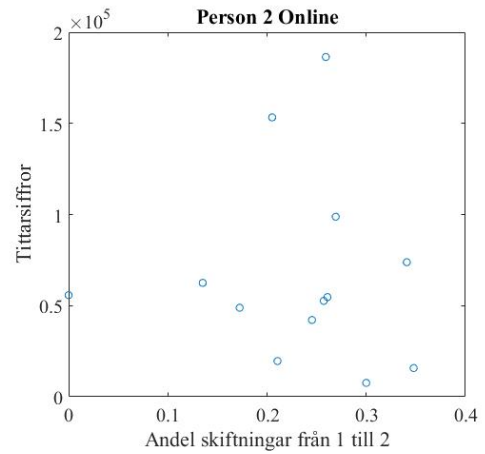
Figur 50: Plot av skiftningarna från stadie 1 till 2 mot broadcast-tittarsiffror.



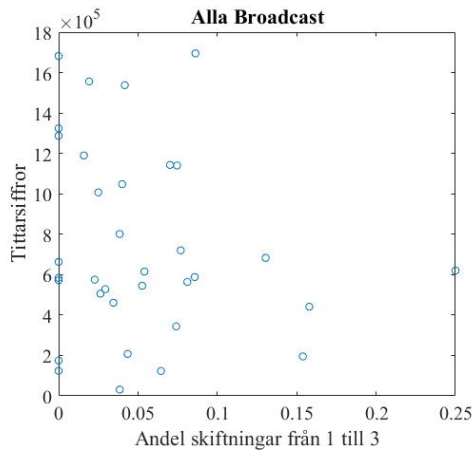
Figur 51: Plot av skiftningarna från stadie 1 till 2 mot online-tittarsiffror.



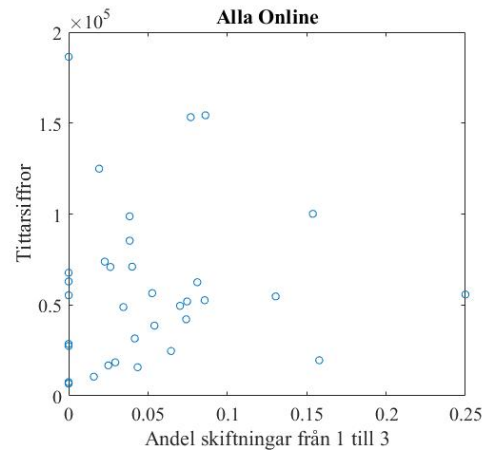
Figur 52: Plot av skiftningarna från stadie 1 till 2 mot broadcast-tittarsiffror.



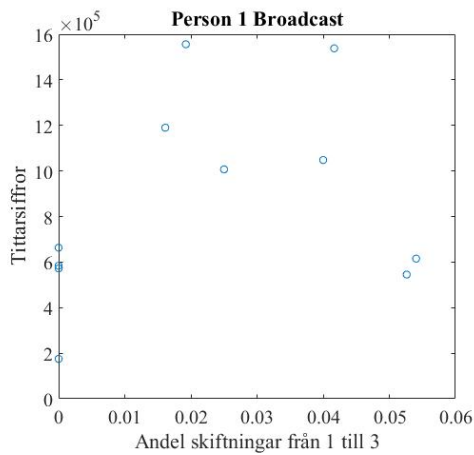
Figur 53: Plot av skiftningarna från stadie 1 till 2 mot online-tittarsiffror.



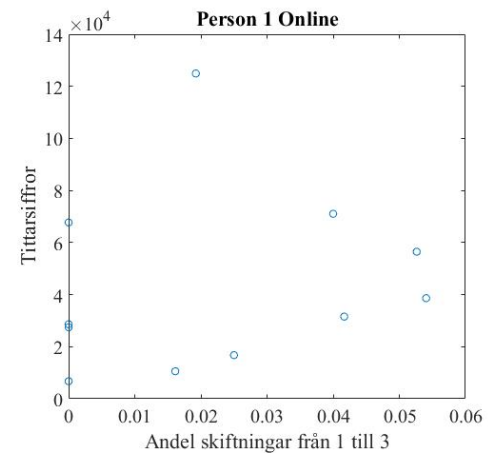
Figur 54: Plot av skiftningarna från stadie 1 till 3 mot broadcast-tittarsiffror.



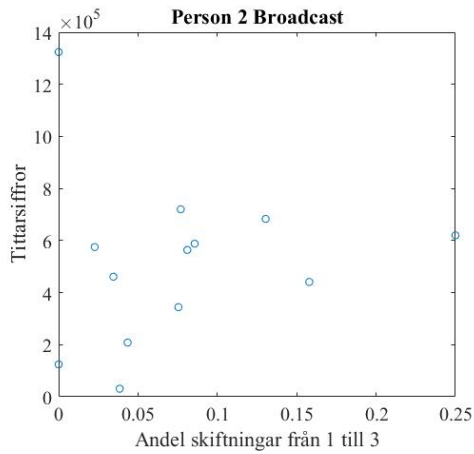
Figur 55: Plot av skiftningarna från stadie 1 till 3 mot online-tittarsiffror.



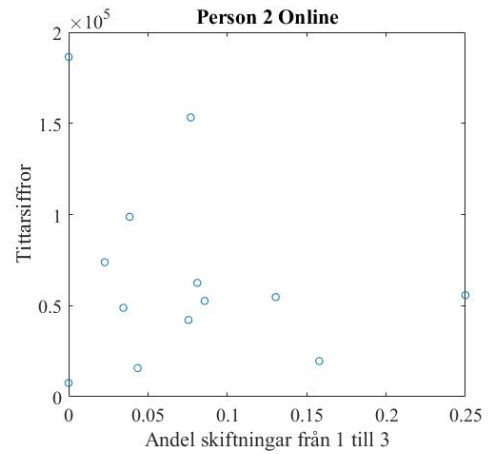
Figur 56: Plot av skiftningarna från stadie 1 till 3 mot broadcast-tittarsiffror.



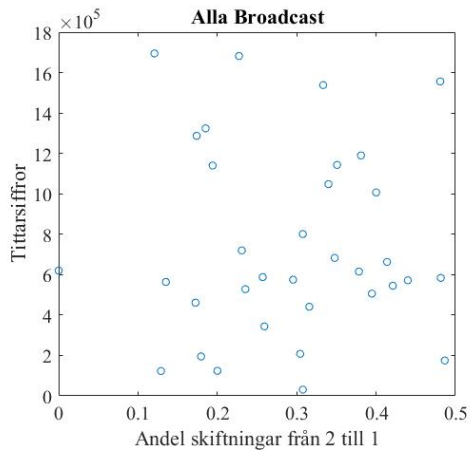
Figur 57: Plot av skiftningarna från stadie 1 till 3 mot online-tittarsiffror.



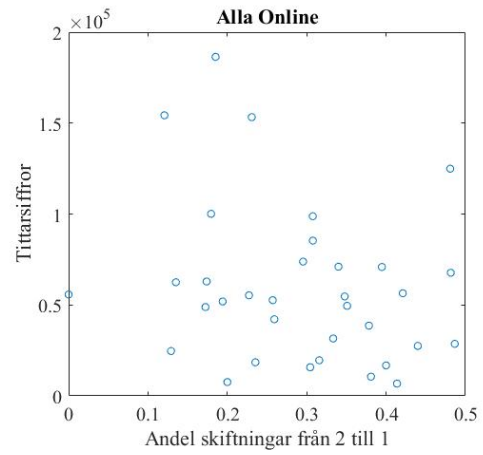
Figur 58: Plot av skiftningarna från stadie 1 till 3 mot broadcast-tittarsiffror.



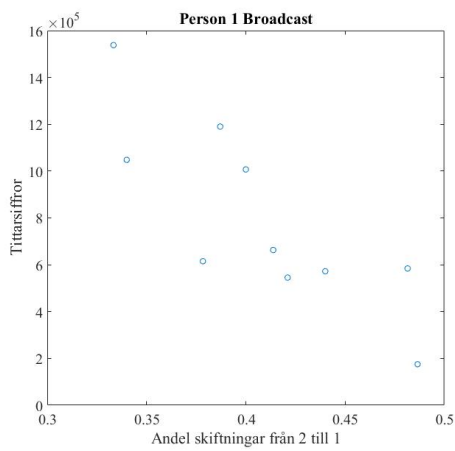
Figur 59: Plot av skiftningarna från stadie 1 till 3 mot online-tittarsiffror.



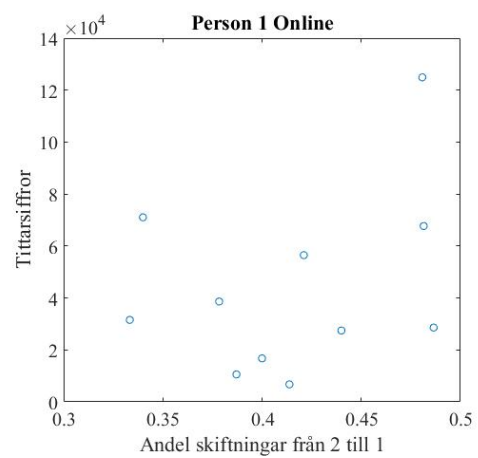
Figur 60: Plot av skiftningarna från stadie 2 till 1 mot broadcast-tittarsiffror.



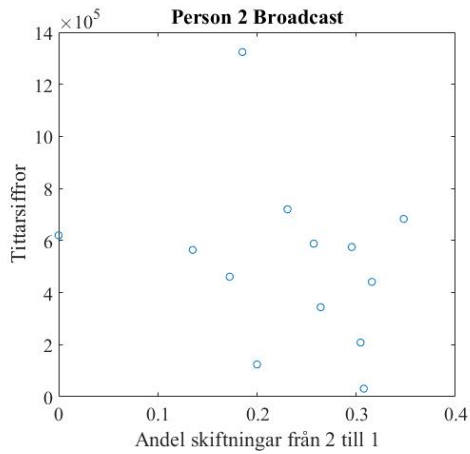
Figur 61: Plot av skiftningarna från stadie 2 till 1 mot online-tittarsiffror.



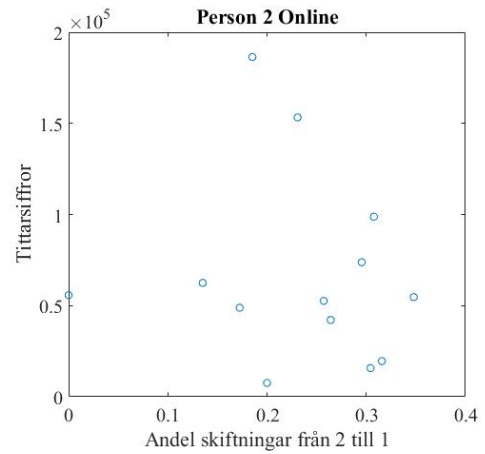
Figur 62: Plot av skiftningarna från stadie 2 till 1 mot broadcast-tittarsiffror.



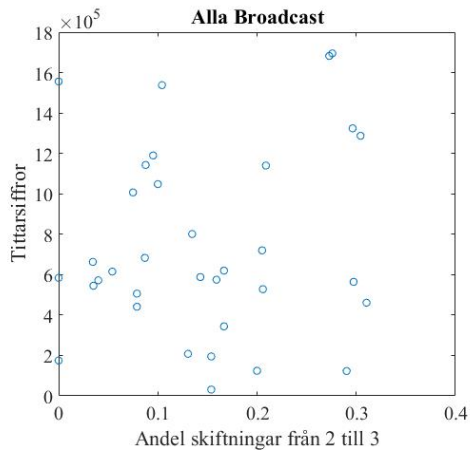
Figur 63: Plot av skiftningarna från stadie 2 till 1 mot online-tittarsiffror.



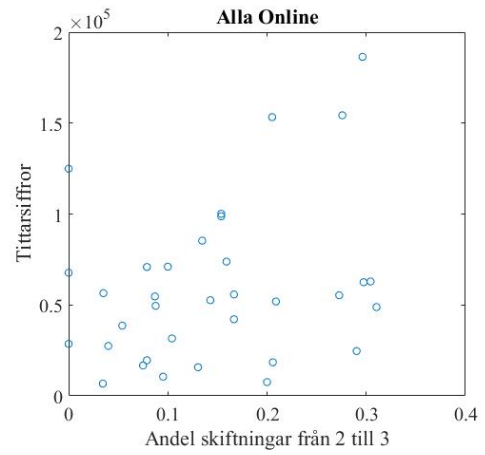
Figur 64: Plot av skiftningarna från stadie 2 till 1 mot broadcast-tittarsiffror.



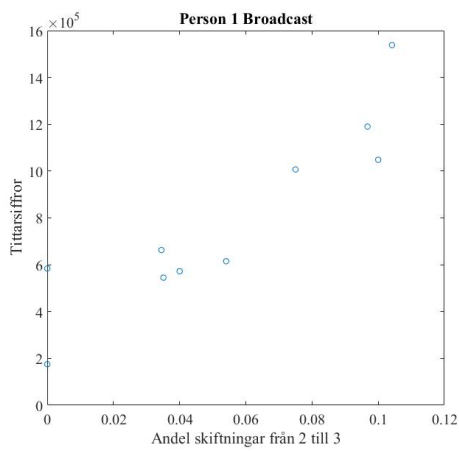
Figur 65: Plot av skiftningarna från stadie 2 till 1 mot online-tittarsiffror.



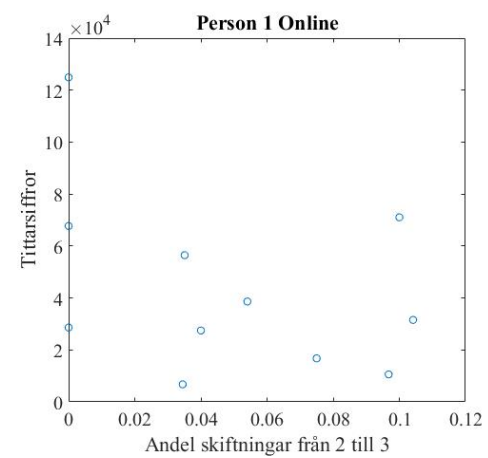
Figur 66: Plot av skiftningarna från stadie 2 till 3 mot broadcast-tittarsiffror.



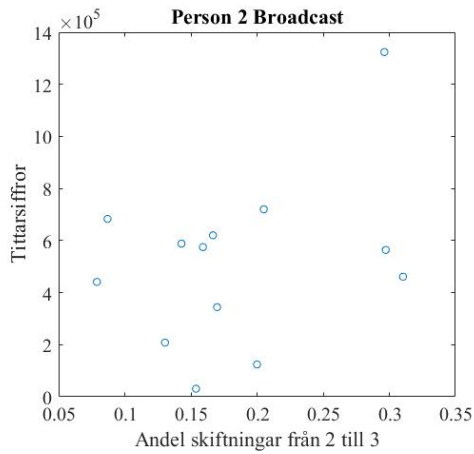
Figur 67: Plot av skiftningarna från stadie 2 till 3 mot online-tittarsiffror.



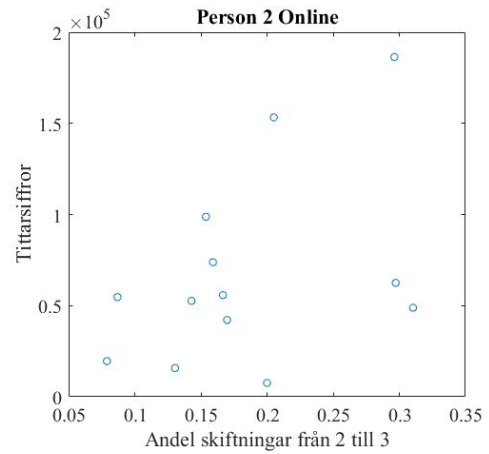
Figur 68: Plot av skiftningarna från stadie 2 till 3 mot broadcast-tittarsiffror.



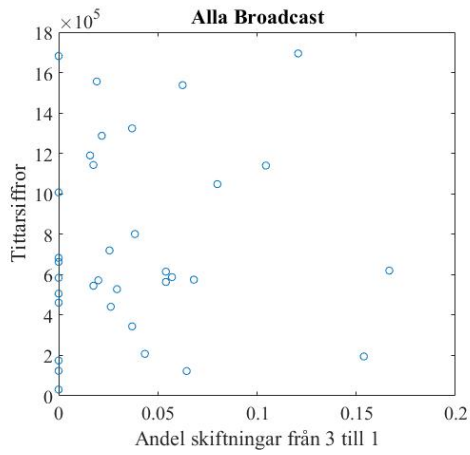
Figur 69: Plot av skiftningarna från stadie 2 till 3 mot online-tittarsiffror.



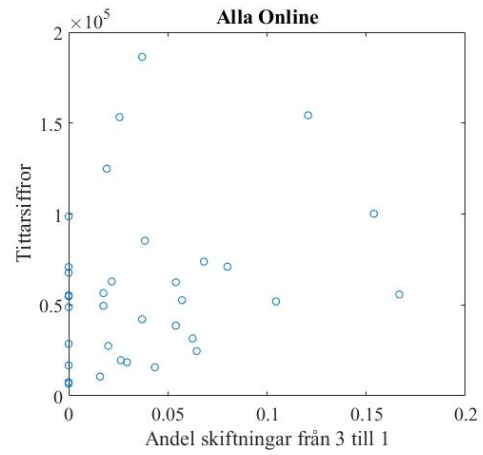
Figur 70: Plot av skiftningarna från stadie 2 till 3 mot broadcast-tittarsiffror.



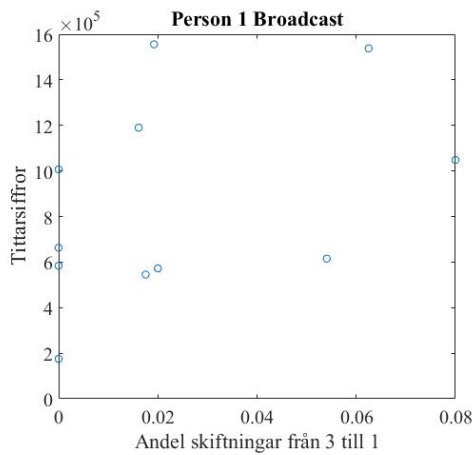
Figur 71: Plot av skiftningarna från stadie 2 till 3 mot online-tittarsiffror.



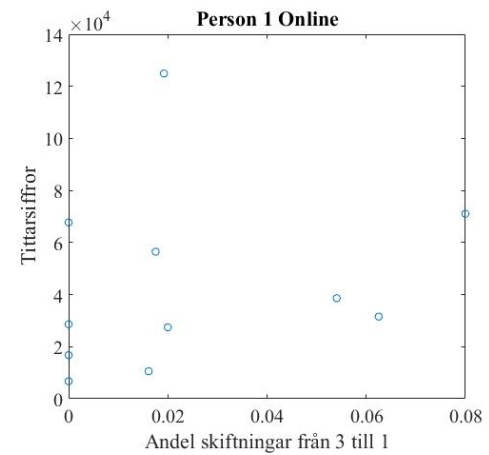
Figur 72: Plot av skiftningarna från stadie 3 till 1 mot broadcast-tittarsiffror.



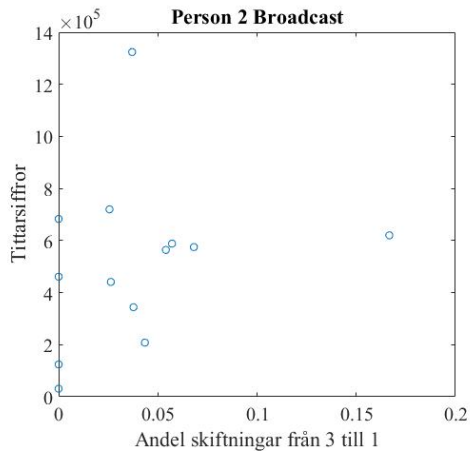
Figur 73: Plot av skiftningarna från stadie 3 till 1 mot online-tittarsiffror.



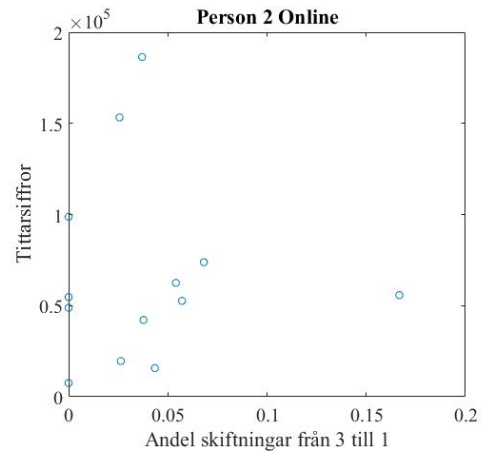
Figur 74: Plot av skiftningarna från stadie 3 till 1 mot broadcast-tittarsiffror.



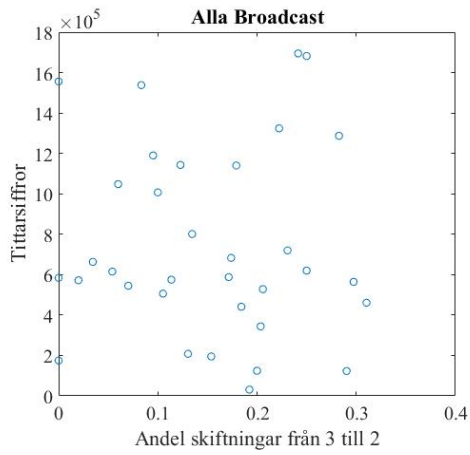
Figur 75: Plot av skiftningarna från stadie 3 till 1 mot online-tittarsiffror.



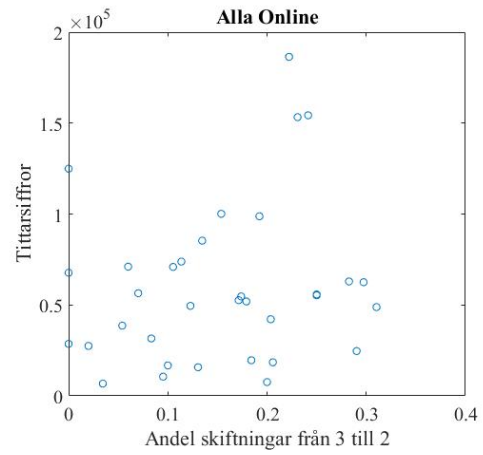
Figur 76: Plot av skiftningarna från stadie 3 till 1 mot broadcast-tittarsiffror.



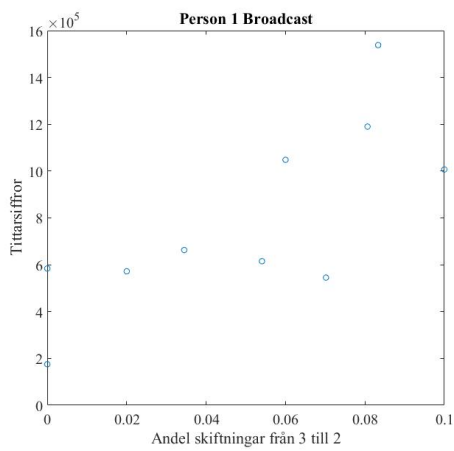
Figur 77: Plot av skiftningarna från stadie 3 till 1 mot online-tittarsiffror.



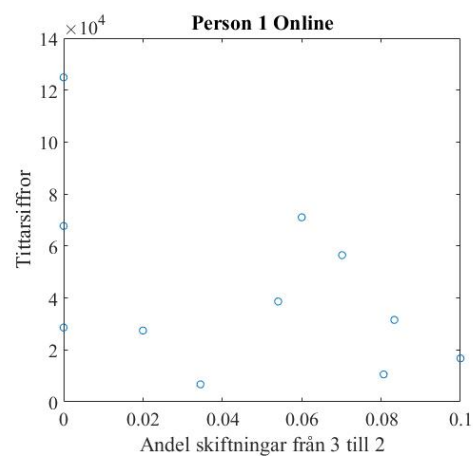
Figur 78: Plot av skiftningarna från stadie 3 till 2 mot broadcast-tittarsiffror.



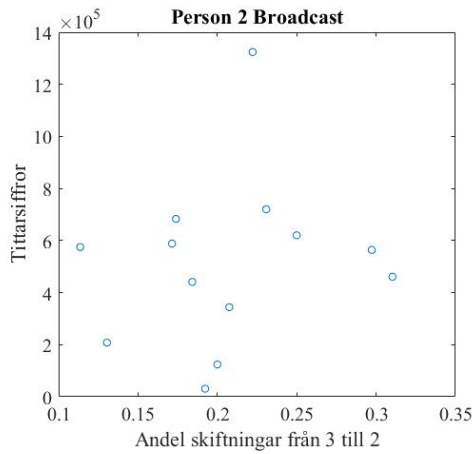
Figur 79: Plot av skiftningarna från stadie 3 till 2 mot online-tittarsiffror.



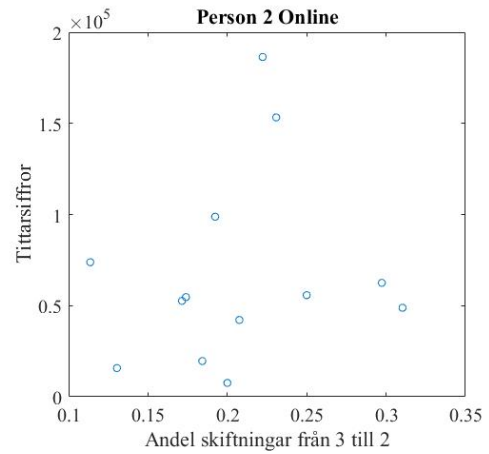
Figur 80: Plot av skiftningarna från stadie 3 till 2 mot broadcast-tittarsiffror.



Figur 81: Plot av skiftningarna från stadie 3 till 2 mot online-tittarsiffror.

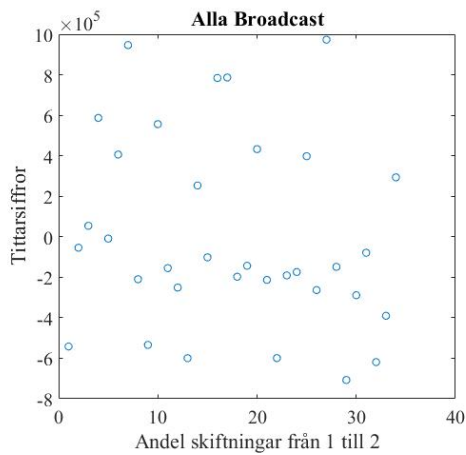


Figur 82: Plot av skiftningarna från stadi 3 till 2 mot broadcast-tittarsiffror.

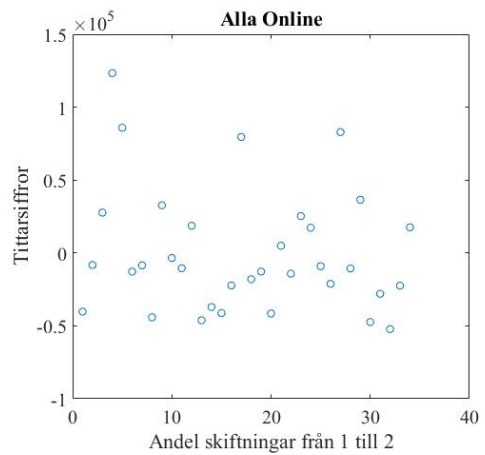


Figur 83: Plot av skiftningarna från stadi 3 till 2 mot online-tittarsiffror.

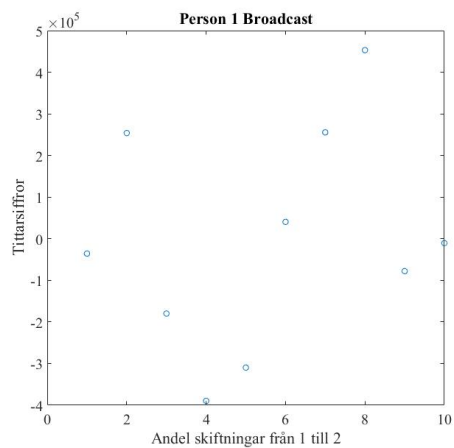
G.17 Residualplottar för de anpassade linjerna för skiftningarna



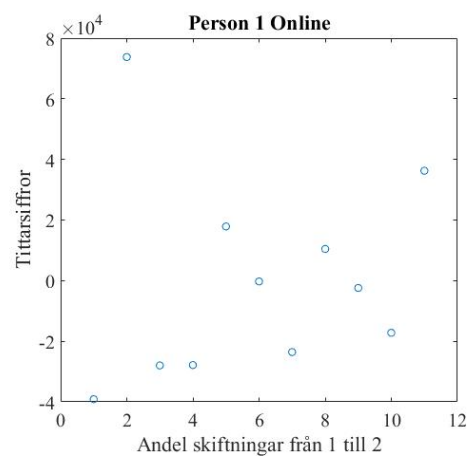
Figur 84: Residualplot för linjen anpassad till data från skiftningarna från stadi 1 till 2 och broadcast-tittarsiffror.



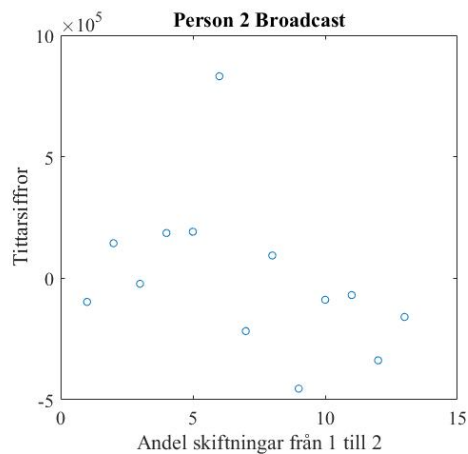
Figur 85: Residualplot för linjen anpassad till data från skiftningarna från stadi 1 till 2 och online-tittarsiffror.



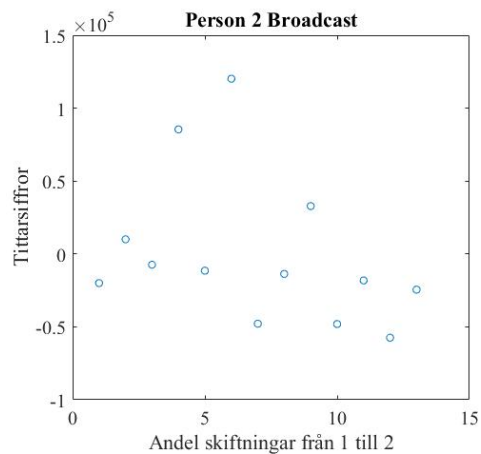
Figur 86: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 2 och broadcast-tittarsiffror.



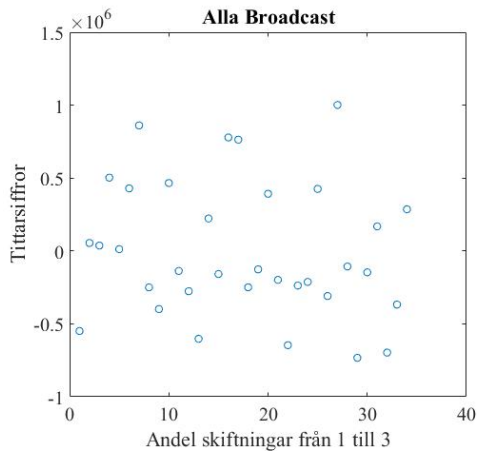
Figur 87: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 2 och online-tittarsiffror.



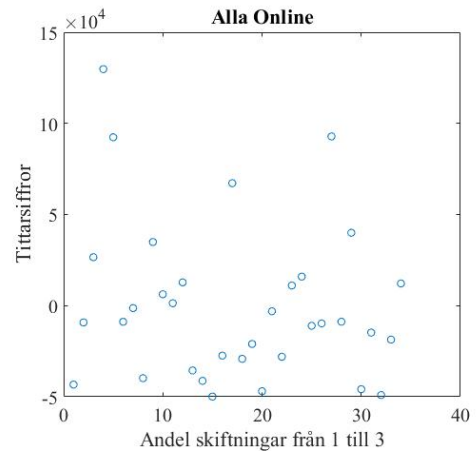
Figur 88: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 2 och broadcast-tittarsiffror.



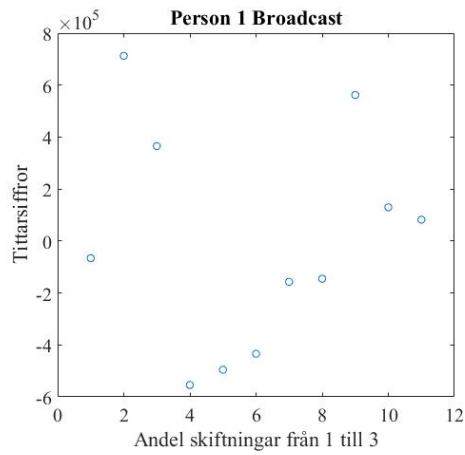
Figur 89: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 2 och online-tittarsiffror.



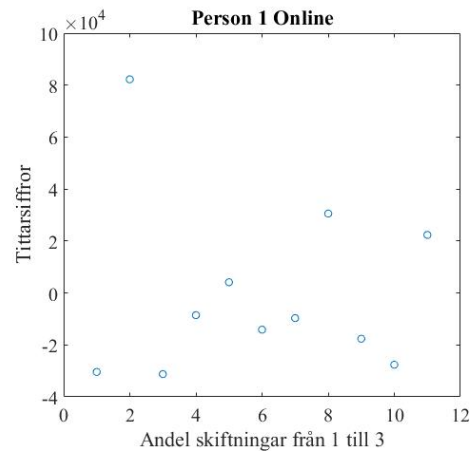
Figur 90: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 3 och broadcast-tittarsiffror.



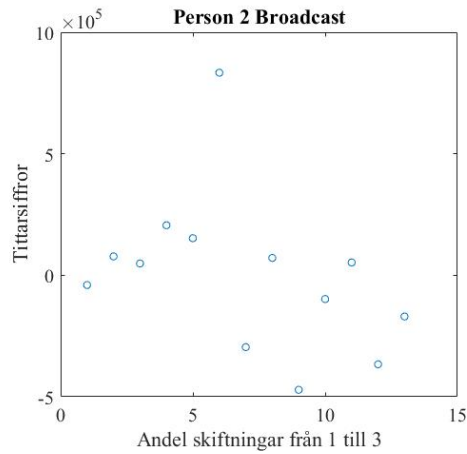
Figur 91: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 3 och online-tittarsiffror.



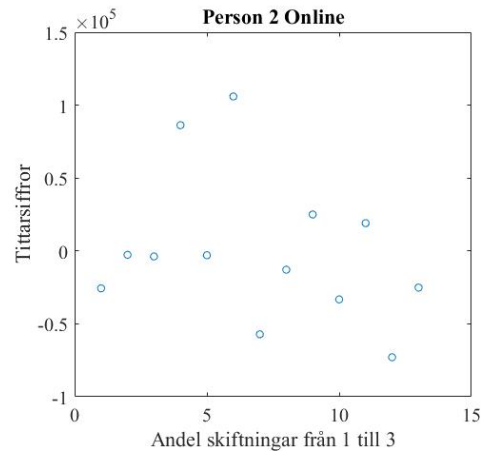
Figur 92: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 3 och broadcast-tittarsiffror.



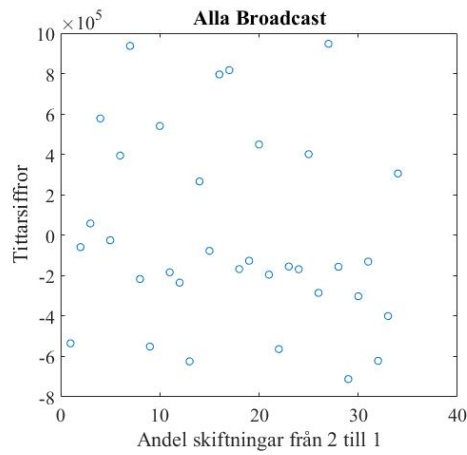
Figur 93: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 3 och online-tittarsiffror.



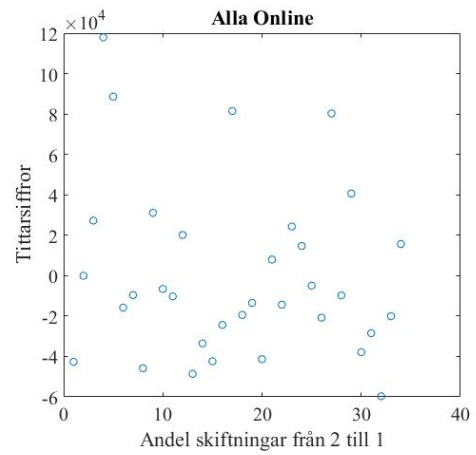
Figur 94: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 3 och broadcast-tittarsiffror.



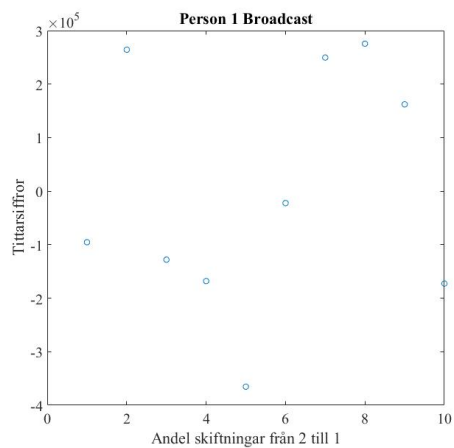
Figur 95: Residualplot för linjen anpassad till data från skiftningarna från stadie 1 till 3 och online-tittarsiffror.



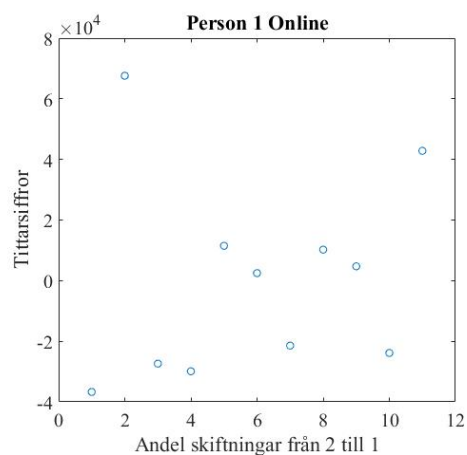
Figur 96: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 1 och broadcast-tittarsiffror.



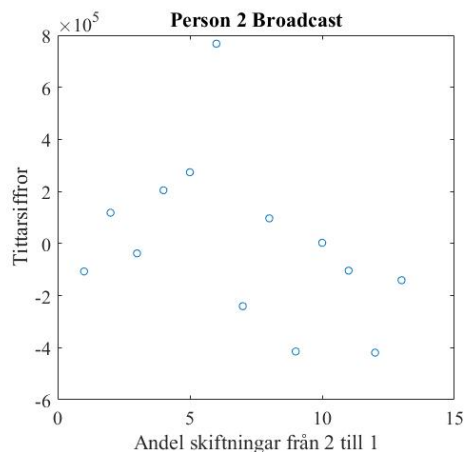
Figur 97: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 1 och online-tittarsiffror.



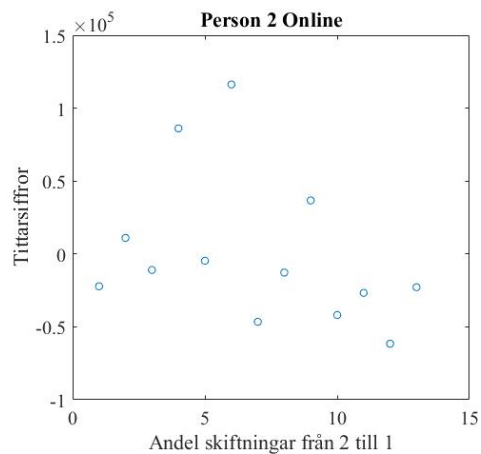
Figur 98: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 1 och broadcast-tittarsiffror.



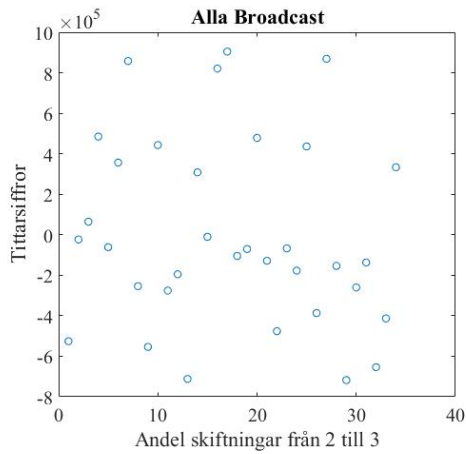
Figur 99: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 1 och online-tittarsiffror.



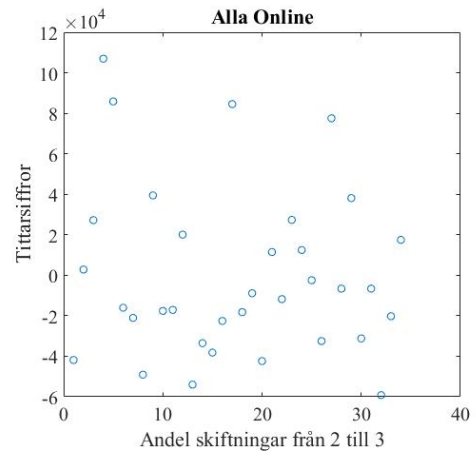
Figur 100: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 1 och broadcast-tittarsiffror.



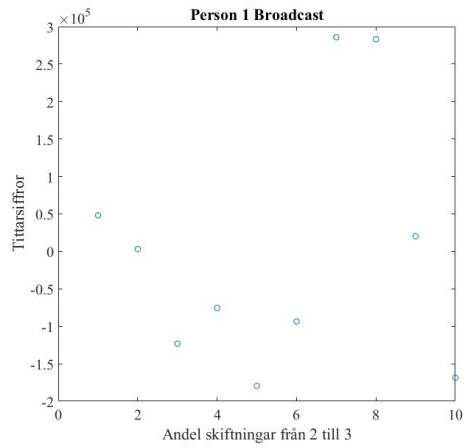
Figur 101: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 1 och online-tittarsiffror.



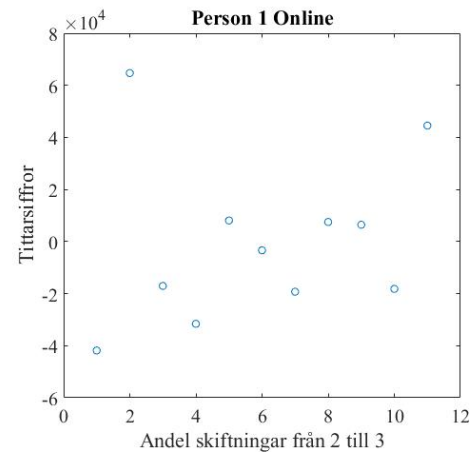
Figur 102: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 3 och broadcast-tittarsiffror.



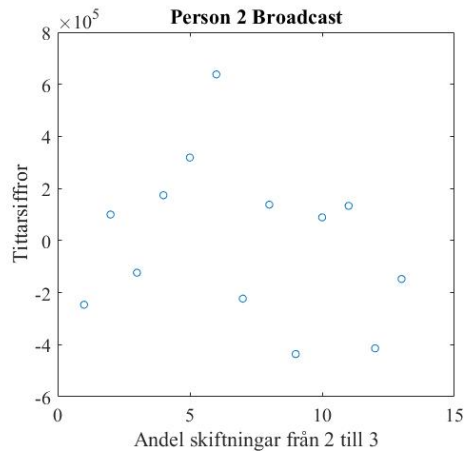
Figur 103: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 3 och online-tittarsiffror.



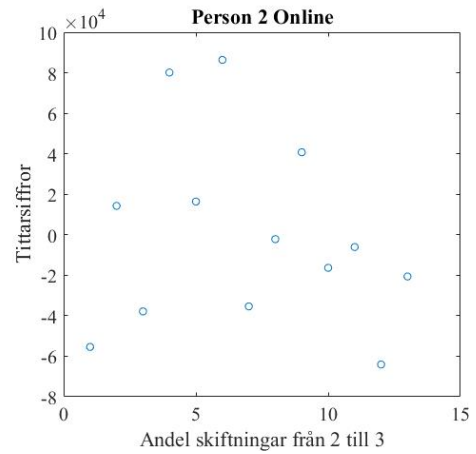
Figur 104: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 3 och broadcast-tittarsiffror.



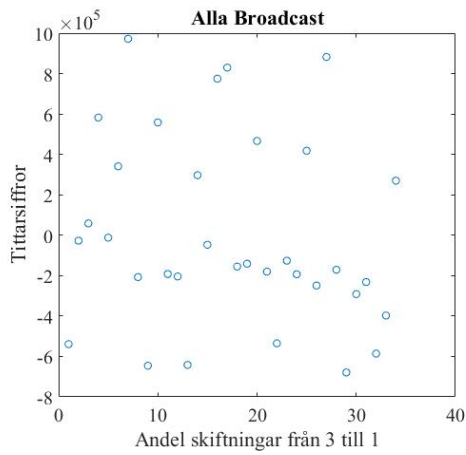
Figur 105: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 3 och online-tittarsiffror.



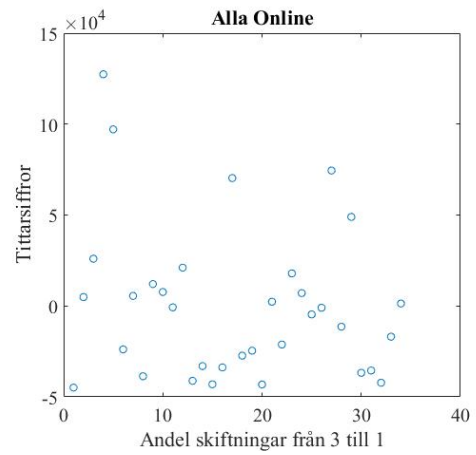
Figur 106: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 3 och broadcast-tittarsiffror.



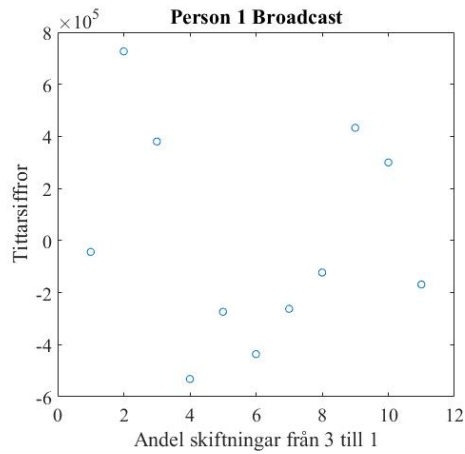
Figur 107: Residualplot för linjen anpassad till data från skiftningarna från stadie 2 till 3 och online-tittarsiffror.



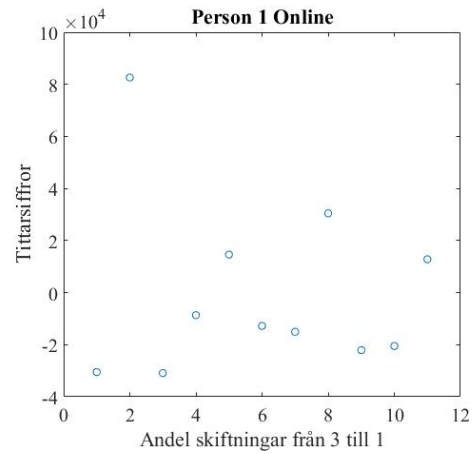
Figur 108: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 1 och broadcast-tittarsiffror.



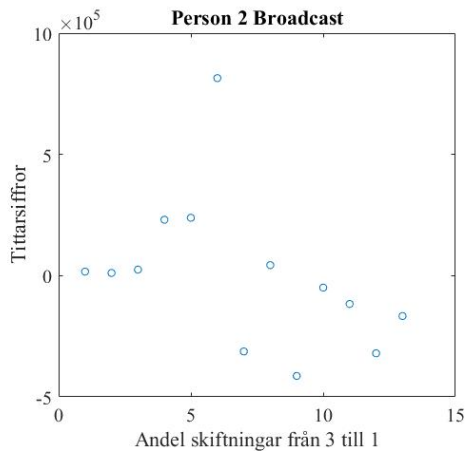
Figur 109: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 1 och online-tittarsiffror.



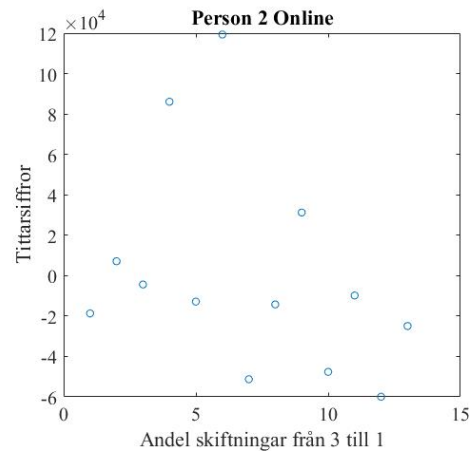
Figur 110: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 1 och broadcast-tittarsiffror.



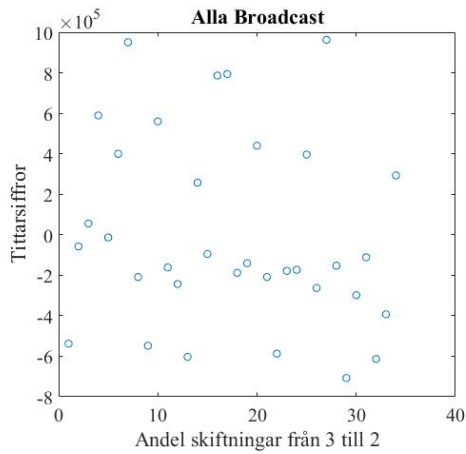
Figur 111: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 1 och online-tittarsiffror.



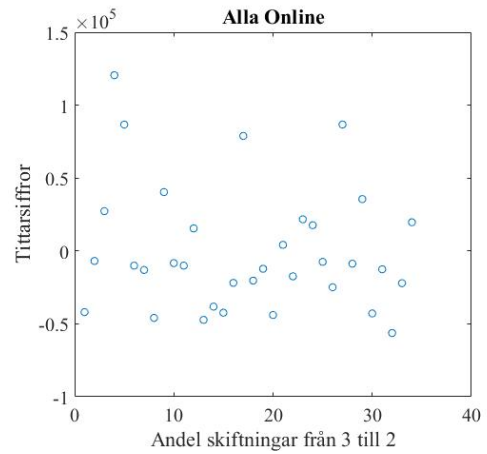
Figur 112: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 1 och broadcast-tittarsiffror.



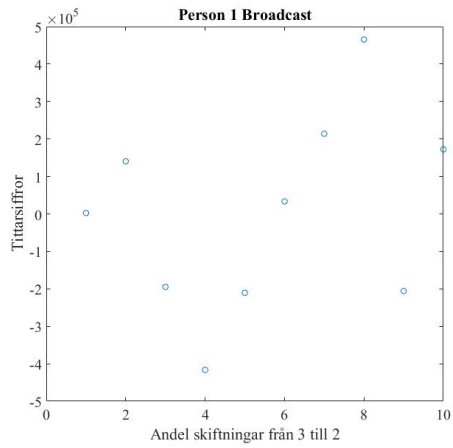
Figur 113: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 1 och online-tittarsiffror.



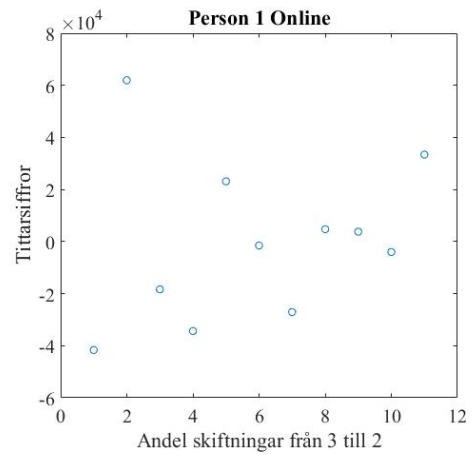
Figur 114: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 2 och broadcast-tittarsiffror.



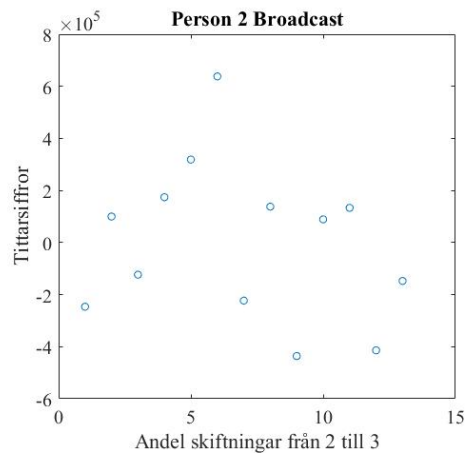
Figur 115: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 2 och online-tittarsiffror.



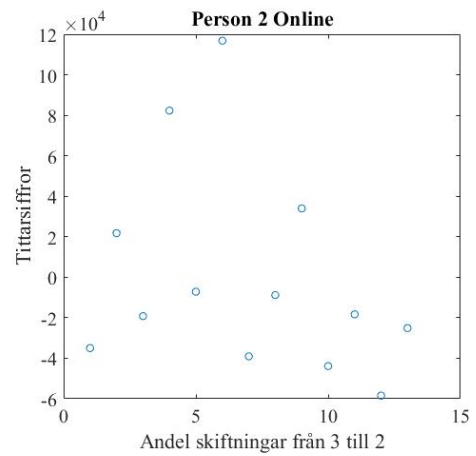
Figur 116: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 2 och broadcast-tittarsiffror.



Figur 117: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 2 och online-tittarsiffror.



Figur 118: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 2 och broadcast-tittarsiffror.



Figur 119: Residualplot för linjen anpassad till data från skiftningarna från stadie 3 till 2 och online-tittarsiffror.