



Noregs miljø- og  
biovitenskaplege  
universitet

**Masteroppgåve 2020 60 stp**

Fakultetet for kjemi, bioteknologi og matvitenskap

## **Kva innsikt kan biometriske målingar gje om studentar si oppleving av statistikkundervisning?**

Can biometric measurements give insight into students' experience of statistics education?

Janne Cathrin Hetle Aspheim

Master i Anvendt Statistikk

## FORORD

---

Når eg i utgangspunktet berre skulle ta eit årsstudium i landmåling, kjennast det verkelegheitsfjernt å skrive forordet til ei masteroppgåve i anvendt statistikk. Når eg først skal skrive eit forord, vil eg nytte høvet til å rette ein takk og ei merksemd mot dei som har bidrege, ikkje berre til sjølve oppgåva, men og på vegen dit. Først respondentane, som har delt av si tid, sine meiningar og sine kjensler. Utan respondentar hadde oppgåva riktignok vore mykje kortare, men ikkje på langt nær så spanande. Vidare må ein ha utstyr og lokale til å utføre forsøka, og Vitenparken har bidrege med begge deler. Når ein først har tilgjengeleg utstyr, er det lite verdt om ein ikkje kan å nytte det. Teknikar hjå Vitenparken, Helge Mathisen har vore til uvurderleg hjelp med opplæring med utstyr og programvare, og med oppbygging av forsøket i iMotions i tillegg til redigering av undervisningsfilmane.

Til mine veiledarar, Hilde Vinje og Kathrine Frey Frøslie, som har heldt meg i akkurat passe stramme taumar, og samstundes gjeitt meg fridomen til å eige prosjektet fullt ut. Sjølv om dykkar kunnskapar har vore til uvurderleg hjelp er det ikkje kva dykk kan, men kven dykk er, som har hatt den største betydninga for meg. Høglytt kakling og latter frå kontoret til Kathrine har ofte vore ein indikasjon på at veiledningsmøte pågjekk, men det er ikkje berre det gode humøret som har hatt betydning for gleda av samarbeidet med Hilde og Kathrine. Det er alle samtalanane som har lært meg mykje om handsaming av forsøk, haldningar til vitskapeleg arbeide, og skilnaden mellom å utføre statistiske testar og å utøve statistikk som fag.

Til Nils Olav Frøysa. Med stort sannsyn er du verdas mest tolmodige sambuar. Då årsstudiet i landmåling skulle syne seg å berre vere starten på mange år med studie, tilpassa du ditt liv til mitt. Det var kun støtte og få når kveldar, helger og feriar i stor grad var forbeholdt skulearbeid. Og sjølv om du med jamne mellomrom vert utsett for tvangslesing av forskjellige oppgåver, kjem du alltid med konstruktive kommentarar som: «Dette må halde eit svært høgt akademisk nivå, eg kjenner eg får hovudverk etter å ha lest tre linjer».

Til dei mange som har bidrege på vegen til denne oppgåva, skal alle nemnast vert forordet lenger en resultatkapittelet. Men dykk veit kven dykk er. Dykk har bidrege med samtalar om «standardavvik og sånn», om forskning, om dykkar erfaringar frå dykkar fagfelt, om livet, og om viktigheita av god kaffi med litt vaspreik attpå. Dykk har bidrege med gode spørsmål, både enkle og vanskelege. Dykk har lytta tolmodig når eg ringer i begeistring for å legge ut i det vide og det breie om resultat som for resten av verda ikkje har noko praktisk betydning. Og dykk har alltid hatt trua.

## SAMANDRAG

---

I dette prosjektet har ein freista å avdekkje samanhengar mellom personlegdomstypar og opplevinga av to undervisningsfilmar om same tema, men med forskjelleg innhald og oppbygning. Alle deltakarane fekk sjå begge filmene i vilkårleggjort rekkefølge. Talfesting av opplevinga vart utført ved spørjeundersøking før og etter kvar film, med kontinuerleg måling av galvanisk hudrespons, blikksporing og bruk av kjenslegjenkjenningssalgoritme samstundes som studentane såg begge filmene. Samanhengane mellom dei forskjellige responsane vart grundig undersøkt, det same vart parvise observasjonar innad i same type respons.

Korkje statistiske testar eller visuell inspeksjon av spredningsdiagram kunne med sikkerheit fastslå samanhengar mellom personlegdomstypar og responsane som skilde seg frå tilfeldige kontrollvariablar. Samanhengar mellom responsar og innad i kvar einskild responstype var kun observert ved sporadiske tilhøve, som og kan skrive seg frå tilfeldige samantreff.

Om det i røynda er svært stor variasjon mellom respondentane si oppleving av forskjellige undervisningsmetodar samstundes som personlegdomstype beskriv svært lite av variasjonen kan vere ei mogelegheit. Ei anna mogelegheit kan vere alle dei svakheitene ved forsøket som har blitt avdekt under arbeidet med prosjektet. I dette einskilde forsøket skulle alle svakheitene syne seg å vere forsøket si største styrke. Komande prosjekt kan trekkje stor lærdom frå dette forsøket, særskilt om korleis ein ikkje skal førebu forsøk når biometriske sensorar skal nyttast. Det vonast og at kvar svakheit avdekkja i dette forsøket kan, saman med implikasjonar for vidare arbeid, inspirere nye og betre forsøk for å belyse studentar sine opplevingar av forskjellige undervisningsmetodar.

I mange høver er studentar si oppleving av forskjellige undervisningsmetodar vel så spanande som deira prestasjonar målt i poeng og karakterar. Den store variasjonen i respondentane sine opplevingar fordrar til variasjon i undervisningsmetodane. Ikkje berre for den einskilde student si oppleving, men og for førebuing mot det komande arbeidslivet der ein med stort sannsyn må kunne forventast å tileigne seg kunnskap og ferdigheiter på fleire forskjellige måtar.

## **ABSTRACT**

---

This project aimed at investigating associations between students' personality characteristics and their experience when exposed to two educational films on the same statistical topic, but with different content and structure. All students watched both films in randomized order. The students' experience was measured by a self-report questionnaire before and after each film, and by continuous measurement of galvanic skin response, eye-tracker and assessment of emotions, through an emotion recognition algorithm. The associations between the different types of responses were thoroughly investigated, along with pairwise observations within the same type of response.

Neither statistical test nor visual inspection of color coded scatterplots revealed associations between the student's personality characteristics and their responses, that exceeded those found with random control variables. The few statistically significant associations which were found, may be attributed to chance alone.

One possible explanation of the lack of significant associations is that the variation in the students' responses to the educational films is large, and that the amount of variation explained by the personality characteristics is small, in comparison. The many weaknesses of the study, that unfolded during the working process, represent another plausible explanation of the lack of positive findings. However, the weaknesses turned out to be the project's biggest strength. Future projects can benefit a lot from this work, especially concerning how not to design educational experiments that rely on output from biometric sensors. Every flaw in the design and conduction of this study, and every consecutive dead end in potential analyses, which were detected and documented, should inspire new and better studies that aims to shed light on students' experiences of different educational methods.

Students' responses to different educational methods are as interesting as their performance measured in points or grades. The large variation in experiences demonstrated in this study can encourage variation in educational methods: Both to enhance students' experiences during their life at the university, but also as a preparation for future jobs where they can be expected to acquire knowledge and skills in several different ways.

## Innhold

Forord .....	i
Samandrag .....	ii
Abstract .....	iii
1 Innleiing .....	1
2 Material og metodar .....	9
2.1 Førebuing av forsøket .....	9
2.1.1 Utstyr, programvare og protokollar .....	9
2.1.2 Utdanningstesten og personlegdomstypar .....	10
2.1.3 Innspeiling av undervisningsfilmar .....	12
2.1.4 Forsøket si oppbygning i iMotions .....	13
2.1.5 Respondentar, personvernhandsaming og samtykkeerklæring .....	15
2.2 Gjennomføring av forsøket .....	15
2.3 Eksport og undersøking av resultatfiler frå iMotion .....	16
2.4 Innleiande utforsking av ferdigprosesserte resultatfiler .....	18
2.5 Samanhengar mellom ferdigprosesserte resultatfiler .....	19
2.6 Samanhengar mellom ferdigprosesserte resultatfiler og personlegdomstypar .....	20
2.7 Statistiske metodar og særskilde hensyn .....	21
3 Resultat .....	25
3.1 Resultat frå spørjeundersøkinga .....	26
3.1.1 Beskrivande statistikk .....	26
3.1.2 Prinsipalkomponentanalyser .....	27
3.1.3 Hierarkisk klynging etter midlare avstand i korrelasjon .....	31
3.1.4 Statistiske testar .....	32
3.2 Resultat frå AFFDEX statistics .....	37
3.2.1 Beskrivande statistikk .....	37
3.2.2 Prinsipalkomponentanalyse .....	42
3.2.3 Statistiske testar .....	45
3.3 Resultat frå GSR summary scores .....	48
3.3.1 Beskrivande statistikk .....	48
3.3.2 Statistiske testar .....	50
3.4 Resultat frå GSR binned peak detection .....	51
3.4.1 Beskrivande statistikk .....	51
3.4.2 Statistiske testar .....	55
3.5 Samanhengar mellom ferdigprosesserte resultatfiler .....	56

3.5.1	Samanhengar mellom spørjeundersøkinga og AFFDEX statistics .....	56
3.5.2	Samanhengar mellom spørjeundersøkinga og GSR summary scores .....	62
3.5.3	Samanhengar mellom spørjeundersøkinga og GSR binned peak detection.....	64
3.5.4	Samanhengar mellom AFFDEX statistics og GSR summary scores .....	67
3.5.5	Samanhengar mellom AFFDEX statistics og GSR binned peak detection.....	70
3.5.6	Samanhengar mellom GSR summary scores og GSR binned peak detection. ..	73
3.6	Samanhengar mellom ferdigprosesserte resultatfiler og personlegdomstypar .....	76
3.6.1	Samanhengar mellom spørjeundersøkinga og personlegdomstypar .....	77
3.6.2	Samanhengar mellom AFFDEX statistic og personlegdomstypar .....	80
3.6.3	Samanhengar mellom GSR summary scores og personlegdomstypar.....	86
3.6.4	Samanhengar mellom GSR binned peak detection og personlegdomstypar.....	89
3.6.5	Samanhengar mellom personlegdomstypar og dei samla responsane.....	92
4	Diskusjon.....	102
4.1	Vidare arbeid .....	111
4.2	Utfordringar som følgje av covid-19 .....	113
5	Oppsummering .....	116
	Kjelder.....	119
	Vedlegg .....	122
	Vedlegg 1. Manual for oppstart av biometrilab og forsøk .....	122
	Vedlegg 2. Protokoll for utføring av forsøk .....	126
	Vedlegg 3. Tabellar og figurar frå undersøking av spørjeundersøkinga.....	128
	Vedlegg 4. Tabellar og figurar frå undersøking av AFFDEX statistics.....	132
	Vedlegg 5. Oversikt over variabelnamn med suffiks og prefiks.....	136

# 1 INNLEIING

---

-Åh, var det berre så enkelt??

Kven kan vel ikkje kjenne seg att i ha stirra på eit eksempel i timesvis utan å sete att med noko anna enn «Hæ?». Eller rekna den same oppgåva fleire gongar og fått forskjellig svar kvar gong, berre for å oppdage at ingen av dei er i nærleiken av fasiten. Og kven har vel aldri lese det same avsnittet oppatt og oppatt, og tenkt for seg sjølve at teksta hadde gjort like mykje meining om ho hadde stått på gamal-egyptisk. Men så kjem redninga i form av ein forelesar, ein medstudent eller ein youtube-video som forklarar problemet på ein slik måte at det som før var uforståeleg med eit vart naturleg og enkelt. Men gleda over å ha meistra problemet kan ofte gli over i undring over kvifor forklaringa i boka eller på tavla førte til forvirring hjå deg sjølve, men vakte begeistring hjå studiekameraten. Og kva med dei som ikkje får den forklaringa som løyser floken, og ikkje får høve til å kome med det etterlengta utsagnet: -Åh, var det berre så enkelt?

Av både idealistiske så vel som kommersielle hensyn har nokre pedagogar og andre med særskild interesse for læring utarbeida teoriar som freistar å beskrive forskjellane i korleis kvart einskild individ lærar, og hevda at læringa blir forbetra di som kvart individ får undervisning tilpassa den einskilde sin læringsstil. Læringsstilane blir bestemt etter kvar modell sitt særskilde rammeverk. Interesse for læringsstilar tok verkeleg til på 70-talet, som ein kan sjå av (Coffield et al., 2004, pp. 166-169) si opplisting av over 70 forskjellige modellar for bestemming av læringsstilar. Fleire av dei forskjellige modellane har nytt stor popularitet sjølv om (Pashler, McDaniel, Rohrer, & Bjork, 2009) og (Rohrer & Pashler, 2012) ikkje har klart å finne tilstrekkeleg bevis for forbetra læring, sjølv etter å omfattande litteraturundersøkingar, og konkluderer med at det ikkje finnast empirisk materiale som støttar at undervisning skreddarsydd til den enkelte student sin læringsstil gjer betre læring. Til tross for dette kunne ei studie (Dekker, Lee, Howard-Jones, & Jolles, 2012) vise at 93% av britiske lærarar var einige i at elevar ville lære betre dersom undervisninga var tilpassa deira læringsstil, og ei studie frå USA rapporterte at 64% av lærarar på høgare utdanning var einige i at å rette læringa mot ein elev sin læringsstil ville forbetre læringa (Dandy & Bendersky, 2014). Ei nyare studie frå Storbritannia (Newton & Miah, 2017), fortel at trua på læringsstilar har avteke noko, men at 32% av lærarar i høgare utdanning ville forsette å bruke læringsstilar sjølv om det manglar bevis som støttar opp om desse teoriane.

Det vil vere naturleg å undre seg over kvifor læringsstilar fortsett vert nytta, sjølv om det ikkje finnast bevis for at undervising retta mot den enkelte student sin læringsstil gjer betre læring. (Newton & Miah, 2017) kunne i si undersøking finne at ei vanleg misoppfatning av å ikkje bruke læringsstilar var det same som at alle studentar burde bli undervist på same måte, eller at ein ikkje skulle variere undervisningsmetodane, til tross for at litteraturen som kritiserer læringsstiler ikkje på nokon måte fremmar dette synet. Å variere undervisinga og å oppfordre studentar til å reflektere over eiga læring er viktig, men ein treng ikkje å få tildelt ein læringsstil frå ei spesifikk klassifisering (Coffield et al., 2004) for å praktisere dette. Likevel har teoriar om læringsstilar gjort sitt inntog, noko som mogelegvis reflektera at undervisarar har tek inn over seg at personar lærer forskjellig og ynskje om å formidle faget sitt på ein forståeleg måte for flest mogeleg.

Å nå inn med budskapet til flest mogeleg studentar, blir vanskelegare dess fleire ein har av dei. Eit av dei større kursa ved Norges Miljø- og Biovitenskapelige Universitet (NMBU), STAT100 – Statistikk, har årleg om lag 500 studentar fordelt over to semester. For dei aller fleste av dei om lag 5000 studentane ved NMBU er STAT100 eit obligatorisk fag, og for nokon; rett og slett eit naudsynt onde. Dei færraste av studentane som tek STAT100 kom til NMBU for å studere statistikk, og mange av dei kan nok kjenne seg att i ein britisk studie (Druggeri, Dempster, Hanna, & Cleary, 2008) som syner at studentane tykte statistikk var skremmande og vanskeleg, og at dei ikkje var førebudde på at statistikk var ein del av studieprogrammet. I følgje ei undersøking gjort av NMBU si eiga studentavis Tuntreet, vart STAT100 nemnd som eit av dei meir ugreie faga, og vart karakterisert som tungt og arbeidskrevjande (Kvam, 2020). Dette er mogelegvis ikkje den mest vitskapelege undersøkinga, men studentaviser kan og gjenspegle ståda. Endring i universitetet si studentmasse og syn på god undervising, samt vilje til omstilling har ført til at faget dei seinare åra har gjennomgått store endringar i korleis det blir gjennomført. STAT100 har sidan hausten 2016 blitt gjennomført etter omvendt klasserom - metoda (Akçayır & Akçayır, 2018), der studentane sjølvstendig ser utvalde filmar om bestemte tema, før dei møtes til organisert og obligatorisk kollokviegruppe for å løyse og levere inn oppgåver. Frå 2017 var det mogeleg for studentar å ta faget som eit individuelt løp med obligatoriske innleveringar, og hausten 2018 vart det og gjett eit prosjektbasert tilbod, STAT101. Alt med det mål for auge å auke læringsutbyttet til studentane, og anerkjenne det store mangfaldet i studentgruppa.



Nettopp det store mangfaldet i studentane sine utdanningsretningar saman med eit stort antal studentar i kvar gjennomføring har gjort STAT100 til eit ynda forsøksobjekt, der det overordna målet for den forskinga som har vore gjort parallelt med undervisinga har vore å finne den gjennomføringa som er best mogeleg for flest mogeleg studentar. Frå 2014 har alle studentar ved grunnleggande emne fått høve til å svare på ei nettbasert spørjeundersøking utvikla av Dr. Helge Brovold for Nasjonalt senter for realfagsrekruttering (Nasjonalt senter for realfagsrekruttering, 2017b). Undersøkinga freistar å kartlegge yrkesinteresser, føretrekt leiingsstil, læringsstil og personlegdomstype (Brovold, 2014). Ei av dei tidlegaste undersøkingane vart utført hausten 2014, der (Sæbø, Almøy, & Brovold, 2015) fann at nokre personlegdomstypar verka til å oppnå betre karakterar enn andre. Det vart argumentert at dette kunne skrivast frå den rigide undervisningsforma. Etter omlegging til den meir studentaktive undervisningsforma omvendt klasserom viste førebelse upubliserte analyser at nokre personlegdomstypar presterar betre på eksamen etter omlegginga enn før, medan dei personlegdomstypane som presterte best før omlegginga fortsatt får tilsvarande gode resultat.

Når studentane som føl kurset STAT100 har blitt undersøkt er det respondentane sine prestasjonar i form av prosentvis score på eksamen, eventuelt omsett til karakterar frå A til F som er den nytta responsen (Sæbø et al., 2015). Noko liknande er og tilhøvet når det har vore undersøkt kvarvidt bruk av læringsstilar har positiv effekt på læringsutbytte, der prestasjonane har vore målt i poeng eller karakterar (Pashler et al., 2009). Det er i studiane undersøkt av (Pashler et al., 2009) gjort lite forsøk på å måle noko anna enn prestasjonar, men det er naturleg å anta at to studentar kan oppnå same karakter og samstundes ha to vidt forskjellige opplevingar av måten emnet vart undervist på. Det kan tenkjast at ein student kan mislike både emnet sitt innhald og undervisningsmetode, og fortsett oppnå ein god karakter. Og i motsett høve, er det ikkje urimeleg og tru at ein student kan synast at faget er kjempespanande og tykkje godt om undervisningsmetoda, men likevel få ein dårleg karakter i emnet. Det er nok ingen som vil nekte for at karakterar er viktig, men ein høyrer ofte utsagnet: -Hugs, karakterar er ikkje alt.

Og kva ligg i dette utsagnet, som mang ein student har fått høyre i eit trøystande tonefall frå vener, studiekameratar eller forelesarar? Det er ikkje utenkeleg at det skriv seg frå kva ein minnst frå studietida. Dei færraste hugsar kva karakterar ein fekk i kva einskild fag, men ein minnst godt det faget der ein sat i fem månadar og følte seg som verdas største idiot. Eller læraren som i si glødande begeistring over kjemiske reaksjonar skoddela ei heil etasje. Og ikkje minst forelesaren som sat ved sida av ein i fleire timar etter end arbeidsdag for å finne

den eine vetle feilen som forpurra eit langt og komplisert programmeringsscript. Og ein minnast dei forelesningane ein gledde seg til, dei gruppearbeida som om lag tok frå ein livsmotet, dei gongane ein kom ut frå auditoriet med lite anna utbytte enn ferdig middagsplan og innkjøpsliste til komande veke, og alle samtaler med medstudentar om kvarvidt forelesninga var god eller ikkje. Og ofte kan to studentar ha motstridande meiningar om kor godt dei likte same forelesinga.

Men kva er det som gjer at studentar som har vore i den same forelesinga kan ha vidt forskjellige opplevingar. Kan personlegdomstype og læringsstil bidra til å forklare denne skilnaden i studentane sine opplevingar? Og korleis kan ein måle opplevingar?

I tillegg til spørjeundersøking, kan biometriske sensorar mogelegvis bidra til å kaste lys over studentane sine opplevingar og kjensler. iMotions biometric platform er døme på ei programvare som kan integrere og synkronisere over 50 forskjellige biometriske sensorar, som til dømes blikksporar, ansiktsuttrykkanalyse og galvanisk hudrespons (iMotions). Med aukande datakraft og stadige teknologiske nyvinningar har fleire fått tilgang til biometriske sensorar, og dei seinare åra har biometriske sensorar gjort sitt inntog i ei rekkje områder som marknad- og forbrukarundersøkingar, tilrettelegging av nettsider, utvikling av spel i tillegg til ei rekkje forskingsområde. Og innanfor utdanning og læring har biometriske målingar nytt stor popularitet dei seinare åra, etter kvart som sensorar har blitt rimelegare og enklare i bruk.

Spesielt blikksporing har vore av spesiell interesse, grunna metoda si sterke tilknytning til lesing allereie frå 1879, då både Hering og Lamare beskreiv augebevegingar under lesing ved hjelp av lydar laga av muskelsamatrekningar i det ytre auget (Wade, 2010). Den første blikksporaren som klarte å registrere augebevegingar ved lesing vart laga av Edmund Huey, som seinare gav ut boka *The psychology and pedagogy of reading* (Huey, 1908). Sidan har arbeidet forsett, og i sin oversiktsartikkel om forskning på blikksporing finn (Alemdag & Cagiltay, 2018) ei hyppig vekst i interessa for blikksporing innan multimedialæring. Denne interessa vert og understøtta av boka *Eye-tracking technology applications in educational research* (Was, Sansosti, & Morris, 2016) der 36 forskarar frå USA og Europa har bidrege til ei bok som spenner over mellom anna metodiske hensyn, språkforståing, læring av matematikk og multimedia-læring.

Innan læring vil ofte blikksporing freiste å måle interaksjonen mellom respondent og stimuli (Horsley, Eliot, Knight, & Reilly, 2013), og er ofte basert på (Just & Carpenter, 1980) si antaking om at det er ein sterk samanheng mellom kva ein ser på og kva ein tenkjer på, kalla

auge-sinn antakinga. Auge-sinn antakinga er omdiskutert og mellom anna (Anderson, Bothell, & Douglass, 2004) finn at det nødvendigvis ikkje er ein samanheng mellom augebevegingar og mentale prosessar, men kan reflektere ein pågåande prosess så sant denne prosessen avheng av dekodning av informasjon. Likevel er auge-sinn antakinga lagt til grunn for ei rekkje studiar, og utviklinga i blikksporingsteknologi har vore særskildt viktig for ei auka forståing av kognitive og visuelle prosessar innan læring nettopp ved å måle kvar respondentane har retta si visuelle merksemd (Was et al., 2016). Kan dette mogelegvis tenkast å ha ein samanheng med det gamle uttrykket «ute av auge, ute av sinn»?

Men sjølv om vi har merksemda retta mot noko, er dette ikkje nok til å avdekkje kva kjensler og assosiasjonar som blir vekka. Frå (Alemdag & Cagiltay, 2018) kan ein finne at det er tre hovudformål for bruk av blikksporing i forskning på multimedialæring; kognitive prosessar som val, organisering og integrering av informasjon, kva faktorar som påverkar desse prosessane, og samanheng mellom kognitive prosessar og læringsutbytte. (Alemdag & Cagiltay, 2018) finn og at dei to faktorane som var minst undersøkt i lys av multimedialæring var metakognisjon og kjensler.

For å avdekkje om kjensler som blir vekka av forskjellige undervisningsmetodar, er forskjellige måling av elektrodermisk aktivitet (EDA) som til dømes galvanisk hudmotstand (GSR) ei mogeleg metode. Allereie i 1879 knytta den franske elektroterapeut Vigouroux EDA til psykologiske faktorar (Boucsein, 2012), og blir i dag nytta til forskning på mellom anna psykopatologi, anvendt psykologi og medisin, neurologi og psykososiale tilstandar. Det er under området psykososiale tilstandar ein finn dømer på bruk av elektrodermiske indikasjonar på opphissing (Boucsein, 2012). Av læreboka *Foundations of Psychology* (Hayes, 2000, pp. 400-401) kan ein sjå at den psykologiske interessa for opphissing ofte omhandlar forholdet mellom ein psykologisk opphissa tilstand og opplevinga av kjenslene frykt eller sinne, og det er antydning at opphissing reagerar refleksivt med desse kjenslene og at dei intensiverar kvarandre. Ein måte å måle opphissing på er GSR, som fangar opp endringar i huda si ledningsevne som følgje av auka svette som inntreff di som vi er engstelege eller opphissa på nokon anna måte (Hayes, 2000). Sidan ei rekkje forskjellige både klimatiske og fysiske faktorar som påverkar huda si ledningsevne (Boucsein, 2012), kan det vere utfordrande å tyde målingane (Matté, 2000). (Hayes, 2000) minner og om at det kun er psykologisk opphissing som blir målt ved GSR, og ein kan ikkje trekke slutningar om kva kjensle som er knytta til opphissinga. For å sjå kva kjensler som blir vekka av forskjellige

undervisningsmetodar er det difor ikkje nok å vite at opphissing inntreff. Det er sjølvklart ikkje likegyldig om ein forelesningsfilm vekker sinne eller begeistring.

Uttrykk for kjensler kan kome i mange formar. Kjensler kan mellom anna uttrykkjast ved det skrivne ord, meir eller mindre overdriven bruk av emojiar, med bevegingar, stemme og ord. Men det er ikkje alltid ein seier kva ein meiner, og det er heller ikkje alltid ein meiner det ein seier. Ein kan trykkje tommel opp på eit facebookinnlegg, sjølv om ein innerst inne veit med seg sjølv at ein ikkje kjem til å reise på den føreslegne dugnaden. Og ein kan klistre på seg eit smil og vinke høfleg til naboen som etter gjentekne anmodingar forset å setje bilen sin frå seg på ein slik måte at ein sjølv ikkje kjem seg ut med sin eigen bil. Å tolke kjensler uttrykt ved ansiktsuttrykk kan by på utfordringar. Kven har vel seiast å aldri ha blitt lurt av eit falsk smil, eller feilaktig anteke at ein anna person er sur eller tverr medan vedkomande berre var djupt konsentert? Ansiktsgjenkjenningssalgoritmer freistar å nytte teknologi til å tyde kjensler, og (Recio & Sommer, 2018) har funne at ein kan skilje ein ekte reaksjon frå ein innøvd reaksjon ved tidsforkjellar, i alle fall i det høvet kjensla er avsky. Ved undersøkingar av kjenslegjenkjenningssalgoritma AFFDEX har (Magdin, Benko, & Koprda, 2019) funne at er fleire tilhøver der AFFDEX ikkje evnar å fange opp eller skilje mellom kjensler, og (Stöckli, Schulte-Mecklenbeck, Borer, & Samson, 2018) i ein større valideringsstudie av to kjenslegjenkjenningssalgoritmer AFFDEX og FACET finn att begge presterar godt når glede skal gjenkjennast, men at AFFDEX ofte ikkje fangar opp eller klassifiserar sinne og frykt riktig. Både (Stöckli et al., 2018) og (Magdin et al., 2019) finn at AFFDEX kan klassifisere prototypiske uttrykk for kjensler med høveleg god grannsemd, men i likheit med oss menneskje vert det vanskelegare å klassifisere meir subtile uttrykk for kjensler. Ved å nytte kjenslegjenkjenningssalgoritmer saman med andre sensorar kan ein moglegvis avdekkje meir av det einskilde menneskje sine kjensler og opplevingar.

Frå nyare litteratur finnast det mange dømer på bruk av biometriske sensorar innan læring, og spesielt blikksporing er mykje nytta (Alemdag & Cagiltay, 2018; Wade, 2010; Was et al., 2016). Det er og gjort forsøk med bruk av kjenslegjenkjenningssalgoritmer innan læring, då mest innanfor spelbasert læring (Andres et al., 2015; Taub, Azevedo, Bradbury, Millar, & Lester, 2018; Taub, Azevedo, & Mudrick, 2018; Yeh, Lai, & Lin, 2016). Det som er felles for alle desse er at forsøka er designa med få forklaringsvariablar og responsar. Til dømes brukar (Taub, Azevedo, & Mudrick, 2018) kun senkning av augebryn(AU4) som biometrisk respons ved undersøking av interaksjon mellom metakognitive reguleringsprosessar og forkunnskapar ved bruk av hypermediabaserte intelligente læringssystem. Når (Taub, Azevedo, Bradbury, et

al., 2018) nytter nivå av kjensler og effektivitet til å undersøke metakognitive reguleringsprosessar og vitenskapeleg undersøking er det kun gjennomsnittet av intensiteten i kjenslene som vert nytta, sjølv om både GSR-målingar og blikksporingsmålingar vert gjort av respondentane. Det er få studiar innan læring som utforskar to eller fleire målingar frå same sensor, og målingar frå forskjellige sensorar. Eksempel på slike studiar frå andre fagfelt kan vere studiar som omhandlar samanhengen mellom spørjing og kjenslegjenkjenningssalgoritmer innan næringsmiddelindustrien (Danner, Sidorkina, Joechl, & Duerschmid, 2014; Leitch, Duncan, O'keefe, Rudd, & Gallagher, 2015; Zhi, Wan, Zhang, & Li, 2018). Samanheng mellom kjenslegjenkjenningssalgoritmer og galvanisk hudmotstand (GSR) kan til dømes vere frå psykologi (Lei, Sala, & Jasra, 2017).

Sidan dei undersøkte kjeldene kvar har fokus på eit fåtal av biometriske responsar og forklaringsvariablar, som det gjerne høver seg for vitenskapelege publikasjonar, har det ikkje til no vore råd å finne kjelder som kombinerar spørjeundersøking, blikksporing, galvanisk hudrespons og ansiktsgjenkjenningssalgoritmer i lys av korkje læring eller undervisningsmetodar. Det er og lite å finne om dei forskjellige responsane si sannsynsfordeling, kva metodar som er mest høvelege til innleiande utforsking og kva statistiske testar som kan gje dei mest pålitelege resultatata då dei relevante undersøkte artiklane i stor grad har mange ( $n > 30$ ) respondentar og dermed nyttar (M)ANOVA og andre metodar som forutset tilnærma normalfordelte responsar. Det er heller ikkje tidlegare utført liknande forsøk ved NMBU.

Mangelen på forsøk der fleire variablar innan kvar type målingar er nytta, samstundes som fleire biometriske responsar og spørjeundersøking sjeldan er knytt saman fordrar til at det første forsøket av sin art på NMBU bør sete fokus på grundig og varsam utforsking av variablar og metodar. Vel så viktig er registrering av relevante resultat, signifikante så vel som ikkje-signifikante, og grundig beskriving av korleis resultatata er framkome. Både for å sikre reproduserbarheit, men kanskje først og fremst for å sikre at dei som skal vidareføre arbeidet kan halde fram der denne tesa stoggar. Difor er eit underliggande mål for tesa å gjere arbeidet lettare for komande prosjekt ved å beskrive variablar frå sensorane ved utforsking og statistiske testar i lys av korleis studentar frå STAT100 med forskjellige personlegdomstypar opplever forskjellige undervisningsmetodar.

Tesa sitt hovudmål er å undersøke om personlegdomstype har samanheng med opplevinga av forskjellige undervisningsmetodar. Vegen ein lyt fylje for å nå dette målet vil både vere svingete og lang, og med fleire bomturar. Dei uunngålege bomturane må utforskast og

beskrivast for å forhindre at komande prosjekt nyttar ressursar på desse, eller i verste fall går seg vill. Ei rekkje delmål må nåast undervegs, om svaret på forskningsspørsmålet skal knyttast lit til.

Opplevinga av forskjellige undervisningsmetodar blir målt ved spørjeundersøking, kjenslegjenkjenning, blikksporing og GSR. Frå iMotions er det mogeleg å eksportere ei rekkje forskjellige ferdigprosesserte resultatfiler frå kjenslegjenkjenning og GSR, i tillegg til spørjeundersøkinga.

Det første delmålet vil vere å undersøke ferdigprosesserte resultatfiler for responstypene kvar for seg. Ved å nytte seg av deskriptiv statistikk, visuelle framstillingar og statistiske metodar vil ein freiste å avdekkje fordelinga til responstypene og samanhengar mellom variablar innad i kvar responstype, i tillegg til å avdekkje om rekkefølga på undervisningsmetodane har effekt på responstypene. Denne undersøkinga vil danne grunnlag for kva variablar om best formidlar budskapet i responstypene, og som kan vere gode kandidatar til variablar når responstypene skal knyttast saman.

Det andre delmålet vil vere å undersøke samanhengar mellom responstypene i dei ferdigprosesserte resultatfilene, særskildt om resultatane frå spørjeundersøkinga har samheng med kjensler og reaksjonar registrert ved GSR. Herunder vil det vere av spesiell interesse å avdekkje om respondentar som brukar ytterkantane av skalaen i spørjeundersøkinga og viser meir kjensler og har fleire utslag på GSR.

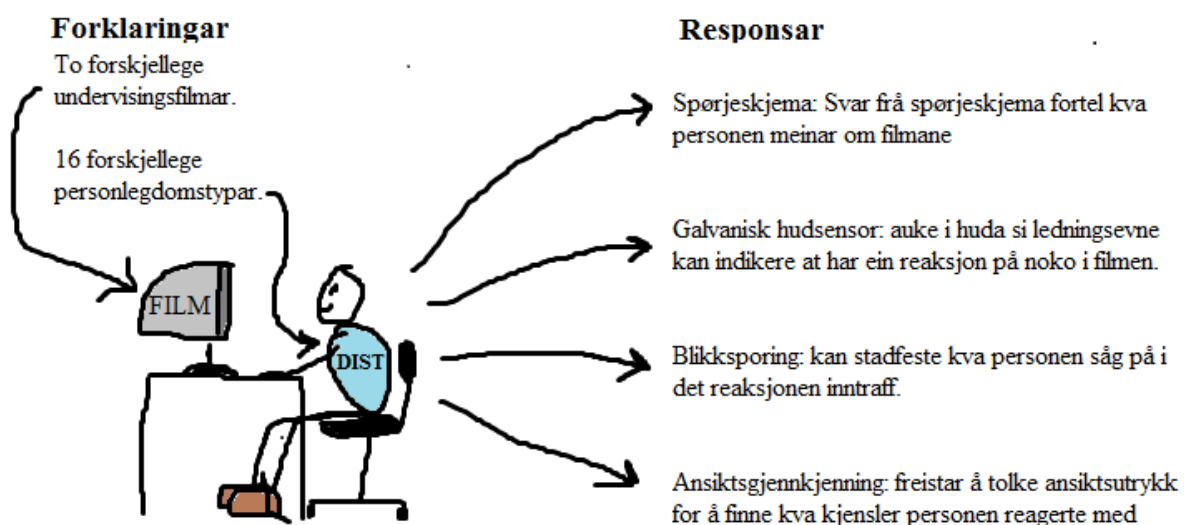
Først når første og andre delmål er nådd, kan ein påbyrje undersøking av tesa sitt forskningsspørsmål.

Det tredje delmålet set fokus på om ferdigprosesserte resultatfiler frå iMotions i dette einskilde tilhøvet er tilstrekkeleg for å svare på forskningsspørsmålet, eller om rådata trass betydeleg større datamengder og behov for etterprosessering likevel vil vere tilrådeleg å nytte.

**Forskningsspørsmål: Har personlegdomstype samanheng med opplevinga av forskjellige undervisningsmetodar målt ved spørjeskjema og biometriske sensorar.**

## 2 MATERIAL OG METODAR

Respondentar frå det grunnleggjande statistikkurset STAT100 ved NMBU fekk sjå to filmar om same tema innan statistikk. Før og etter filmane fekk respondentane avkryssings spørsmål om haldningar til statistikk og kva dei tykte om filmane. Medan respondentane besvarte spørsmåla og såg filmane, var dei tilkopa blikksporar og galvanisk hudresponsensor (GSR) samstundes som ansikta deira vart filma og kjensler klassifisert av ansiktsgjenkjenningsalgoritma AFFDEX som vist i Figur 1. Data frå spørjeundersøkinga, blikksporaren, GSR og estimerte kjensler frå AFFDEX vart analysert kvar for seg, saman og til slutt satt i samanheng med resultatata frå utdanningstesten.



Figur 1 Enkel teikning som syner forsøket på eit overordna konseptuelt nivå.

### 2.1 FØREBUING AV FORSØKET

#### 2.1.1 Utstyr, programvare og protokollar

iMotions, sensorar og datamaskin med tilhøyrande utstyr vart stilt til rådighet av Vitenparken. Alle tilgjengelege sensorar vart nytta i forsøket. Til forsøket vart det nytta ei ASUS ROG stasjonær datamaskin med ein i7-8700 3.2ghz prosessor. Datamaskina hadde 32gb Ram og eit Nvidia GTX 1080 grafikkort. Skjermen som vart nytta var av typen HP243. Kamera nytta til å filme respondentane var av typen Logitech c920. Blikksporinga vart utført med ein Tobii x2-30 og GSR vart målt med Shimmer 3 iMotions GSR kit. iMotions versjon 8.0 vart nytta under forsøka. Statistiske analyser vart gjort i Rstudio versjon 1.1.463.

Sidan forsøka avhang av at ei rekkje sensorar verka saman, vart det ved prøve og feile – metoda funne ei rekkefølgje for oppkobling og oppstart som gav lite problem og feilmeldingar. Det vart utvikla ein enkel manual som vart nytta i både dette og andre forsøk, og som kan nyttast ved seinare høve. Sjå Vedlegg 1. Sidan forsøket vart gjennomført i to omgangar, Haust 2019 og Vår 2020, vart det på forhånd utarbeidd ein protokoll for korleis respondentane skulle takast imot og i kva rekkefølgje sensorane skal tilpassast, sjå vedlegg 2. Ein skulle då kunne sikre at respondentane fekk ei mest mogeleg lik mottaking, sjølv om det var seks månadar mellom kvar datafangstperiode.

### **2.1.2 Utdanningstesten og personlegdomstypar**

Utdanningstesten har, i to forskjellige versjonar, vore nytta ved NMBU til forskning på store, grunnleggjande fag som STAT100. Sidan 2017 har den forkorta versjonen vore nytta, og blir og nytta i dette forsøket. Utdanningstesten er ein nettbasert test utvikla for Nasjonalt senter for realfagsrekruttering av psykolog Dr. Helge Brovold, og er ein forkorta versjon av testbatteriet utvikla under Brovold si doktorgradsavhandling (Brovold, 2014). Utdanningstesten består av 5 delar. Del 1 omhandlar personlegdomstype og består av 4 sett med to påstandar der ein vel den påstanden som høver best. Del 2 omhandlar yrkesinteresser og består av 5 sett med 4 påstandar der ein vel den som høver best og den som høver dårlegast. Del 3 omhandlar korleis ein lærer best og består av 3 sett med 4 påstandar der ein vel den som høver best og den som høver dårlegast. Del 4 omhandlar faginteresser og består av 6 utsegn som rangerast frå 1 (høver dårlegast) til 6 (høver best). Del 5 omhandlar kjensler om matematikk og består av 2 sett med 5 påstandar der ein vel den påstanden som høver best. Det tek om lag 10 minutt å svare på utdanningstesten, og respondenten får attende ei beskriving av kva personlegdomstype vedkomande har, kva fagområde som kan vere høveleg og kva andre med same svar arbeidar med og kva undervisningsform som mest sannsynleg vil høve bra.

Dei resultata som i dette høvet er tilgjengelege frå utdanningstesten er personlegdomstypen til kvar einskild respondent. Personlegdomstypen blir gitt i ein kode på fire bokstavar, og det er 16 moglege kombinasjonar. Kvar av dei fire bokstavane vart gjett som ein av dei fire personlegdomskaraktistikkane beskrive av (Myers, 1980), J (Judgement) mot P (Perception), E (Extraversion) mot I (Introversion), F (Feeling) mot T (Thinking) og N (iNtuitive) mot S (Sensing). I høve utdanningstesten vart J mot P erstatta av D (Digital) mot C (kontekstuell), sjå (Brovold, 2014). Frå resultat av Utdanningstesten (Nasjonalt senter for realfagsrekruttering, 2017b) er følgande forklaringar gjett dei forskjellige personlegdomskaraktistikkane, sjå oversikt i Tabell 1.



Tabell 1 Forklaring til personlegdomstypene i Utdanningstesten, henta frå resultat av utdanningstesten (Nasjonalt senter for realfagsrekruttering, 2017b).

<p>Introvert (I): Jeg er stille og rolig, og liker godt å arbeide konsentrert på egenhånd, f.eks med leksene mine. Jeg trivest best i mindre grupper, og jeg unngår ofte å ta ordet i større forsamlinger eller i klasserommet. Jeg kan bli sjenert viss eg får mye oppmerksomhet, og kun mine nærmeste venner vet mye om meg.</p>	<p>mot</p>	<p>Ekstrovert (E): Jeg er utadvendt, pratsom og lett å bli kjent med, og jeg trives med mye folk rundt meg – både på skolen, hjemme og i gruppearbeid. Jeg kan godt presentere noe for hele klassen, og jeg gjør det gjerne med intensitet, godt humør og engasjement.</p>
<p>Fantasibasert (N): Jeg er kreativ og engasjert, og liker og gjøre ting på min egen måte. Tankene mine vandrer lett avsted, og jeg undrer ofte på både store og små ting i livet. Jeg liker å filosofere, og lar meg lett inspirere av kunst og kultur.</p>	<p>mot</p>	<p>Sansebasert (S): Jeg er praktisk, ærlig og jordnær, og er opptatt av det jeg faktisk kan oppfatte gjennom sansene mine. Jeg trives godt med oppgaver som krever aktivitet og jeg finner lett løsning på praktiske utfordringer. Jeg liker å se at det jeg gjør har en nytteverdi.</p>
<p>Rasjonell (T): Jeg oppleves som seriøs, fornuftig, faktaorientert og rettferdig, og jeg stoler oftere på min logiske sans enn på magefølelsen. Jeg uttrykker meg gjerne på en ærlig og direkte måte.</p>	<p>mot</p>	<p>Følsom (F): Mange opplever meg som mild, snill og tillitsfull, og jeg er ofte opptatt av å skape god stemning og harmoni både i klasserommet og hjemme. Jeg er sympatisk, og ofte svært medfølende og lyttende når folk rundt meg har problemer.</p>
<p>Se i sammenheng (C): Jeg er impulsiv og fleksibel, liker best variasjon og er meir opptatt av å se og forstå sammenhengene og helheten enn å fokusere på detaljene. Jeg kan ofte utsette ting f.eks. skolearbeidet, men jobber godt når jeg først kommer i gang og det er travelt og hektisk rundt meg.</p>	<p>mot</p>	<p>Se detaljer (D): Jeg oppfattes ofte som nøyaktig og strukturert, og trives best når jeg har kontroll også på detaljene. Jeg liker å legge planer, som jeg gjennomfører i jevnt, godt tempo – uten stress eller forstyrrelser underveis.</p>

### 2.1.3 Innspeling av undervisningsfilmar

I januar 2019 vart det produsert to undervisningsfilmar om temaet variasjon i gjennomsnittet. Dei to filmene vart produsert av Hilde Vinje og Kathrine Frey Frøslie, som begge er ansatt ved forskningsgruppa bioinformatikk og anvendt statistikk på NMBU, og har begge inngåande kjennskap til STAT100. Hilde og Kathrine hadde ansvar for innspeling av sine respektive filmar, og dei sjølve forelas kvar sin film. Begge filmene vart laga i greenscreenstudio med hjelp av Læringscenteret, og hadde same bakgrunn og omgjevningar. Hilde sin film hadde ei varigheit på 9 minutt og 50 sekund, medan Kathrine sin film hadde ei varigheit på 9 minutt og 36 sekund. Sjølv om dei to filmene fannar same tema og har mange likskapar, er både innhald og struktur forskjellig.

Hilde hadde i sin film vald ei tradisjonell oppbygning, der temaet først blir presentert og formalar blir vist og utleia før filmen vert avslutta med eit praktisk rekneeksempel. Kathrine hadde valt å starte med eit visuelt og intuitivt eksempel før formalar blir utleia og vist. Filmene er ikkje forskjellige berre i oppbygning og i innhald, men og i måten innhaldet vert presentert på. Eigne preferansar om læring vil bevisst eller ubevisst vere viktige for den som undervisar. Hilde og Kathrine har begge eit sterkt engasjement for formidling, og nyttar dette på ein utadvendt måte når dei undervisar. Samstundes som dei kan framstå som noko like i framtoning, har dei gjennom diskusjonar kome fram til at dei tileignar seg kunnskap svært forskjellig. Dette indikerast og ved at dei scorar til dels likt på utdanningstesten. Begge heller mot å vere ekstroverte (E) og følelsesbaserte (N). Dei scorar og til dels svært ulikt på utdanningstesten, og desse ulikheitene spelar ei vesentleg rolle for måten dei undervisar på. Medan Hilde har ei klar digital (D) og sansebasert (S) tilnærming, har Kathrine ein tilsvarande kontekstuell (C) og fantasibasert (N) profil. Den yrkesmessige bakgrunnen er og forskjellig og forsterkar ytterlegare skilnadane i tilnærming til undervisning. Hilde er utdanna lektor og har arbeidd som matematikklærer på vidaregåande skule. Ho har difor mykje erfaring med å undervise og forklare eit avgrensa pensum på ein strukturert og korrekt måte. Kathrine har på si side lang erfaring frå veiledning av medisinske forskarar og er vant til pragmatisk bruk av statistiske testar. Ho er og røynd i å improvisere og forenkla forklaringar og nytte eksemplar og metaforar for å forklare statistiske prinsipp. Ulikheitene i personlegdom og yrkeserfaring gjennomsyra dei to filmene. Sidan dette var eit nybrottsarbeid for alle involverte partar vart det lagt vekt på at Hilde og Kathrine skulle nytte sine bakgrunnar og innfallsvinklar på best mogeleg måte for å lage to ulike filmar, heller enn å planlegge filmene ut frå kva ein kunne forvente å få ut av dei biometriske sensorane.

#### **2.1.4 Forsøket si oppbygning i iMotions**

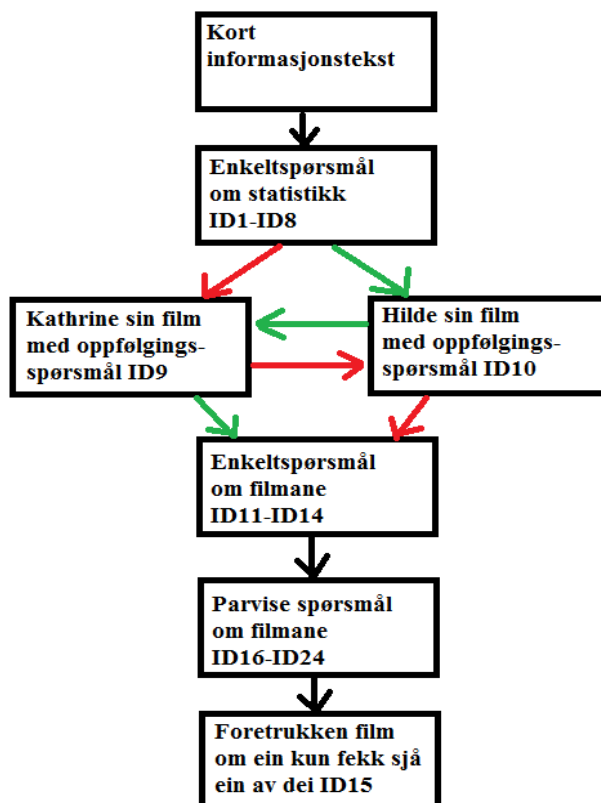
Programvara iMotions er ei programvare spesielt utvikla for å håndtere ei rekkje biometriske sensorar samstundes. iMotions gjer ein høve til å designe heile forsøket i same plattform, og det er nær sagt uante mogelegheiter for forskjellige oppsett og strukturering av forskjellige typar stimuli. Ein kan anvende ei rekkje forskjellige stimuli, som bilete, filmar, og lydar. Spørjeundersøkingar kan opprettast direkte i forsøket, og ein kan velje mellom mange forskjellige utformingar på spørsmål og svaralternativ. Ein kan vilkårleggjere heile eller delar av forsøket etter eige ynskje. Figur 2 syner korleis forsøket vart sett opp i iMotions.

Sidan antal respondentar og fordeling av personlegdomstypar var uvisst fram til etter andre runde av datainnsamling var gjennomført, fall valet på å la respondentane sjå begge filmane. Fordelen med at respondentane såg begge filmane er at ein ikkje risikerte å kome i ein situasjon der det er stor overvekt av ein personlegdomstype som såg til dømes Hilde sin film. Ein fekk og høve til å gjere parvise samanlikningar mellom filmane. Ulempa er at det måtte takast høgde for at det respondenten lærte og opplevde frå den første filmen, kunne påverke korleis den andre filmen vart opplevd. Desse effektane kunne til ei viss grad modellerast inn i forsøka, og effektane målast. Dette endra dog ikkje på den kjensgjerning at respondentane såg ein film først og ein etterpå. Hadde ein på førehand visst antal respondentar og deira personlegdomstype hadde dette åpna for heilt andre typar forsøksdesign. Til dømes kunne ein gjort antakingar om kva personlegdomstypar ein trudde ville føretrekt kva film, og tildelt halvparten av respondentane den filmen ein antok dei ville føretrekke, og oppnådd ein tydelegare forsøksdesign.

Når det først vart bestemt at respondentane skulle sjå begge filmane i vilkårleg rekkefølge, måtte det finnast ein høveleg måte for respondentane å skilje filmane på. Sjølv om å kalle filmane henholdsvis Hilde sin film og Kathrine sin film kunne framkalle assosiasjonar hjå respondentane då dei fleste har kjennskap til både Hilde og Kathrine frå før, vil ein kunne anta at dette likevel vil kunne inntreffe sidan Hilde og Kathrine har spelt inn sine eigne filmar. Det vart ansett som viktigare at respondentane skilde riktig mellom filmane, i alle fall sidan dei vart synt i vilkårleg rekkefølge. Det kan og vurderast om ein kunne skjerma for forutinntattheit ved å til dømes la ein for respondentane ukjend person spele inn begge filmane og navngje filmane på ein ikkje-ordnande måte. Det kan og tenkjast at ein annan person ikkje vil klare å formidle budskapet med same begeistring, men ein vil oppnå at ein ikkje måler person-effektar heller enn undervisningsmetode-effektar. Men igjen, så kan det

tenkjast at person-effekt og undervisningsmetode-effekt begge saman har sin eigenverdi i bevisst eller ubevisst bruk av naturlege pauser, ekstra trykk på særskild viktige ord, seinare taletempo for viktige definisjonar og andre subtile verkemidlar som medverkar til formidling av budskapet.

Rekkefølga på filmene til Hilde og Kathrine, med tilhøyrande oppfølgingsspørsmål vart vilkårleggjort av iMotions. Spørsmåla var ei blanding av enkeltspørsmål og parvise spørsmål sjå vedlegg 3, Tabell 3-A. Dei parvise spørsmåla vart nytta for å få ei samanlikning av dei to filmene, og samtidig måle i kor stor grad studenten er einig, ueinig eller nøytral i karakteristikkar av filmene. Til dømes vart studentane spurt om kor einige dei er i at filmen var frustrerende, på ein skala frå 0 til 10. Studentane fekk då svare på spørsmålet både for Hilde sin film og Kathrine sin film kvar for seg. Ein hadde då høve til å avdekkje ikkje berre kvifor ein av filmene dei tykte var mest frustrerende, men og i kva grad dei tykte filmene var frustrerende eller ikkje i det heile. Val av Likert-skala (Likert, 1932) mellom 0 og 10 verka rimeleg då skalaen 0 til 10 ville gje ei tilstrekkeleg differensiering og i tillegg er mykje nytta i daglegtale så vel som i ei rekkje populære idrettar og kunne difor antakast å vere kjend for respondentane.



Figur 2 Visuell oversikt over forsøket si oppbygning. Forsøket har to moglege rekkeføljer på undervisningsfilmene. Grøn løype syner oppsettet for dei som ser Hilde sin film først. Raud løype syner oppsettet for dei som ser Kathrine sin film først.

### **2.1.5 Respondentar, personvernhandsaming og samtykkeerklæring**

STAT100 er obligatorisk for svært mange studieretninger ved NMBU, og vert gjennomført kvart semester. Respondentane var alle studentar ved NMBU, og oppmeld i faget STAT100 – Statistikk hausten 2019 eller våren 2020. Respondentane vart rekruttert gjennom forelesingar og e-læringsplattforma Canvas, og kunne sjølve melde seg på forsøket via eit delbart elektronisk rekneark (Google sheet). For å motivere flest mogeleg studentar til å delta i forsøket, vart det i første omgang trekt ut fem respondentar som fekk kvar sitt par med heimestrikka raggsokkar. I andre omgang fekk kvar respondent ein bong som kunne vekslast inn i eit valfritt bakverk på Vitenparken.

Ein kunne og ha vurdert fleire metodar for rekruttering av respondentar, men ein må og akseptere at i ein travel studiekvardag kan studentane finne det vanskeleg å sete av tid til anna en kva som er naudsynt. I alle høver må ein hugse at motivering for deltaking vil påverke kven som deltok og kvifor. I dette forsøket kunne deltakarane tykkje forsøket var spanande og ynskja å bidra til forskning. Men dei kunne like gjerne vore svoltne på eit gratis bakverk, eller ha lyst på eit par raggsokkar. Dei kunne og ha kome for å få høve til å sjå dei to undervisningsfilmene, noko som kun var mogeleg for dei som deltok i forsøket. Kva som gjorde at dei som deltok valde å gjere nettopp det, og kva som gjorde at så få ynskja å delta er uvisst, men fordrar til ettertanke.

Det var frivillig å delta i forsøket, og respondentane sitt samtykke vart innhenta i tråd med Norsk senter for forskningsdata sine retningslinjer for handsaming av persondata.

## **2.2 GJENNOMFØRING AV FORSØKET**

Forsøket vart gjennomført i to omgangar. Første omgang var i STAT100 haust 2019, andre omgang var i STAT100 vår 2020. I begge høva fann forsøket stad frå to til fire veker inn i kursa, opp til tidspunktet då temaet i undervisningsfilmene vart gjennomgått. Alle enkeltforsøk vart gjennomført i tråd med vedlegg 2, protokoll for mottaking av respondentar. Like etter forsøket vart respondentane spurt om kva film dei vart synt først, som ein ekstra kontroll på at dei skilde filmene rett frå kvarandre.

Til saman 44 respondentar deltok i forsøket, 25 hausten 2019 og 19 våren 2020. Av dei 44 respondentane var 36 av dei mellom 19 og 25 år, og 8 av dei var over 25 år. 8 av respondentane var menn, 36 var kvinner. 41 av forsøka vart gjennomført utan manglar i spørjeundersøkinga, derav eit forsøk mangla registrering av blikksporing og eit forsøk mangla registrering av GSR. 38 av forsøkspersonane svarte på utdanningstesten, og antal komplette

forsøk bestående av alle ferdiprosesserte resultatfiler og resultat frå utanningstesten er 34. I alle høver vert det nytta flest moglege forsøk til analyser.

### **2.3 EKSPORT OG UNDERSØKING AV RESULTATFILER FRÅ iMOTION**

Eksport av resultatfiler vart gjort etter at respondentane vart anonymiserte. Resultatfilene vart lagra på ekstern hardisk i tillegg til skylagring. Forsøket med alle tilhøyrande data vart og lagra lokalt på Vitenparken si datamaskin. Frå iMotions er det moglege å eksportere ei rekkje forskjellige resultatfiler, både rådata frå sensorane og tilrettelagde ferdigprosesserte filer. Dei ferdigprosesserte filene vart åpna i excel rekneark for inspeksjon og sletting av unødig informasjon. For dei tilrettelagde ferdigprosesserte filene det fleire tilhøver der variablane vart dikotomisert, eller terskla om ein vil. I dei høva terskling var bakgrunn for resultata i filene, vart iMotions sine fabrikkinnstillingar av terskelverdiar nytta der endringar var moglege, sjå Tabell 2.

Tabell 2 Oversikt over innhald i tilrettelagde filer frå iMotions. Oversikta fortel kva resultatfila heiter, kva terskling som er nytta og ei kort beskriving av kva fila inneheld og måten ho er oppbygd på.

Namn	Innhald og oppbygning	Terskling
Survey matrix	Inneheld resultat frå spørjeundersøkinga. Fila er ei enkel tekstfil, med spørsmåla sine namn som kolonnenamn. Det er ei rad for kvar respondent, som inneheld svara frå undersøkinga.	Nei
AFFDEX statistics	Inneheld resultat frå AFFDEX si ansiktsgjenkjenningalgoritme. Fila er ei enkel tekstfil der ein for kvar kjensle får ei kolonne med ein terskelverdi, ei kolonne med antal scener som har kjensler over terskelverdien, og kor stor prosent av filmen si varigheit ei gitt kjensle vart opplevd. Ei rad for kvar respondent per stimuli.	Ja. For kvar scene blir kvar kjensle uttrykt ved ein skala frå 0 (ikkje tilstade) til 100 (fullt tilstade). Førehandsinnstilling i iMotions er ein terskel på 50.
GSR summary scores	Inneheld resultat frå galvanisk hudrespons. Fila er ei enkel tekstfil med antal gongar huda si ledningsevne overskrid ein gitt terskelverdi (toppar), samt antal toppar per minutt. Ei rad for kvar respondent per stimuli.	Ja. Det er satt ein terskel på 800 av iMotions. Det er tatt høgde for forseinking ved eit fem sekundars flyttbart vindauge, som flyttar seg eit sekund i gongen.
GSR binned peak detection	Inneheld resultat frå galvanisk hudrespons. Fila er ei enkel tekstfil med ei kolonne for kvart fem sekunds intervall i filmane. For kvart intervall er verdien 1 om minst ein topp har inntruffe i løpet av dei fem sekunda, 0 disom ikkje. Det vert laga ei fil per stimuli, kalla M2M_stimuliinavn ved eksport.	Ja. Resultata i GSR peak detection er bakgrunn for å avgjere om ein topp har inntruffe.

## 2.4 INNLEIANDE UTFORSKING AV FERDIGPROSESSERTERTE RESULTATFILER

Dei ferdigprosesserte resultatfilene frå Tabell 2 vart undersøkt kvar for seg, i tillegg vart det undersøkt om rekkefølga filmene blir vart synt i hadde samanheng med resultatane. Resultatane frå dei ferdigprosesserte resultatfilene sin varierte natur oppmoda til forsiktigheit omkring antakingar om fordeling, og fordra til grundig og varsam utforsking før sjølve analysene tok til. For å seinare kunne bestemme kva statistiske testar som høvde best, danna enkle og grunnleggjande utforskande teknikkar gje eit godt grunnlag for slike avgjerdsler.

Oppsummeringstabellar, frekvensdiagram og spredningsdiagram er enkle, men verknadsfulle metodar for utforsking av variablane i seg sjølve og deira samanhengar. Pearsons korrelasjonskoeffesient vert berekna i dei tilhøva spredningsdiagramma antyda ein lineær samanheng og ein skulle ynskje å talfeste denne. I fleire høver kan og prinsipalkomponentanalyse og klynging gje auka innsikt i samanhengar mellom variablane.

Høvelege statistiske testar for dei einskilde datatypene vart nytta for å undersøkje om kva film respondentane såg først har ein samanheng med responsane, og for parvise samanlikningar mellom variablar der dette var relevant. Einskilde variablar kunne og vurderast til å ha evna til å forklare ein stor del av variasjonen i datasetta, og dei som viste seg å vere av særskild interesse vart testa som mogelege forklaringsvariablar for utforsking av samanhengar mellom variablar frå forskjellige ferdigprosesserte resultatfiler. Erfaringane frå dei innleiande undersøkingane vart nytta når data frå sensorane skulle knyttast saman med både kvarandre og personlegdomstypene.

Oversikt over statistiske metodar og særskilde hensyn, sjå 2.7.



## 2.5 SAMANHENGAR MELLOM FERDIGPROSESSERTER RESULTATFILER

Det var av vesentleg interesse å avdekkje om det fantes samanhengar mellom dei ferdigprosesserte resultatfilene, og om responstypane støttar opp om kvarandre.

For å avdekkje eventuelle trendar og sjå etter høvelege modellar vart spredningsdiagramm mellom alle kombinasjonar av forklaringsvariablar og responsvariablar inspisert. Der ingen særskilde antydningar av andre, meir kompliserte modellar vart funne, vart den enkle lineære modellen nytta i statistiske testar. Det vart undersøkt både opprinnelege variablar, prinsipalkomponentar og skilnaden mellom parvise variablar.

For å avdekkje meir ein kva dei statistiske testane kunne rå med, vart visuell inspeksjon mykje nytta. Fargekoding kunne moglegevis syne samanhengar og grupperingar som spredningsdiagram og lineære samanhengar ikkje kunne evne å avdekkje. Å nytte fargekoding etter øvre eller nedre kvartil av ein variabel i eit spredningsdiagram mellom to variablar frå ein anna responstype var eit godt hjelpemiddel for auka heiheitsforståing av samhengane, og ein var og fri for mange antakingar. Eksempel på dette kunne vere når ein ville undersøkje om respondentar med høge kjensleverdiar nyttar ytterkantane av skalaen i spørjeundersøkinga meir enn dei øvrige respondentane. Då kunne visuell inspeksjon av spredningsdiagram frå spørjeundersøkinga der respondentane vart fargekoda etter kvartilverdiane i enkeltvariablane i AFFDEX statistics gje ny innsikt. Ein hadde då høve til å sjå om respondentar med høge kjensleverdiar låg meir i utkantane av spredningsdiagramma enn dei øvrige, utan å måtte ta avgjersler på kva som er ein ytterkant, og kva som er ein høg kjensleverdi i kvart einskild tilhøve. Ein hadde og høve til å fange opp fenomen som ikkje framkom ved å nytte kjensledifferansane, sidan to respondentar kunne ha ein låg kjensleforskjell men vise stor grad av kjensla i begge filmene.

## **2.6 SAMANHENGAR MELLOM FERDIGPROSESSERTER RESULTATFILER OG PERSONLEGDOMSTYPAR**

Sidan forsøket var av ei slik oppbygning at undersøkt litteratur ikkje kunne nyttast til direkte samanlikning, var det stor uvisse om kva resultat som kunne forventast. Det var og stor uvisse omkring kor stor grad av variasjon personlegdomskaraktistikkane burde kunne forklare, før ein hadde høve til å seie med sikkerheit at det i røynda var eit skilje i oppleving forklart av personlegdomstypen. For å ha noko samanlikningsgrunnlag, vart tilfeldige binære kontrollvariablar oppretta. Om det i røynda var samheng mellom oppleving og dei fire personlegdomskaraktistikkane, burde desse syne meir og tydelegare skilje enn tilfeldige variablar.

Det vart oppretta fire kontrollvariablar for å kunne gje eit betre heilheitsinntrykk enn kun ein einskild kontrollvariabel, sidan det kunne tenskjast og finnast kontrollvariablar som gav både betre og dårlegare resultat enn personlegdomskaraktistikkane. Det vart ikkje sett bestemningar for balansering av klassane i kontrollvariablane, sidan personlegdomskaraktistikkane og har synt forskjellig grad av ubalansert fordeling. For å ivareta reproduserbarheit og redelegheit var initialiseringsverdi for utrekka av kontrollvariablane bestemt av eit tilfeldig positivt tosifra heiltal gitt frå veileidar på forespørsel. Kvart siffer, summen av siffera og produktet av siffera vart dei fire brukte initialiseringsverdiane. Dei fire kontrollvariablane, K1 til K4 vart nytta på nett same måte som personlegdomskaraktistikkane.

Sidan det er visse begrensingar i kva statistiske testar som kunne nyttast grunna datasetta si beskjedne storeleik, og den komplekse naturen kjensler og meiningar gjerne har, var visuell inspeksjon eit svært nyttig hjelpemiddel. Difor vart det og gjort mykje undersøkingar av spredningsdiagram og scoreplot av data frå responstypene der respondentane vart fargekoda etter personlegdomskaraktistikkar og kontrollvariablar. Desse vart undersøkt for trendar og grupperingar, og var eit godt supplement til heilheitsforståinga av datasetta i lys av personlegdomskaraktistikkane.

For å gje samhengane mellom personlegdomskaraktistikkane og dei ferdigprosesserte resultatfilene som heilheit best mogeleg tilhøve til å tre fram, vart det utført ein partiell minste kvadraters diskriminantanalyse, PLS-DA. Alle dei undersøkte variablane frå dei ferdigprosesserte resultatfilene vart samla i ei matrise, og personlegdomskaraktistikkane og kontrollvariablane vart nytta som responsar etter tur. Det vart utført 10-fold kryssvalidering

med tilfeldig uttrekk av foldar, dette vart gjenteke 50 gongar for å sikre god beskriving av feilklassifiseringsrate og variasjon mellom kvart tilfeldig uttrekk av foldar.

## **2.7 STATISTISKE METODAR OG SÆRSKILDE HENSYN**

Ei kort redegjering av statistiske metodar nytta i tesa, og kva særskilde hensyn som måtte takast ved val av høvelege statistiske metodar.

Multivariat variansanalyse, MANOVA, vart nytta der ein hadde fleire responsvariablar, og ein ikkje kan anta at responsvariablane er uavhengige av kvarandre. MANOVA (Rencher, 1995) gjer dei same mogelegheitene som ANOVA, men responsen er ikkje lenger ein einskild variabel. Ein forutset fortsatt uavhengigheit mellom observasjonane, normalfordelte restfeil og lik varians. Ei svært hendig anvending av metoda i Rstudio er om ein vil teste same forklaringsvariabel på fleire responsar. I staden for å lage ein og ein modell, så vart det nytta ei matrise av dei forskjellige variablane som respons for ein einskild modell. Når ein då fekk oppsummering av modellen, vart kvar enkelt respons skriven ut som ein enkelt lineær modell og ein kunne lett lese ut p-verdi, stigningstal og  $R^2$  for kvar respons. Når ein inspiserte resultatata frå MANOVA, kunne ein finne ut om forklaringsvariablane hadde samanheng med responsane som heilheit sjølv om forklaringsvariabelen ikkje hadde samanheng med kvar einskild respons.

Prinsipalkomponentanalyse (PCA). PCA (Mardia, Kent, & Bibby, 1979) er ein uovertaka teknikk for utforsking av eit datasett. Datasettet si kovariansmatrise let seg dekomponere i eigenverdiar med tilhøyrande eigenvektorar. Som eit resultat får ein nye variablar som er lineærkombinasjonar av dei opprinnelege variablane og som står ortogonalt på kvarandre, kalla prinsipalkomponentar. Den første prinsipalkomponenten, PC1, forklarar mest av variasjonen i datasettet. Den andre prinsipalkomponenten, PC2 forklarar nest mest av variasjonen i datasettet, og vidare til all variasjonen er forklart. PCA kan gje ny innsikt i samhengane mellom variablane, og visuell inspeksjon av figurar der respondentane og variablane representert ved prinsipalkomponentar kan vere svært nyttig i utforskingaugemed.

PoissonPCA er ei metode av noko nyare dato, som estimerer prinsipalkomponenter med utgangspunkt i ein transformasjon av dei underliggende poissonmiddela (Kenney, Huang, & Gu, 2019) og testa mot liknande metodar har vist betre evne til å finne dei prinsipalkomponentane som beskriv mest variasjon og er i tillegg raskare i bruk og meir robust mot grove feil.

Prinsipalkomponentregresjon (PCR) er når alle eller helst nokre av prinsipalkomponentane frå PCA blir nytta som responsar eller forklaringsvariablar i staden for dei opprinnelege variablane. I dei fleste høver nyttar ein desse som forklaringsvariablar, og ein vel det antalet variablar som i dei einskilde høva blir ansett som tilstrekkeleg for å forklare variasjonen i dei opprinnelege variablane. Det er særst viktig å merkje seg at dei prinsipalkomponentane som forklarar mest av variasjonen i forklaringsvariablane heller ikkje nødvendigvis evnar å forklare noko i responsvariablane, og varsam utforsking av datasettet er difor ein føresetnad for bruk av PCR.

McNemars test høver for parvise binære data, og testar i motsetning til kjikvadrattabellar ikkje for uavhengigheit, men for konsistens i respons over to variablar (McNemar, 1947).

Invertert beta-binomisk (Pham & Jimenez, 2012) test vart nytta for å avgjere om det er skilnadar mellom parvise tellevariablar. Dette gjer og høve til samanlikning med para t-test.

Sign test (Hogg, McKean, & Craig, 2005, p. 540) av parvise variablar er ein ikkje-parametrisk test som kun forutset at observasjonane er uavhengige, tilfeldige og at dei to variablane er para. Kort fortalt vil testen sjå på om skilnaden mellom to parvise variablar har ein median lik 0. Sign test vil generelt ha ein noko lågare teststyrke enn para t-test, men har høgare teststyrke der forskjellen mellom para ikkje er normalfordelt men heller har tyngre halar. Tidlegare nevnte McNemars test nyttar det same rammeverket, men dog kun for binære variablar.

Generaliserte lineære modellar (GLM), og herunder poisson-regresjon og logistisk regresjon er i dei tilhøva responsane i lineære modellar eksempelvis er binomisk- eller poissonfordelt (Dobson & Barnett, 2018).

Rang-regresjon (Kloke & McKean, 2012) gjev moglegheita til å avdekkje lineære samanhengar mellom rangerte variablar heller enn deira opprinnelege verdiar, og vil vere meir robust for ekstreme verdiar.

For undersøking av klynger, vart det foretatt ei agglomerativ hierarkisk klynging (Hastie, Tibshirani, & Friedman, 2009, p. 523) med utgangspunkt korrelasjonsmatrisa og vidare midlare avstandar i samanhengar mellom spørsmåla. Dette er ein uovervaka metode som nyttar avstand i korrelasjon til å klynge variablane saman, og framstilling i dendrogram gjer høve til visuell inspeksjon av samanhengar.

Partiell minste kvadraters diskriminantanalyse, PLS-DA. PLS-DA freistar å maksimere samvariasjonen mellom forklaringsvariablar og respons, med enklare ord, lage nye lineærkombinasjonar som forklarar mest mogeleg av variasjonen mellom forklaring og respons (Barker & Rayens, 2003). For diskriminantanalyser vert PLS-DA ansett som gullstandarden, og vert nytta som samanlikningsgrunnlag for nye metodar frå maskinlæring (Mendez, Reinke, & Broadhurst, 2019).

Sjølv om ein kan etter vanlege tommelfingerreglar i statistikk byrje å sjå vekk frå fordeling når antal målingar er større enn 30, vart det under visuell inspeksjon av variablane klart at fleire av sensortypene i si opprinnelege form er tellevariablar, og ofte har låge førekomstar av fenomenet ein ynskjer å registrere. Som førebuing av komande prosjekt, samt best mogeleg val av variablar og metodar når responstypene skal sjåast i samanheng med kvarandre og med personlegdomstypar vart det utført både testar som forutset tilnærma normalfordelte responsar, testar spesielt for ein viss type fordeling som til dømes binomisk eller poisson og testar som ikkje forutset fordeling i det heile. Ein fekk då høve til samanlikning, og mogelegheit til å gjere eit grunna val om ein vil nytte eit velkjend rammeverk som forutset tilnærma normalfordeling av restfeila og konstant variasjon, eller om ein fann at å nytte andre metodar er verdt den ekstra ressursbruken og mogelege tap av teststyrke det kunne medføre.

Vidare måtte det takast bestemmingar om korleis ein skulle måle pålitelegheita i resultatane frå forsøket. Det aller beste er di som ein får høve til å legge til sides ein del av respondentane, og nytte desse som eit heilt uavhengig test-datasett. Dette vil i stor grad vere avhengig av kor mange respondentar som deltek i forsøket og som svarar på utdanningstesten. Sidan første omgang av datainnsamling gav 22 komplette enkeltforsøk derav 18 som svarte på utdanningstesten, vart bestemminga utsett til etter andre omgang av datainnsamling vart gjennomført. Då denne resulterte i 19 komplette enkeltforsøk derav 17 som svarte på utdanningstesten, var det naturleg at alle respondentane samlast til eit datasett og nyttast i analysene. Før denne bestemminga vart teken, vart det av metodiske hensyn nytta eit eige datasett til utforsking av datatypar og variablar. Eit eige, men identisk forsøk i iMotions vart nytta til å framskaffe testsettet. Som respondent nytta forfattern seg sjølv i fire tilhøve. Mellom anna vart prototypiske ansiktsuttrykk og framprovosering av GSR nytta for å auke kjennskapen til variablane frå dei forskjellige responstypene. Dette testsettet gav høve til å klargjere programmeringsskript som rensa og strukturerte dei forskjellige responstypene utan at datasetta frå forsøket vart røyvd.

Antakingar om variablane sine fordelingar vart tufta på heilheitsvurderingar av visuell inspeksjon av frekvens- og spredningsdiagram, kjennskap til målemetodar og kontroll av residual- og qq-plott der dette er relevant.

For alle statistiske testar vart det nytta signifikansnivå 0.05, sjølv om ein er medveten om at oppgåva i stor grad er utforskande, og at ein form for korreksjon av signifikansnivå er naturleg når så mange testar utførast.

### 3 RESULTAT

---

Det vil først bli presentert nærare beskriving av kvar responstype som syner kva den einkilde resultatfila inneheld, med tilhøyrande beskrivande statistikk. Samanhengar innad i kvar responstype vert forsøkt avdekkja, og parvise samanlikningar vert gjort der det er høveleg. Vidare vil samanhengar mellom responstypene bli utforska. Det vil vere av særskild interesse å avdekkje om dei forskjellige responstypene evnar å støtte opp om kvarandre, og dermed gje høve til meir pålitelege svar på samhengane mellom opplevinga av undervisningsfilmane og personlegdomstypene. Til sist vert samhengane mellom personlegdomstypene og dei forskjellige responstypene undersøkt, både kvar for seg og samla.

Som det vil framkome av resultata er det utført mange statistiske testar supplert med visuell inspeksjon av spredningsdiagram med eller utan fargekoding av respondentane etter forskjellige variablar. Det er svært få statistisk signifikante resultat, og det blir difor synt eit utval av dei ikkje-signifikante resultata. Dei ikkje-signifikante resultata er presentert for å gje komande arbeid både samanlikningsgrunnlag og lærdom, for å syne grunnlaget for heilheitsvurderingar av resultata og ikkje minst fordi den overveldande mengda ikkje-signifikante resultat i seg sjølve er eit spanande resultat.

## 3.1 RESULTAT FRÅ SPØRJEUNDERSØKINGA

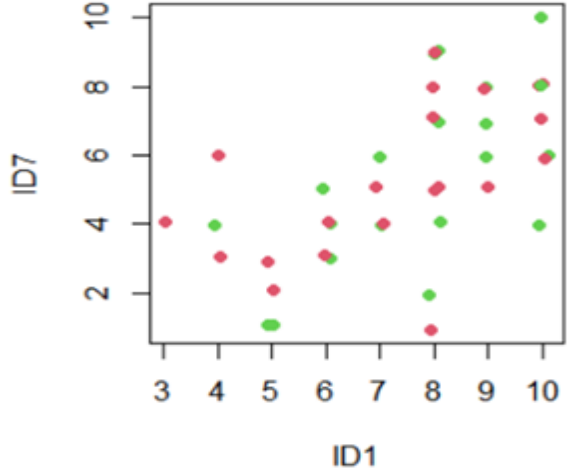
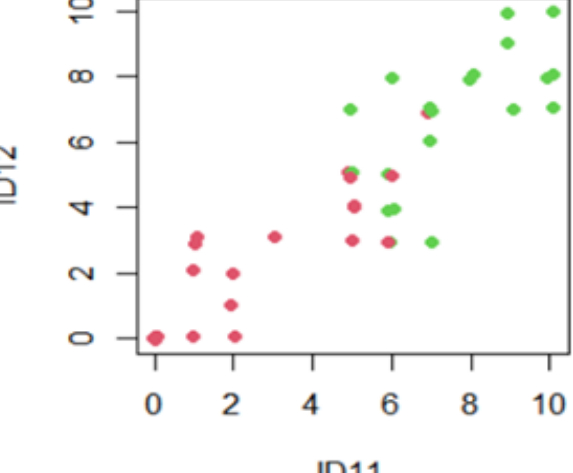
### 3.1.1 Beskrivande statistikk

Det første ein merka seg ved inspeksjon av nøkkeltala frå spørjeundersøkinga i vedlegg 3, Tabell 3-A, er at respondentane i stor grad er einige i at statistikk er nyttig i framtidige studiar og i arbeidslivet. Respondentane var i mindre grad einige i positivt lada utsegn om filmene, til dømes at filmene var engasjerande, at dei forsto det som vart presentert, og at dei oppfatta temaet som relevant. Respondentane var derimot meir einige i at filmene var frustrerande. Sjølv om det fantast hovudtrekk ved korleis dei fleste respondentane svarte på spørjeundersøkinga, synte ho og ei stor spreing blant respondentane sine meiningar om kor nyttig eller vanskeleg statistikk er, og kva dei tykte om filmene. Dette syner og att i frekvensdiagramma i vedlegg 3, Figur 3-A

Det vart berekna Pearsons korrelasjonskoeffesient ( $r$ ) for alle spørsmål i vedlegg 3, Tabell 3-A, Særskilt interessante resultat kan vere dei presentert i Tabell 3. Det er stor lineær samanheng ( $r = 0.79$ ) mellom kor nyttig dei meiner statistikk er (ID1), og kor vanskeleg dei tykkjer det er (ID7). Det er og høg korrelasjon ( $r=0.9$ ) mellom kva film dei sjølve tykte dei lærte mest av (ID11), og kva film dei trur deira medstudentar vil få best utbytte av (ID12). Det er og verdt å merke seg at kva film dei ville valt om dei berre fekk sjå ein film i STAT100 verkar etter visuell inspeksjon ikkje til å ha noko betydning for kor viktig og vanskeleg dei tykkjer statistikk er. Men dette er ikkje tilhøvet for kva film dei lærte mest av og kva film dei trur deira medstudentar vil få best utbytte av. Her er det eit visuelt markant skilje mellom svara på ID11 og ID12 sett i samanheng med kva film respondentane ville sett i STAT100 om dei berre kunne velje ein. Dette skiljet er og ved nokre høve tilstade ved visuell inspeksjon av spredningsdiagram der parvise spørsmål er plotta mot kvarandre, sjå vedlegg 3, Figur 3-B. ID15 og ID11 verkar til kunne avdekkje to forskjellige typar preferansar som mogelegvis tolkast til kva film dei ville ha høve til og sjå igjen, og kva film dei tykte dei lærte mest av. Ei anna moglegeheit kunne vere at svara frå ID15 kan reflektere kva film dei tykte dei forsto minst, og difor kunne tenkje seg å få sjå heile eller delar av filmen på nytt. Skiljet mellom ID15 og ID11 vart difor ytterlegare utforska ved statistiske testar.



Tabell 3 Tabell som viser to særskilt interessante samanhengar mellom spørsmål med tilhøyrande spredningsdiagram. Punkta i grøn farge syner dei som ville valt Hilde sin film om dei kun hadde fått sett ein av dei, medan punkta i raud farge syner dei som ville valt Kathrine sin film.

SpørsmålsID og kort beskriving av spørsmålet.	Korrelasjonskoeffesient	Spredningsdiagram. Grøn = Hilde. Raud = Kathrine.
<p>ID1: Kor nyttig statistikk vil vere for framtidige studier.</p> <p>ID7: Både konseptene og utregningene er vanskelige.</p>	0.79	
<p>ID11: På en skala fra film Hilde (0) til film Kathrine (10), hvilken av videoene følte du du lærte mest av.</p> <p>ID12: På en skala fra film Hilde (0) til film Kathrine (10), hvilken av videoene tror du studenter flest får best utbytte av i stat100.</p>	0.90	

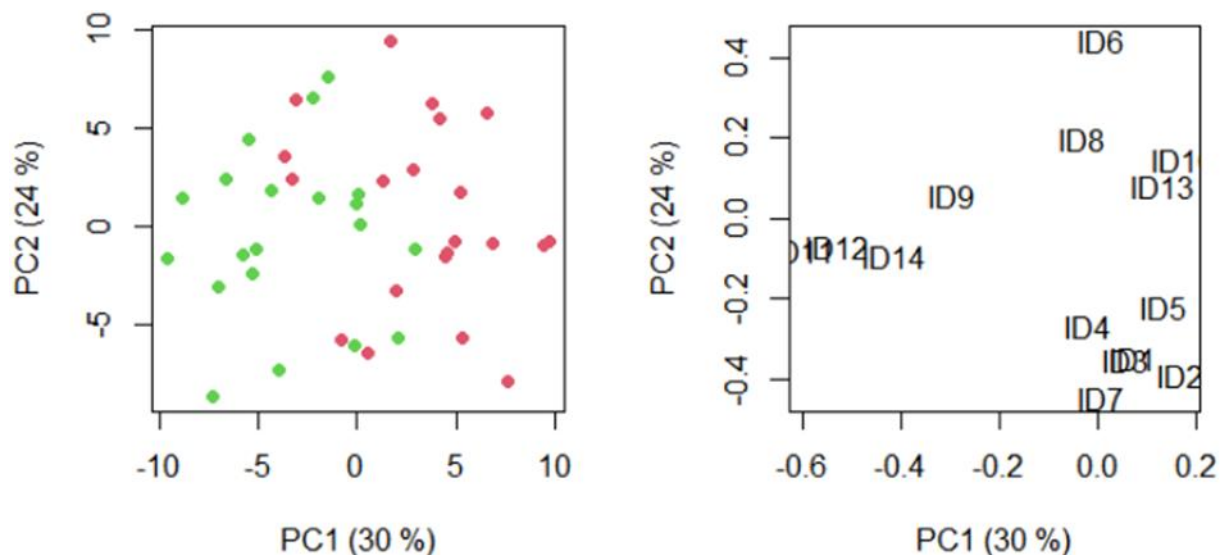
### 3.1.2 Prinsipalkkomponentanalyser

For å få eit nærare innblikk i samanhengane i spørjeundersøkinga vart datasettet delt i enkeltspørsmål og parvise spørsmål. Desse vart analysert kvar for seg og til slutt saman ved hjelp av PCA. Prinsipalkkomponentar frå enkeltspørsmåla får suffiks ES og prinsipalkkomponentar frå parvise spørsmål får suffiks PS. Prinsipalkkomponentane frå den fulle spørjeundersøkinga unnateke dei særskild interessante variablane ID15 og ID11 får suffiks SU. Alle spørsmål var på same skala, og skalering av komponentane var difor ikkje naudsynt.

### Enkeltspørsmål ID1 – ID14

Første og andre prinsipalkomponent (PC1\_ES og PC2\_ES) forklarte henholdsvis 30% og 24% av variasjonen i enkeltspørsmåla. PC3\_ES forklarte 14% av variasjonen, og forklaringsgrada minka jamt vidare. Ein måtte nytte til og med PC8\_ES for å forklare 95% av variasjonen i enkeltspørsmåla.

Av scoreplottet i Figur 3 ser ein lite antyding til tydelege geografiske mønster og klynger i respondentane fordelt ved dei to prinsipalkomponentane. Først når ein ser på fargelegginga kunne det sjå ut som PC1\_ES bidrog til å forklarar skiljet mellom dei som føretrekk Hilde sin film og dei som føretrekk Kathrine sin film. Ei meir tydeleg geografisk fordeling såg ein i ladningsplottet, til høgre i Figur 3. PC1\_ES såg ut til å halde informasjon om meiningar om filmene. Det verkar til at PC2\_ES skil spørsmål om nyttingheit og kva som er vanskeleg frå spørsmål om arbeidsmengde og kva som er enkelt.



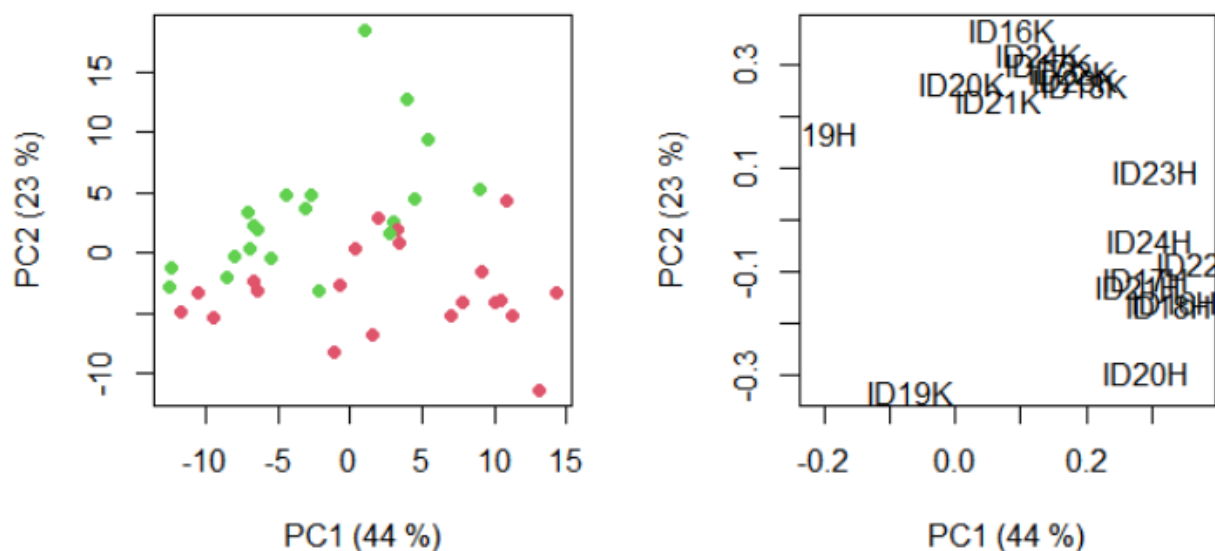
Figur 3 Prinsipalkomponentanalyse av enkeltspørsmål, spørsmål ID1 – ID14. T.v. Scoreplot som viser fordelinga av respondentar representert ved PC1 og PC2. T.h. Ladningsplot som viser fordelinga av variablar representert ved PC1 og PC2. Punkta i grøn farge syner dei ville valt Hilde sin film om dei kun hadde fått sett ein av filmene, medan punkta i raud farge syner dei som ville valt Kathrine sin film om dei kun hadde fått sett ein av dei.

### Parvise spørsmål ID16 – ID24

PC1\_PS og PC2\_PS forklarte henholdsvis 44% og 23% av variasjonen i dei parvise spørsmåla. PC3\_PS forklarte 9%, og PC4\_PS forklarar 5%. Vidare avtok forklaringsgrada sakte. Ein må nytte til og med PC10\_PS for å forklare 95% av variasjonen i dei parvise spørsmåla.

Ein ser av Figur 4 at skilje mellom respondentane etter kva film dei ville sett i STAT100 om dei berre kunne velje ein av filmene var meir markert for dei parvise spørsmåla enn enkeltspørsmåla, og kjem no til uttrykk av PC2\_PS. I tillegg ser ein at spreinga PC1\_PS og PC2\_PS aukar for høgare verdiar.

Ein kan sjå at spørsmåla om kvar einskild film klyngar seg saman, nesten heilt innad i kvar sin kvadrant. Ein kunne og sjå at ID19, om filmen var frustrerende, låg for seg sjølve i motståande kvadrant i forhold til dei øvrige spørsmåla i same film. Det skuldast at dette spørsmålet var det einaste spørsmålet der ein høg score er negativt. Ein observerte og at spørsmåla om Hilde sin film varierte mest i utstrekning langs PC2\_PS, og spørsmåla om Kathrine sin film varierte mest i utstrekning langs PC1\_PS. Ei mogleg tolking kunne vere at PC1\_PS og PC2\_PS saman synte føretrakt film og spreing i meiningar om kvar einskild film.



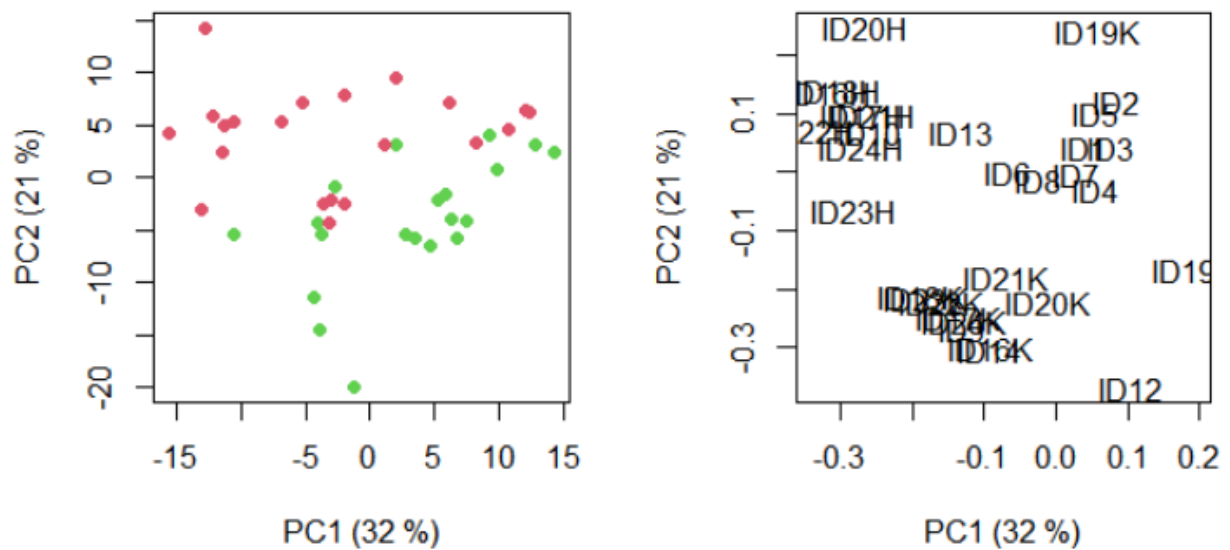
Figur 4 Prinsipalkomponentanalyse av parvise spørsmål, spørsmål ID16 – ID24. T.v. Scoreplot som viser fordelinga av respondentar representert ved PC1 og PC2. T.h. Laddingsplot som viser fordelinga av variablar representert ved PC1 og PC2. Punkta i grøn farge syner dei ville valt Hilde sin film om dei kun hadde fått sett ein av filmene, medan punkta i raud farge syner dei som ville valt Kathrine sin film om dei kun hadde fått sett ein av dei.

### Samla PCA for spørjeundersøkinga unnateke ID11 og ID15

PC1\_SU og PC2\_SU forklarar henholdsvis 32% og 21% av variasjonen i spørsmåla. PC3\_SU forklarar 10%, og PC4\_SU forklarar 5%. Vidare avtek forklaringsgrada sakte. Ein må nytte til og med PC15 av dei 31 prinsipalkomponentane for å forklare 95% av variasjonen i spørsmåla.

Av Figur 5 ser ein naturleg nok att mykje av tendensane frå prinsipalkomponentanalysene av enkeltspørsmåla og dei parvise spørsmåla kvar for seg. I tillegg kan ein merke seg at skiljet mellom respondentane som ville valt Hilde sin film, og respondentane som ville valt Kathrine sin film verkar til å vere noko mindre visuelt tydelegare for den fulle undersøkinga.

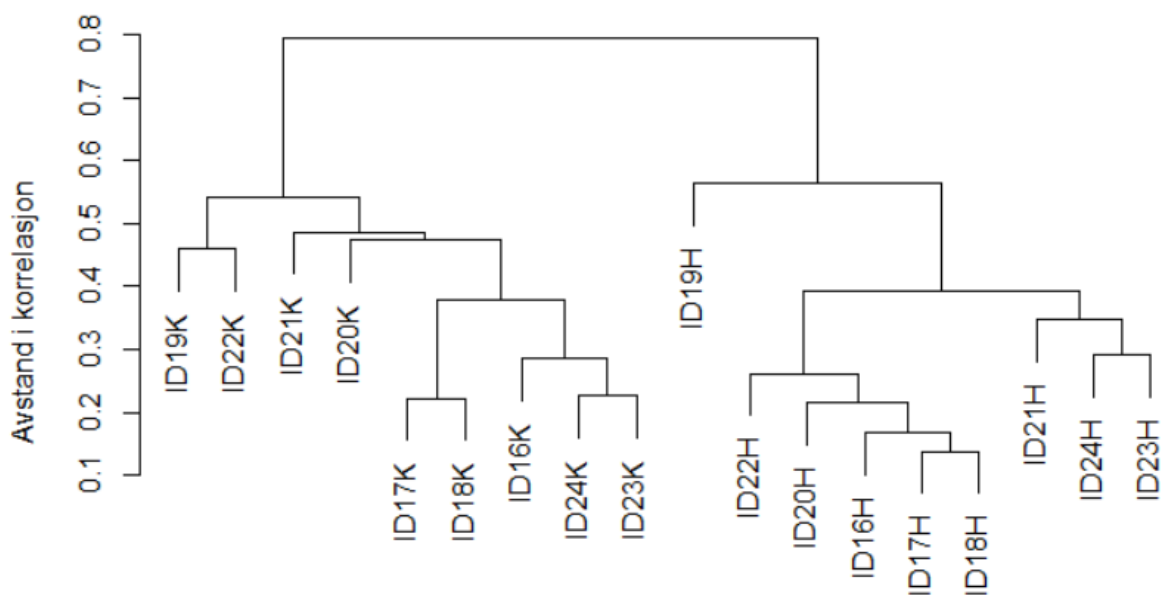
Enkeltspørsmåla, spørsmåla om Hilde sin film og spørsmåla om Kathrine sin film dannar og tre hovudklynger, dog med nokre variablar sprett utanfor klyngene. Det verkar til at PC1\_SU og PC2\_SU i samhandling beskriv føretrekt film og meiningar om statistikk. PC2\_SU kan halde noko meir informasjon om meiningar om statistikk, medan kva film respondentane føretrekk vert skild ved både PC1\_SU og PC2\_SU. PC1\_SU verkar og til ei viss grad å omhandle filmene for verdiar lågare enn 0, og meiningar om statistikk for verdiar høgare enn 0. Det er, i lys av PC1\_SU, ikkje utenkjeleg at respondentane si einigheit om at filmene var frustrerende og kan reflektere deira haldning til faget i seg sjølve.



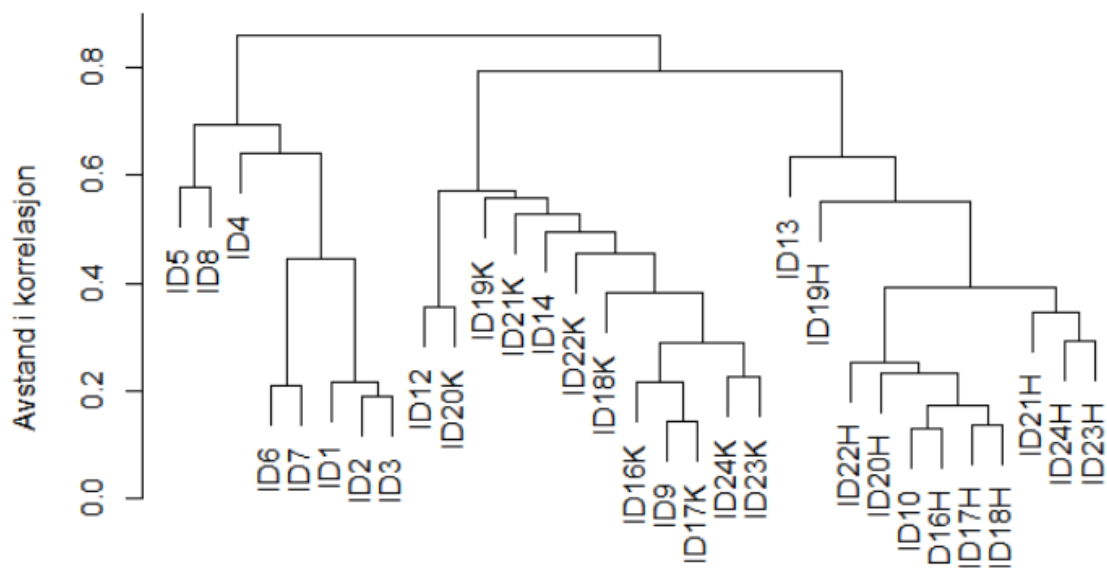
Figur 5 Prinsipalkomponentanalyse av alle spørsmål unnateke ID11 og ID15. T.v. Scoreplot som viser fordelinga av respondentar representert ved PC1 og PC2. T.h. Laddingsplot som viser fordelinga av variablar representert ved PC1 og PC2. Punkta i grøn farge syner dei ville valt Hilde sin film om dei kun hadde fått sett ein av filmene, medan punkta i raud farge syner dei som ville valt Kathrine sin film di som dei kun hadde fått sett ein av dei.

### 3.1.3 Hierarkisk klynging etter midlare avstand i korrelasjon

For ytterlegare utforsking vart det utført ei agglomerativ hierarkisk klynging. Av dendrogrammet i Figur 6 kan ein sjå tydeleg gruppering av spørsmåla om Hilde og Kathrine sine filmar. For Figur 7 kan ein merke seg at ID10 som er oppfølgingsspørsmålet til Hilde sin film er klynga saman med dei parvise spørsmåla som omhandlar Hilde sin film. ID9 som er oppfølgingsspørsmålet til Kathrine sin film er på same viset klynga saman med spørsmåla om Kathrine sin film. Dendrogramma har likskap med observasjonane frå prinsipalkomponentanalyse grunna bruk av korrelasjon som avstandsmål, og denne visualiseringa er eit godt supplement til ladningsplotta der mange av variablane kan vere oppteikna over kvarandre.



Figur 6 Dendrogram som syner hierarkisk klynging av parvise spørsmål etter midlare avstand i korrelasjon..



Figur 7 Dendrogram som syner hierarkisk klynging av alle spørsmåla samstundes etter midlare avstand i korrelasjon.

### 3.1.4 Statistiske testar

#### **Har rekkefølga filmene vart synt i samanheng med kva film respondentane hadde valt di som dei berre fekk sjå ein av dei i STAT100 (ID15)?**

Frå McNemars test finn ein at  $p$ -verdi = 0.83. Det er ingen samanheng mellom kva film respondentane såg først, og kva film dei hadde valt om dei berre fekk sjå ein av dei i STAT100.

#### **Har rekkefølga filmene vart synt i samanheng med spørjeundersøkinga?**

Frå MANOVA,  $p$ -verdi = 0.497. Rekkefølga filmene vart synt i hadde ingen samanheng med på spørjeundersøkinga eller nokon av enkeltpørsmåla.

#### **Har ID15, kva film respondentane ville valt di som dei kun fekk sjå ein film i STAT100 samanheng med spørjeundersøkinga?**

Frå MANOVA,  $p$ -verdi = 0.0109. Kva film respondentane ville valt di som dei kun fekk sjå ein film i STAT100 hadde samanheng med spørjeundersøkinga som heilheit. Nyttå forklaringsvariabel i den lineære regresjonen er Hilde = 1 og Kathrine = 0.

#### **Har ID11, kva film respondentane lærte mest av samanheng med spørjeundersøkinga?**

Frå MANOVA,  $p$ -verdi =  $<0.001$ . Kva film respondentane ville lært mest av frå en skala frå Hilde (0) til Kathrine (10) har samanheng med spørjeundersøkinga som heilheit.

Kvar enkeltrespons vart inspisert og spørsmål ID15 og ID11 hadde samanhengar med er samla i Tabell 4 og Tabell 5 der ein kan observere at for fleire av dei parvise spørsmåla var begge responsane signifikante på eiga hand. Der kun ein av dei var signifikant, var dette i alle høve gjeldande for Hilde sin film. Det er og verdt å merke seg at ingen av spørsmåla som vart stilt før filmene vart synt hadde samanheng med ID15, så kor nyttig eller vanskeleg respondentane verkar ikkje til å ha noko samanheng med kva film respondentane ville valt om dei kun fekk sjå ein film i STAT100. Ved samanlikning av Tabell 4 og Tabell 5 ser ein at det er noko skilje mellom kva responsar ID15 og ID11 har samanheng med. Ein kan likevel observere at det i begge høver er motsett forteikn på stigningstala i kvart parvise spørsmål. For Hilde er det i begge høver negativt forteikn, og positivt forteikn for Kathrine. I lys av ID11, viser dette at dess høgare score, dess meir vil scoren auke på spørsmål om Kathrine sin film. Ein høg score på ID11 vil føre til ei minsking av score på spørsmåla om Hilde sin film. Dette er naturleg nok omvendt på ID19, filmen er frustrerande. ID15 og ID11 har og oftare samanheng med spørsmål om Hilde sin film enn om Kathrine sin film. Auka i  $R^2$  for dei aller fleste spørsmåla ved å nytte ID11 som respons er og verdt og merke seg.

Tabell 4 Responsar der ID15, kva film respondentane hadde valt om dei kun fekk sjå ein av filmene i STAT100, hadde samanheng med spørsmåla i spørjeundersøkinga.

Respons	stigningstal	R <sup>2</sup>	p-verdi
ID9 Utbytte av Kathrine sin film	1.85	0.16	0.009
ID10 Utbytte av Hilde sin film.	-2.19	0.21	0.002
ID11 Hvilken av videoene følte du du lærte mest av. *	4.83	0.58	<0.001
ID12 Hvilken av videoene tror du studenter flest får best utbytte av i stat100. *	4.42	0.52	<0.001
ID16H Filmen engasjerte meg -Hilde	-2.76	0.22	0.002
ID16K Filmen engasjerte meg -Kathrine	2.16	0.19	0.004
ID17H Jeg forsto det som ble presentert -Hilde	-2.04	0.15	0.012
ID17K Jeg forsto det som ble presentert -Kathrine	1.64	0.13	0.019
ID18H Det var greit å følge med på den tekniske notasjonen -Hilde	-2.24	0.15	0.013
ID19H Filmen var frustrerende -Hilde	2.01	0.10	0.039
ID20H Jeg likte oppbygningen av filmen -Hilde	-3.69	0.40	<0.001
ID20K Jeg likte oppbygningen av filmen -Kathrine	2.19	0.32	<0.001
ID21H Eksempelet var interessant -Hilde	-3.2	0.31	<0.001
ID22H Jeg forsto hvorfor variasjonen til gjennomsnittet var mindre enn variasjonen til enkeltverdiene -Hilde	-2.45	0.15	0.014
ID24H Filmen fikk meg til å forstå den praktiske relevansen til dette temaet -Hilde	-2.04	0.13	0.023

\*Svara blir gjedd på ein skala frå Hilde (0) til Kathrine (10)

Tabell 5 Responsar der ID11, kva film respondentane følte dei lærte mest av på ein skala frå Hilde (0) til Kathrine (10), hadde samanheng med spørsmåla i spørjeundersøkinga.

Respons	stigningstal	R <sup>2</sup>	p-verdi
ID9 Utbytte av Kathrine sin film	0.37	0.27	0.006
ID10 Utbytte av Hilde sin film.	-0.39	0.28	<0.001
ID12 Hvilken av videoene tror du studenter flest får best utbytte av i stat100. *	0.84	0.80	<0.001
ID13 Hvilken av videoene var lengst *	-0.24	0.10	0.043
ID14 Jobber mye med faget	0.48	0.20	0.003
ID16H Filmen engasjerte meg -Hilde	-0.53	0.33	<0.001
ID16K Filmen engasjerte meg -Kathrine	0.43	0.30	0.002
ID17H Jeg forsto det som ble presentert -Hilde	-0.39	0.23	0.002
ID17K Jeg forsto det som ble presentert -Kathrine	0.32	0.20	0.003
ID18H Det var greit å følge med på den tekniske notasjonen -Hilde	-0.49	0.29	<0.001
ID19H Filmen var frustrerende -Hilde	0.54	0.31	<0.001
ID20H Jeg likte oppbygningen av filmen -Hilde	-0.66	0.56	<0.001
ID20K Jeg likte oppbygningen av filmen -Kathrine	0.39	0.41	<0.001
ID21H Eksempelet var interessant -Hilde	-0.32	0.12	0.023
ID22H Jeg forsto hvorfor variasjonen til gjennomsnittet var mindre enn variasjonen til enkeltverdiene -Hilde	-0.41	0.16	0.008
ID24H Filmen fikk meg til å forstå den praktiske relevansen til dette temaet -Hilde	-0.30	0.11	0.034
ID24K Filmen fikk meg til å forstå den praktiske relevansen til dette temaet -Kathrine	0.29	0.15	0.012

\*Svara blir gjedd på ein skala frå Hilde (0) til Kathrine (10)



### **Er det forskjell mellom Hilde og Kathrine sin film for kvart par av spørsmål?**

Kun tre av ni parvise spørsmål viste signifikant skille mellom Hilde sin film og Kathrine sin film, sjå Tabell 6. Om det tydar at filmene ikkje er forskjellige nok til å gje ein statistisk signifikant forskjell, eller om respondentane ikkje har så sterke meiningar om undervisningsfilmar om grunnleggande statistikk kan vere ei mogelegheit. Det verkar dog meir sannsynleg at variasjonen i meiningar er så stor i forhold til forskjell i meiningar, at skilnadane mellom filmene har vanskeleg for å kome fram. Dette vert understøtta av nøkkeltala for spørjeundersøkinga, sjå vedlegg 3 Tabell 3-A, som syner eit standardavvik mellom 2 og 3.2 for dei parvise spørsmåla.

*Tabell 6 Resultat frå para t-test, der det er signifikant forskjell mellom Hilde og Kathrine sine filmar innad i kvart av dei parvise spørsmåla ID16-ID24*

Spørsmål	p-verdi
ID21 Eksempelet var interessant	<0.01
ID22 Jeg forsto hvorfor variasjonen til gjennomsnittet var mindre enn variasjonen til enkeltverdiene	0.039
ID24 Filmen fikk meg til å forstå den praktiske relevansen til dette temaet	0.041

### **Multivariat prinsipalkomponentregresjon**

Prinsipalkomponentane frå PCA\_SU vart nytta som responsvariablar. Kva film som vart synt først, kva film respondentane ville valt om dei berre fekk sjå ein av filmene i STAT100 (ID15) og ID11, kva film respondentane lærte mest av på ein skala frå Hilde (0) til Kathrine (10) vart nytta som forklaringsvariablar etter tur.

### **Har rekkefølja filmene vart synt i samanheng med prinsipalkomponentane?**

Det var signifikant samanheng mellom kva film respondentane såg først og prinsipalkomponentane PC8\_SU og PC11\_SU, som kvar forklarar henholdsvis 0.29% og 0.16% av variasjonen i spørjeundersøkinga.

### **Har kva film respondentane ville valt om dei kun fekk sjå ein film i STAT100 samanheng med prinsipalkomponentane?**

Av prinsipalkomponentane hadde ID15 signifikant samanheng med PC1\_SU, PC2\_SU og PC14\_SU. Desse forklarar henholdsvis 32 %, 21 % og 0.12 % av variasjonen i datasettet. ID15 gjev ein  $R^2$  på 0.13 for PC1\_SU, 0.38 for PC2\_SU og 0.09 for PC14\_SU.

**Har kva film respondentane følte dei lærte mest av på ein skala frå Hilde (0) til Kathrine (0) samanheng med prinsipalkomponentane ?**

ID11 hadde signifikant samanheng med PC1\_SU og PC2\_SU som responsar. Desse forklarar henholdsvis 32 % og 21 % av variasjonen i datasettet. ID11 gjev ein  $R^2$  på 0.126 for PC1\_SU og 0.68 for PC2\_SU.

Dette er eit godt eksempel på at den prinsipalkomponenten som beskriv mest av variasjonen i datasettet ikkje nødvendigvis fangar opp mest av variasjonen som let seg forklare ved valt modell. Både PC1\_SU saman med PC2\_SU, eller ID11 kan vere særskildt gode kandidatar som variablar når spørjeundersøkinga skal koblast saman med andre responsar eller med personlegdomstypene.

## 3.2 RESULTAT FRÅ AFFDEX STATISTICS

### 3.2.1 Beskrivande statistikk

AFFDEX statistics er ei ferdig generert resultatfil som inneheld resultata frå ansiktsuttrykkanalysa. Fila inneheld informasjon om respondentane, deira tildelte nummer samt alder og kjønn. Det finnast og informasjon om kva stimuli målingane angår og stimuliet si varigheit i millisekund og scener. Det er og fleire variablar og for kvar variabel er det beskrive kva terskel som er nytta, antal gongar terskelen er overskride og kor stor prosent av tida terskelen er overskride. Variabeltypane kjensler, engasjement og valens er kalkulert ut frå variabeltypen ansiktsuttrykk, og er nærare beskrive i Tabell 7.

Tabell 7 Kort beskriving av variabeltypane i den ferdigprosesserte resultatfila AFFDEX statistics

Type variablar	antal	beskriving
Kjensler	7	Beskriv dei sju kjenslene anger, sadness, disgust, joy, surprise, fear, contempt.
Ansiktsuttrykk	20	Ofte kalla action units (AU) og beskriv ansiktsuttrykk som smil, heving eller rynking av augebryn, hakeslepp og vidare.
Engasjement (Engagement)	1	Eit mål på ansiktsmuskulaturens aktivering og derav respondenten si grad av ansiktsuttrykking.
Valens	3	Eit mål på respondenten si oppleving. Valensen målast på ein skala frå -100 til 100, og uttrykkast ved variablane positiv, negativ og nøytral.
Merksemd (Attention)	1	Mål på respondenten sin fokus basert på hovudet si orientering.

I undersøkinga av AFFDEX statistics blir det nytta kjensler, engasjement, valens og merksemd, vidare samla omtalt som kjensler. Sidan Hilde sin film er 14 sekund lengre enn Kathrine sin film, vert kor stor prosent av tida ein variabel overskrid terskelen, heller enn talet på scener der ein variabel overskrid terskelen. I analyser blir dei opprinnelege engelske namna frå resultatfila nytta for enklare samanlikning med andre studiar og for å forenkle vidare arbeid.

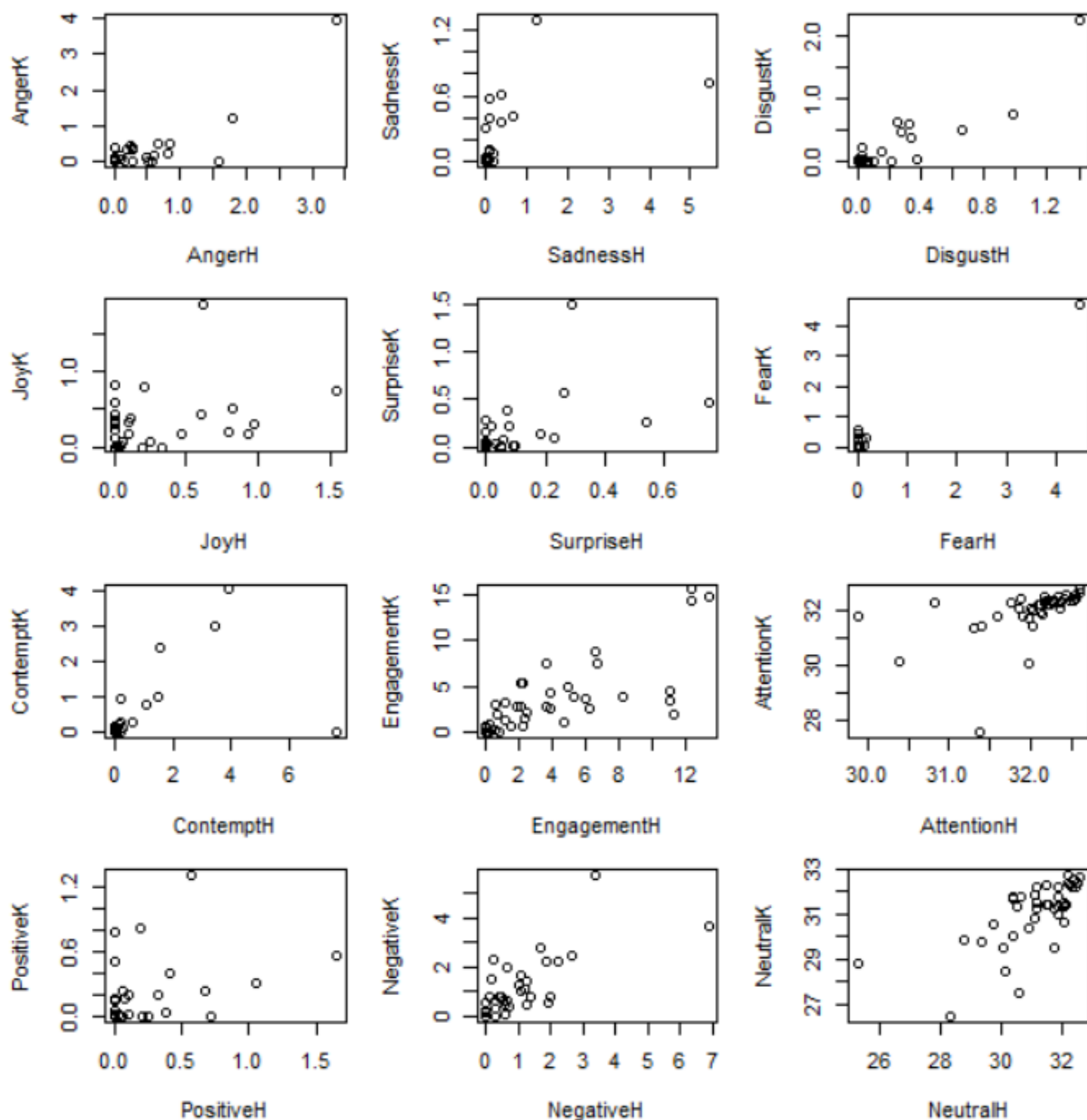
Innleiande undersøking byrja med inspeksjon av middel og standardavvik for kvar variabel i kvar av filmane, sjå Tabell 8. Frå Tabell 8 kan ein merke seg at det er til dels lite skilnad mellom Hilde sin film og Kathrine sin film om ein ser kun på middel og standardavvik. Dette syner ikkje heile sanninga, og om ein inspiserer vedlegg 4, Tabell 4-A og Tabell 4-B som syner for kvar respondent kor stor prosent av tida filmen varte kvar variabel oversteig terskelen. Der kan ein sjå ein stor variasjon både innad i kvar respondent og innad i filmane.

Frå kjennskap til variablane, veit ein at det er tellevariablar sjølv om dei vert uttrykt som prosent, og variablane blir ytterlegare undersøkt ved oppsummeringstabellar og frekvensdiagram. For frekvensdiagram for alle variablar, sjå vedlegg 4 Figur 4-A. Av Figur 8, samt vedlegg 4 Tabell 4-A og Tabell 4-B, ser ein at alle variablane unnateke neutral og attention har høg andel av 0 (kjensla har ikkje overkride terskelen ved nokre tilhøver) som indikerer poisson-fordeling. Dette vert og styrka ved å inspisere samanheng mellom middel og variasjon, og desse er tilnærma høveleg like, som og indikera at ein med rimeleg sikkerheit kan anta at variablane i Tabell 8 med unnatak av neutral og attention kjem frå ei poissonfordeling.

*Tabell 8 Tabell som syner gjennomsnitt og standardavvik i for Hilde sin film og Kathrine sin film for kvar av dei nytta variablane frå AFFDEX statistics.*

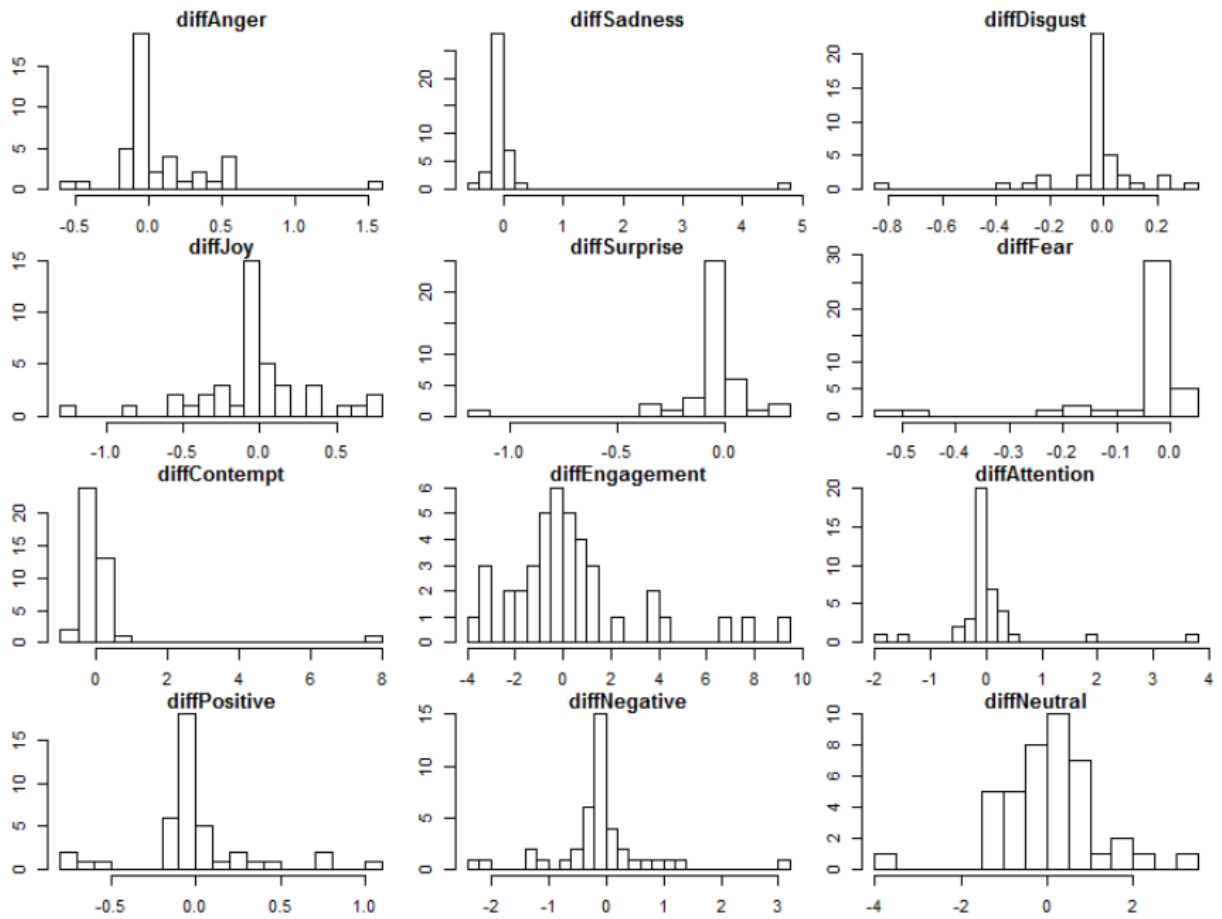
	anger	sadness	disgust	joy	surprise	fear	contempt	engagement	attention	positive	negative	neutral
Middel Hilde	0,32	0,22	0,13	0,20	0,07	0,12	0,52	3,84	32,00	0,17	0,89	31,14
S.avvik Hilde	0,64	0,87	0,29	0,36	0,15	0,70	1,41	4,03	0,57	0,34	1,27	1,39
Middel Kathrine	0,22	0,12	0,16	0,22	0,11	0,16	0,34	3,46	31,99	0,16	0,99	31,09
S.avvik Kathrine	0,64	0,27	0,39	0,36	0,26	0,73	0,85	3,92	0,89	0,28	1,18	1,41

Felles for mange av variablane for dei to filmene når dei blir satt saman til spredingdiagram for i Figur 8 er naturleg nok eit tyngdepunkt om 0, og ei tilnærma vifteforma utbeining frå opprinninga. Ein kan og merke seg at det for alle variablane er ei hovedklynge av respondentar, og det for mange av variablane er eit fåtal av respondentar som spreiar seg ut frå hovedklynga. Frå visuell inspeksjon av Figur 8 ser ein og at oppsummering av samanhengen mellom variablane med ein enkelt korrelasjonskoeffesient vil vere vanskeleg, og i nokre høver direkte missvisande. Eksempelvis vil variabelen Fear ha ein  $r = 0.987$ , grunna ei tett klynge nær origo og ein einskild ekstremverdi som for begge filmene er nær åtte gongar høgare enn hovedklynga sine verdiar. Denne lærdomen vil vere av interesse når responstypene skal sjåast i samanheng med kvarandre og personlegdomstypene, og påverknaden av val av undersøkingsmetodar kjem fram allereie ved samanlikning av ordinær PCA og poissonPCA i neste delkapittel.



Figur 8 Spredningsdiagram som syner dei parvise kjenslevariablane frå AFFDEX statistics. Variablar med suffiks H er frå Hilde sin film, variablar med suffiks K er frå Kathrine sin film.

Forskjell mellom i variablar mellom filmane vart funne ved å trekkje AngerK frå AngerH, og resultatane blir visualisert ved frekvensdiagramma i Figur 9. Ein kan merkje seg at alle differansane har som venta eit tydeleg tyngdepunkt om 0, og det er ofte nokre få observasjonar som skil seg betydeleg frå tyngdepunktet i anten eine eller begge retningar. Dette gjenspeglar og godt inntrykket frå spredningsdiagramma i Figur 8. Forskjell i variablar kan vere moglege kandidatar når responstypane skal knytast saman, då hovedansamlingane av observasjonar antek ei tilnærma normalfordeling. Ein må likevel utvise stor varsemnd grunna dei få observasjonane som ligg langt unna tyngdepunktet.



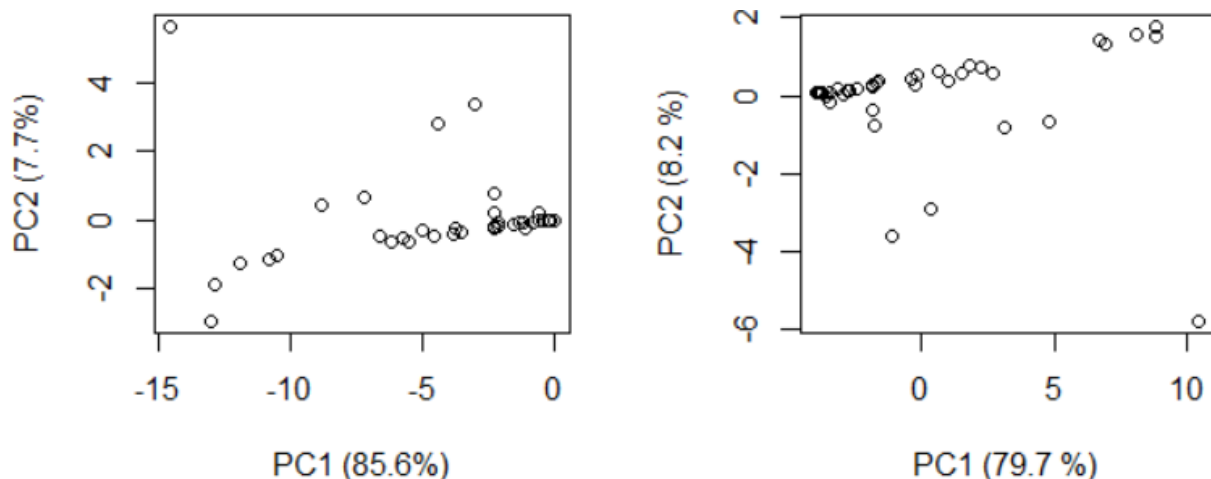
Figur 9 Forskjell i parvise kjenslevariabler frå AFFDEX statistics. Forskjellen er framkome ved å trekkje variablane frå Kathrine sin film frå variablane frå Hilde sin film.

### 3.2.2 Prinsipalkomponentanalyse

For alle PCA er Attention og Engagement er utelatt sidan dei er på ein betydeleg høgare skala, har eit lågt standardavvik, og ikkje syner same tendens til poissonfordeling som dei øvrige variablane. Dette gjev betre grunnlag for direkte samanlikning, og variablane vil vere på om lag same skala. For samanlikningsaugemed er det nytta både poissonPCA og PCA ved alle høver. Filmene blir analysert kvar for seg, så saman. Prinsipalkomponentane blir tildelt suffiks etter kva film som er nytta, H eller K. Prinsipalkomponentar der begge filmene er nytta får suffiks AS (AFFDEX statistics). Det vert nytta prefiks p for poissonkorrigerte prinsipalkomponentar.

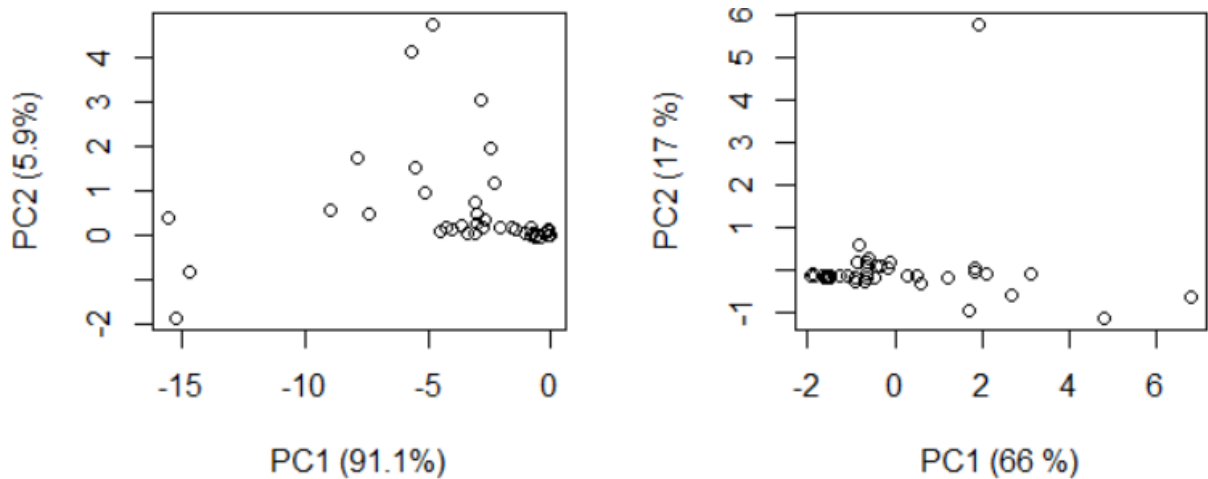
#### Prinsipalkomponentanalyser av filmene kvar for seg.

Frå scoreplotta i Figur 10 og Figur 11 ser ein at sjølv om punkta er spegla og rotert, er fasongen på utbreiinga ved dei to metodane forholdsvis like einannan. Det er tydeleg ansamling av respondentar om 0 for alle dei fire tilhøva av PC2, men ein kan observere at nokre respondentar skil seg mykje frå hovudansamlingane. Både pPC1\_H og pPC1\_K nyttar ein vesentleg større skala enn PC1\_H og PC1\_K i begge filmene. Dette er særskild tydeleg når ein samanliknar pPC1\_K og PC1\_K, og det er stor skilnad i andel forklart variasjon. Det er og vesentleg skilnad i spreing mellom p\_PC2\_K og PC2\_K, samanlikna med skilnad i spreing mellom p\_PC2\_H og PC2\_H



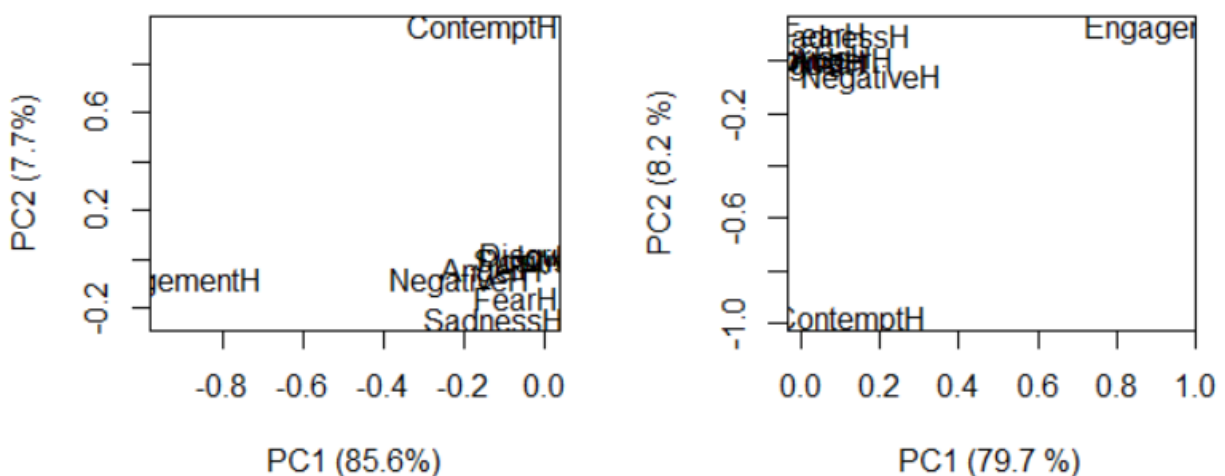
Figur 10 Scoreplot som syner respondentane representert ved PC1 og PC2 frå Hilde sin film. PoissonPCA t.v., PCA t.h.



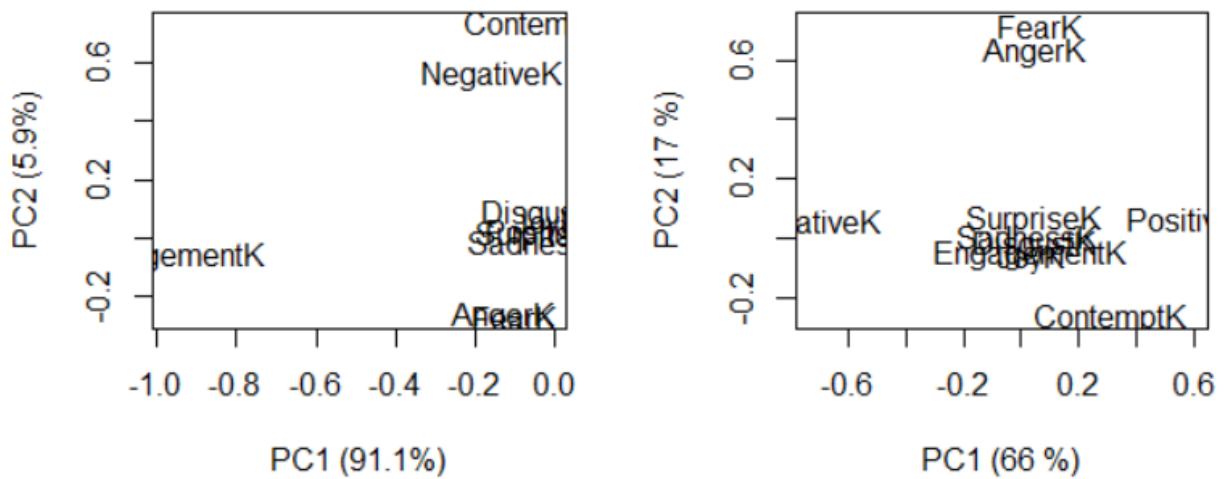


Figur 11 Scoreplot som syner respondentane representert ved PC1 og PC2 frå Kathrine sin film. PoissonPCA t.v., PCA t.h.

Frå Figur 12 kan ein sjå at utanom spegling om aksane er ladningsplotta for pPCA\_H og PCA\_H sin film svært like. Det kan verke til at fordelinga av respondentar langs pPC2\_H og PC2\_H vert i stor grad styrt av variabelen contemptH, medan fordelinga av respondentar langs pPC1\_H og PC1\_H vert styrt av engagementH. Den same klåre utskiljinga av to variablar er ikkje tilstades i begge ladningsplotta for Kathrine sin film synt i Figur 13. For begge metodane har Kathrine sin film den same hovudklynga i origo som Hilde sin film, men fleire av variablane i Kathrine sin film riv seg laus frå klynga, særskilt for PCA\_K. Utbreiing av respondentar langs PC1\_K verkar til å i stor grad bli styrt av negativeK og positiveK, som gjer høve til ei logisk og anvendeleg tolking. pPC1\_K ser i stor grad ut til å halde informasjon om grad av engagementK, som ikkje seier noko om kva respondenten meiner om filmen i motsetnad til PC1\_K. PC2\_K får sine sterkaste føringar frå angerK og fearK for verdiar høgare enn 0, og for contempt ved lågare verdiar enn 0. Det kan vere ei mogeleg tolking at pPC2\_K og PC2\_K i stor grad held variasjon i negative kjensler.

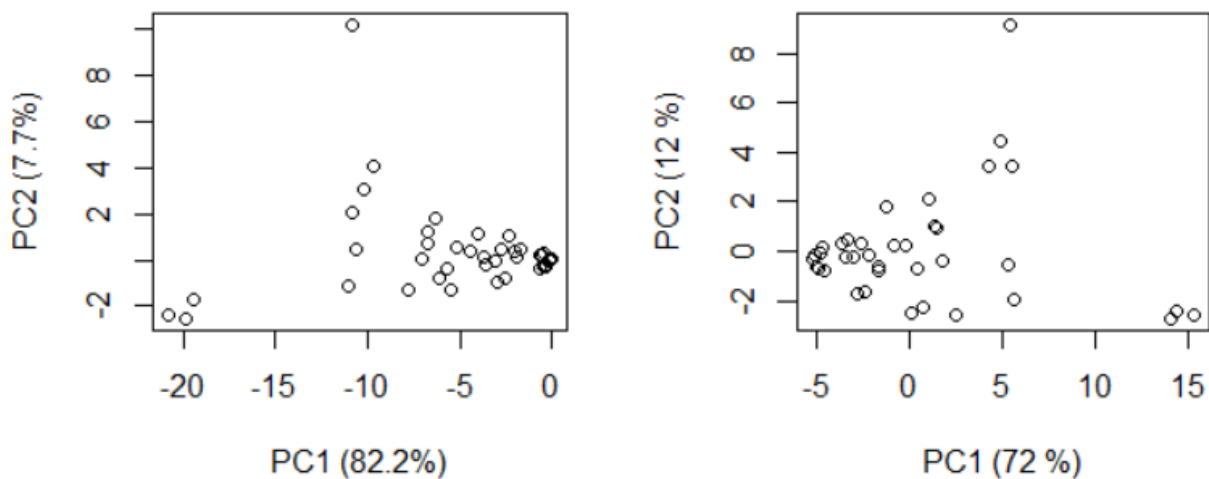


Figur 12 Ladningsplot som syner variablane representert ved PC1 og PC2 frå Hilde sin film. PoissonPCA t.v., PCA t.h.

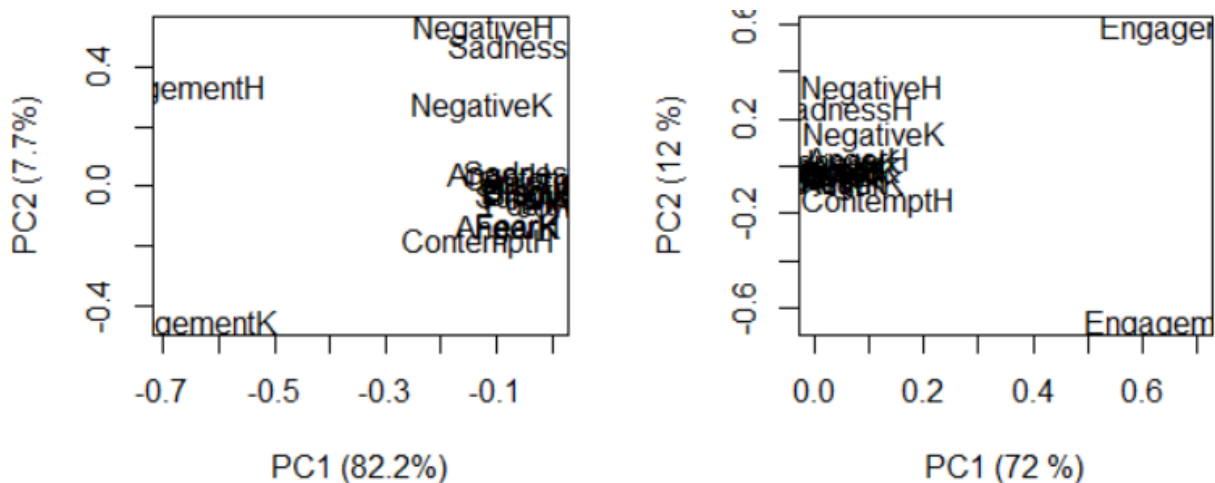


Figur 13 Ladningsplot som syner variablane representert ved PC1 og PC2 frå Kathrine sin film. PoissonPCA t.v., PCA t.h.

Ved samanlikning av scoreplotta i Figur 14, er fasongen på fordelinga av respondentar lik, men ein kan merke seg at hovedtyngda av respondentar er tettare samla ved pPCA\_AS enn ved PCA\_AS, og at pPCA\_AS ser ut til å ha hovedtyngda meir diskinkt plassert i origo enn PCA\_AS. For ladningsplotta i Figur 15 kjem spreieing langs PC2\_AS noko betre fram i pPCA\_AS, og i begge høver er engagement for dei to filmene for seg sjølve i kvar sin kvadrant. Det verkar til at pPC1\_AS og PC1\_AS i stor grad forklarar engagement, medan pPC2\_AS og PC2\_AS forklarar skilnad i engagement mellom filmene, og spreieing i negative kjensler.



Figur 14 Scoreplott som viser respondentane representert ved PC1 og PC2 for begge filmene. PoissonPCA t.v. PCA t.h.



Figur 15 Ladnings plott som viser variablane representert ved PC1 og PC2 for begge filmene. PoissonPCA t.v., PCA t.h.

Felles for alle prinsipalkomponentanalysene var at pPCA i alle høver forklarte meir variasjonen i datasetta enn pPCA. Men det er og nokre skilnadar ein skal merkje seg. Noko av det viktigaste lærdomen frå dei to forskjellige metodane som vart nytta, var at det var lite skilnad mellom pPCA og PCA for Hilde sin film, men stor skilnad mellom metodane for Kathrine sin film. Sjølv om pPCA\_K var den metoda som forklarte mest av variasjon i datasettet, var det PCA\_K som anspora til klåraste og mest logiske tolkinga av dei første prinsipalkomponentane. Dette vitnar om at metodane kan, men treng ikkje å gje samanfallande resultat. Det tydar nødvendigvis ikkje at den eine metoda var betre enn den andre, men at dei kunne syne forskjellige framferder. Samanstilling kan gje auka innsikt i datasettet, men fordrar til varsam utforsking og val av metodar.

### 3.2.3 Statistiske testar

#### Test for forskjell i variablar mellom begge filmene.

For samanlikning vart tre forskjellige metodar nytta for å avdekkje om det er forskjellar mellom variablane for dei to filmene. Invertert beta-binomisk test høver for teljevariablar, para t-test for tilnærma normalfordelte variablar og para sign test som ikkje antek noko fordeling. Dette er testar med forskjellig teststyrke, og er nærare omtala kapittel 2.7. Tabell 9 syner p-verdiar frå dei forskjellige testane, og det var kun eit høve der forskjellen er signifikant. Den signifikante forskjellen var kun for Fear, og kun for den para t-testen. Det ver og verdt å merke seg at for para sign test var den lågaste p-verdien funne for den parvise variabelen som har høgast p-verdiar frå både invertert beta binomisk test og para t-test. At dei forskjellige testane gjer svært forskjellige resultat tyder ikkje at ein var betre enn dei andre, men at tre testar som alle burde kunne passe til formålet ikkje einast om resultatata. Med

bakgrunn i ei heilheitleg vurdering av testane og kunnskapen frå beskrivande statistikk er det stort sannsyn for at det i røynda ikkje var ein forskjell mellom kor mykje kvar kjensle vart opplevd i filmene.

Tabell 9 Parvise testar for å avdekkje forskjell i variablar mellom filmene. Resultat der p-verdi < 0.05 er merka med \*.

Variabel	invertert beta binomisk test p-verdi	para t-test p-verdi	para sign test p-verdi
Anger	0.51	0.07	0.69
Sadness	0.51	0.42	1
Disgust	0.84	0.36	1
Joy	0.89	0.75	1
Surprise	0.73	0.22	0.52
Fear	0.79	0.04*	0.42
Content	0.31	0.33	0.59
Engagement	0.37	0.84	1
Attention	0.99	0.90	0.053
Positive	0.94	0.85	1
Negative	0.67	0.48	0.08
Neutral	0.97	0.79	0.64

### Har kva film respondentane såg først samanheng med AFFDEX statsitics?

Frå MANOVA, p-verdi = 0.65. Kva film respondentane såg først har ikkje signifikant samanheng med kjenslene som heilheit. Rekkefølga på filmene hadde ingen signifikant samanheng med nokon av enkeltvariablane. Kva film som vart synt først hadde heller ingen

samanheng med nokre dei ordinære prinsipalkomponentane eller dei poissonkorrigerede prinsipalkomponentane.

Frå poissonregresjon av kvar enkeltvariabel finn ein at kva film som vart synt først hadde kun samanheng med to enkeltvariablar. Desse var contempt-variablane for både Hilde (p-verdi  $<0.01$ ) og Kathrine (p-verdi = 0.043) sin film. Det var meir opplevd contempt for begge filmene om respondentane såg Hilde sin film først, noko som kunne antyde at den tradisjonelle oppbygninga av Hilde sin film vekker assosiasjonar til vanlege forelesningar. Dette synte nødvendigvis ikkje at vanlege forelesningar er korkje bra eller dårlege, men kunne vere ein liten indikasjon på at respondentane mogelegvis forventar noko nytt og spanande som dei ikkje var vande med frå før. Ein måtte likevel ta atterhald om at dette var 86 andre tilhøve der kva film respondentane såg først ikkje hadde nokon samanheng med kjenslene. Derav er det uvisst om den statistisk signifikante samanhangen mellom kva film respondentane såg først er tilfeldig, eller om Hilde sin meir tradisjonelle film verkeleg var noko anna enn det respondentane forventar og at dette kjem til uttrykk gjennom contempt. I alle høve er dette resultatet ei god påminning om kor lett ein kan finne plausible forklaringar til resultatata, sjølv om det i røynda ikkje treng vere nokon slik samanheng.

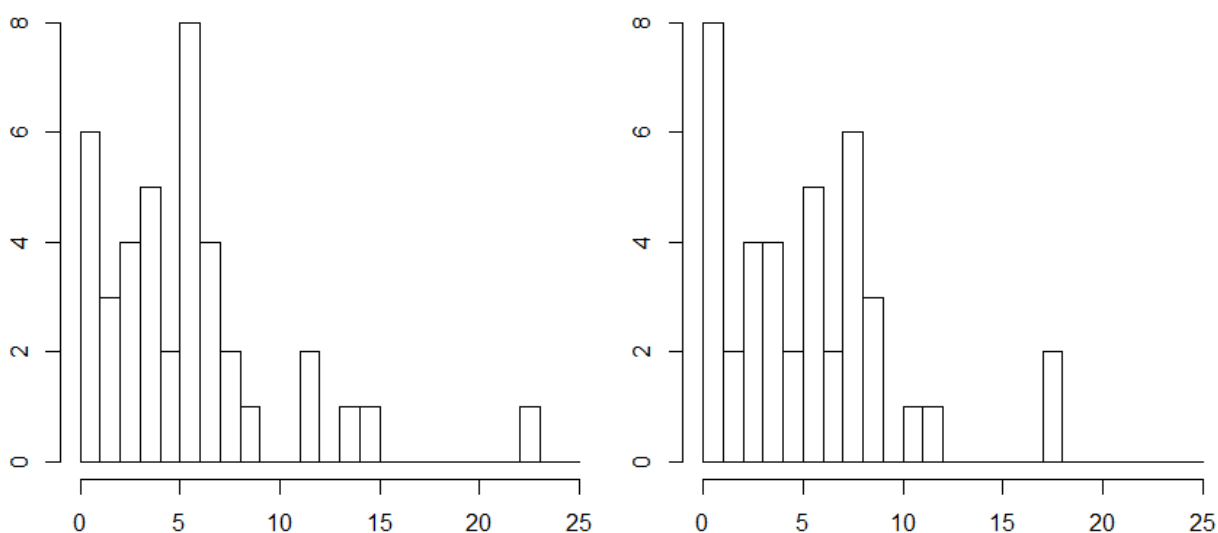
### 3.3 RESULTAT FRÅ GSR SUMMARY SCORES

#### 3.3.1 Beskrivande statistikk

GSR summary scores held, for kvar respondent og for kvart stimuli, informasjon om respondenten, om toppar har forekome i stimuliet, antal toppar per stimuli, og antal toppar per minutt per stimuli (TPM). Sidan filmene er av forskjellig lengde, blir kun toppar per minutt per film nytta i vidare analyser. Toppar per minutt er og telldata, så eksakte testar og testar som forutset tilnærma normalfordeling blir nytta og samanlikna. Tabell 10 syner nøkkeltal frå undersøkinga. Det er vel verdt å merke seg at det for begge filmene var respondentar som ikkje hadde nokon toppar i det heile. Middela er henholdsvis 13% og 3% høgare enn medianen, som kan indikere skyvning mot høgre eller at nokre særskild høge verdiar får stor innverknad på middelet, særskild for Hilde sin film. Frå Figur 16 kan ein merke seg at dei aller fleste respondentane har 10 eller færre TPM, og at dei respondentane som har færre enn 10 TPM fordelar seg ganske jamt. Ein kan sjå antyding til poissonfordeling, med få hendingar av mange moglege, men sidan variansen vil vere om lag 3 gongar større enn middelet må ein vere merksam ovanfor overdispersjon.

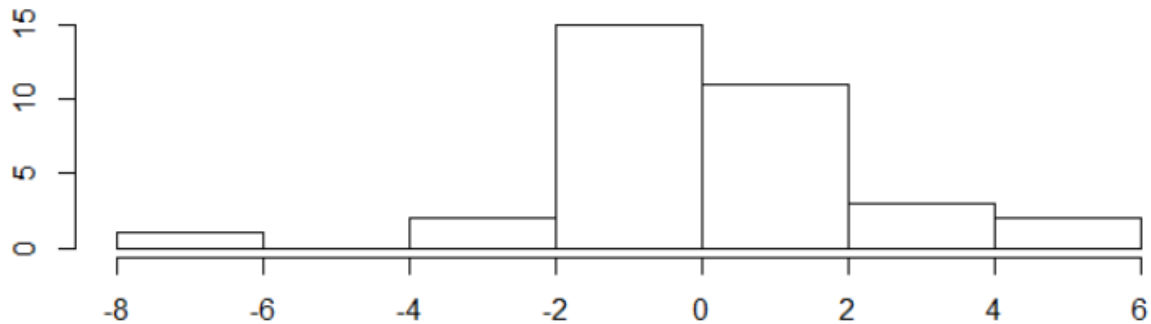
Tabell 10 Nøkkeltal frå GSR summary scores, toppar per minutt for kvar av filmene.

Film	Min	Maks	Median	Middel	St.avvik
Hilde	0.00	22.06	4.63	5.23	4.44
Kathrine	0.00	17.93	4.95	5.11	4.65



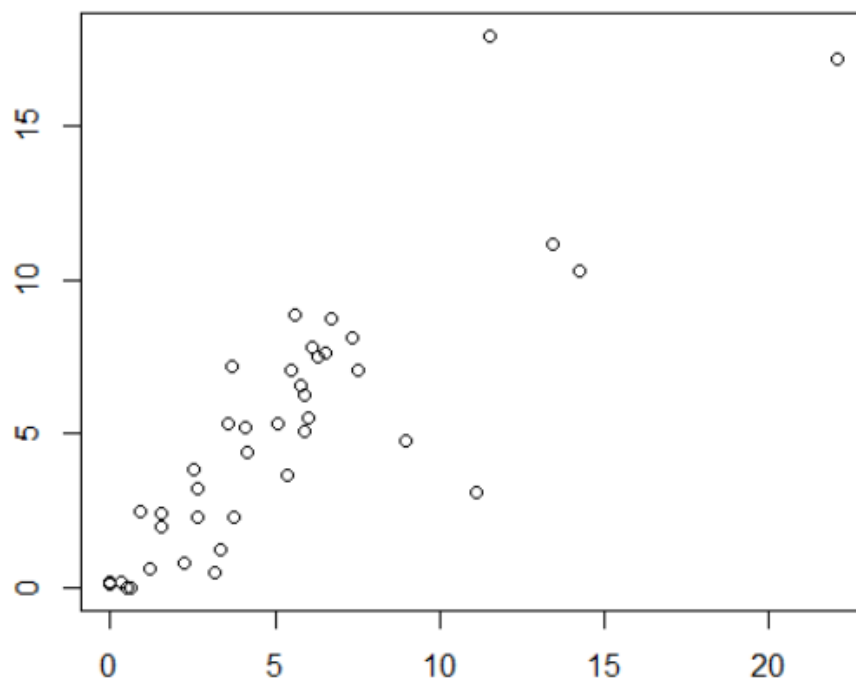
Figur 16 Frekvenshistogram som syner fordelinga av respondentar (y) etter antal toppar per minutt (x). T.v. for Hilde sin film, t.h. for Kathrine sin film.

Når skilnadane (diff TPM) mellom TPM ved å trekkje verdiane for Kathrine sin film (TPM\_K) frå verdiane for Hilde sin film (TPM\_H) blir inspisert i form av frekvensdiagram, Figur 17, ser ein at desse skilnadane antek ei tilnærma normalfordeling om 0. Denne skilnaden kan vere ein god kandidat for statistiske testar for å utforske samanhengar mellom toppar per minutt og både andre ferdigprosesserte resultatfiler og personlegdomstypane.



Figur 17 Frekvensdiagram som syner fordeling av skilnad i toppar per minutt ( $x$ ) og antal respondentar ( $y$ ). Skilnaden er berekna ved å trekkje toppar per minutt i Kathrine sin film frå toppar per minutt i Hilde sin film.

Lineær samanheng mellom Hilde og Kathrine sine filmar var 0.84 talfesta ved pearsons korrelasjonskoeffesient, og Figur 18 syner ein sterk lineær samanheng mellom TPM for dei to filmane, i alle fall for respondentar med få TPM. Ein skal i alle høve vere svært merksam på den store spreininga blant respondentar med fleire enn 10 TPM, og at dei kan ha stor påverknad på eit fenomen som fram til 10 TPM verkar i dette høvet til å ha ein enkel, lineær oppførsel.



Figur 18 Spredningsdiagram som syner respondentane representert ved toppar per minutt for Hilde sin film ( $x$ ) og toppar per minutt for Kathrine sin film ( $y$ ).

### 3.3.2 Statistiske testar

#### Er det forskjell mellom toppar per minutt for filmene?

Frå para t-test, p-verdi = 0.75. Frå invertert beta binomisk test, p-verdi = 0.81. Frå sign test, p-verdi = 0.63. Det er ikkje signifikant forskjell i toppar per minutt for dei to filmene.

#### Har kva film som vart synt først samanheng med toppar per minutt?

Frå MANOVA, p-verdi = 0.86. For enkeltresponsane er p-verdiane 0.63 og 0.58 for henholdsvis TPM\_H og TPM\_K.

Frå poissonregresjon er p-verdiane 0.35 og 0.30 for henholdsvis TPM\_H og TPM\_K.

Kva rekkefølge filmene vart synt i har ikkje signifikant samanheng med toppar per minutt. Frå inspeksjon av residual- og qq-plott for begge modellane byrjar antakingane om konstant variasjon og normalfordelte restfeil å kome på gyngande grunn, og hadde p-verdiane byrja å nærme seg forkastningsgrensene måtte ytterlegare undersøkingar gjerast.



### 3.4 RESULTAT FRÅ GSR BINNED PEAK DETECTION

#### 3.4.1 Beskrivande statistikk

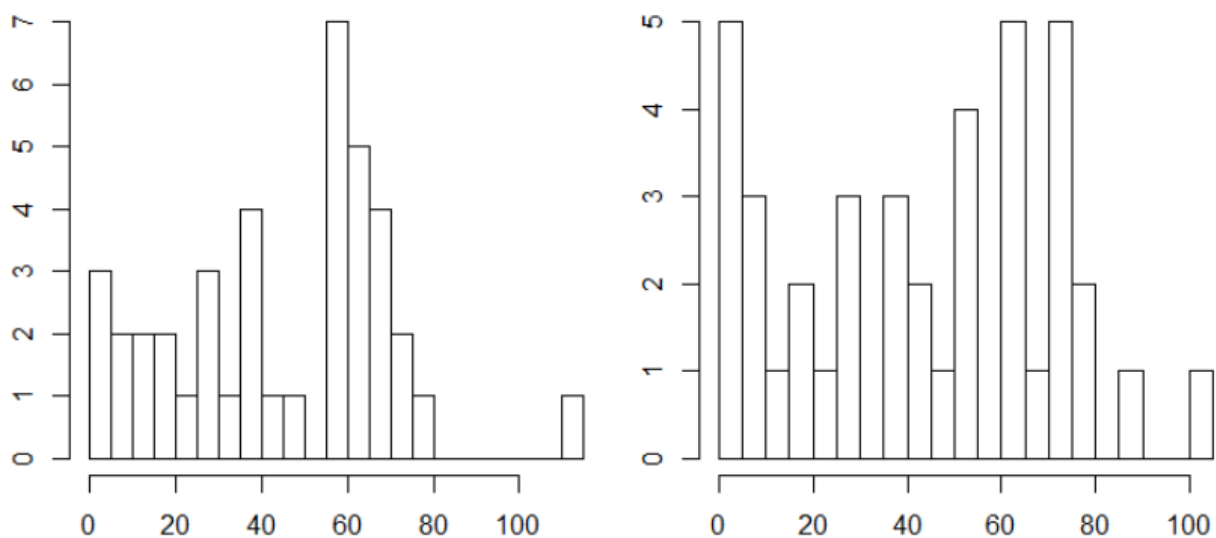
Frå iMoitions får ein for kvart stimuli ei tekstfil som viser om terskelen for reaksjon har blitt overstige minst ein gong innad i eit forhåndsdefinert intervall. I dette høvet er det nytta fabrikkinnstilling som set tidsintervallet til 5 sekund. Mogelege verdiar er 0 (ingen toppar), eller 1 (minst ein topp har inntruffe).

GSR binned peak detection gir informasjon om når kvar respondent overstig terskelen på 800 minst ein gong innad i eit tidsintervall. Ein får då høve til å telle kor mange av intervalla i filmene som hadde toppar (SBP). Resultata blir oppsummert i tabell, ved frekvenshistogram og framstilt som binære bilete.

Frå samanlikning mellom Tabell 11 og Figur 19 er det først og fremst den store spreinga i SBP ein kan merke seg. Det var grovt sett om lag same antal respondentar for kvart tjuelfold med SBP, som viser ei forholdsvis jamn fordeling av respondentar, i alle høver blant dei som har under 80 SBP.

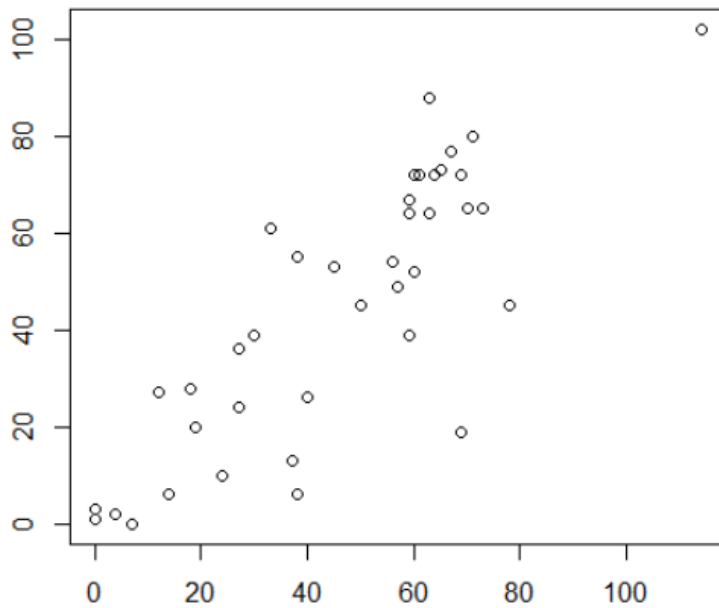
Tabell 11 Nøkkeltal frå summering av intervall med minst ein topp for begge filmene.

Film	antal intervall	min	maks	median	middel	standard-avvik
Hilde	119	0	114	53	45.2	25.9
Kathrine	116	0	102	47	43.7	28.3



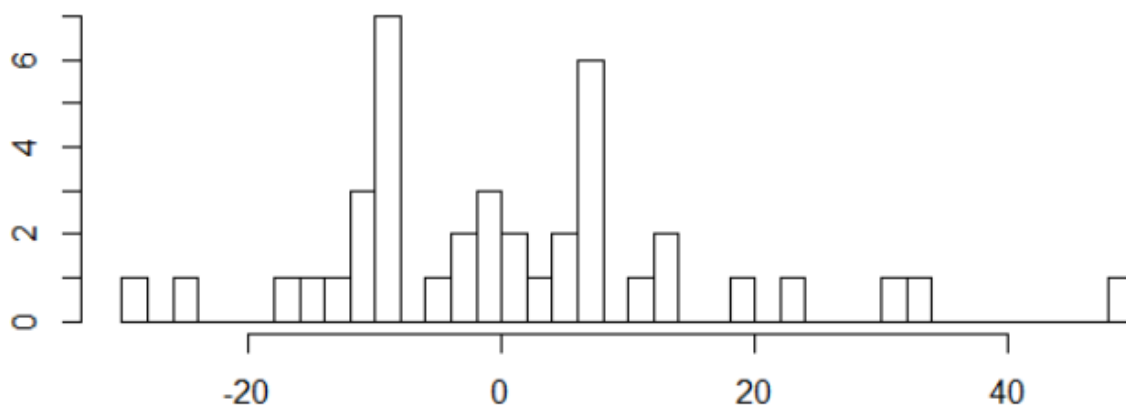
Figur 19 Frekvenshistogram som syner antal respondentar (y) fordelt på antal intervall med minst ein topp (x). T.v. Hilde sin film, t.h. Kathrine sin film.

Pearsons  $r$  for SBP for dei to filmane var 0.84, og stemmer godt overeins med inntrykket frå spredningsdiagrammet i Figur 20. Det er ikkje korrigert for at Hilde sin film har 3 fleire SBP enn Kathrine sin film sidan korreksjonen vil bli svært liten og vil ikkje ha noko praktisk betydning. Ein enkel lineær regresjon der SBP\_H vart forklart av SBP\_K gav ein p-verdi på  $<0.001$ . Stigningstalet for modellen var 0.77, og 71 % av variasjonen i SBP\_H lar seg forklare av modellen. Inspeksjon av residualplot og qq-plott avslørte ingen brot mot antaking om konstant varians og normalfordelte restfeil.



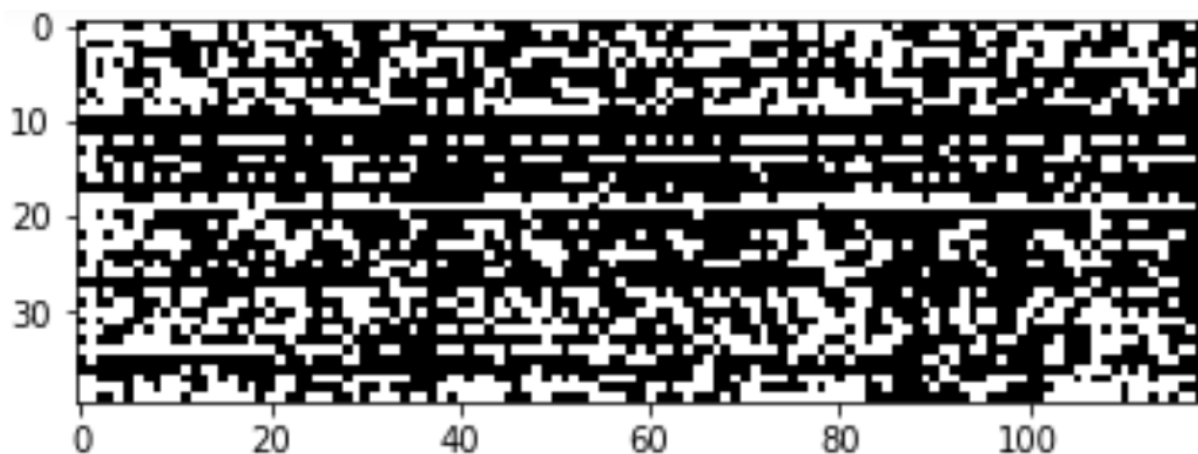
Figur 20 Spredningsdiagram som syner respondentane representert ved antal intervall med toppar for Hilde sin film ( $x$ ) og Kathrine sin film ( $y$ ).

Forskjell i antal intervall med minst ein topp, diffSBP var inspisert ved hjelp av frekvensdiagrammet i Figur 21, og ein kan observere at fasongen i frekvensdiagrammet viser antydning til eit tyngdepunkt om 0. Ein kan skimte antydning til ei tilnærma normalfordeling, men med svært tunge halar.

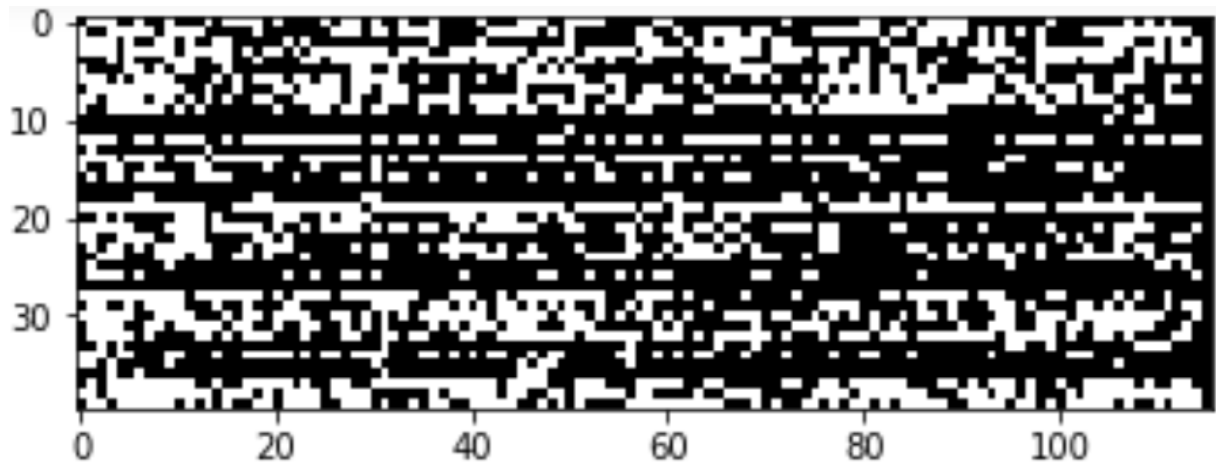


Figur 21 Frekvenshistogram som syner fordelinga av forskjell i antal intervall med minst ein topp (x) etter antal respondentar (y). Forskjellen er berekna ved å trekkje antal intervall med minst ein topp i Kathrine sin film frå antal intervall med minst ein topp i Hilde sin film.

Ved å samanlikne Figur 22 og Figur 23 ser ein at ein vanskeleg kan peike ut ansamlingar av intervall som har minst ein topp. Men ved samanlikning av respondentane ser ein at nokre har fleire intervall med minst ein topp enn andre, og at det er til dels stor skilnad i kva intervall respondentane syner minst ein topp. Nokre har ikkje intervall med toppar i det heile, men nokre har toppar i nesten alle intervalla. Figur 22 og Figur 23 syner og at spørsmåla om kva som vekker reaksjonar i respondentane ikkje kan besvarast av summering åleine, men må sjåast i samanheng med tidsintervalla i filmene. Begrensingar i så måte omtalast vidare under diskusjon.



Figur 22 Binært bilde som syner i kva intervall (x) kvar respondent (y) har hatt minst ein topp i Hilde sin film. Kvite pixlar viser at respondenten har hatt minst ein topp i intervallet..



*Figur 23 Binært bilde som syner i kva intervall (x) kvar respondent (y) har hatt minst ein topp i Kathrine sin film. Kvite pixlar viser at respondenten har hatt minst ein topp i intervallet..*

### 3.4.2 Statistiske testar

#### **Er det forskjell i antal intervall med minst ein topp for dei to filmene?**

Frå para t-test, p-verdi = 0.54. Frå sign test, p-verdi = 0.87. Det er ikkje signifikant skilnad mellom antal intervall med minst ein topp i dei to filmene.

#### **Har kva film som vart synt først samanheng med antal intervall med minst ein topp?**

Frå MANOVA, p-verdi = 0.32. For enkeltresponser var p-verdi på 0.18 og 0.13 for antal intervall med minst ein topp i henholdsvis Hilde og Kathrine sine filmar. Det er ikkje signifikant samanheng mellom antal intervall med minst ein topp og kva film respondentane såg først.

## 3.5 SAMANHENGAR MELLOM FERDIGPROSESSERTER RESULTATFILER

### 3.5.1 Samanhengar mellom spørjeundersøkinga og AFFDEX statistics

#### Har spørsmåla i spørjeundersøkinga samanheng med kjensledifferansane?

Spredningsdiagram av alle mogelege kombinasjonar av spørsmål mot kjensledifferansar, totalt 384 spredningsdiagram, vart visuelt inspisert utan at nokre trendar kunne seiast å merkje seg ut. Sidan spørsmåla i spørjeundersøkinga kan ha samanheng med kjensledifferansane som heilheit, vart det likevel utført MANOVA med dei tolv kjensledifferansane frå Figur 9 som respons, og eit og eit spørsmål frå spørjeundersøkinga som forklaringsvariablar.

Frå MANOVA er det kun 2 av 32 spørsmål som hadde signifikant sammenheng kjensledifferansane som heilheit. ID4, utregningene går bra, men det er vanskeleg å forstå konseptane med  $p$ -verdi = 0.043 og ID8, jobber mye med faget med  $p$ -verdi = 0.021. Fleire av spørsmåla hadde signifikant samanheng med ein eller to enkeltresponsar, totalt 16 tilhøver der det er signifikant samanheng mellom eit spørsmål og ei kjensledifferanse. Merk at det er 16 tilhøver av 396 mogelege, og det er på trass av at spredningsdiagramma ikkje antydgar noko særskilde samanhengar mellom kjensledifferansar og spørsmål. Felles for alle spørsmål og enkeltresponsar er at  $R^2$  er ofte lågare enn 0.05, og høgaste observerte  $R^2$  var på 0.14.

Av særskild interesse var om kva film respondentane følte dei lært mest av (ID11), hadde samanheng med kjensledifferansane, noko som ikkje var tilhøve. Av enkeltresponsane var det kun for kjensledifferansa anger at samanhengen var statistisk signifikant med ein  $p$ -verdi på 0.044 og  $R^2$  på 0.1. Igjen er konstantleddet om lag 0, og stigningstalet er 0.033. Stigningstalet fortel at for kvar auke i poeng i ID11, bevegar kjensledifferansa seg med 0.033. Ei praktisk tolkning kan vere at når ein går frå Hilde mot Kathrine, vil dei respondentane som syner mest anger i Hilde sin film føle dei lærer mest av Kathrine sin film. Og omvendt. ID11 har frå tidlegare undersøkingar synt seg å vere eit spørsmål som beskriv mykje av variasjonen i spørjeundersøkinga, og at det kun er eit tilhøve av samanheng mellom ID11 og kjensledifferansane kan vere ein indikasjon på at dei statistisk signifikante resultatane kan vere tilfeldig, eller at spørjeundersøkinga og AFFDEX statistics ikkje støttar særskildt godt opp om kvarandre.

### **Har prinsipalkomponentane frå spørjeundersøkinga samanheng med kjensledifferansane?**

Det var 17 tilhøver av statistisk signifikante samanhengar mellom PCA\_SU og kjensledifferansane, av 384 mogelege. Dei statistisk signifikante samanhengane inntreff frå PC7\_SU og lågare, og PC7\_SU forklarar 3 % av variasjonen i spørjeundersøkinga. Samanhengane ser og ut til å inntreffe vilkårleg.

### **Har prinsipalkomponentane frå spørjeundersøkinga samanheng med ordinære prinsipalkomponentar frå AFFDEX statistics?**

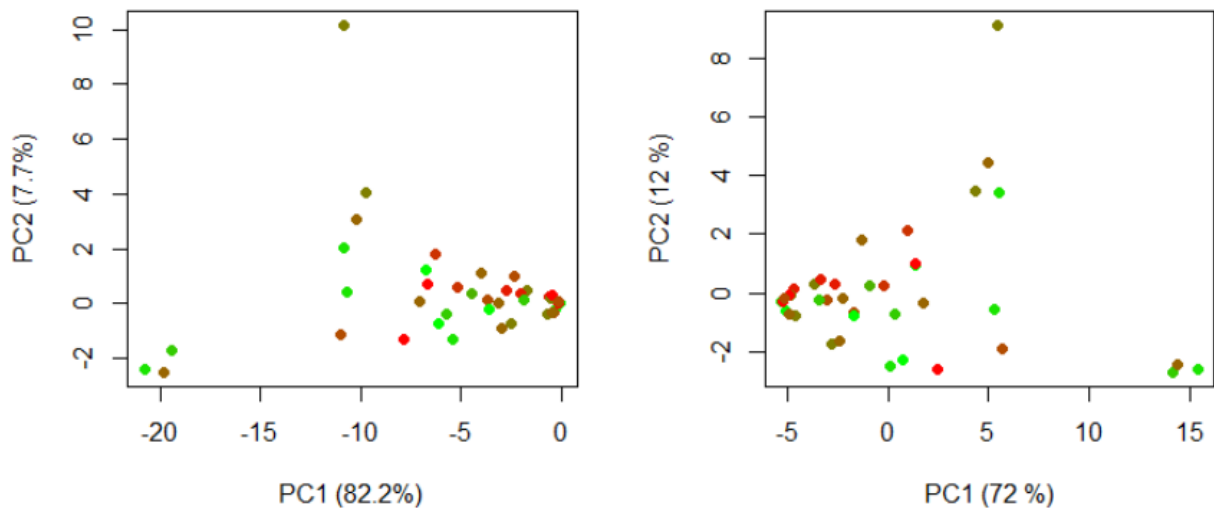
PC1\_SU og PC1\_AS hadde ein statistisk signifikant samanheng, med p-verdi = 0.018, og gjer ein  $R^2$  på 0.14. Det er totalt 28 tilhøver av statistisk signifikante samanhengar mellom prinsipalkomponentar av i alt 640 mogelegheiter. I alle høver er stigningstala  $<0.01$ , og  $R^2$  reint unntaksvis over 0.1. Det verkar heller ikkje til å vere noko mønster for kvar dei signifikante samanhengane inntreff.

### **Har prinsipalkomponentane frå spørjeundersøkinga samanheng med poissonprinsipalkomponentar frå AFFDEX statistics?**

pPC1\_SU og pPC1\_AS hadde ein statistisk signifikant samanheng, med p-verdi = 0.017, og  $R^2$  på 0.14. Det er signifikant samanheng mellom totalt 22 kombinasjonar av prinsipalkomponentar av 608 mulige. Største observerte  $R^2$  er 0.23, og det er fleire høver der  $R^2$  overstig 0.1. Konstantledd og stigningstal ligg i dei fleste høver mellom -1 og 1, men med nokre få unntak. Det verkar det heller ikkje til å vere noko mønster i kvar dei signifikante samanhengane inntreff.

### **Scoreplot frå ordinær PCA og poissonPCA frå AFFDEX statistic med fargekoding.**

ID11, kva film respondentane følte dei lærte mest av på ein skala frå Hilde (0) til Kathrine (10) kan vere ein særskild interessant kandidat for samanknytning av responstypar, då ID11 verkar til å halde mykje av informasjonen i spørjeundersøkinga og samstundes er lett å tolke. Både pPC1\_AS og PC1\_AS forklarar i begge høve mykje av variasjonen i kjenslene, og eventuelle samanhengar kan difor vere av betydning. Som ein kan observere av Figur 24 er det ingen indikasjonar på at scoreplot frå pPCA\_AS og PCA\_AS syner trendar og grupperingar når respondentane vert fargelagt etter ID11.



Figur 24 Scoreplott fargelagt etter kva film respondentane følte dei lærte mest av frå Hilde i klar grønfarge til Kathrine i klar raudfarge. PoissonPCA\_AS t.v., PCA\_AS t.h.

### Visuell inspeksjon av spredningsdiagram frå spørjeundersøkinga med fargekoding av respondentar.

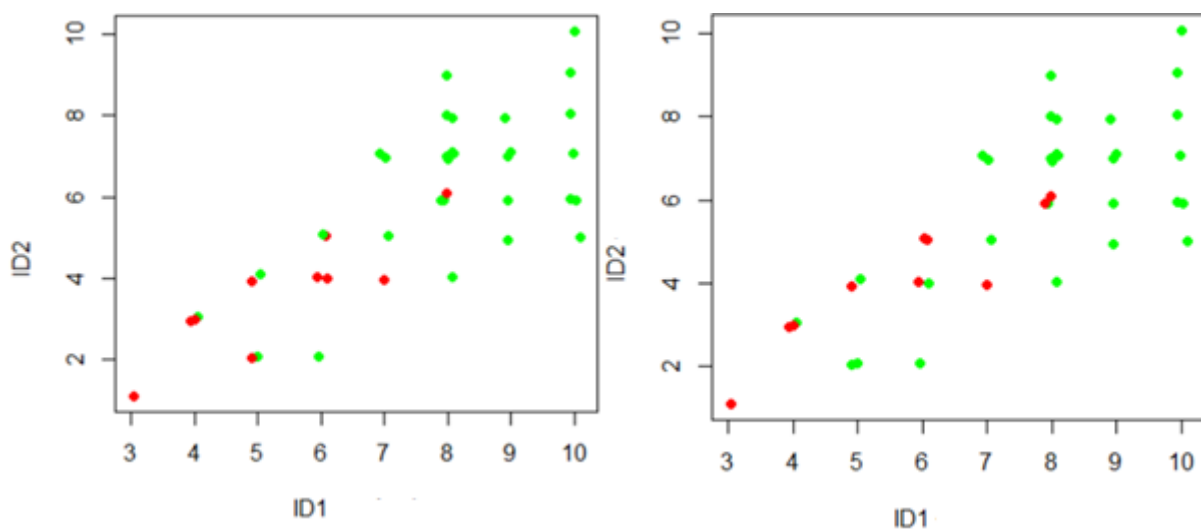
Spredningsdiagram for par av spørsmål for spørjeundersøkinga der respondentane vart fargekoda etter kor stor andel av tida dei gav uttrykk for kvar av dei til saman 24 kjenslevariablane som syner kor stor prosent av tida kvar kjensle oversteig terskelverdien. Respondentar i fjerde kvartil vart farga raude, øvrige vart farga grøne. Dette gav 16 forskjellige spredningsdiagram fargelagt etter 12 variablar for kvar film, i alt 384 forskjellige kombinasjonar av spredningsdiagram og fargelegging. Av desse 384 forskjellige tilhøva, var det 20 fargekoda spredningsdiagram der ein best kunne sjå mogelege grupperingar av raude punkt. I desse 20 spredningsdiagramma kunne ein ikkje sjå nokon samanfallande trekk ved måten grupperingane opptrer på. Dei kan vere over eller under ein viss poengsum, vere samla i ei sentrert klynge eller i ein gitt kvadrant eller ha noko spreing mot kantane av punkta i spredningsdiagrammet.

Særskild merknad kan viast at det kun er to tilhøve der ein finn grupperingar av respondentar med høgare andel av opplevde kjensler i same kjensle for begge filmene og for same spredningsdiagram. Dette inntreff for spredningsdiagramma av ID1, ID2 (Nyttigheit av statistikk) og ID9, ID10 (Utbytte av Kathrine sin film og utbytte av Hilde sin film) for kjensla sadness. Av spredningsdiagram med gruppering av respondentar frå fjerde kvartil av kjenslevariablane er det fire tilhøver for ID21, eksempelet var interessant og tre tilhøver for ID19, filmen var frustrerande. 10 av 16 spredningsdiagram viste tilhøve av gruppering, og grupperingar inntraff for 10 av 24 mogelege kjenslevariablar. Sadnessvariablane hadde i



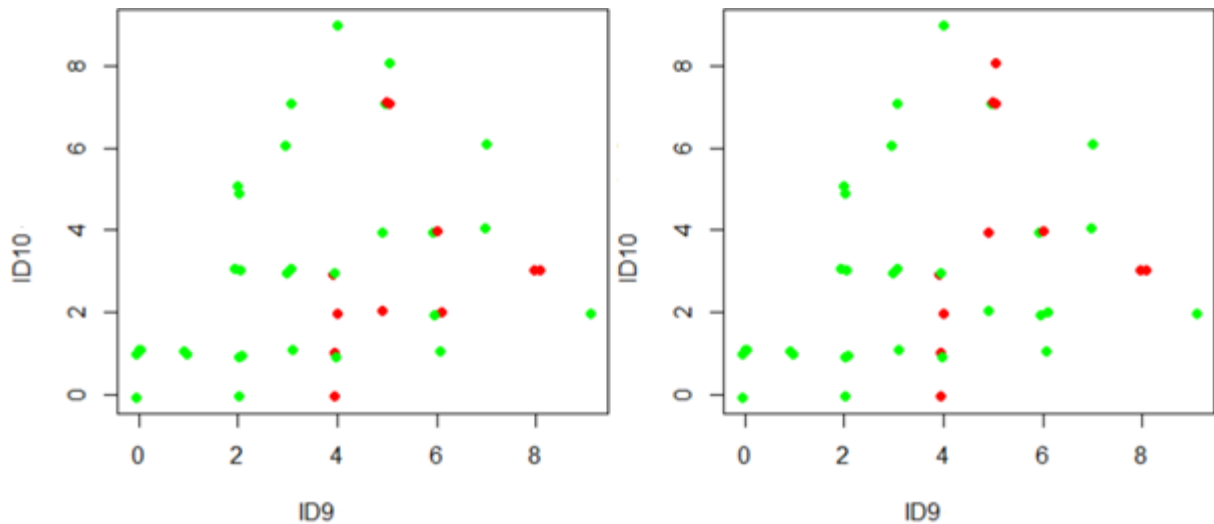
begge høver fire tilhøver av grupperingar, og anger for Kathrine sin film med tre tilhøver. Ein kan og merke seg at for positive kjenslevariablar er det kun for Hilde sin film at det er tilhøver av grupperingar av respondentar i øvre kvartil.

Nærare inspeksjon og samanstilling av dei 8 spredningsdiagramma som syner aller mest antyding til gruppering, og ein kan merkje seg at 5 av dei framkjem når kjensla sadness vert nytta til fargekoding. I desse høva kan ein og sjå at det ofte er dei same respondentane som er i fjerde kvartil av dei nytta kjenslene. Figur 25 syner at respondentar som ligg i øvre kvartil av andel oppleving av kjensla sadness verkar til å tykkje statistikk er mindre nyttig i framtidige studiar og i dagleglivet, noko som verkar rimeleg.



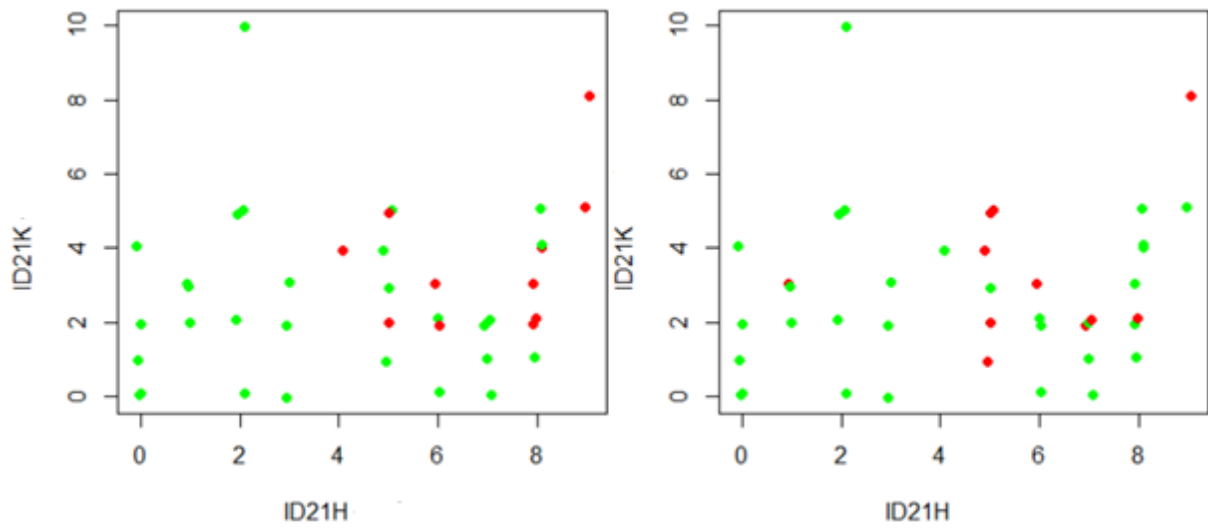
Figur 25 Spredningsdiagram frå ID1 og ID2 med respondentar frå øvre kvartil i raud farge. T.v fargelagt etter sadness for Hilde sin film, t.h. fargelagt etter sadness frå Kathrine sin film.

Frå Figur 26 ser ein at respondentar som opplevde større andel sadness for både Hilde og Kathrine sin film set utbyttet av Kathrine sin film (ID9) til 4 eller høgare, medan mange av dei set utbyttet av Hilde sin film til 4 eller lågare. Dette kan tyde at respondentar som syner mykje av kjensla sadness, følte at dei fekk eit betre utbytte av Kathrine sin film. Dette kunne ha ein samanheng med at Kathrine sin film skil seg frå dei meir tradisjonelle forelesingsfilmane, og kan indikere at dei som generelt syner mykje sadness omkring grunnleggjande statistikk finn at Kathrine si meir intuitive tilnærming gjer dei eit betre utbytte av filmen.

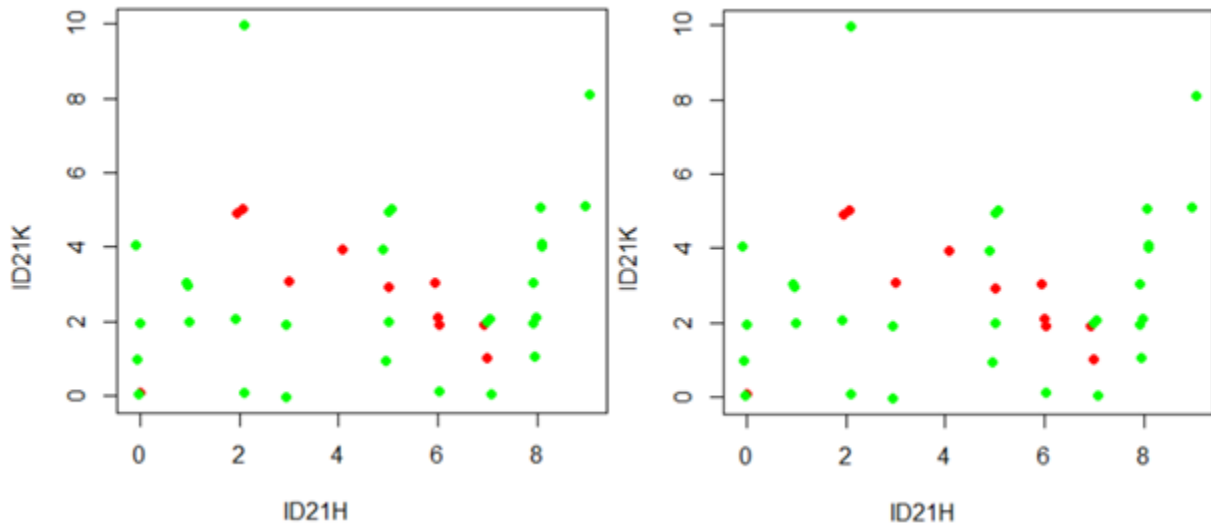


Figur 26 Spredningsdiagram frå ID9 og ID10 med respondentar frå øvre kvartil i raud farge. T.v fargelagt etter sadness for Hilde sin film, t.h. fargelagt etter sadness frå Kathrine sin film.

Det parvise spørsmålet ID21, om eksempelet var interessant er det og nokre moglege grupperingar ein kan merkje seg. Frå Figur 27 kan ein observere at det kan finnast grupperingar for anger og sadness frå Kathrine sin film. Det er svært få respondentar som er einige i at eksempelet til Kathrine var interessant. Dei respondentane som syntest mest av kjenslene anger og sadness tykkjer og eksempelet frå Hilde sin film var meir interessant, med ein score på 4 eller høgare. Ei mogleg tolking kunne ha ein samanheng med at Hilde sitt eksempel er eit vanleg rekneeksempel som synar framgongsmåte og konkret tolking. Eit slikt eksempel kan vere viktig for forståinga til respondentane. Respondentane som syner mykje sadness og anger kan finne Hilde sitt eksempel meir klårgjerande og forståeleg enn animasjonen nytta i Kathrine sin film som freistar å auke forståinga på eit meir konseptuelt nivå. For same parvise spørsmål er det grupperingar for joy og positive for Hilde sin film, sjå Figur 28. Her burde ein kanskje kunne forvente at dei som uttrykkjer joy for Hilde sin film ville uttrykkje dette meir i spørjeundersøkinga, og vere meir einige i at eksempelet var interessant. Det er i det heile tankevekkjande at for respondentar som syner negative kjensler som sadness og anger syner dette betre att i spørjeundersøkinga enn positive kjensler, sjølv om det er svært få tilhøve som syner grupperingar i det heile.



Figur 27 Spredningsdiagram for det parvise spørsmålet ID21 med respondentar frå øvre kvartil i raud farge. T.v fargelagt etter Anger for Kathrine sin film, t.h. fargelagt etter sadness fra Kathrine sin film.



Figur 28 Spredningsdiagram for det parvise spørsmålet ID21 med respondentar frå øvre kvartil i raud farge. T.v fargelagt etter Joy for Hilde sin film, t.h. fargelagt etter positive fra Hilde sin film.

### **3.5.2 Samanhengar mellom spørjeundersøkinga og GSR summary scores**

Alle spredningsdiagram som var visuelt undersøkt for trendar og grupperingar, totalt 94, kunne ikkje vise noko særskilde trendar eller antydningar til potensielle modellar. For å fange opp samanhengar som strekkjer seg over fleire variablar, vart det likevel forsett med ytterlegare undersøkingar.

#### **Statistiske testar.**

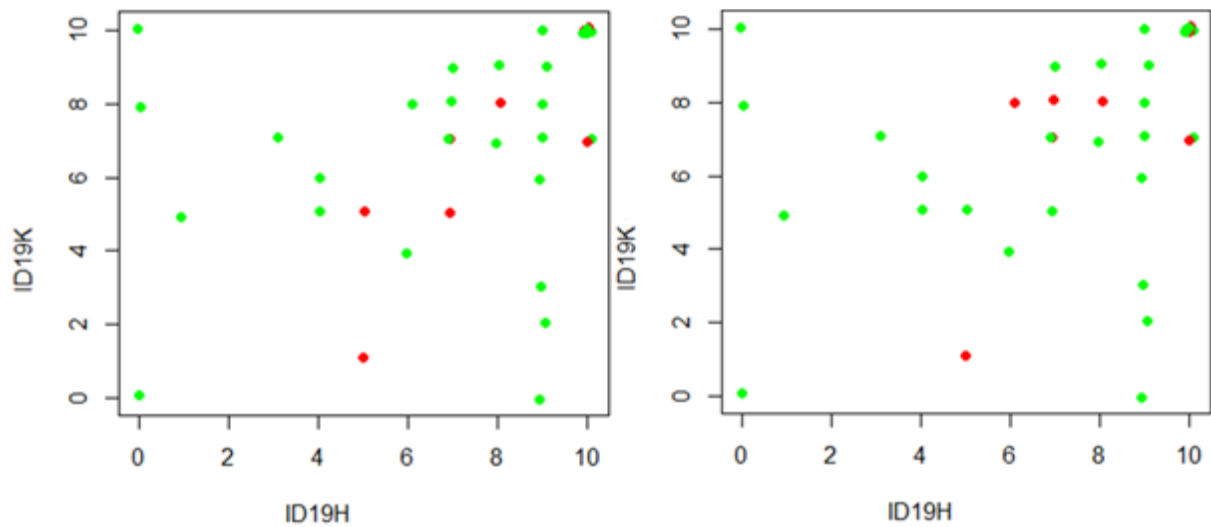
For TPM\_H, TPM\_K og diffTPM vart det etter tur utført MANOVA med spørjeundersøkinga som respons. Dette vart gjenteke med PCA\_SU. Kvar enkeltrespons vert inspisert. Av totalt 201 statistiske testar var det 4 tilhøve av statistisk signifikans, og det var kun for enkeltresponsar. Ein av desse var samanhengen mellom TPM\_K og ID9, utbytte av Kathrine sin film. Denne samanhengen kan og ha ei fornuftig tolking, der eit negativt stigningstal kan bety at dess fleire toppar per minutt, dess lågare utbytte av Kathrine sin film.  $R^2$  var reint unntaksvis over 0.05, og høgast observerte  $R^2$  var 0.14. Det må og takast atterhald for at dei få signifikante resultatata kan vere tilfeldige, som heng godt saman med intrykka frå visuell inspeksjon av spredningsdiagram.

#### **Visuell inspeksjon av scoreplot og spredningsdiagram frå spørjeundersøkinga med fargekoding av respondentar.**

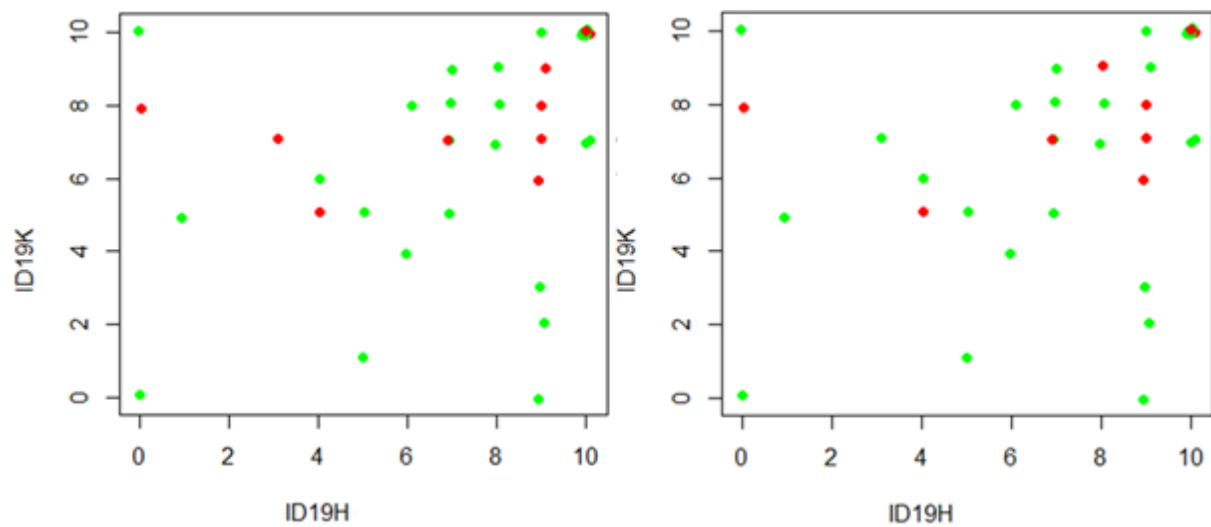
Scoreplotta frå PCA\_SU med respondentane fargelagt med grøn og raud palett frå få til mange TPM for dei to filmene kunne ikkje syne noko antyding til grupperingar.

Totalt 96 spredningsdiagram var inspisert visuelt for å sjå etter mogelege grupperingar av respondentar. TPM\_H, TPM\_K og diffTPM vart nytta til fargekodinga. Respondentar vart farga raude i både øvre og nedre kvartil etter tur. Det var få tilhøve der ein kunne sjå antyding til gruppering i det heile. Dei aller beste tilhøva inntraff for ID19, om filmen var frustrerande. Frå Figur 29 kan ein merkje seg at respondentar i øvre kvartil verkar til å finne filmene noko meir frustrerande, og det kan sjå ut som dei respondentane som syner flest TPM\_K finn filmene noko meir frustrerande enn dei som synt flest TPM\_H. Når ein samanliknar nedre kvartil av TPM for Hilde og Kathrine sine filmar, som synt i Figur 30, kan ein observere at dei respondentane som ligg i nedre kvartil av TPM verkar til å bruke skalaen noko meir for Hilde sin film enn for Kathrine sin film når dei blir spurt om filmen var frustrerande. Om ein samanliknar Figur 29 og Figur 30, ser ein at øvre kvartil har ei aning meir vertikal fordeling, medan nedre kvartil har ei aning meir horisontal fordeling. Men når ein veit at desse eksempla er dei som syner absolutt mest antyding til gruppering, og at

hovudinntrykket er at dei aller fleste av dei raudfarga respondentane ikkje syner noko særskild forskjell mellom kva kvartil som er nytta, så kan ein med rimeleg stor sikkerheit at det er få og utydelege samanhengar å finne mellom spørjeundersøkinga og GSR summary scores.



Figur 29 Spredningsdiagram for det parvise spørsmålet ID19 med respondentar frå øvre kvartil i raud farge. T.v fargelagt etter toppar per minutt for Hilde sin film, t.h. fargelagt etter toppar per minutt for Kathrine sin film.



Figur 30 Spredningsdiagram for det parvise spørsmålet ID19 med respondentar frå nedre kvartil i raud farge. T.v fargelagt etter toppar per minutt for Hilde sin film, t.h. fargelagt etter toppar per minutt for Kathrine sin film.

### 3.5.3 Samanhengar mellom spørjeundersøkinga og GSR binned peak detection

Ingen av dei totalt 96 inspiserte spredningsdiagramma viste indikasjon på trendar eller grupperingar. Det vart likevel forsett med MANOVA for å kunne fange opp eventuelle signifikante samanhengar på tvers av variablar og enkeltvariablane inspisert for å undersøkje om det finnst samanhengar som den visuelle inspeksjonen ikkje oppfatta.

#### Statistiske testar

For SBP\_H, SBP\_K og diffSBP vart det etter tur utført MANOVA med spørjeundersøkinga som respons, og kvar enkeltrespons vart inspisert. Dette vart gjenteke med PCA\_SU. Av totalt 201 statistiske testar var det 9 tilhøve av statistisk signifikans, og det var kun for enkeltresponsar. Ein av desse var samanhengen mellom SBP\_K og ID9, utbytte av Kathrine sin film. Denne samanhengen kan og ha ei fornuftig tolking, der eit negativt stigningstal kan bety at dess høgare SBP\_K, dess lågare utbytte av Kathrine sin film.  $R^2$  var ved svært få høver over 0.05, og høgaste observerte  $R^2$  var på 0.15. Dei høgaste observerte  $R^2$  fann ein ved statistisk signifikante samanhengar mellom dei seinare prinsipalkomponentane. Desse beskriv så lite av variasjonen i datasetta at samanhengar her ikkje kan seiast å vere av noko praktisk betydning. Det må og takast atterhald for at dei få signifikante resultata kan vere tilfeldige utslag, noko som heng godt saman med intrykka frå visuell inspeksjon av spredningsdiagram.

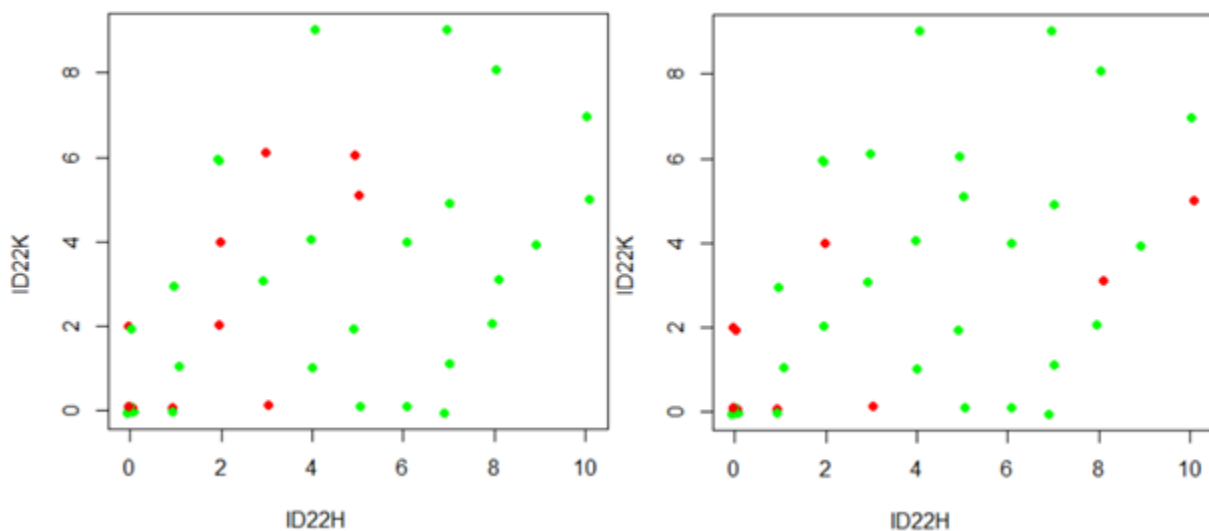
#### Visuell inspeksjon av scoreplot og spredningsdiagram frå spørjeundersøkinga med fargekoding av respondentar.

Scoreplotta frå PCA\_SU med respondentane fargelagt med grøn og raud palett frå få til mange SBP for dei to filmane kunne ikkje syne noko antydning til grupperingar.

Totalt 96 spredningsdiagram var inspisert visuelt for å sjå etter moglege grupperingar av fargekoda respondentar. SBP\_H, SBP\_K og diffSBP vart nytta til fargekoding. Respondentar i øvre og nedre kvartil vart farga raude etter tur. Det var få tilhøve der ein kunne sjå antydning til gruppering. I alt 16 spredningsdiagram synte antydning til moglege gruppering av respondentar i øvre eller nedre kvartil. 7 av tilhøva i øvre kvartil, dei resterande 9 i nedre kvartil. Øvre kvartil for diffSBP var utan spredningsdiagram med antydning til grupperingar. I nedre kvartil for diffSBP var det 4 tilhøver der en kunne sjå antydning til gruppering. For SBP\_K er det berre eit tilhøve der fargekoding etter nedre kvartil syner antydning til gruppering, for spørsmål ID19, om filmen var frustrerende. Om filmen var frustrerende gjer

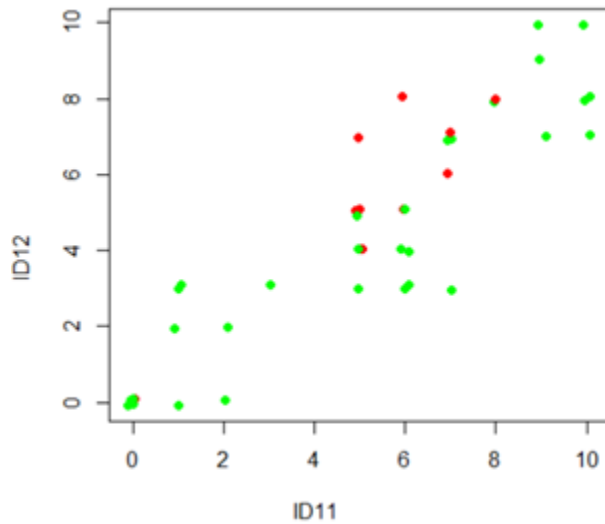
som einaste spørsmål fire tilhøver av grupperingar, for andre spørsmål med antyding av grupperingar er det anten eit eller to tilhøve.

Dei 3 tilhøva som synte klårast gruppering vert presentert under. Figur 31 syner to av tilhøva, der SBP\_H og SBP\_K syner antyding til gruppering. Desse framkjem når respondentar i øvre kvartil blir farga raude for det parvise spørsmålet ID22, om ein forstod kvifor variasjonen til gjennomsnittet var mindre enn variasjonen til enkeltobservasjonen. Når ein freistar å sjå dei to spredningsdiagramma i samanheng, kan det sjå ut som respondentar som syner flest SBP\_H er mindre einige i at dei forstod temaet frå Hilde sin film. Dette gjentek seg for dei respondentane som hadde flest SBP\_K. Dette kan vere ein mogeleg indikasjon på at for nokre få respondentar kan høg SBP vere ein indikasjon på at dei tykkjer dei har ei svak forståing av temaet og difor vert forkava. Ein kan og merkje seg at nokre respondentar som er svært einig i at ein forstod temaet for Hilde sin film er i øvre kvartil av SBP\_K. Dette kunne tolkast som som kan bety at dei tykte dei fekk ei svært god forståing av temaet frå Hilde sin film, men at ein låg score og høg SBP for Kathrine sin film indikerar at den høge SBP\_K kan skrive seg frå ein negativ reaksjon på Kathrine sin film.



Figur 31 Spredningsdiagram for det parvise spørsmålet ID22 med respondentar frå nedre kvartil i raud farge. T.v. antal intervall med minst ein topp i Hilde sin film, t.h. antal intervall med minst ein topp i Kathrine sin film.

Frå Figur 32 kan ein sjå at respondentar i øvre kvartil frå SBP\_H verkar til å klynge seg saman for spredningsdiagrammet frå ID11, kva film respondentane følte ei lærte mest av på ein skala frå Hilde (0) til Kathrine (10) og ID12, kva film respondentane trudde folk flest i STAT100 ville ha best utbytte av. Ein finn ikkje att den same tendensen for SBP\_K, som anten kan ha ei betydning, eller kan vere ein indikasjon på at dei få antydingane til grupperingar er tilfeldige.



Figur 32 Spredningsdiagram for ID11 og ID12 som syner respondentar i øvre kvartil i raudt for antal intervall med minst ein topp i Hilde sin film.

Ein må merkje seg at desse spredningsdiagramma var dei som synte mest antyding til moglege grupperingar, og at det i alle høver ikkje skal meir til enn nokre få raude punkt utanfor dei moglege klyngene før ein ikkje har noko antyding i det heile. Der ein har respondentar i raud farge er det vel så mange respondentar i nærleiken som har grøn farge, som fordrar til varsemd om ein skal nytte desse figurane åleine som bakgrunn for beslutningar.



### **3.5.4 Samanhengar mellom AFFDEX statistics og GSR summary scores**

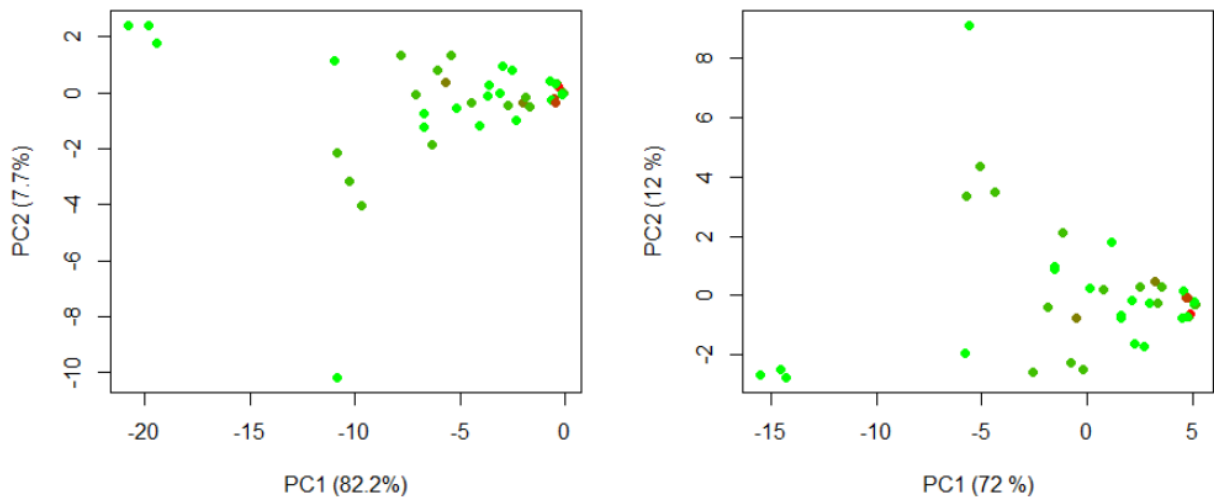
Frå den visuelle inspeksjonen av dei 36 spredningsdiagramma var det ingen særskilde trendar, ei heller antydning til grupperingar. For å avdekkje samanhengar som ikkje vart oppfatta av den visuelle inspeksjonen vart det utført ei rekkje statistiske testar.

#### **Statistiske testar.**

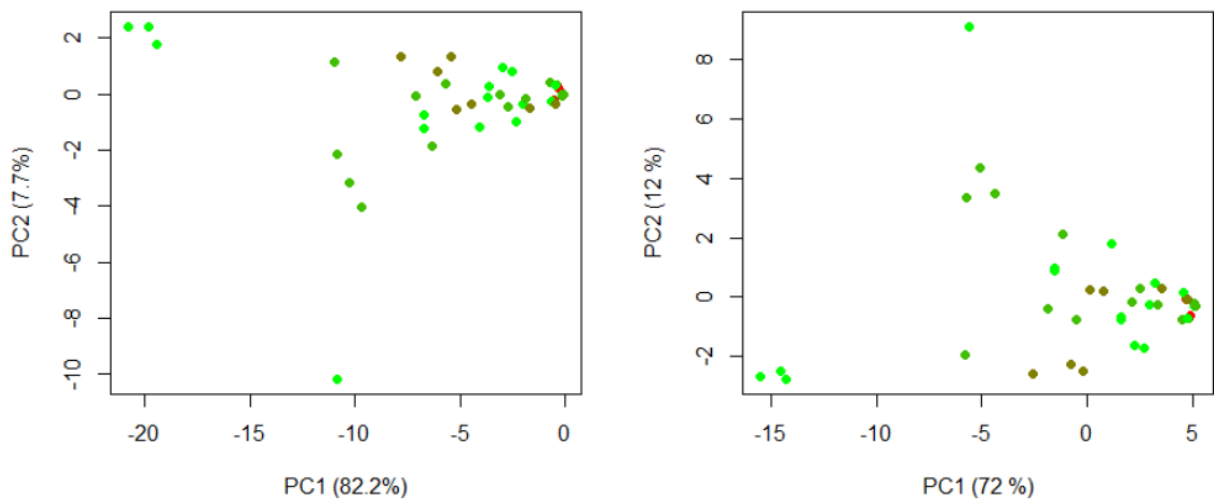
MANOVA vart nytta med kjensledifferansane som respons og TPM\_H, TPM\_K og diffTPM som forklaringsvariablar etter tur. Enkeltresponsane vart og inspisert. Dette vart gjenteke pPCA\_As og PCA\_AS som responsar. Det var kun statistisk signifikant samanheng mellom pPC1\_AS, PC1\_AS og TPM for begge filmene, til saman 4 tilhøver av totalt 126 statiske testar. I desse høva var  $R^2$  mellom 0.13 og 0.14. I dei høve testane ikkje synte statistisk signifikant samanheng var  $R^2$  svært sjeldan over 0.05. Det kan merkast at dette er eit av dei svært få tilhøva der det er samanheng mellom forklaringsvariablar og første prinsipalkomponent. Dette kunne tyde på at TPM har ein samanheng med den prinsipalkomponenten som beskriv over 70% av variasjonen i AFFDEX statistics, som var eit ynskjeleg høve. Men det må takast atterhald av betydinga av funnet, sidan PC1\_AS og pPC1\_AS i stor grad forklarar engagement, ikkje nødvendigvis skilnadar mellom filmene. Det er heller ikkje er noko særleg skilnad i samhengane mellom ordinære og poissonkorrigerde prinsipalkomponentar.

#### **Visuell inspeksjon av scoreplot frå PCA\_SA og pPCA\_SA med respondentane fargelagt etter GSR summary scores**

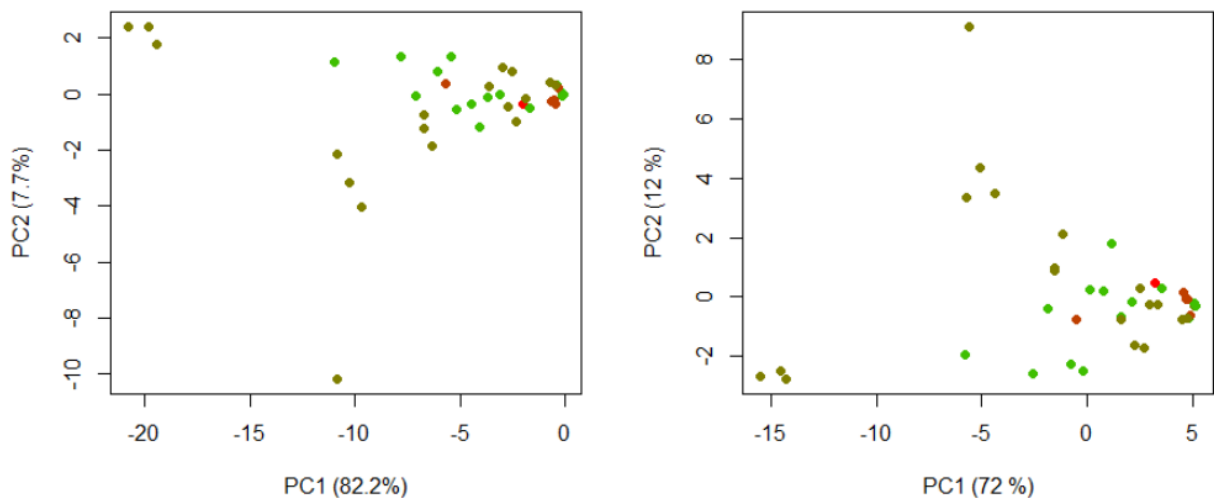
Ved inspeksjon av TPM\_H i Figur 33, TPM\_K i Figur 34 og diffTPM i Figur 35, er det ikkje nokre klare antydingar til grupperingar av respondentar. Ein kan dog observere at dei forskjellige variablane kan avdekkje forskjellige fenomen, og gjerne kan samantillast for auka innsikt i datasettet.



Figur 33 Scoreplot fargelagt etter toppar per minutt frå Hilde sin film. Fargane går frå få toppar (klar grønfarge) til mange toppar (klar raudfarge). PoissonPCA\_AS t.v., PCA\_AS t.h.



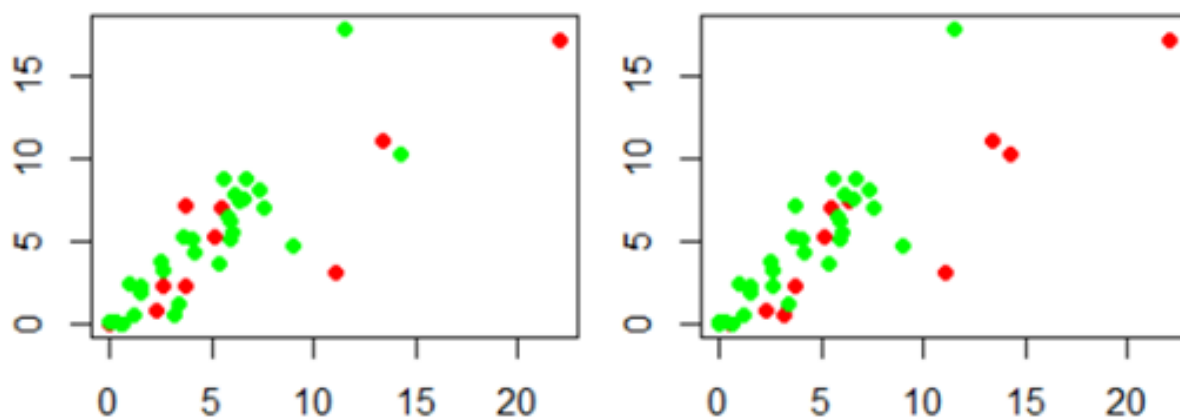
Figur 34 Scoreplot fargelagt etter toppar per minutt frå Kathrine sin film. Fargane går frå få toppar (klar grønfarge) til mange toppar (klar raudfarge). PoissonPCA\_AS t.v., PCA\_AS t.h.



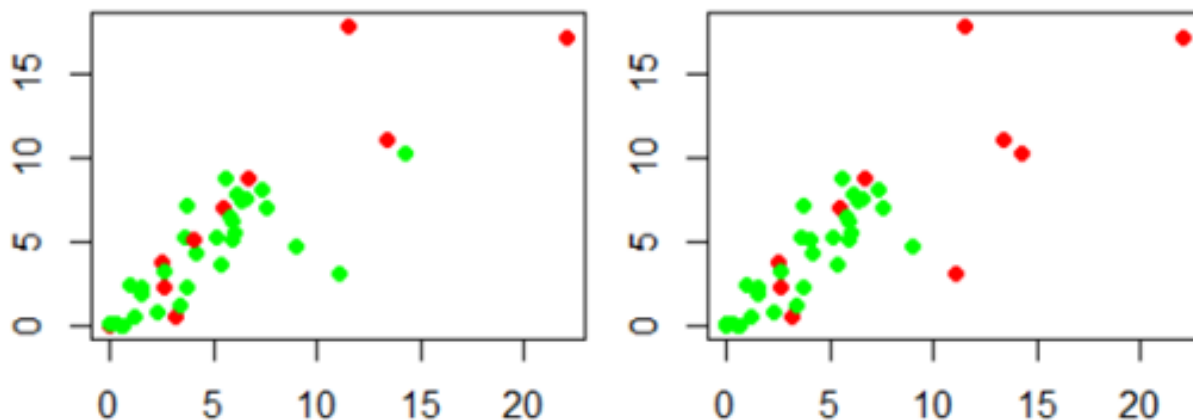
Figur 35 Scoreplot som syner respondentane fargelagt etter skilnad i antal toppar per minutt. Klar grønfarge er respondentar som har flest toppar for Kathrine sin film, medan klar raudfarge er respondentar som har flest toppar for Hilde sin film. PoissonPCA\_AS t.v., PCA\_AS t.h.

## Visuell inspeksjon av spredningsdiagram frå TPM fargelagt etter respondentane sine kjensler.

Av totalt 48 inspiserte spredningsdiagram med respondentane fargekoda etter kjensler og forskjell i kjensler var det ingen teikn til grupperingar av respondentar. Kun ved tilhøva Attention og Neutral, var nokre av respondentane som hadde flest TPM markert til å vere i øvre kvadrant, både for Hilde sin film og Kathrine sin film, sjå Figur 36 og Figur 37. Og i desse høva, som ved alle andre inspeksjonar av spredningsdiagram med respondentar fargekoda etter kvartil, må ein merkje seg at dette var dei tilhøva som synte aller mest antyding til gruppering, og det gjentek seg og i dette høvet at antydinga er svært svak. I tillegg er mogelegvis ikkje kor mykje attention respondentane syner i eit lite rom med få andre distraksjonar, eller kor stor del av tida dei hadde nøytrale kjensler som er dei mest relevante kjenslene i ein slik samanheng.



Figur 36 Spredningsdiagram som syner toppar per minutt for Hilde sin film (x) og Kathrine sin film (y). Respondentar i raudt tilhøyrer øvre kvartil av variabelen Attention, for Hilde til film t.v. og for Kathrine sin film t.h.



Figur 37 Spredningsdiagram som syner toppar per minutt for Hilde sin film (x) og Kathrine sin film (y). Respondentar i raudt tilhøyrer øvre kvartil av variabelen Neutral, for Hilde til film t.v. og for Kathrine sin film t.h.

### 3.5.5 Samanhengar mellom AFFDEX statistics og GSR binned peak detection

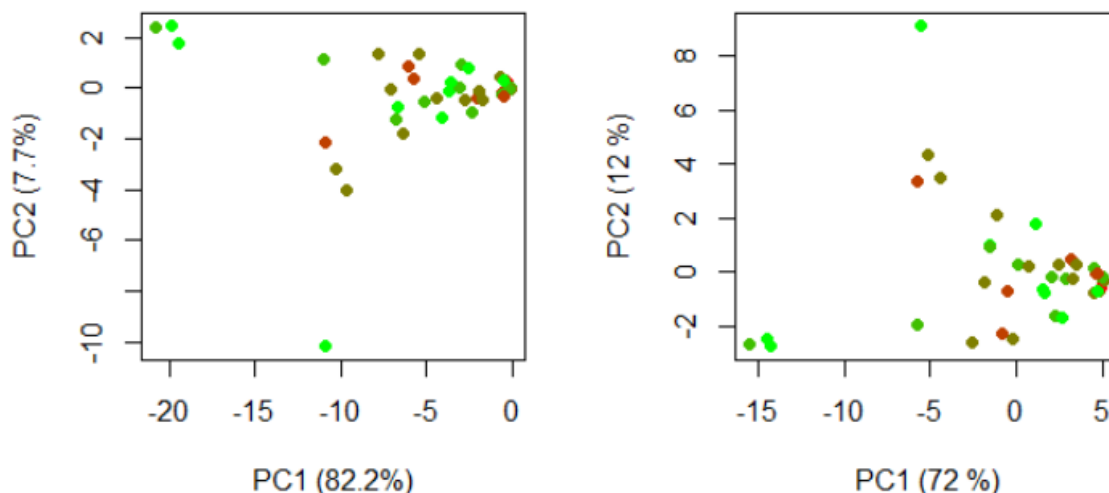
Frå visuell inspeksjon av til saman 24 spredningsdiagram av kjensler mot SBP kan det ikkje seiast å vere antyding til distinkte trendar eller grupperingar. Det var og tilhøvet når forskjel i kjensler og forskjel i antal intervall med minst ein topp vart plotta mot kvarandre.

#### Statistiske testar.

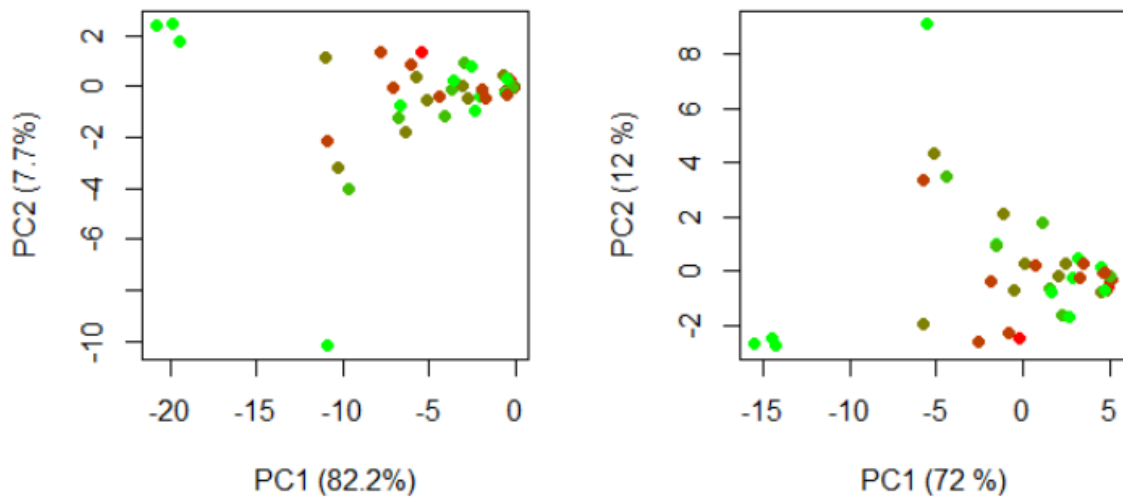
MANOVA vart nytta med kjensledifferansane som respons SBP\_H, SBP\_K og diffSBP som forklaringsvariablar etter tur. Enkeltresponsane vart og inspisert. Dette vart gjenteke med både pPCA\_AS og PCA\_AS som responsar. Det var statistisk signifikant samanheng mellom pPC1\_AS, PC1\_AS og SBP for begge filmene, men i dette høvet var det ytterlegare 5 tilhøve av signifikante samanhengar mellom forklaringsvariablane og enkelte prinsipalkomponentar. I dei 5 tilhøva gjaldt det seinare PCar, 3 tilhøve for dei PCA\_AS og 2 tilhøve for pPCA\_AS.

#### Visuell inspeksjon av scoreplot frå PCA\_AS og pPCA\_AS med respondentane fargelagt etter antal intervall med minst ein topp for Hilde sin film, Kathrine sin film og skilnad i intervall med minst ein topp.

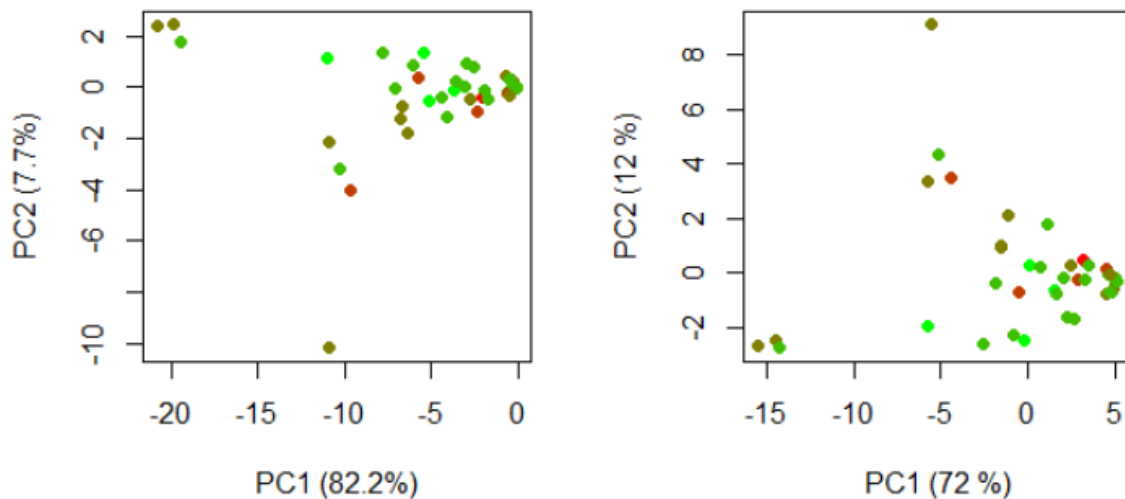
Frå Figur 38 og Figur 39 kan ein observere lite antyding til noko særskilde grupperingar av respondentar, og det verkar heller ikkje til at prinsipalkomponentane evnar å skilje mellom få eller mange SBP, korkje for Hilde eller Kathrine sin film. DiffSBP verkar heller ikkje til å syne grupperingar eller trendar i fargelegging av scoreplotta, sjå Figur 40.



Figur 38 Scoreplot fargelagt etter antal intervall med misnt ein topp frå Hilde sin film. Fargane går frå få intervall (klar grønfarge) til mange intervall (klar raudfarge). PoissonPCA\_AS t.v., PCA\_AS t.h.



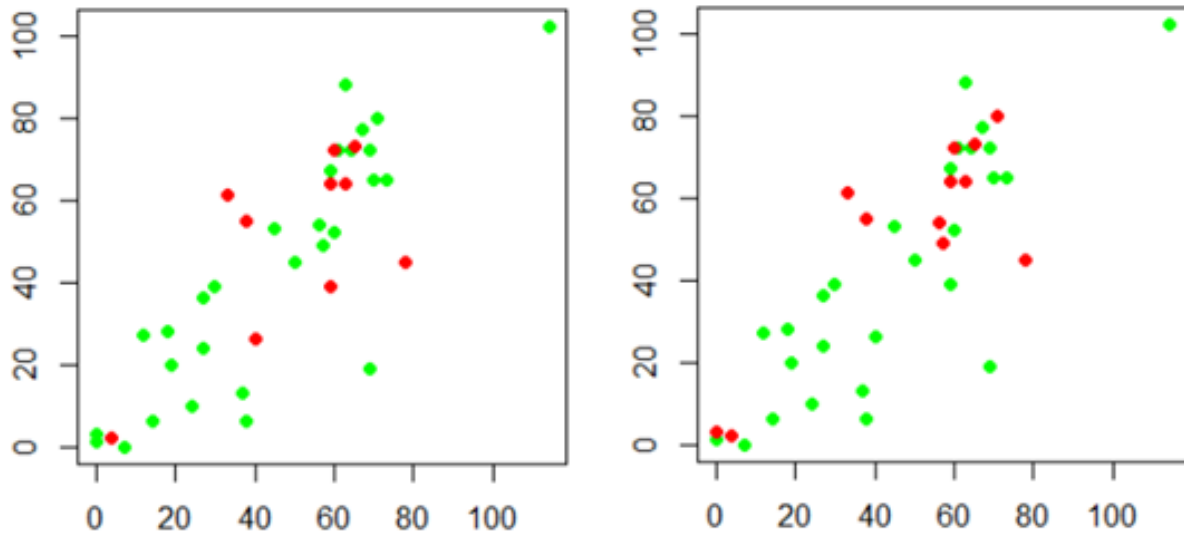
Figur 39 Scoreplot fargelagt etter antal intervall med misnt ein topp frå Kathrine sin film. Fargane går frå få intervall (klar grønffarge) til mange intervall (klar raudffarge). PoissonPCA\_AS t.v., PCA\_AS t.h.



Figur 40 Scoreplot som syner respondentane fargelagt etter skilnad i antal intervall med minst ein topp. Klar grønffarge er respondentar som har flest intervall for Kathrine sin film, medan klar raudffarge er respondentar som har flest intervall for Hilde sin film. PoissonPCA\_AS t.v., PCA\_AS t.h.

### Visuell inspeksjon av spredningsdiagram frå antal intervall med minst ein topp fargelagt etter respondentane sine kjensler.

Av totalt 48 inspiserte spredningsdiagram med respondentane fargekoda etter kjensler og forskjell i kjensler, var det kun i det høvet at øvre kvartil av kjensla disgust var nytta til fargekoding av respondentane at ein kunne sjå antyding av gruppering, sjå Figur 41. Denne grupperinga er ikkje særskild spanande i seg sjølve, men det kan merkast at den aller beste antydinga er svært svak, og at det ikkje skal meir til enn eitpar raude respondentar mellom 20 og 40 SBP før det ikkje lenger er antyding til skilje. Dette understøttar intrykka frå dei statistiske testane og dei øvrige fargekoda spredningsdiagramma.

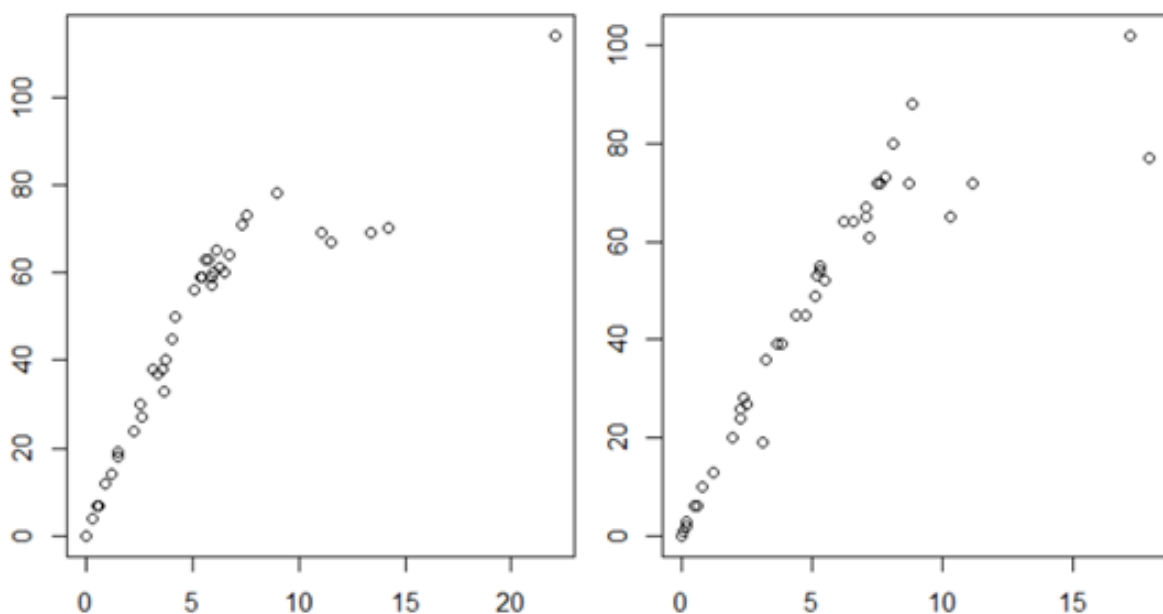


Figur 41 Spredningsdiagram av antal antal intervall med minst ein topp for Hilde sin film (x) og Kathrine sin film (y). Respondentar i raud farge er i øvre kvartil av kjensla disgust for Hilde sin film (t.v.) og Kathrine sin film (t.h.)

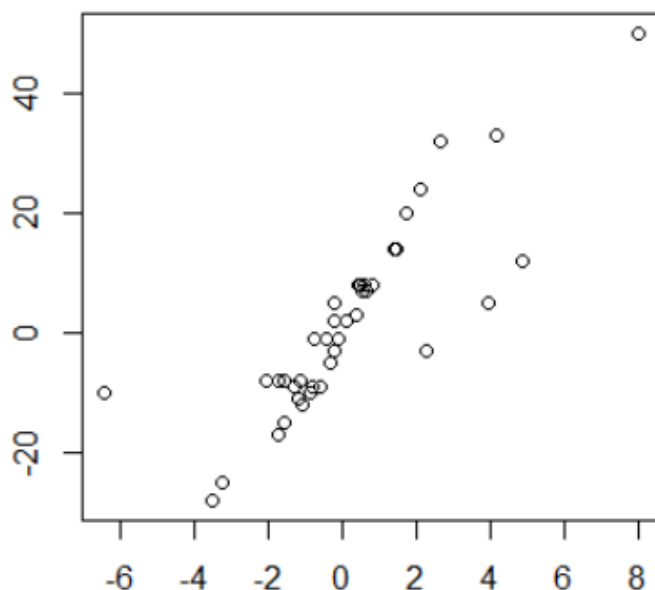
### 3.5.6 Samanhengar mellom GSR summary scores og GSR binned peak detection.

Når TPM og SBP blir samanstillt, ser ein av Figur 42 at det er ein sterk lineær samanheng fram til om lag 10 TPM og 60 SBP. I begge høver skjer ei avbøying, dog noko meir markert for Hilde sin film. Det verkar til vere konstant variasjon fram til avbøyinga, men når ein får mange TPM og SBP verkar det til å vere vanskeleg å skilje mellom få og korte toppar, og lange toppar. Som ein har observert når GSR summary scores og GSR binned peak detection har vore inspisert kvar for seg, er det ofte ein sterk lineær samanheng mellom same typar variablar, heilt til ein bikkar eit visst antal TPM eller SBP. Dei få respondentane dette gjeld får då stor påverknad på eventuelle estimerte parameter.

Når diffTPM og diffSBP blir samanstillt, sjå Figur 43, er det samanstimling av respondentar som ligg om 0 i skilnad, og noko meir konstant variasjon enn når ein ser på variablane i Figur 42. Desse kan då begge vere gode kandidatar til samanknytting av variablar, men om ein ser dei i lys av filmene til Hilde og Kathrine så strekkjer respondentar med flest TPM\_H og SBP\_H seg lenger vekk frå 0 enn dei som har flest TPM\_K og SBP\_K. Stilt saman i Figur 43 ser det likevel ut til å vere ein jamnare samanheng og variasjon når ein nyttar differnasane i staden for dei opprinnelege variablane.



Figur 42 Spredningsdiagram som syner samanhengen mellom antal toppar per minutt (x) og antal intervall med minst ein topp (y) for Hilde sin film (t.v.) og Kathrine sin film (t.h.)



Figur 43 Spredningsdiagram mellom forskjell i antal toppar per minutt (x) og forskjell i antal intervall med minst ein topp (y). Respondentar i kvadranten nede til venstre har flest toppar og intervall med minst ein topp for Kathrine sin film, respondentar i kvadranten oppe til høgre har flest toppar og intervall med minst ein topp i Hilde sin film.

### Samanheng mellom variablane utforska ved ordinær og rang-regresjon.

Samanheng mellom TPM og SBP vart utforska ved ordinær og rang-regresjon, med resultat som synt i Tabell 12. Som venta etter inspeksjon av spredningsdiagramma klarar modellane å fange opp mellom 78% og 93% av variasjonen i responsvariablane, og at rang-regresjon fangar opp meir enn ordinær regresjon. Rang-regresjon gjer og noko lågare stigningstal, og ein kan merke seg at stigningstalet for SBP\_K blir merkbar høgare ved rang-regresjon.

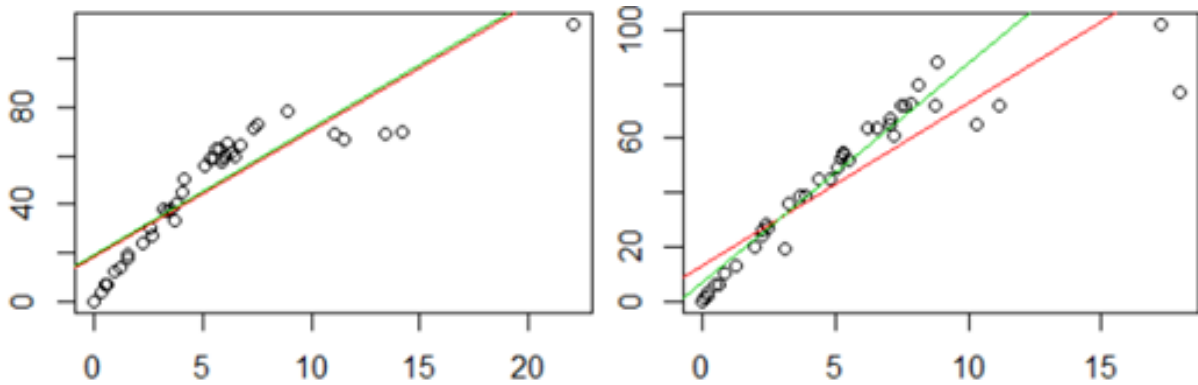
Tabell 12 Resultat frå regresjonsanalyser av samanhengen mellom toppar per minutt og intervall med minst ein topp.

Metode	Forklaring	Respons	Stigningstal	R <sup>2</sup>	P-verdi
ordinær	TPM_H	SBP_H	5.18	0.78	<0.01
ordinær	TPM_K	SBP_K	6.03	0.81	<0.01
rang	TPM_H	SBP_H	5.21	0.88	<0.01
rang	TPM_K	SBP_K	8.08	0.93	<0.01

Dei store andelane forklart variasjon betyr i midlertid ikkje at ein har ein god modell som er svært høveleg for denne type data, sjå Figur 44. For Kathrine sin film verkar rang-regresjonslinja i grønt til å til å fange opp hovudretninga respondentane fordelar seg i, men for



Hilde sin film er ikkje nokre av modellane særskild høvelege. Inspeksjon av residualplott og qqplott viser og at antakinga om konstant variasjon og normalfordelte restfeil ikkje held. Forsøk med andre- og tredjegradspolynom og logtransformasjon av TPM gav heller ikkje noko betre resultat. I alle høver syner dette at ein vanskeleg kan bestemme ein modell som vil høve for klarlegging av samanhengen mellom desse variablane, grunna spreinga av respondentar med høgare verdiar, og ein må vise varsemd ved behandling av desse responstypene.



Figur 44 Spredningsdiagram mellom antal toppar per minutt (x) og antal intervall med minst ein topp (y). Hilde sin film t.v. og Kathrine sin film t.h. Raud linje er ordinær regresjon og grøn linje er rang-regresjon.

### **Samanheng mellom diffTPM og diffSBP.**

Det er ein statistisk signifikant samanheng mellom diffTPM og diffSBP, p-verdi = <0.05 og  $R^2 = 0.71$ . Forutsetningar om konstant variasjon og normalfordelte restfeil verkar til å vere ivaretekne.

Undersøkinga av samanhengar mellom dei to responsane frå GSR synar at akkurat desse to responstypene, så sant ein nyttar dei med varsemd, er i stand til å støtte opp om kvarandre. Aller best er det å nytte skilnadane mellom dei to filmene der dette let seg gjere. Når ein undersøker samanhengar mellom desse to responstypene og andre forklaringsvariablar bør GSR summary scores og summar frå GSR binned peak detection gje om lag same resultat. I alle høve sidan dei kjem frå same sensor burde moglegvis resultatata vere ennå likare, men det er forsatt uvisse om korleis ein topp er definert.

### 3.6 SAMANHENGAR MELLOM FERDIGPROSESSERTER RESULTATFILER OG PERSONLEGDOMSTYPAR

38 av 44 personar har tatt utdanningstesten, desse fordeler seg som vist i Tabell 13. Av desse var det 35 som hadde resultat frå spørjeundersøkinga og AFFDEX statistics og det var totalt 34 tilhøver der respondentane hadde data frå alle ferdigprosesserte resultatfiler. Sidan det var svært få av nokre av personlegdomstypane, vil det av hensyn til statistisk styrke og personvernmessige hensyn kun bli sett på personlegdomskaraktistikkane E/I, N/S, F/T og C/D.

*Tabell 13 Fordeling av antal respondentar innad i kvar personlegdomskaraktistikk for dei totalt 38 respondentane som svarte på utdanningstesten.*

Personlegdoms- karaktistikk	E	I	N	S	F	T	C	D
Antal	19	19	11	27	25	13	14	24

### 3.6.1 Samanhengar mellom spørjeundersøkinga og personlegdomstypar

Det er total 35 respondentar som har resultat frå spørjeundersøkinga og utdanningstesten.

Fordeling av personlegdomskararakteristikkar er synt i Tabell 14.

Tabell 14 Fordeling av personlegdomskararakteristikkar for respondentar med resultat frå både spørjeundersøkinga og utdanningstesten.

Personlegdoms- karakteristikk	I	E	N	S	F	T	C	D
Antal	17	18	10	25	23	12	12	23

Frå Tabell 15 er det lite forskjell på kva film respondentar innad i kvar personlegdomskarakteristikk ville valt om dei kun fekk sjå ein av filmene i STAT100, ID15. Innad i personlegdomskarakteristikkane er det berre ein eller to respondentar i forskjell på kva film dei ville valt, unnateke dei med F (følsom) i personlegdomstypen, der 6 fleire respondentar foretrakk Kathrine sin film. Frå McNemars test er det ikkje signifikant samanheng mellom kva film respondentane ville valt og personlegdomskarakteristikkane I/E, N/S, F/T, og C/D. Det var heller ikkje signifikant samanheng mellom ID15 og dei fire kontrollvariablane K1-4.

Tabell 15 Fordelinga av personlegdomskararakteristikkar etter kva film respondentane ville valt om dei kun fekk sjå ein film i STAT100.

	I	E	N	S	F	T	C	D
Hilde	9	8	6	13	7	10	6	11
Kathrine	8	10	4	12	13	10	6	12

Frå den visuelle inspeksjonen kan ein spesielt merkje seg at spørjeundersøkinga som til no har vore den responsen som har gjeitt lettast tolkbare svar samanstillt med dei andre responstypar, ikkje kunne vise til noko grupperingar når spredningsdiagram og scoreplott vart fargelagt etter respondentane sine personlegdomskararakteristikkar. Spørjeundersøkinga har, når respondentane er fargekoda etter dei øvrige responstypane og særskildt for AFFDEX statistics, klart å syne nokre antydingar til grupperingar. Når personlegdomskarakteristikkane ikkje evnar å syne grupperingar i nokre av totalt 68 mogelegheiter, kan ein undrast om dette er den sanne ståda, eller om spørjeundersøkinga ikkje var eigna til å skilje

personlegdomskaracteristikkane. Til samanlikning var det heller ikkje nokre tilhøve der dei tilfeldige kontrollvariablane kunne seiast å syne særskild klåre grupperingar.

### **Statistiske testar**

Når MANOVA vert nytta med spørjeundersøkinga som respons og personlegdomskaracteristikkane og kontrollvariablane som forklaringsvariablar etter tur er det kun ved eit einskild tilhøve at ein har signifikant samanheng mellom ein forklaringsvariabel og spørjeundersøkinga som heilheit. Dette inntraff for den tilfeldige kontrollvariabelen K2.

Det var og 4 av 128 tilhøver der inspeksjon av enkeltresponsar kunne avdekkje signifikant samanheng mellom personlegdomskaracteristikkane og spørsmåla frå spørjeundersøkinga. I alle dei fire tilhøva var  $R^2$  observert til å vere mellom 0.12 og 0.21 for statistisk signifikante samanhengar, og typisk lågare enn 0.05 for dei øvrige. To av tilhøva var for spørsmålet om respondentane hadde lest om eller høyrte om temaet før forsøket. Ein kan likevel ikkje umiddelbart konstatere at ekstroverte og følsomme respondentar møter meir forberedt enn introverte og sansebaserte respondentar, sidan studentane kun visste at dei skulle få sjå to undervisningsfilmar om statistikk og slik ikkje hadde høve til å forberede seg til eit visst tema. Spørsmålet vart og stilt etter filmene, og med den hensikt å kunne sjå om forkunnskap hadde samanheng med dei øvrige responsane, som ikkje har vore tilhøve. Dei to andre tilhøva var for oppbygginga av Hilde sin film, og om Hilde sin film fekk dei til å sjå den praktiske relevansen til temaet. I begge høver verkar fantasibaserte framfor sansebaserte respondentar til å føretrekkje Hilde sin film. At Hilde sin film, med si tradisjonelle oppbygging, skulle som einaste tilhøver som omhandlar filmene, falle meir i smak hos dei fantasibaserte respondentane enn dei sansebaserte var noko uventa i lys av beskrivinga av personlegdomskaracteristikkane frå 2.1.2, Tabell 1. Til samanlikning var det 7 av 128 tilhøver av signifikante samanhengar mellom ein kontrollvariabel og eit spørsmål.  $R^2$  var observert til å ligge mellom 0.11 og 0.18, ikkje så ulikt  $R^2$  for motlpolane.

Når PCA\_SU vert nytta som responsar og personlegdomskaracteristikkane og kontrollvariablane vert nytta som forklaringsvariablar etter tur er det totalt 12 tilhøver av statistisk signifikante samanhengar av 256 mogelege. Desse er alle mellom PC6\_SU eller seinare prinsippkomponentar, og  $R^2$  er observert til å ligge mellom 0.12 og 0.22. Høgaste observerte  $R^2$  inntreff for K2, og 5 av 12 signifikante samanhengar inntreff for kontrollvariablane.



### 3.6.2 Samanhengar mellom AFFDEX statistic og personlegdomstypar

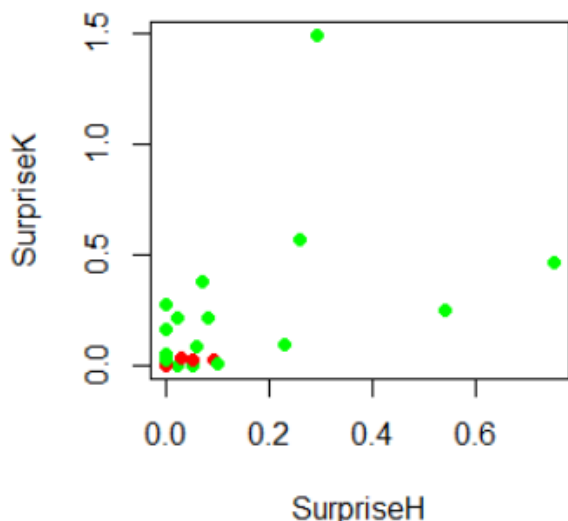
Det er total 35 respondentar som har resultat frå AFFDEX statistics og utdanningstesten.

Fordeling av personlegdomskararakteristikkar er synt i Tabell 16.

Tabell 16 Fordeling av personlegdomskararakteristikkar for respondentar med resultat frå både AFFDEX statistics og utdanningstesten.

Personlegdoms- karakteristikk	I	E	N	S	F	T	C	D
Antal	17	18	10	25	23	12	12	23

Når kjenslene og scoreplotta vert inspisert og forskjellig skala langs aksane vert teke høgde for, er det kun for personlegdomskararakteristikkane N/S er det antydning til trendar når variablane surprise og negative blir samanstillt for dei to filmane, sjå Figur 45. Det kan verke som dei fantasibaserte respondentane i mindre grad en dei sansebaserte gjer utrykk for kjensla surprise, for begge filmane. Ein skal likevel vakte seg for å legge for mykje i dette, då det er over dobbelt så mange sansebaserte som fantasibaserte respondentar, som kan påverke den visuelle inspeksjonen. Om det er tilfeldig at akkurat desse 10 fantasibaserte respondentane ikkje let seg like lett overraske som dei 25 sansebaserte er uvisst. Det er heller ikkje råd å antyde om overrasking er positivt eller negativt.



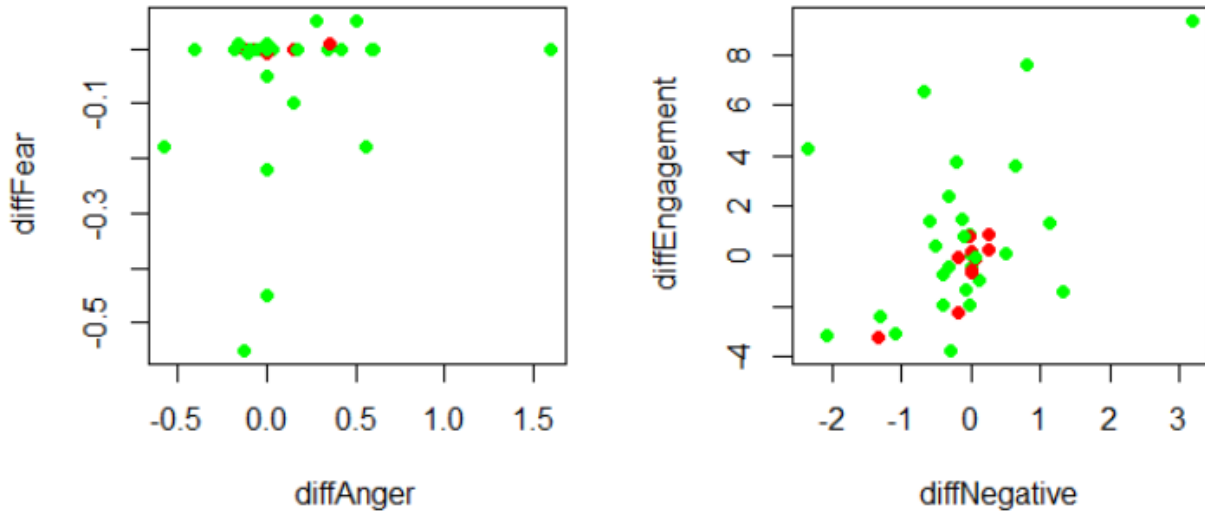
Figur 45 Spredningsdiagram over surprise frå Hilde sin film (x) og Kathrine sin film (y) som syner respondentane fargelagt etter personlegdomskararakteristikk. Respondentar i grøn farge har S i personlegdomstypen, medan respondentar med raud farge har N i personlegdomstypen. Surprise t.v. og negative t.h.

Kjensledifferansane frå AFFDEX statistics verkar til å kunne avsløre noko meir om respondentane sine kjensler i lys av filmene, enn kjenslene i seg sjølve. I dei høva det er grupperingar skriv det seg i stor grad frå at nokre personlegdomskaraktistikkar syner mindre spreieing enn andre, heller enn kvarvidt ein av personlegdomskaraktistikkane syner meir kjensler i favør av ein av filmene. Det er i seg sjølve interessant at personlegdomskaraktistikkane i nokre høver verkar til å ha forskjellig spreieing i kjensledifferansar, men ein må og ha med seg at for dei aller fleste tilhøve er det ikkje noko teikn til gruppering. Det verkar det til at personar med S (sansebasert) i personlegdomstypen har ei noko større utbreieing i forskjell i kjenslene anger, fear, engagement og negative, og at respondentar med N (fantasibasert) i personlegdomstypen har lita utbreieing i dei same kjenslene, sjå Figur 46.

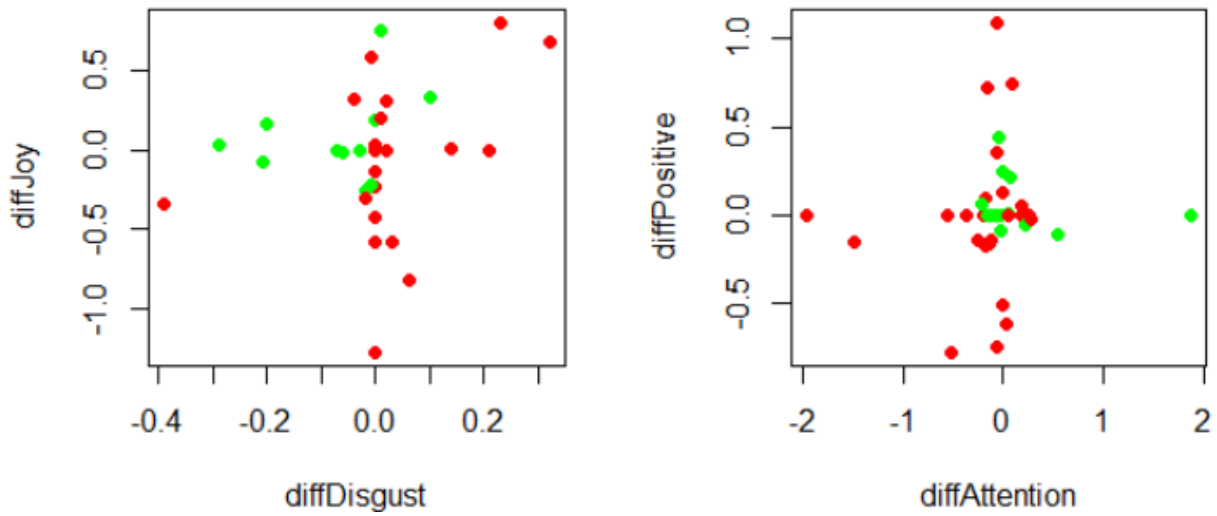
Det høvet som kan syne antyding til at personlegdomskaraktistikkane til respondentane har betydning for skilnad i kjensler i favør av ein av filmene er for personlegdomskaraktistikkane F/T for skilnad i disgust og joy, sjå Figur 47. At dei rasjonelle (T) respondentane kan vise noko meir avsky for Kathrine sin film då denne avvik frå tradisjonell oppbygning, er ikkje overraskande då desse i følgje beskrivinga av personlegdomskaraktistikkane frå utdanningstesten i Tabell 1 ofte er faktaorienterte og heller stolar på sin logiske sans enn på magesfølelsen. Ein bør likevel merke seg at skilnaden er liten, og at desse respondentane fortsett kan ha sterk avsky for statistikk, berre ei aning mindre for Hilde sin film. Desse respondentane ser og ut til å syne noko meir glede for Hilde sin film enn for Kathrine sin film, som kan antakast som fornuftig sidan dei faktaorienterte respondentane kan kunne tenkjast å foretrekke ein meir tradisjonelt oppbygd undervisingsfilm.

For personlegdomskaraktistikkane C/D verkar kjensledifferansane til å i nokre høver vise noko forskjell mellom respondentane, som ein kan sjå av Figur 48. Særskildt for forskjellane i anger, disgust og surprise verkar dei digitale respondentane (D) til å vise mykje forskjell i kjensler for dei to filmene. Dette kan høve med beskrivinga av personlegdomskaraktistikkane frå Tabell 1, då digitale respondentar ofte kan oppfattast som strukturerte og gjerne vil ha kontroll på detaljar og planar. At dei digitale respondentane syner stor spreieing i kjensler mellom ein tradisjonell undervisingsfilm og ei meir intuitiv tilnærming der detaljane kjem heilt til sist, vil då verke rimeleg. At dei kontekstuelle syner vesentleg mindre spreieing i forskjell i kjensler kan og verke rimeleg då dei er vande med tradisjonelle

undervisningsfilmer frå før, men er i følge Tabell 1 meir opptekne av å sjå heilheiter samstundes som dei er fleksible og likar variasjon.

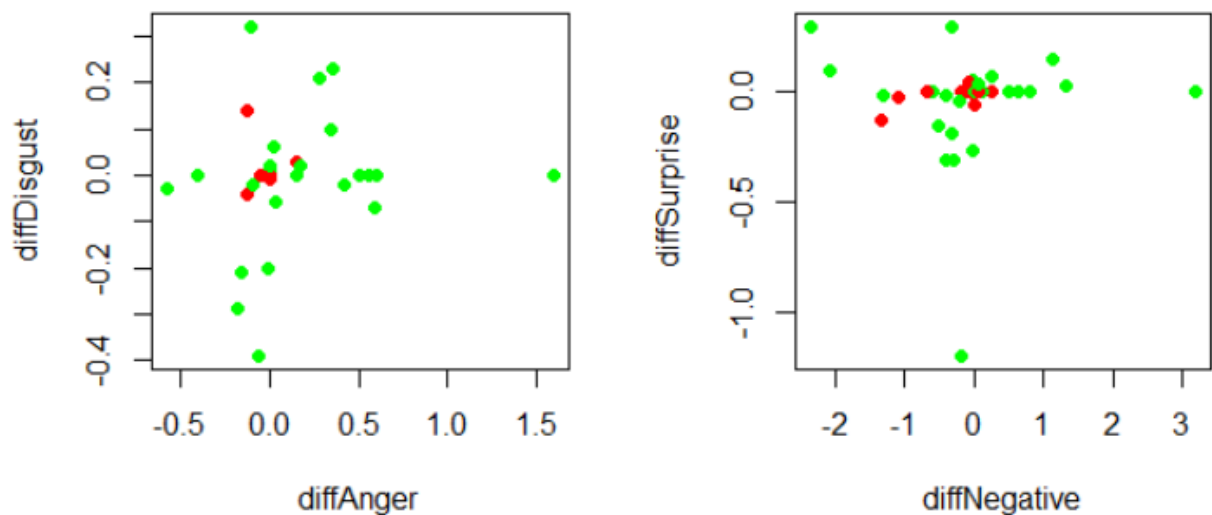


Figur 46 Spredningsdiagram mellom kjensledifferansar, som syner respondentane fargelagt etter personlegdomstype. Respondentar i grøn farge har S i personlegdomstypen, medan respondentar med raud farge har N i personlegdomstypen.



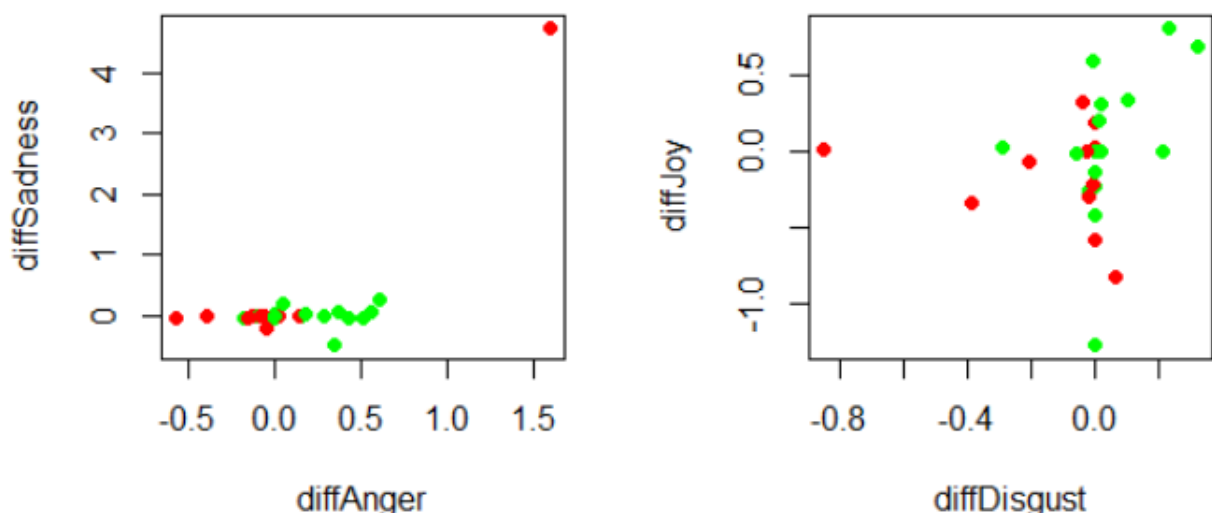
Figur 47 Spredningsdiagram mellom kjensledifferansar, som syner respondentane fargelagt etter personlegdomstype. Respondentar i grøn farge har T i personlegdomstypen, medan respondentar med raud farge har F i personlegdomstypen.



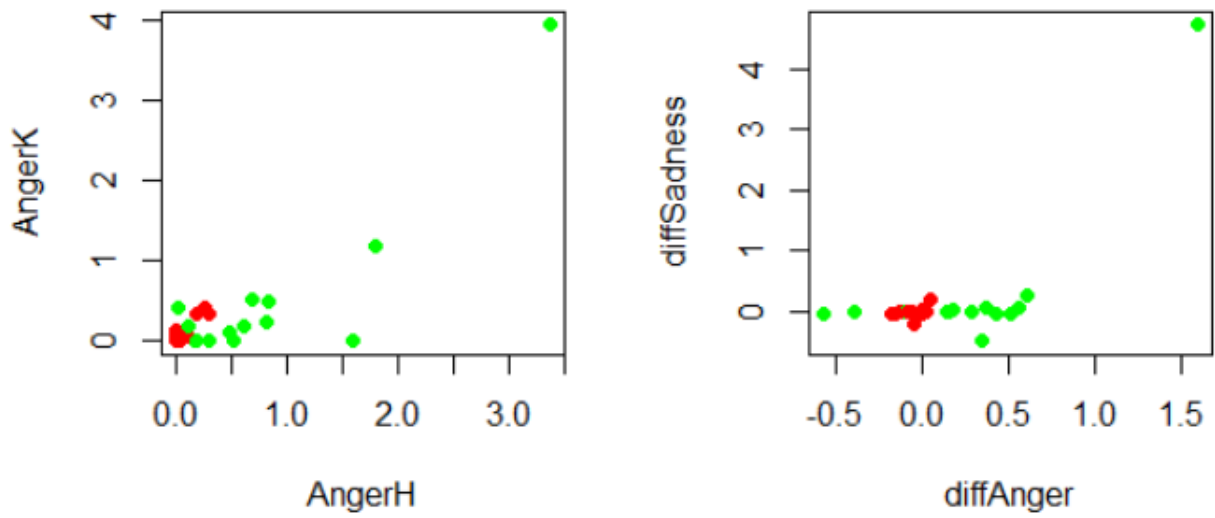


Figur 48 Spredningsdiagramm mellom kjensledifferansar, som syner respondentane fargelagt etter personlegdomstype. Respondentar i grøn farge har D i personlegdomstypen, medan respondentar med raud farge har C i personlegdomstypen.

Figur 49 og Figur 50 syner eksemplar på at dei tilfeldige kontrollvariablane ved fleire høver evna å gje betre skilje mellom respondentane enn personlegdomskaracteristikkane. Til dømes vil K3 gje svært tydeleg skilnad i diffAnger, som lett kunna la seg tolke som at raude respondentar synte mykje mindre skilnad i anger enn grønne. Det held i seg sjølve inga hensikt å tolke fleire av desse resultatata i lys av kontrollvariablane, men at dei kan vere lettare tolkbare og syne tydelegare skilje enn personlegdomskaracteristikkane frå utdanningstesten er eit faktum det er svært viktig å belyse. Det indikera eit stort sannsyn for at funna frå inspeksjonen av spredningsdiagram med respondentane fargelagt etter personlegdomskaracteristikkane og kan vere tilfeldige.



Figur 49 Spredningsdiagram frå kjensledifferansar som syner respondentane fargelagt etter den tilfeldige kontrollvariabelen K2.



Figur 50 Spredningsdiagram som syner respondentane fargelagt etter den tilfeldige variabelen K3.

## Statistiske testar.

Det var ingen tilhøve der MANOVA kunne syne samanhengar mellom personlegdomskaraktistikkane og kjensledifferansane som heilheit. Det var 3 tilhøve av signifikante samanhengar mellom ein personlegdomskaraktistikk og ei kjensledifferanse, og 2 tilhøve av signifikante samanhengar mellom dei tilfeldige kontrollvariablane og ei kjensledifferanse, og gjer 5 statistisk signifikante resultat av 99 mogelege. I alle desse tilhøva ligg  $R^2$  mellom 0.11 og 0.12.

Når PCA\_AS vart nytta som responsar, hadde desse som heilheit signifikant samanheng med N/S. For pPCA\_AS var det nær signifikant samanheng i same tilhøve med p-verdi på 0.052. Det var totalt 15 tilhøve av samanhengar mellom begge sortar av prinsipalkomponentar og forklaringsvariablar, jamt fordelt med 8 tilhøve for personlegdomskaraktistikkane og 7 tilhøve for kontrollvariablane. Dette gjer 17 statistisk signifikante resultat av 315 mogelege. Alle tilhøva inntraff for PC6\_AS eller seinare. Om samanhengen mellom PC6\_AS og forklaringsvariablane er signifikant i ein meir praktisk forstand, er dog uvisst sidan PC6\_AS forklarar mindre enn 1% av variasjonen i kjenslene. For PCA\_AS gjer og K2 signifikant samanheng med PC8\_AS, med ein  $R^2$  på 0.22, som er den høgaste observerte  $R^2$  for alle statistiske testar frå principalkomponentane. At høgaste oppnådde  $R^2$  kjem frå å nytte ein tilfeldig kontrollvariabel som forklaringsvariabel kan vere ein indikasjon på at dei statistisk signifikante samanhengane kan vere tilfeldige.

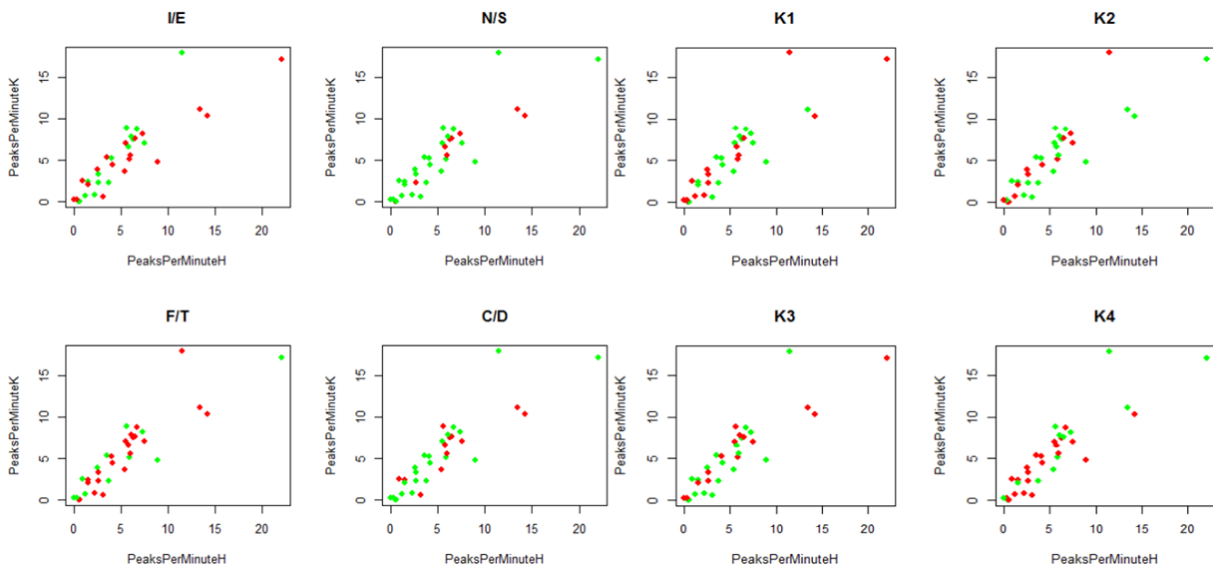
### 3.6.3 Samanhengar mellom GSR summary scores og personlegdomstypar

Det er total 34 respondentar som har resultat frå GSR summary scores og utdanningstesten. Fordeling av personlegdomstypar er synt i Tabell 17.

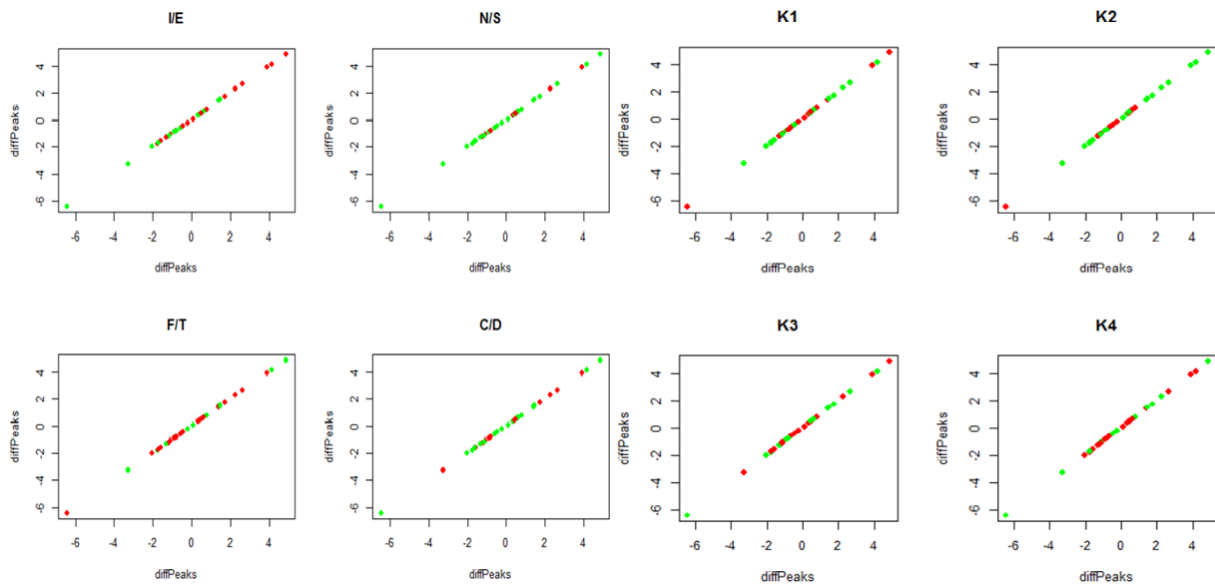
Tabell 17 Fordeling av personlegdomskaraktistikkar for respondentar med resultat frå både GSR summary scores og utdanningstesten.

Personlegdoms- karaktistikk	I	E	N	S	F	T	C	D
Antal	16	18	10	24	22	12	12	22

Frå Figur 51 kan ein sjå at det ikkje verkar til å vere særskilde grupperingar av personlegdomstypar når respondentane vert representert ved TPM, dog med eit mogeleg unnatak for respondentar med C og N i personlegdomstypa som verkar til å mogelegvis ha noko fleire TPM i begge filmene. Respondentar med N i personlegdomstypa verkar og til å ha noko mindre spreiding i antal TPM. Ved inspeksjon av spredningsplott over diffTPM, kan ein frå Figur 52 at det er lite antyding gruppering blant dei forskjellige personlegdomskaraktistikkane. I tilhøvet I/E og kan det vere antyding til at ekstroverte respondentar syner nokre fleire TPM\_H enn TPM\_K, og at introverte respondentar syner nokre fleire TPM\_K enn TPM\_H.



Figur 51 Spredningsdiagram som syner respondentane representert ved toppar per minutt for kvar av dei to filmene. Respondantar med I, S, T eller D i personlegdomstypa er farga grøne. Respondantar med E, N, F eller C i personlegdomstypa er farga raude.



Figur 52 Spredningsdiagram som syner respondentane representert ved skilnad i toppar per minutt og fargelagt etter personlegdomskararakteristikkane og kontrollvariablane. Respondentar med I, S, T eller D i personlegdomstypa er farga grøne. Respondentar med E, N, F eller C i personlegdomstypa er farga raude.

## Har skilnad i toppar per minutt samanheng med personlegdomstypa?

### For personlegdomskararakteristikkane E/I

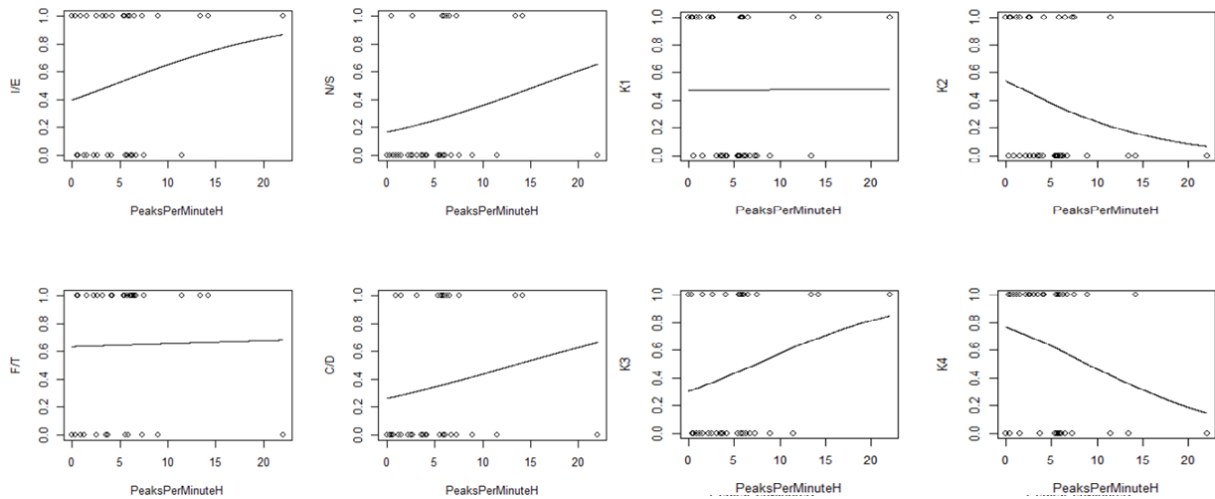
Frå lineær regresjon, p-verdi = 0.048. Det er statistisk signifikant samanheng mellom diffTPM for introverte og ekstroverte respondentar. Ekstroverte respondentar har eit stigningstal på 1.45 diffTPM i forhold til dei introverte.  $R^2$  var observert til å vere 0.12.

For dei øvrige 3 personlegdomskararakteristikkane og 4 kontrollvariablane var det ingen signifikant samanheng mellom desse når diffTPM vert nytta som responsvariabel. I alle tilhøve var forutsetninga om konstant variasjon og normalfordelte restfeil ivareteken.

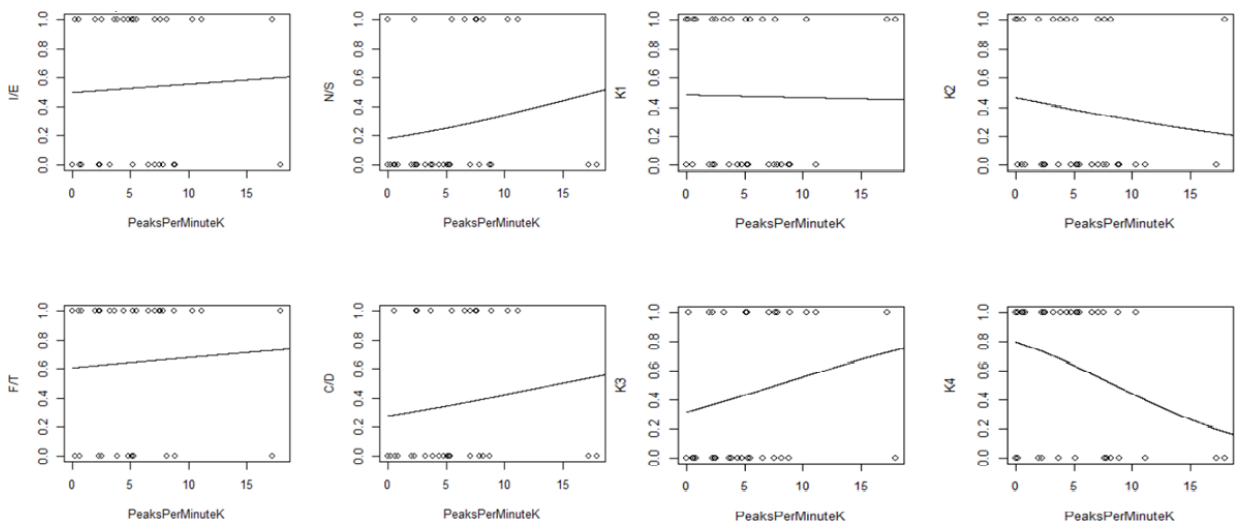
### Har toppar per minutt samanheng med personlegdomstypa, og samantatt resultat mellom enkel regresjon, poissonregresjon og logistisk regresjon?

Enkel regresjon, poissonregresjon og logistisk regresjon kunne frå ei heilheitleg vurdering av toppar per minutt vere tre moglege høvelege metodar for utforsking og samanlikning. Frå totalt 48 statistiske testar er lineær regresjon og logistisk regresjon samantatt, men poissonregresjon skil seg ut særskildt for TPM\_H. Sjølv om det er faktorar som antydgar poissonfordeling, kan overdispersjon kunne vere eit mogleg problem som påpeika i 3.3. Det er difor rett urovekkjande at poissonregresjon evnar å gje 8 signifikante resultat av 16 moglege, 5 av dei for dei tilfeldige kontrollvariablane, og dei resterande 3 når TPM vert nytta som respons.

Frå Figur 53 og Figur 54 kan ein sjå den tilpassa logistiske regresjonslinja for kvar av personlegdomskaracteristikkane for begge filmane, og ein kan observere at denne er svært flat, noko som bekreftar at personlegdomstypar ikkje har ein signifikant samanheng med TPM. I tillegg fordrar desse resultatata til varsemd ved bruk av statistiske metodar og tolking av resultat.



Figur 53 Figurar som syner den tilpassa logistiske regresjonslinja for kvar av personlegdomskaracteristikkane for toppar per minutt i Hilde sin film.



Figur 54 Figurar som syner den tilpassa logistiske regresjonslinja for kvar av personlegdomskaracteristikkane og kontrollvariablane for toppar per minutt i Kathrine sin film.

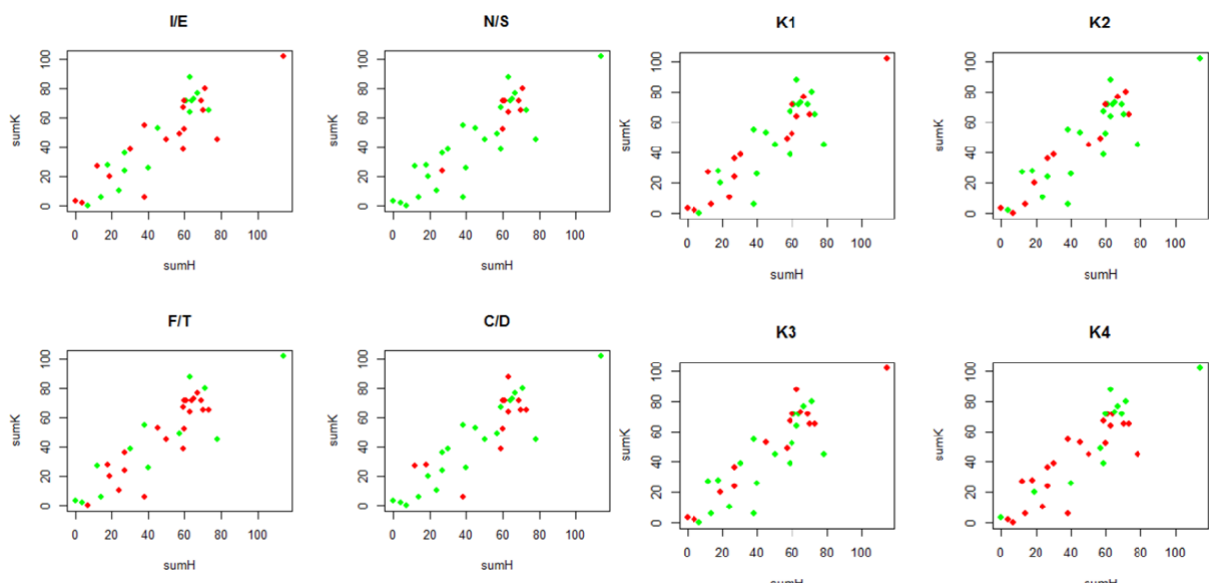
### 3.6.4 Samanhengar mellom GSR binned peak detection og personlegdomstypar

Det er total 34 respondentar som har resultat frå GSR binned peak detection og utdanningstesten. Fordeling av personlegdomstypar er synt i Tabell 18.

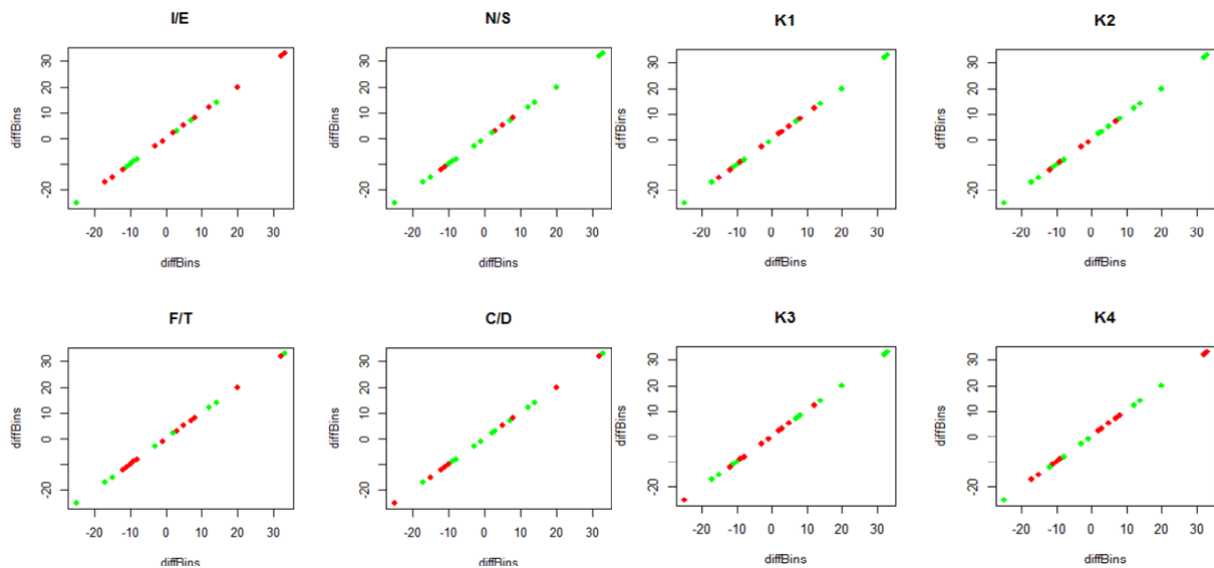
Tabell 18 Fordeling av personlegdomskaraktistikkar for respondentar med resultat frå både GSR binned peak detection og utdanningstesten.

Personlegdoms- karaktistik	I	E	N	S	F	T	C	D
Antal	16	18	10	24	22	12	12	22

Frå inspeksjon av spredningsdiagramma i Figur 55 kan ein sjå antydning til at fantasibaserte (N) respondentar verkar til å ha fleire SBP enn dei sansebaserte (S) respondentane, medan frå Figur 56 verkar desse respondentane til å ha ei lågare spreieing i diffSBP. Dei kontekstuelle (C) respondentane verkar og til å vise litt SBP enn dei digitale (D), men for diffSBP verkar respondentar med C i personlegdomstypen til å ha stor spreieing i skilnad. Det er ellers lite antydning til trendar og grupperingar i fordeling av personlegdomskaraktistikkar representert i spredningsdiagram.



Figur 55 Spredningsdiagram for antal intervall med minst ein topp for dei to filmene. Respondentane er fargelagt etter personlegdomskaraktistikane og kontrollvariablane. Respondentar med I, S, T og D i personlegdomstypen er farga grøne. Respondentar med E, N, F og C i personlegdomstypen er farga raude.



Figur 56 Spredningsdiagram som syner respondentane representert ved skilnad i antal intervall med minst ein topp for kvar av dei to filmene. Respondentane er fargelagde etter personlegdomskaraktistikkane og kontrollvariablane. Respondentar med I, S, T eller D i personlegdomstypa er farga grøne. Respondentar med E, N, F eller C i personlegdomstypa er farga raude.

### Har skilnad i antal intervall med minst ein topp samanheng med personlegdomstypa?

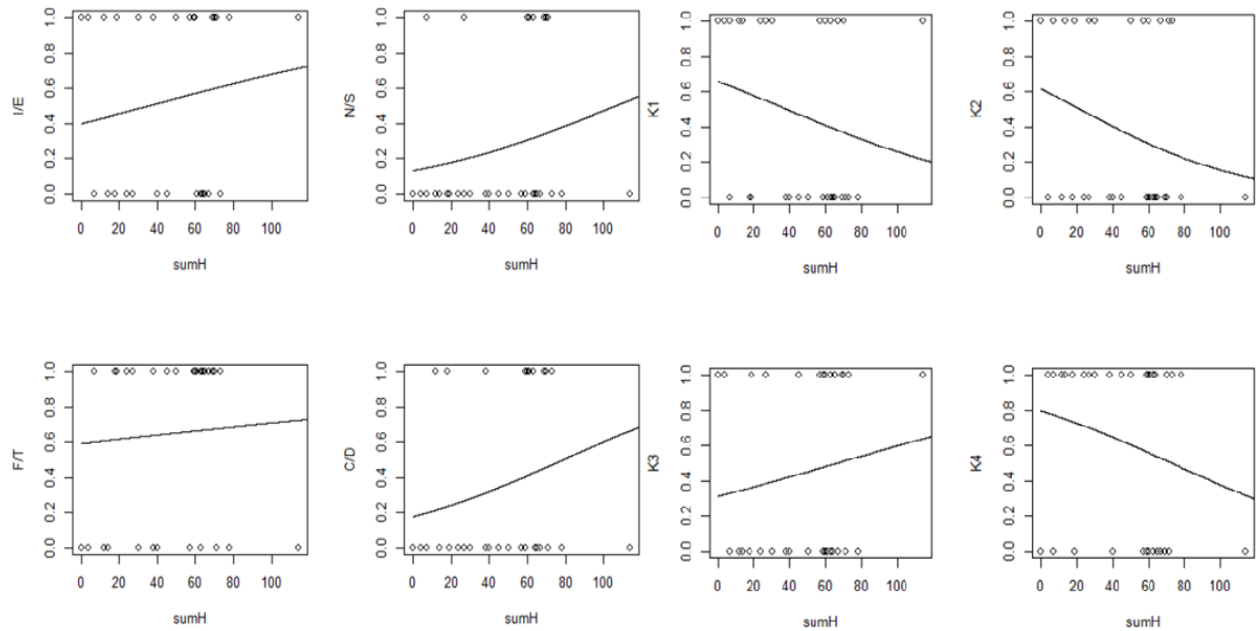
Frå totalt 8 utførte enkle lineære regresjonar med diffSBP som respons og personlegdomskaraktistikkane og kontrollvariablane som forklaringsvariablar etter tur, var dei ikkje i nokre tilhøver statistisk signifikant samanheng. I alle tilhøva er antaking om konstant variasjon og normalfordelte restfeil godt ivareteken.

### Har antal intervall med minst ein topp samanheng med personlegdomstypa, og samantatt resultat mellom enkel regresjon og logistisk regresjon?

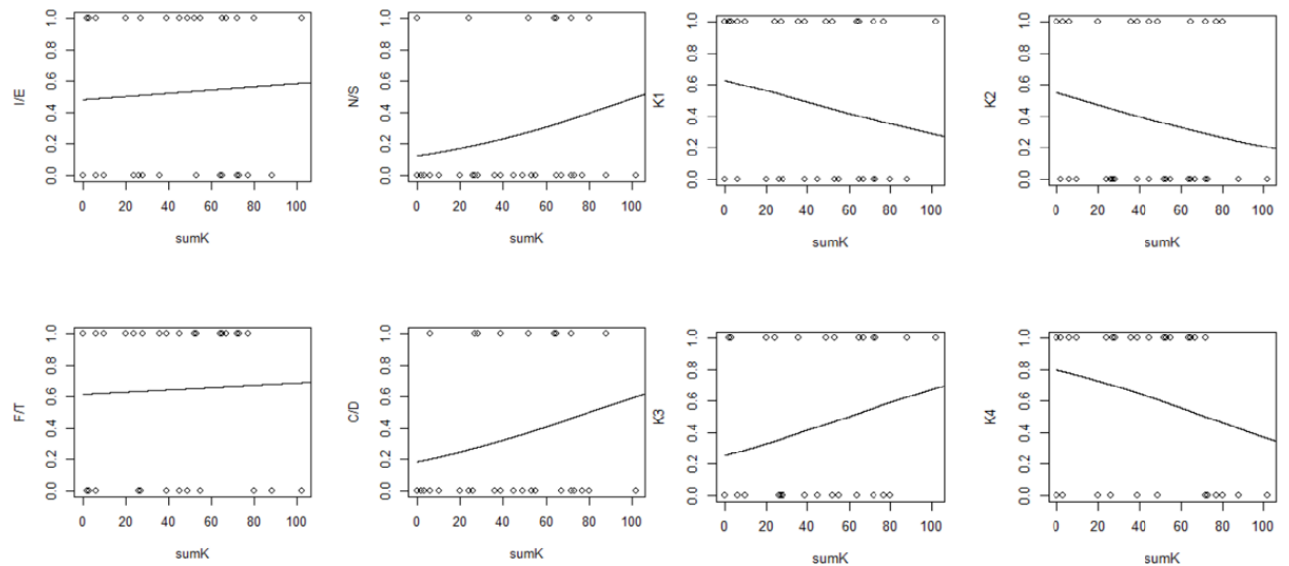
Dei totalt 32 statistiske testane syner ingen samanhengar mellom intervall med minst ein topp for dei to filmene og personlegdomskaraktistikkane eller kontrollvariablane. Resultata mellom dei to metodane samantatt godt. Enkel lineær regresjon gjer ved fleire høver stigningstal opp til nær 15 SBP, utan at resultata vert statistisk signifikante. Dette indikerer stor variasjon mellom respondentane innad i kvar personlegdomskaraktistikk, og mellom responsane som heilheit. Inspeksjon av diagnostiske plot gjer ikkje indikasjonar på brot på antakingar, og forhold mellom residualdeviasjon og friheitsgrader gjer ikkje indikasjon på ueigna tilpassing av den logistiske regresjonsmodellen. Figur 57 og Figur 58 som syner dei logistiske regresjonslinjene understøttar inntrykket frå dei statistiske testane. Ein kan og merkje seg at ved samanlikning av logistiske regresjonslinjer for



personlegdomskaracteristikkane og kontrollvariablane, at kontrollvariablane verkar til å i fleire høver produsere brattare regresjonslinjer enn personlegdomskaracteristikkane.



Figur 57 Figurar som syner den tilpassa logistiske regresjonslinja for kvar av personlegdomskaracteristikkane og kontrollvariablane for antal intervall med minst ein topp i Hilde sin film.



Figur 58 Figurar som syner den tilpassa logistiske regresjonslinja for kvar av personlegdomskaracteristikkane og kontrollvariablane for antal intervall med minst ein topp i Kathrine sin film.

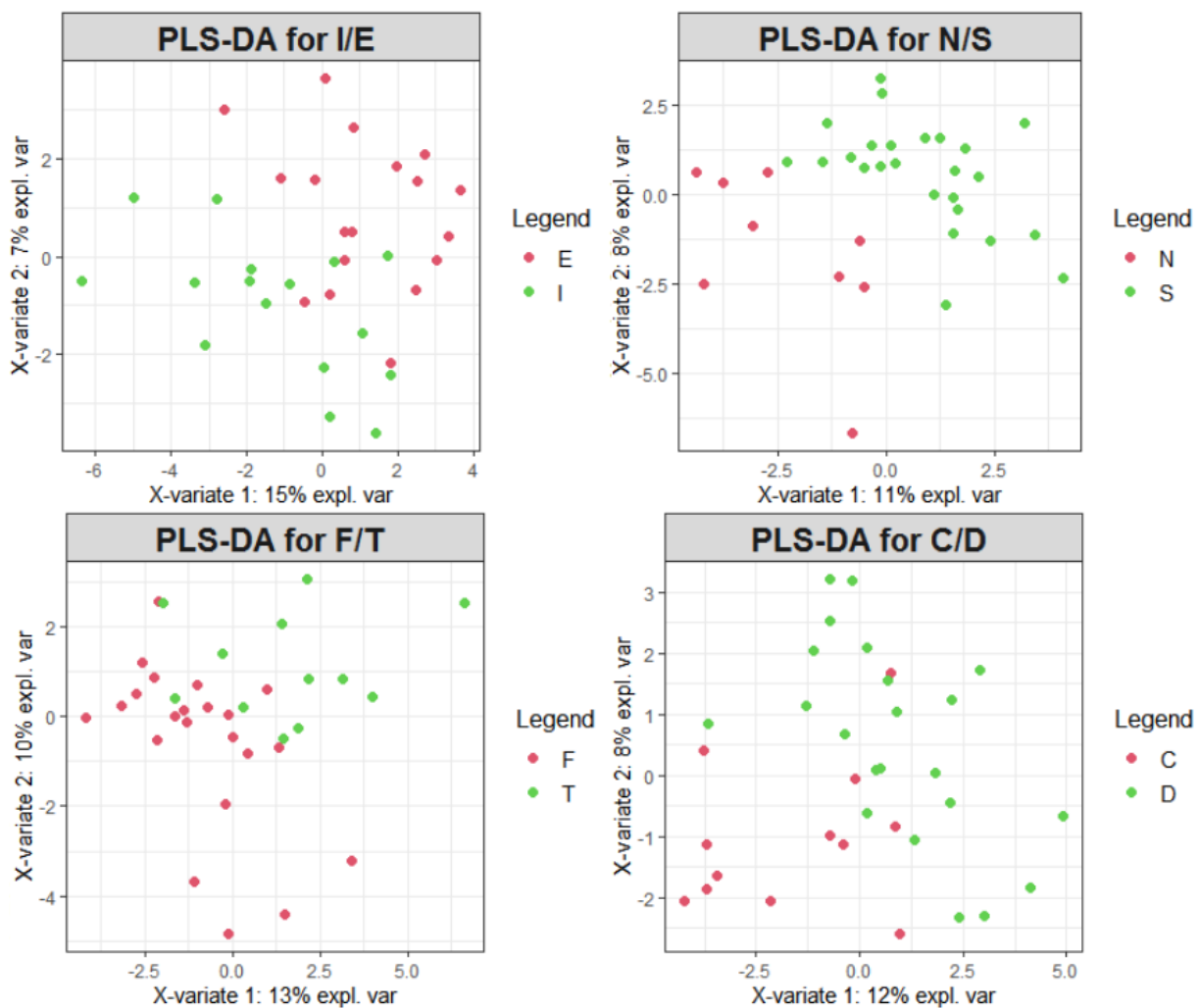
### **3.6.5 Samanhengar mellom personlegdomstype og dei samla responsane**

Frå PLS-DA får ein ei rekkje forskjellige resultat, og fleire av dei er samanfatta i resultatfigurar. For alle høve er variablane skalerte.

Av særskild interesse er kor godt personlegdomskaraktistikkane let seg klassifisere ved å nytte variablane frå alle dei fire responstypene frå dei ferdigprosesserte resultatfilene i Tabell 2.

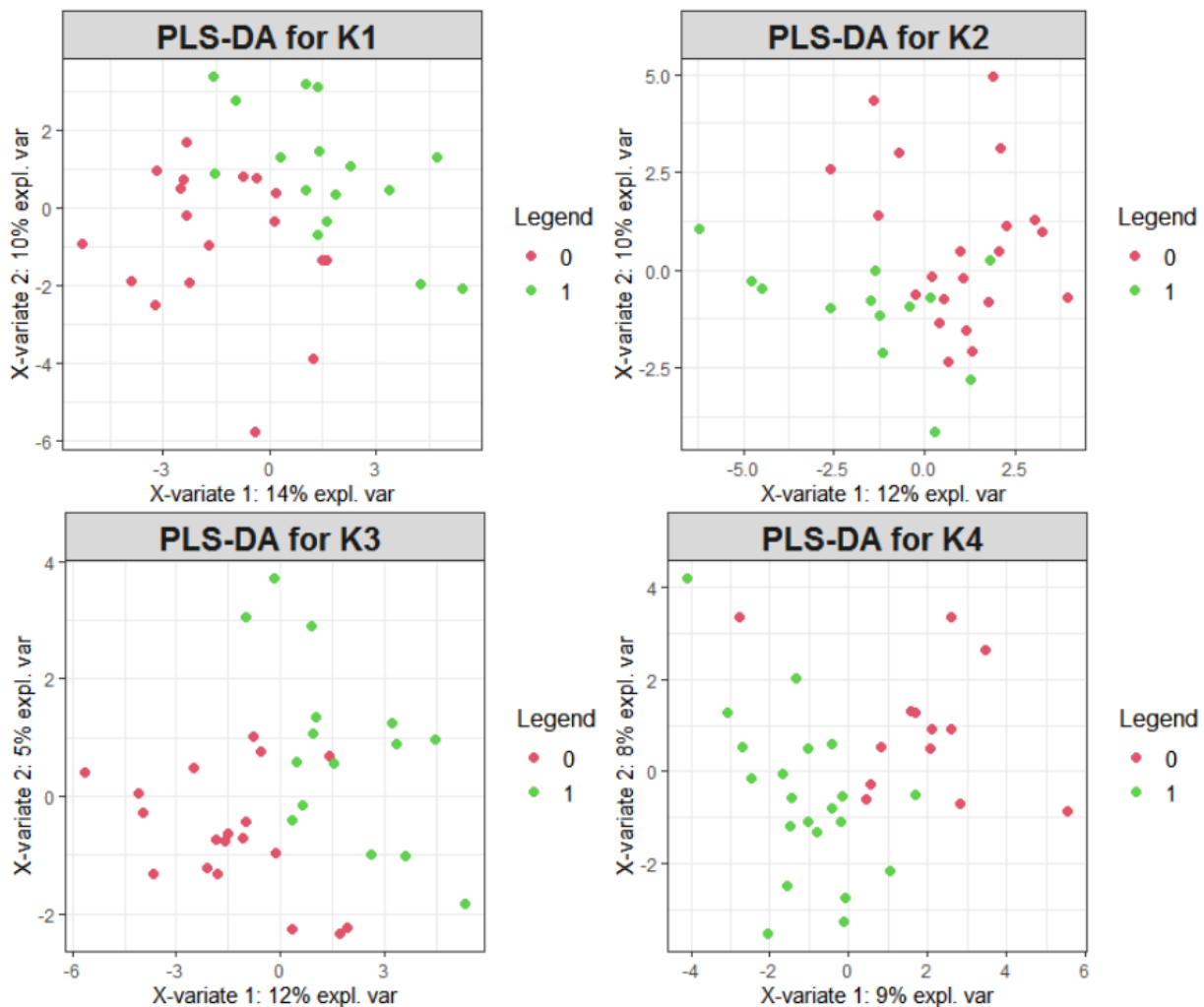
Tabell 2. Variablane nytta er alle spørsmåla frå spørjeundersøkinga, kva film respondentane såg først, dei 12 kjenslene frå Tabell 7, antal toppar per minutt og antal intervall med minst ein topp for dei to filmene. Dette gjer ei matrise av dimensjon 34x62.

Personlegdomskaracteristikkane og kontrollvariablane vert nytta som responsar etter tur. Frå Figur 59 syner korleis PLS-DA har skilt respondentane etter personlegdomskaracteristikkane. Det er ikkje råd å sjå eit heilt tydeleg skilje, med unntak for N/S.



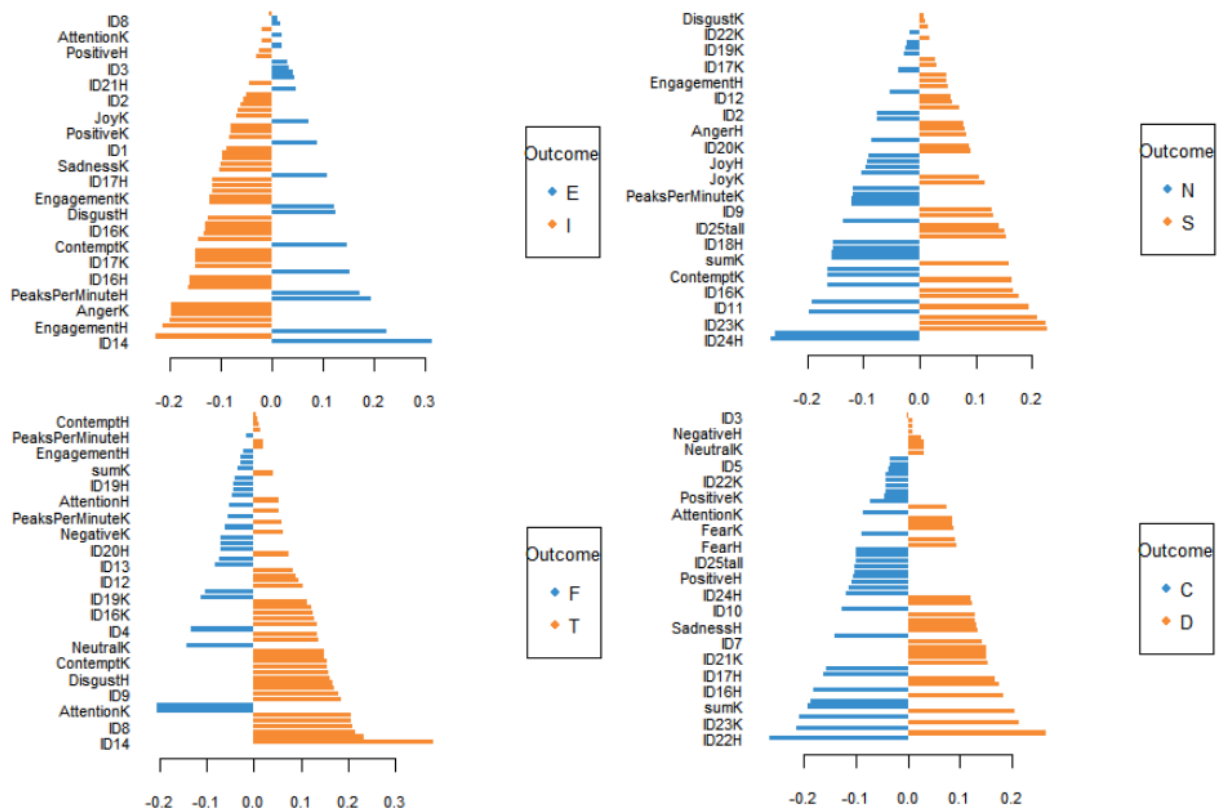
Figur 59 Scoreplott som syner respondentane representert ved dei to komponentane frå PLS-DA som beskriv mest samvariasjon mellom personlegdomskaracteristikkane og respontypane.

For samanlikning vert dei fire kontrollvariablane K1-K4 nytta som responsar, og Figur 60 syner at det er lite skilje mellom respondentane og i dette høvet. Det verkar til å vere ei aning høgare forklart samvariasjon mellom forklaring og respons for personlegdomskaracteristikkane enn for kontrollvariablane.

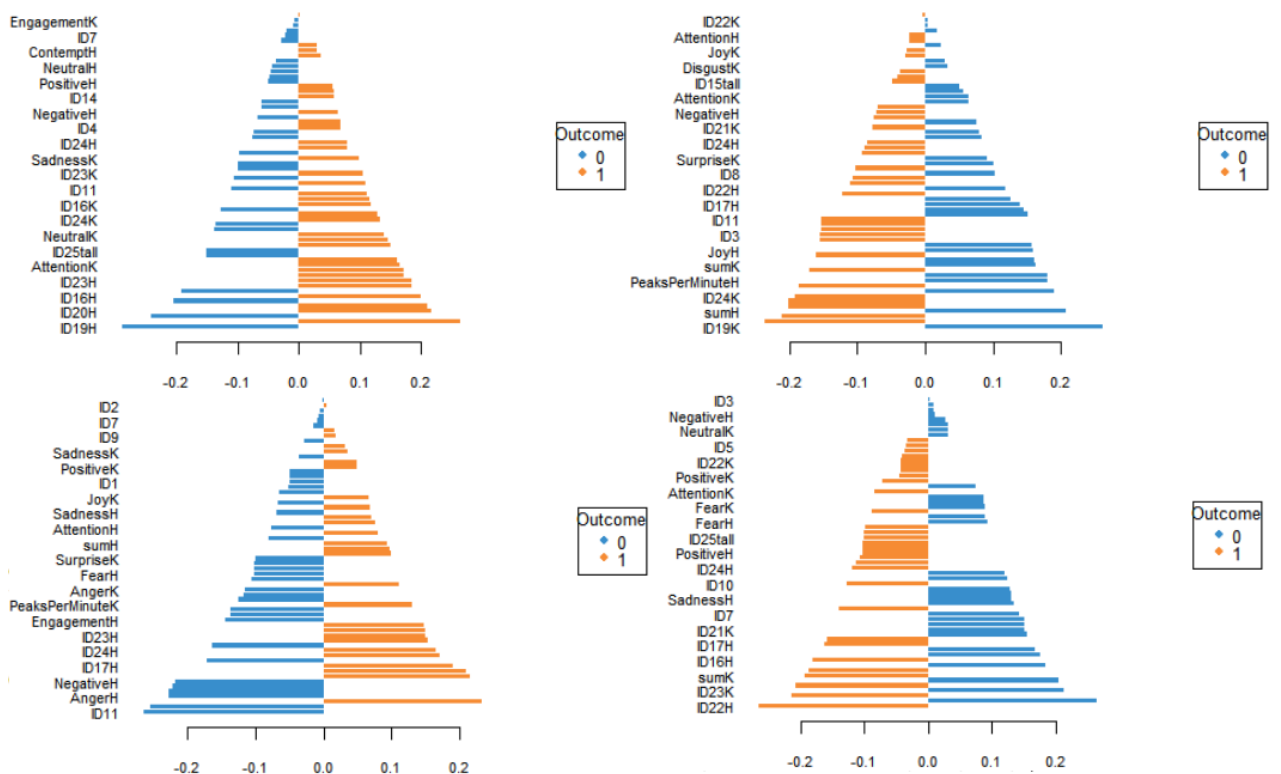


Figur 60 Scoreplott som syner respondentane representert ved dei to komponentane frå PLS-DA som beskriv mest samvariasjon mellom kontrollvariablane og responstypene.

Ladningsdiagramma i Figur 61 syner kva variablar som har størst påverknad på den første komponenten. Ein kan merkje seg at for alle høver er det ei blanding av variablar frå alle responstypene, og det er skilnad i kven av personlegdomskaracteristikkane variablane bidreg til mest til skilje. Ein kan og observere at i tilhøva I/E og N/S er ID14, om respondentane hadde lest om eller høyrte om temaet før forsøket som er den variabelen som bidreg mest. Det er ingen variablar som skil seg særskilt ut med svært mykje bidrag til den første PLS-komponenten. Ladningsdiagramma frå kontrollvariablane i Figur 62 skil seg ikkje noko særskilt frå ladningsdiagramma til personlegdomskaracteristikkane i fasing og uttrykk.



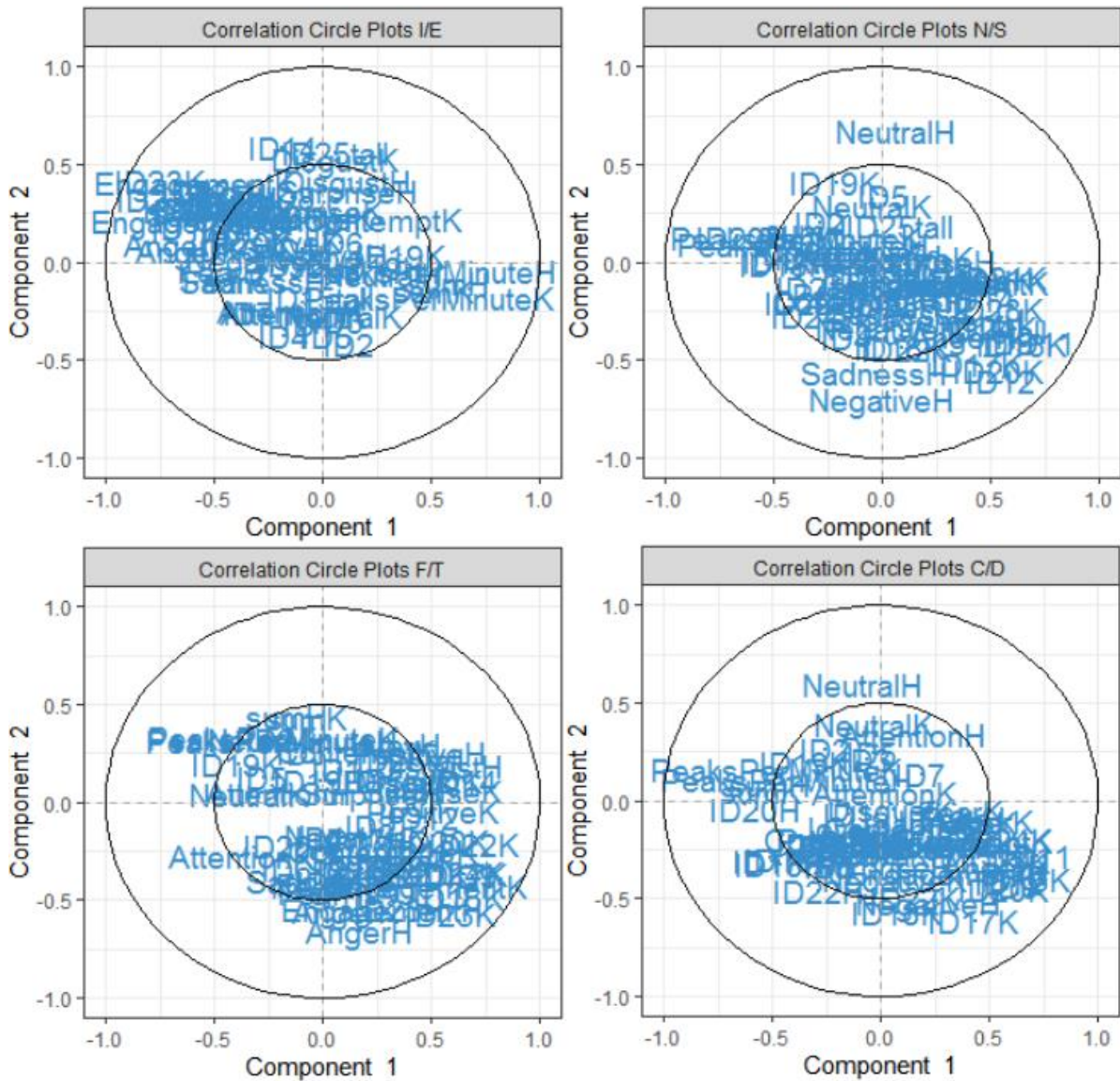
Figur 61 Laddingsdiagram som syner kva variablar som bidreg mest i første komponent for personlegdomskaracteristikkane.



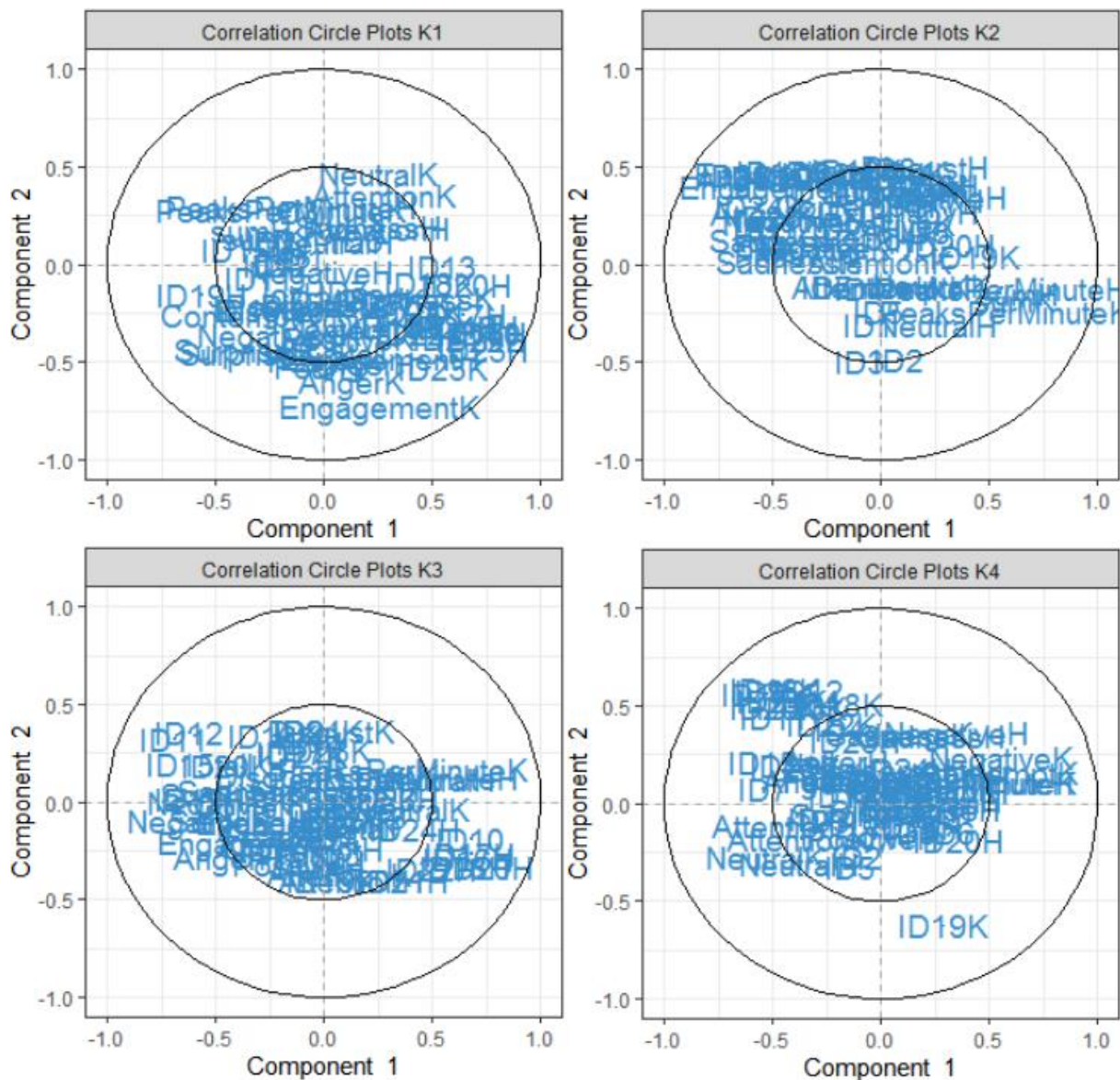
Figur 62 Laddingsdiagram som syner kva variablar som bidreg mest i første komponent for kontrollvariablane.

Variablane kan og representerast ved korrelasjonssirkelplott, for å sjå om dei på nokon måte gruppera seg. Figur 63 syner korrelasjonssirkelplotta frå personlegdomskaracteristikkane,

medan Figur 64 syner korrelasjonssirkelplodda frå kontrollvariablane. Det er i dette høvet og lite skilnad mellom personlegdomskararakteristikkane og kontrollvariablane når variablane vert representert ved første og andre komponent. Det er fleire høver der einssilde variablar trekkjer seg ut av hovedklynga, men dette er tilhøve både for personlegdomskararakteristikkane og personlegdomstypane. Hovudklynga av variablar verkar i begge høve til å ha noko større utbreiing langs den første komponenten.



Figur 63 Korrelasjonssirkelplodda som syner variablane representert ved komponent 1 og komponent 2 for personlegdomskararakteristikkane.

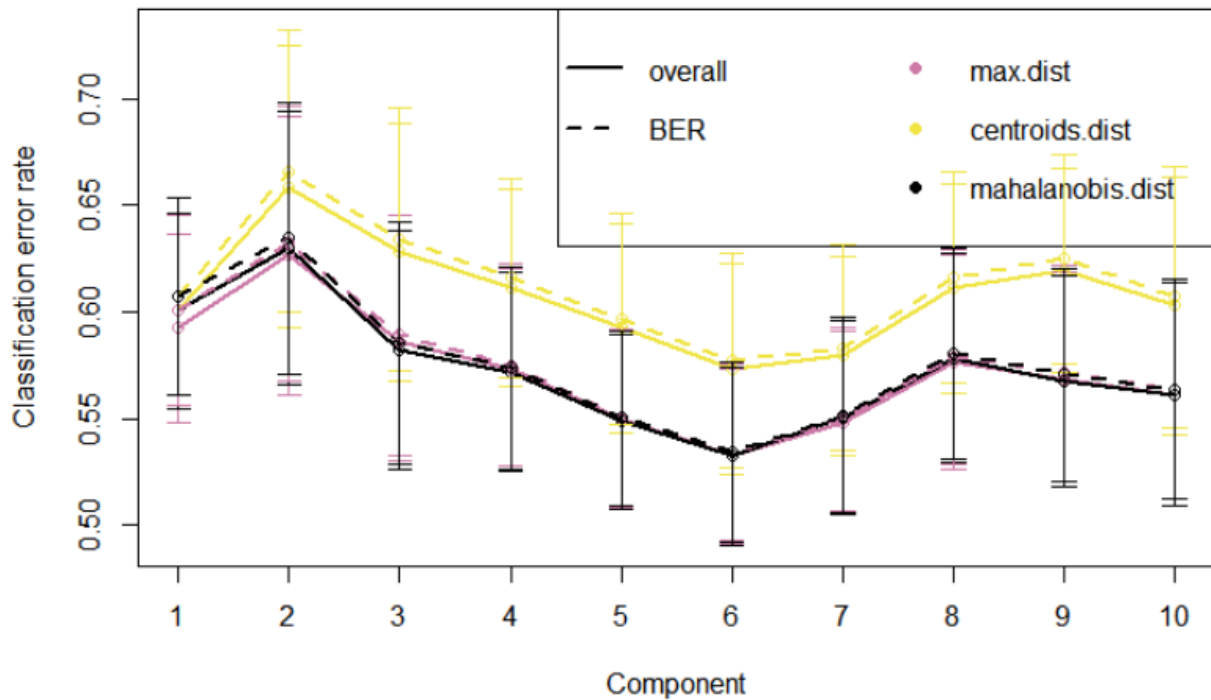


Figur 64 Korrelasjonssirkelplokk som syner variablane representert ved komponent 1 og komponent 2 for kontrollvariablane.

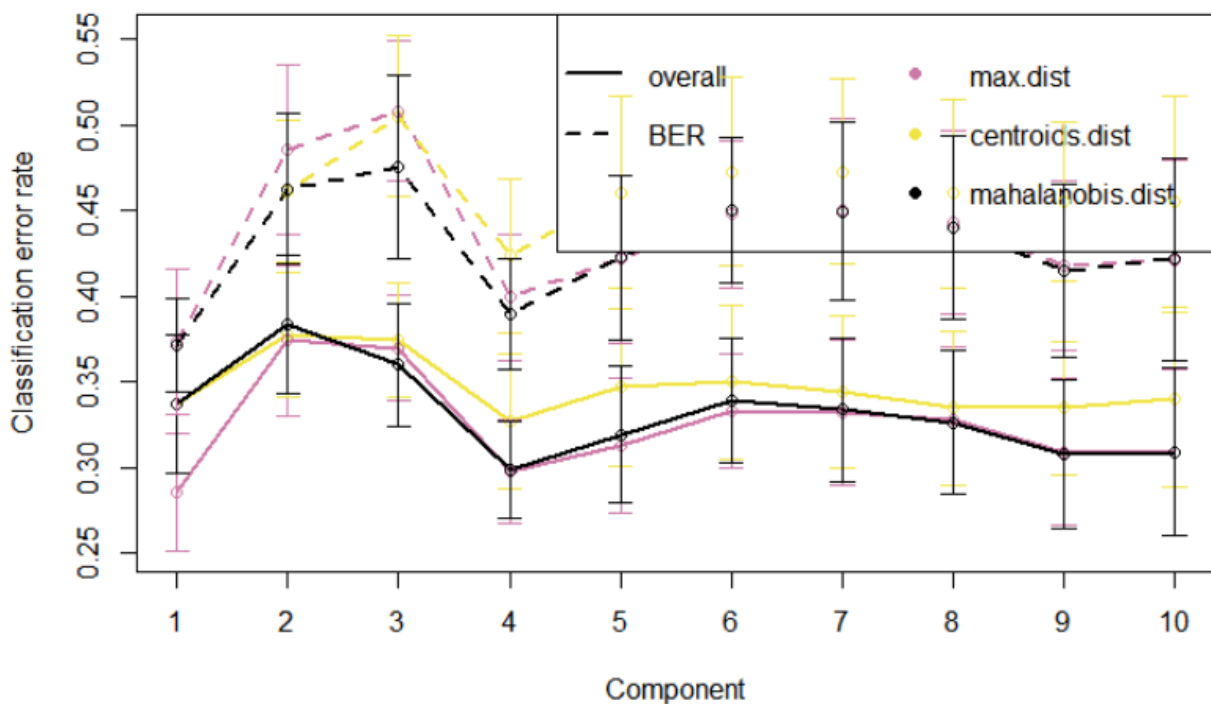
For å danne eit bilete av godheita til modellen, vart det ved nytta ei 10-fold kryssvalidering med tilfeldig uttrekk av foldar. For å få eit innblikk i variasjonen mellom utrekka vart det gjennomført 50 repetisjonar. Sidan scoreplotta i Figur 59 syner at lite av variasjonen som vert forklart i dei første to komponentane, og for å få eit innblikk i utviklinga i feilklassifiseringsrata, vart det nytta 10 komponentar. Skilnad mellom feilklassifiseringsrate og balansert feilklassifiseringsrate (BER), samt avstandsmål vert og inspisert.

Ved inspeksjon av Figur 65 til og med Figur 72 er det svært lite som tydar på samanhengar mellom personlegdomskaraktistikane og dei samla variablane frå forsøket. Med unntak av Figur 65 er det auke i feilklassifiseringsrate ved auke i inkluderte komponentar. I dei tilhøva responsane er mest balanserte, som ved I/E, syner feilklassifiseringsrata og BER same resultat. Ved ubalanserte responsar er BER noko høgare. Ein høgare BER kjem særst godt til

syne for personlegdomskararakteristikkane N/S, som og er den minst balanserte responsen. Ved inspeksjon av avstandsmål, er det ingen av dei som verkar til å gje jamt over betre resultat enn dei øvrige. Ved inspeksjon av BER, kjem denne reint unntaksvis under ei feilklassifiseringsrate på 0.5. Dette inntreff for personlegdomskararakteristikken N/S, men det inntreff og for kontrollvariablane K1, K3 og K4.

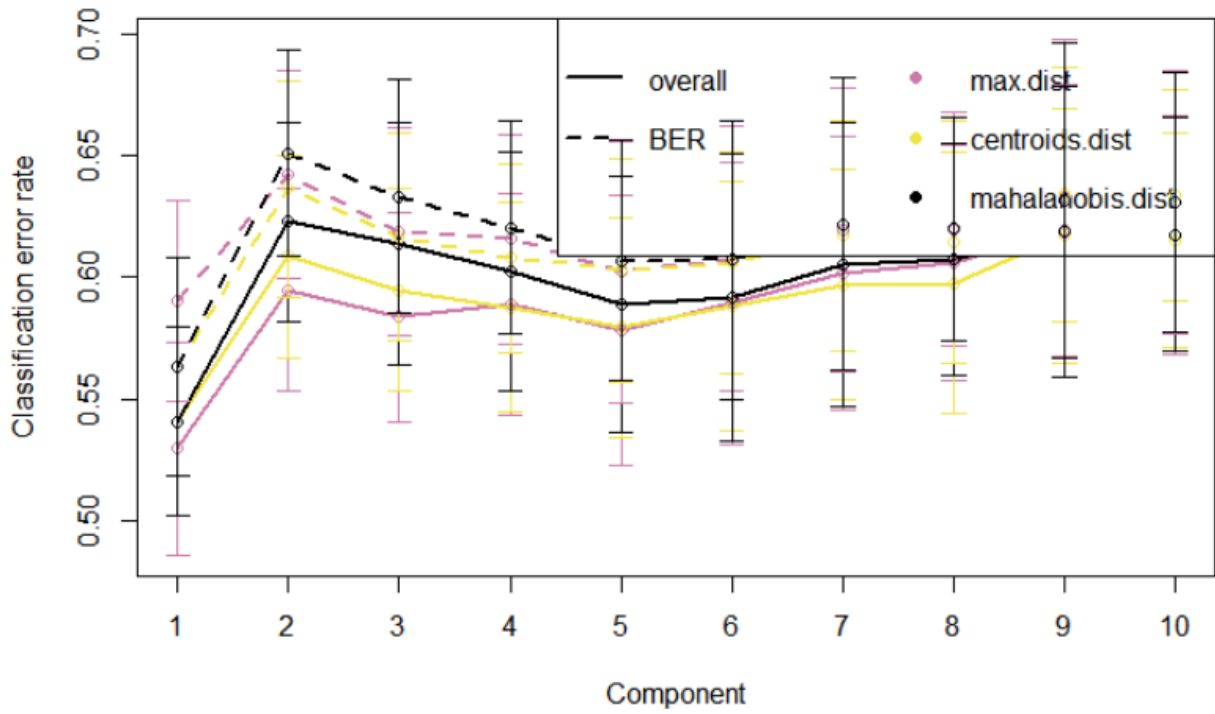


Figur 65 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for personlegdomskararakteristikkane I/E.

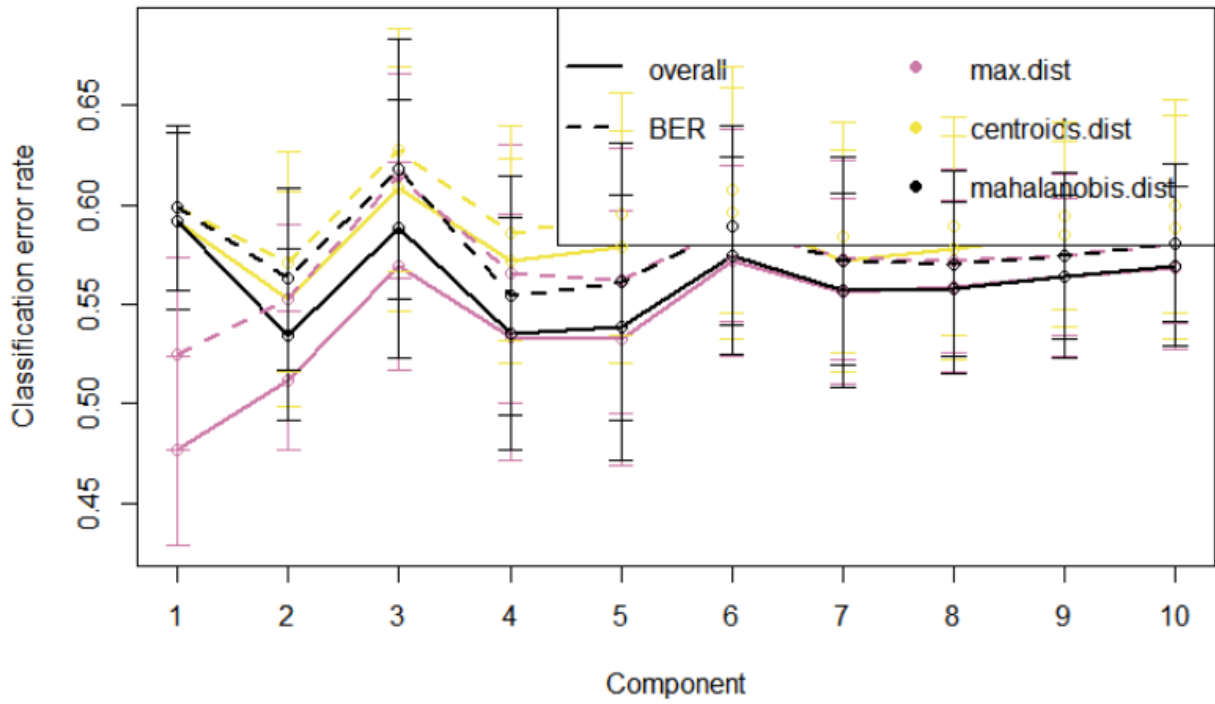


Figur 66 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for personlegdomskararakteristikkane N/S.

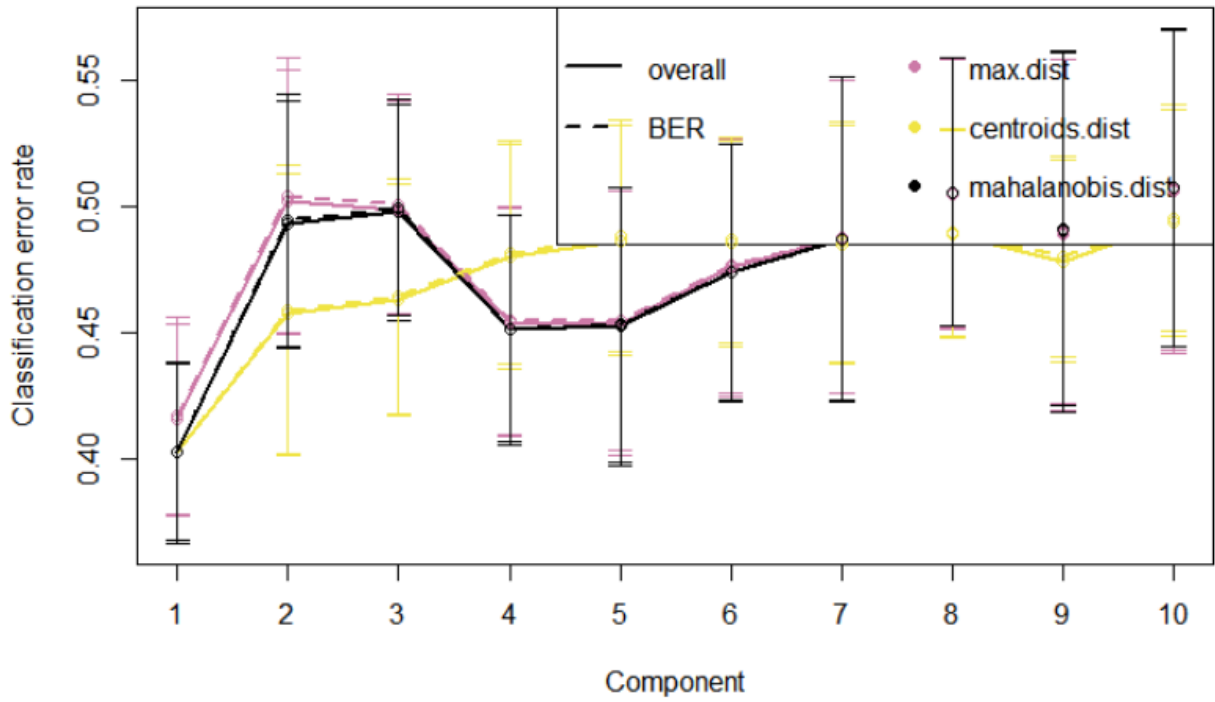




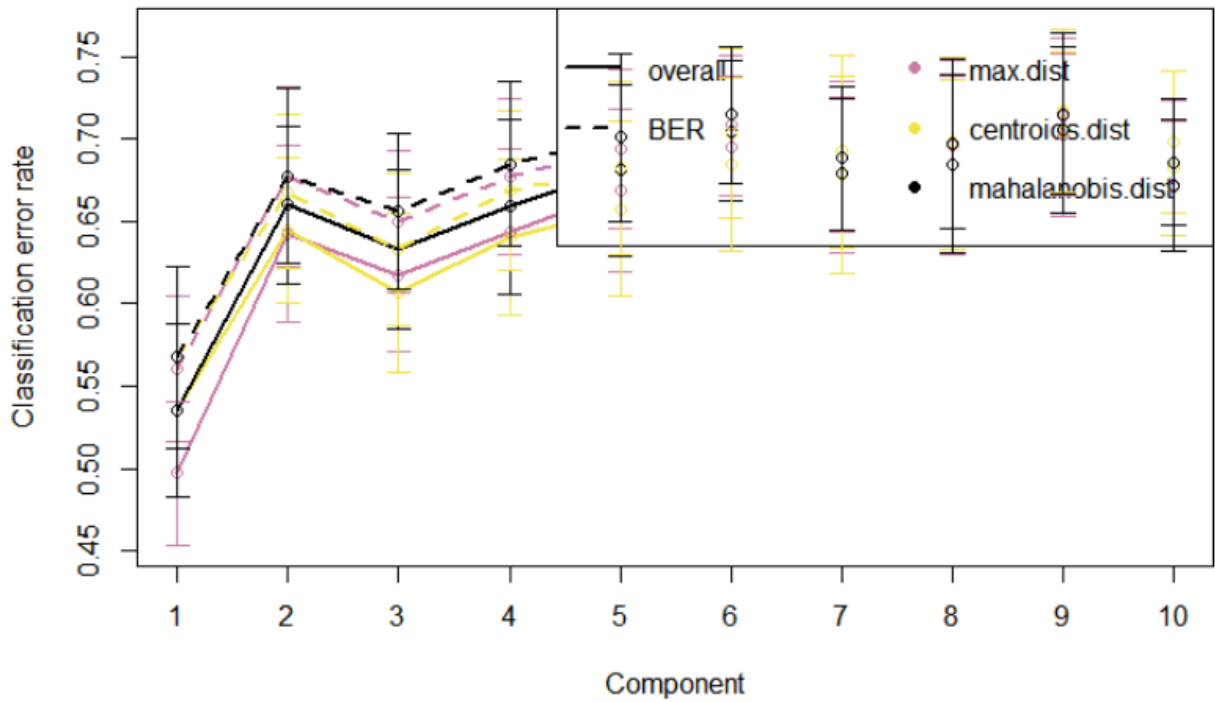
Figur 67 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for personlegdomskaraktistikkane F/T.



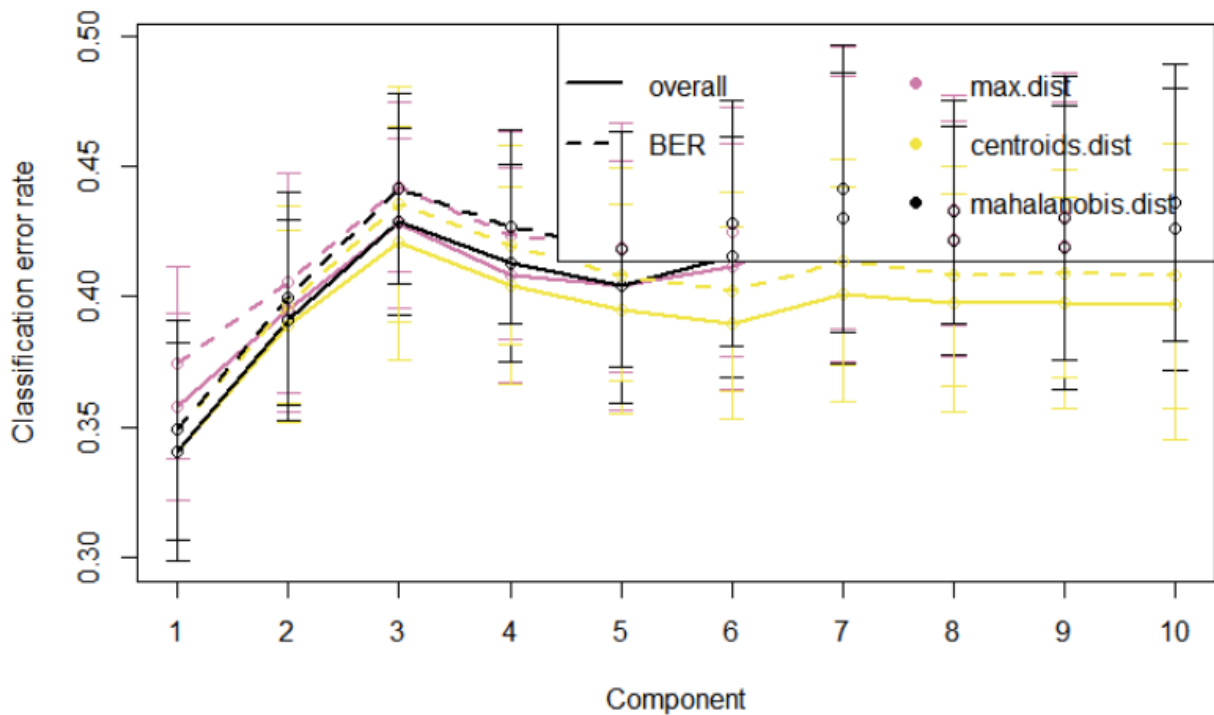
Figur 68 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for personlegdomskaraktistikkane C/D.



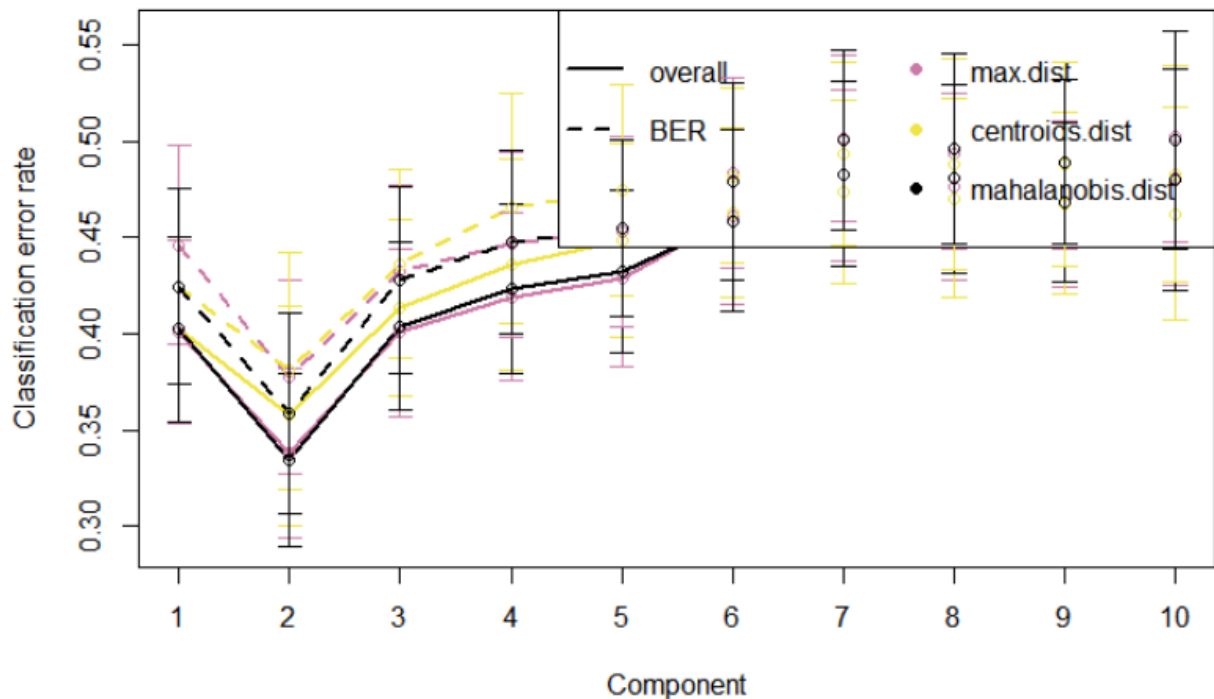
Figur 69 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for kontrollvariabelen K1.



Figur 70 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for kontrollvariabelen K2.



Figur 71 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for kontrollvariabelen K3.



Figur 72 Diagram som syner utvikling av feilklassifiseringsrate over ti komponentar for kontrollvariabelen K4.

Om dei overståande 8 figurane som syner utviklinga av feilklassifiseringsrate, kan det og påpeikast at dei ikkje liknar noko på eksempelfigurar i litteratur om PLS-DA, eller figurar frå tidlegare erfaringar med metoda, som syner ei synkende feilklassifiseringsrate ved å inkludere fleire komponentar. Ein kan undrast om det skuldast feil i utvinninga i resulata, men tidlegare undersøkingar styrkar vissa om at dette er den sanne ståda i datasettet.

## 4 DISKUSJON

---

### **Forskningsspørsmål: Har personlegdomstypen samanheng med opplevinga av forskjellige undervisningsmetodar målt ved spørjeskjema og biometriske sensorar?**

Etter ei heilheitleg vurdering av ei stor mengde diagram og statistiske testar, der dei fire binære karakteristikkane av personlegdomstypene er nytta som forklaringsvariablar for mange forskjellige responstypar, er det svært lite som antydner ein samanheng mellom personlegdomstypene og opplevinga av dei to undervisningsfilmane studentane vart eksponert for. Det er kun ved nokre få sporadiske tilhøve at personlegdomskaraktistikk har signifikant innverknad på ein respons eller at visuell framstilling syner antyding til grupperingar som følgjer personlegdomskaraktistikkane. Sidan dei tilfeldige kontrollvariablane gjer like mange statistisk signifikante resultat og grupperingar, kan ein med høveleg god sikkerheit konkludere med at i dette enskilde forsøket, med desse respondentane ikkje finnast anna enn enskilde og sporadiske samanhengar mellom personlegdomstypar og opplevingar målt ved spørjeskjema og biometriske sensorar. Det er ikkje gjort nokon formell korreksjon av signifikansnivå som følgjer av at eit stort antal testar er utført, og ein må difor ta høgde for at desse få sporadiske samanhengane kan vere tilfeldige.

Fråvær av samanhengar mellom personlegdomskaraktistikkane og responstypar kan vere vel så betydningsfullt som resultat i seg sjøve. Spørjeundersøkinga aleine held variablar som evnar å beskrive mykje av variasjonen i undersøkinga som heilheit, har synt seg å gje flest signifikante parvise skilnadar og har gjeitt resultat som kan vere lette å tilskrive ei logisk tolking. Av dei resultata som kan vere av relevant betydning når ein samanliknar responstypar, kjem over halvparten av dei frå spørjeundersøkinga. Det var difor overraskande at det var forsvinnande få statistisk signifikante samanhengar mellom spørjeundersøkinga og personlegdomskaraktistikkane, og ingen antyding til grupperingar av personlegdomskaraktistikkar korkje i spredningsdiagram eller scoreplott av data frå spørjeundersøkinga. Av dei statistisk signifikante samanhengane, er det kun to tilhøve der det er ein fornuftig mogeleg samanheng mellom personlegdomskaraktistikkane og spørjeundersøkinga, medan dei tilfeldige kontrollvariablane generera 7 signifikante samanhengar. Når tilfeldige kontrollvariablar evnar å gje nærast dobbelt så mange signifikante samanhengar, fordrar dette til varsemnd og kan indikere at dei observerte samanhengane mellom personlegdomskaraktistikkane og spørsmåla i spørjeundersøkinga like gjerne kan vere tilfeldige. Om dette syner den sanne ståda, eller om spørjeundersøkinga

ikkje er egna til å beskrive opplevingar i lys av personlegdomstypar er uvisst. I alle høver må komande prosjekt merkje seg og ta lærdom av denne uvissa.

Ein anna viktig lærdom ein kan trekkje frå undersøkinga av samanhengar mellom personlegdomskaraktistikkane og responstypane er at responstypar som burde støtte opp om kvarandre, men som ikkje gjer det, og kan gje nyttig innsikt. Eit godt døme på dette er når to datasett frå same sensor, målingane av galvanisk hudrespon (GSR), ikkje einast. Same observasjon, når overstiging av ein terskel indikera opphissing, blir representert både som antal GSR-toppar per minutt og antal tidsintervall med minst ein GSR-topp. Skilnad i GSR-toppar per minutt og skilnad i antal intervall med minst ein GSR-topp gjer svært sprikande resultat som responsar for dei introverte og dei ekstroverte respondentane (personlegdomskaraktistikken I/E). I/E hadde, som einaste høve, signifikant samanheng med skilnad i GSR-toppar per minutt. Når ein frå innleiande undersøkingar veit at det er godt samsvar mellom skilnad i GSR-toppar per minutt og skilnad i antal intervall med minst ein GSR-topp, vil det ikkje vere urimeleg å anta at ein burde kunne observere samanfallede resultat. Det var ikkje tilhøve. For skilnad i GSR-toppar per minutt og I/E er det signifikant samanheng og ein  $R^2$  på 0.12. For skilnad i antal intervall med minst ein GSR-topp og I/E er det ikkje signifikant samanheng og ein  $R^2$  på 0.03. Om den eine er meir riktig en den andre er uvisst, men denne samanlikninga bidreg til å syne kor vart og kjensleg dette datasettet er, og kor liten skilnaden mellom signifikans og ikkje signifikans kan vere.

At skilnaden mellom signifikans og ikkje signifikans kan vere svært liten, fordrar til stor varsemd ved val av statistiske metodar. Utforskinga av samanhengar mellom GSR summary scores og personlegdomstypane demonstrera dette på ypparleg vis. Når tre potensielt høvelege metodar vert utprøvd og samanstillt, er det signifikant samanheng mellom GSR-toppar per minutt for Hilde sin film og tre av fire personlegdomskaraktistikkar når poissonregresjon vert nytta. Det er og statistisk signifikant samanheng mellom GSR-toppar per minutt og kontrollvariablane ved fem tilhøve ved poissonregresjon. Resultata frå både lineær og logistisk regresjon, visuell inspeksjon av fargekoda spredningsdiagram og vurdering av dei logistiske regresjonslinjene syner tydeleg at det finnast svært lite samanheng mellom GSR-toppar per minutt og personlegdomskaraktistikkane. Dette er berre eit av mange eksempel på at forskjellige statistiske testar ikkje einast om resultatata. Det er diverre sjeldan så enkelt at den metoda som gjev lågast p-verdi gjer det riktigaste resultatet. Dette fordrar ytterlegare til varsamheit kring val av statistiske metodar og viktigheita av heilheitlege vurderingar av resultat.

Når brotstykket av datasettet vert sett saman til eit heile ved hjelp av PLS-DA, kan ein med rimeleg stor sikkerheit sei at det er svært lite skilje mellom personlegdomskaraktistikkane i lys av opplevingar målt med biometriske sensorar og spørjeundersøking. Det er og tilhøvet når dei tilfeldige kontrollvariablane vert nytta, dog med marginalt forskjellige resultat. Dette betyr ikkje nødvendigvis at resultatane frå tilfeldige kontrollvariablar er betre enn resultatane frå utdanningstesten til å forklare kva som gjer at respondentar reagera forskjellig på undervisningsfilmene. Det er fleire viktige faktorar som tilseier at ein, på trass av ei mengde statistiske testar og figurar som hevdar det motsette, ikkje bombastisk kan hevde at personlegdomstypene frå utdanningstesten ikkje har nokon samanheng med korleis STAT100-studentar opplev to forskjellige undervisningsfilmar.

Ein faktor er det låge talet på respondentar og korresponderande låg statistisk styrke. I tillegg til låg statistisk styrke er det og eit grunnleggjande problem med utvalet i forsøket, sidan ein ikkje veit kvifor akkurat desse respondentane valde å delta i forsøket. Sjølv om desse respondentane har delt sine meiningar og kjensler med oss, og kvar observasjon har derav sin eigenverdi, veit vi ingenting om dei vel 500 andre STAT100-studentane. Ein faktor er bruk av personlegdomskaraktistikk heller enn den fulle personlegdomstypen. Sjølv om statistisk styrke og personvernmessige hensyn ikkje tillot bruk av dei 16 personlegdomstypene, kan det ikkje neglisjerast at kvar einskild personlegdomskaraktistikk er ein del av eitt heile. Ein anna faktor er ei ofte lite omtalt begrensing som sit 30 cm frå PC-skjermen. Statistikk utan kontekst er berre tal på ein skjerm, og er nærast verdilause utan tolking og drøfting. Sjølv kor mykje ein enn freistar å unngå det, så vil eins person, kunnskapar og bakgrunn i større eller mindre grad farge tolkinga og drøftinga av resultat, så vel som påverke kva statistiske metodar som skal nyttast.

Vidare måtte resultatane frå utdanningstesten vurderast nærare. Som det framkjem av beskrivinga av utdanningstesten i 2.1.2, så består utdanningstesten av få spørsmål og er meir tilpassa ungdom i ungdoms- og vidaregåande skule. Rådgevarveileiing for utdanningstesten (Nasjonalt senter for realfagsrekruttering, 2017a) påpeikar at utdanningstesten grunna for få spørsmål ikkje kvalifiserer til å vere ein valid psykologisk test. Kvar personlegdomskaraktistikk framkjem kun frå eit einskild spørsmål med to svaralternativ, og kvart svaralternativ består av fleire setningar og påstandar, som kan gjere det vanskeleg å velje kva svaralternativ som passar ein best. Med dette er det dermed ikkje sagt at utdanningstesten gjer ei dårleg bestemming av personlegdomstypene som beskrive av Tabell 1. Men utan kalibrering mot ein lengre psykologisk test er det hefta usikkerheit til desse

resultata. For å forbedre dette kunne ein eksempelvis bedt respondentane og svare på den større utgåva nytta i (Sæbø et al., 2015) for samanlikning.

Sjølv om utdanningstesten si evne til å bestemme personlegdomstype er noko uviss, og det er respondentane si personlegdomstype som var utgangspunktet for dette forsøket, er det nokre fleire aspekt ved utdanningstesten som bør trekkjast fram. Vilkårleg om respondentane kjenner seg att i beskriving av personlegdomstype, vil prosessen med å besvare testen og lese resultata med stort sannsyn oppmode til refleksjon omkring ein sjølv, sine personlegdomstrekk og korleis ein lærer best. Alle som svarar på Utdanningstesten får attende eit omfattande resultatskriv. Resultatskrivet held ein gjennomgåande oppmuntrande og lite bastant tone som reflekterar uvissa i resultata, og syner mange mogelegheiter for utdanning og arbeid innan realfag og teknologi som ungdomar kanskje ikkje veit finnast. Måten resultata er beskrive på oppmodar til diskusjon og refleksjon omkring eins eigne personlegdomstrekk og læringspreferansar, både i einsam overveining og i samtalar med andre som og har tatt utdanningstesten. Difor vil noko av det overordna formålet ved prosjektet truleg dekkast, ved at studentane som inkluderast sjølv vil byrje å reflektere over korleis dei lærer best.

Eit av delmåla ved prosjektet var å undersøkje om responstypene frå Tabell 2 ville støtte opp om kvarandre. Formålet med å nytte fleire forskjellige sensorar i tillegg til spørjeundersøking er å nytte deira eigenskapar på eit slikt vis at fenomenet ein ynskjer å studere trår klarare fram når ein kan samanstille både bevisste og ubevisste reaksjonar hjå respondentane. Både for forskingsspørsmålet og for prosjektet som heilheit vil ei grundig undersøking av samanhengar mellom responsane belyse styrkar og svakheiter ved forsøket si oppbygning og responsane.

Den samla vurderinga av samanhengar mellom data i ferdigprosesserte resultatfiler skriv seg frå resultat frå nær 3000 statistiske testar, mange av dei med tilhøyrande inspeksjon av residual- og qqplot. Der variablane si utbreiing har vist seg eigna for særskilde metodar har dette i stor grad vore hensynsteke, og der utbreiinga ikkje syner noko form for trendar har velkjende og kraftige verktøy som MANOVA vore nytta for å freiste å fange opp samanhengar på tvers av variablar. Til tross for dette er det berre 3.8 % av resultata som er statistisk signifikante, sjølv om ein i det minste burde kunne forvente 5% tilhøve i tråd med det signifikansnivået som vart valt<sup>1</sup>.

---

<sup>1</sup> Som ein liten digresjon kan det nevast at ved ei feiltaking vart respondentnummera inkludert i framskaffing av prinsippalkomponentane frå spørjeundersøkinga nytta i alle PCR mellom spørjeundersøkinga og AFFDEX statistics. Når ein får færre signifikante samanhengar og lågare  $R^2$  ved å rette opp i feilen og nytte riktig datasett, kan det vere ein sterk indikasjon på at det rett og slett ikkje er mykje å finne.

Eit anna trekk ved resultatane frå dei statistiske testane av samanhengar mellom responstypane er at  $R^2$  for ikkje-signifikante samanhengar svært ofte er lågare enn 0.05. Ved signifikante samanhengar finn ein  $R^2$  til å i dei fleste høver ligge mellom 0.09 og 0.15, og reint unntaksvis høgare enn 0.15. Dette gjer god anledning til å diskutere kva som er ein låg  $R^2$  for generelle tilhøver. I nokre høver er ein  $R^2$  på 0.1 svært lågt, medan det er nærast rekord i andre høver. Kva som er ein høg eller låg  $R^2$  vil vere svært domeneavhengig, og vil vere basert på dels erfaring frå liknande forsøk og på samanlikning med resultat frå liknande studiar. I dette høvet er oppbygginga av forsøket av ein slik art at samanlikning med andre studiar vil vere uråd. For å finne eit klart svart på kva som i desse tilhøva er ein høg eller låg  $R^2$ , må eit utføre fleire forsøk. Den det kan tenkjast at noko så komplekst som eit menneskje si oppleving av ein undervisningsfilm vil variere i ei slik grad at noko som i fleire høver vil vere ein låg  $R^2$  i røynda kan vere eit svært godt resultat. Noko sikrare antaking vil vere uråd å befeste utan bakgrunn i litteratur eller erfaring frå fleire liknande forsøk.

Av dei totalt 114 statistisk signifikante samanhengane, vil nærare inspeksjon avsløre at dei aller fleste av dei tilfører lite anna enn ei gylden anledning til å diskutere den flyktige, men akk så viktige skilnaden mellom statistisk signifikans og praktisk signifikans. Eit av mange eksempel på dette kan vere samanhengar mellom dei lågare prinsipalkomponentane frå spørjeundersøkinga og AFFDEX statistics. Sjølv om det riktignok er ein statistisk signifikant samanheng mellom til dømes PC30\_SU og PC17\_AS tydar det kun at noko som forklarar under 0.01% av spørjeundersøkinga har samanheng med noko som forklarar under 0.01% av AFFDEX statistics. Sjølv om det er ein statistisk signifikant samanheng mellom desse variablane, kan ein ikkje neglisjere det faktum at den praktiske signifikansen neppe kan seiast å vere av betydning.

Frå dei 114 statistisk signifikante samanhengane mellom responstypane, var det eit fåtal av dei som vart vurdert til å vere både statistisk signifikante og samstundes holde ei antyding til ei praktisk og relevant tolking. Når så få av samanhengane kan seiast å ha både statistisk og praktisk relevans, vil det vere naturleg å vende merksemda mot metodane nytta for å undersøkje desse samanhengane. Ein kan, med rette, undrast over kvarvidt dei nytta metodane er velegna for formålet. Den samla erfaringa frå inspeksjon av 700 spredningsdiagram og 600 spredningsdiagram fargelagde etter utvalde variablar tilseier at dei statistiske testane stemmer med inntrykket frå den visuelle inspeksjonen. Det er svært lite å finne av trendar og grupperingar. I dei få tilhøva der fargekoding av respondentane har vist antyding til klynging, var det framleis vanskeleg å finne noko naturleg skilje. Dette syner svært stor variasjon



mellom respondentane, lite samanheng mellom responstypene og understøttar den samla vurderinga av dei statistiske testane. Det er difor uvisst om dei samhengane som mogelegvis kunne tenkjast å ha ei logisk forklaring, i røynda har betydning. I dette høvet har dei absolutt si betydning ved å tydeleg demonstrere kor enkelt det kan vere å tillegje tal meiningar og årsaka, når det i røynda ikkje treng å finnast korkje meining eller årsak.

Eit av delmåla for oppgåva var å utforske kvar einskild responstype. Sidan det er svært få samhengar mellom responstypene, og mellom responstypene og personlegdomstypene, vil det vere naturleg å undersøkje om årsaka til dette kunne finnast her. Hovudinntrykket frå den samla undersøkinga av responstypene kvar for seg, er at dei nærast utløyser fleire spørsmål enn dei gjer svar på. Først og fremst kan ein merkje seg det store spennet i både responstype, skala og fordeling. Datasetta favnar mellom anna binære, tilnærma normalfordelte og poissonfordelte variablar, samt variablar som ikkje kan seiast å ha noko tydeleg fordeling i det heile. Ein har og både einskilde og parvise variablar samt faktorar om einannan. For å kunne svare best mogeleg på forskningsspørsmålet, og for å lette vidare arbeid måtte alle variablane i kvar einskild responstype utforskast med varsemd.

Eit av måla for utforskinga av responstypene i seg sjølve og oppgåva som heilheit, var å finne variablar eller lineærkombinasjonar av variablar som kan vere særers hensiktsmessige å nytte når variablar frå to eller fleire sensortypar skal nyttast saman, eller knytast opp mot personlegdomstypene. To potensielt gode variablar frå spørjeundersøkinga er ID15, kva film respondentane hadde valt om dei kun fekk sjå ein film i STAT100, og ID11, kva film respondentane følte dei lærte mest av på ein skala frå Hilde (0) til Kathrine (10). Desse i seg sjølve beskriv mykje av variasjonen i spørjeundersøkinga som heilheit. Frå AFFDEX statistics, GSR summary scores og GSR binned peak detection kan nye lineærkombinasjonar som syner skilnad mellom observasjonar frå Hilde og Kathrine sine filmar i mange høver vere gode kandidatar til responsvariablar når responsane skal knytast saman eller når dei skal analyserast i samanheng med personlegdomstypene. Fordelen med å nytte skilnad i parvise observasjonar er at desse skilnadane ofte er tilnærma normalfordelte med eit tyngdepunkt om null, som tilet bruk av statistiske testar som forutset normalfordelte restfeil og konstant variasjon og derav kan vere lettare å anvende statistiske testar på. Ulempa er at det er kun skilnad mellom variablane som kjem fram, ikkje variabelen i seg sjølve.

Sjølv om hovudfokus ligg på samhengane mellom personlegdomstypene og responsane, opnar inspeksjonen av responstypene for parvise samanlikningar mellom reaksjonane på dei to filmane. Ei rekkje statistiske testar for parvise variablar var utført for å avdekkje om det

finnast statistisk signifikant skilnad mellom reaksjonane på dei to filmene. Av 49 forskjellige testar var det statistisk signifikant skilnad ved 4 tilhøve, 3 av dei frå spørjeundersøkinga og ein av dei frå AFFDEX statistics. At så få tilhøve av parvise testar syner signifikant skilnad støttar opp om hovudintrykket frå utforsking av samanhengar mellom responstypene og mellom responstypene og personlegdomstypene. Når parvise testar innad i ein responstype ikkje evnar å skilje opplevinga av dei to filmene, er det lite sannsyn for at korkje personlegdomstypene eller dei andre responstypene evna og skilje filmene. I lys av antalet ikkje signifikante resultat vil det vere naturleg å vende merksemda mot utforminga av forsøket og sensorane i seg sjølve.

Sensorane nytta i forsøket har sine svakheiter og moglege feilkjelder. Temperatur og luftfuktigheit er to av fleire faktorar som kan ha påverknad på GSR (Boucein, 2012), og i retrospekt kunne ein gjerne ha ein målar for temperatur og luftfuktigheit i rommet der forsøka blir utført. Ein hadde då høve til å halde oppsyn med ståda og å kunne modellere inn desse effektane ved særskilde høver. Det er og ein del faktorar ein ikkje har høve til å gjere noko med, men som påverkar huda si ledningsevne, som huda sin temperatur, om ein er dehydrert eller ikkje og kor tørr ein er i huda (Boucein, 2012). Sjølv om ein ikkje kan rå over desse faktorane, er kjennskapen til dei viktig. Det er difor viktig å sjå målingane til kvar respondent under eit, for å fanne om desse forskjellane. Om resultatata frå kalibrerte rådata eller tilrettelagde resultatfiler tek høgde for individuelle forskjellar eller tek utgangspunkt i ein gjennomsnittsrespondent er diverre uvisst, og av svært stor betydning. Til dømes vil det vere uråd å vite sikkert om ein respondent med mange toppar per minutt let seg engasjere heilt enormt av undervisningsfilmar om statistikk, eller om respondenten rett og slett berre er veldig klam på hendene til vanleg.

For å avgjerde om ein respondent var klam på hendene eller svært engasjert i undervisningsfilmene vart AFFDEX statistics nytta, men det er og nokre feilkjelder knytt til denne sensoren. Kjenslegjenkjenningsalgoritma AFFDEX evnar å klassifisere kjensler frå bilete av menneskje som syner prototypiske kjensler om lag like bra som eit vanleg menneskje (60-80% treffsikkerheit), men berre ved å erstatte bileta med forsøkspersonar som blir bedde om å syne prototypiske kjensler dett treffsikkerheita til 55% (Stöckli et al., 2018), og det er forskjell mellom treffsikkerheita for kvar kjensle. Når respondentar blir synt positive, negative og nøytrale bilete under antakinga at positive bilete vekker positive kjensler og at negative bilete vekker negative kjensler, klarte AFFDEX å klassifisere riktig i berre 21% av tilhøva, og dei korrekte klassifiseringane skriv seg i stor grad frå at AFFDEX

klassifiserer 70% av tilhøva som nøytrale og vil difor i stor grad treffe riktig på nøytrale kjensler (Magdin et al., 2019). AFFDEX si evne til å fange opp subtile kjensler frå naturlege ansikt er fortsett ukjend, og er nok ein faktor som fordrar til varsemnd i tolking av resultatata.

Sjølv om ei felles plattform for samanknyting av sensorar og synkronisering av resultat er svært nyttig og arbeidssparande, det har gjennom heile arbeidet med prosjektet vore ei utfordring å tyde innhaldet i resultatfiler frå iMotions. Særskilt i ferdigprosesserte filer og for nokre rådata er det få beskrivingar av kva variablane er, og kva eining målingane har. Det er og nokre variablar som rett og slett ikkje har noko eining og nokre einingar som mistenkjast og for å vere direkte feil. Det er til dømes vanskeleg å forestilje seg at ein respondent sit 593 centimeter frå skjermen. Det er og vanskeleg å finne forklaringar til og grunnlag for val av terskelverdiar. Det er og lite informasjon å finne om kva kalibrering som blir gjort, og om det blir korrigert for at nokon er meir responsive en andre. Sjølv om iMotions er ei kommersiell bedrift som er behjelpelege med støtte til bruk av programvara og analyser av data mot betaling, ville enkelt tilgjengelege forklaringar til algoritmar, variablar og tersklar gjort forskning enklare. Resultat av forskning vil og vere meir påliteleg med auka kjennskap til korleis kjenslene vert klassifisert, noko (Stöckli et al., 2018) påpeikar i sin omfattande valideringsstudie av AFFDEX.

Eit forsøk er aldri betre enn si førebuing, og ei årsak til at dette forsøket gjer svært få og utydelege svar kan like gjerne skrive seg frå førebuingane, eksempelvis frå filmane eller oppbygning av forsøket. Om ein skal freiste å nytte sensorane og spørjeundersøkinga for å utforske korleis respondentar opplev forskjellige undervisningsmetoda, må forsøket førebuast på ein slik måte at sensorane sine styrkar og svakheiter vert ivaretekne. Av ei rekkje årsaker er dette ikkje tilhøvet for dette forsøket, og den mest framtrudande av desse årsakene er undervisningsfilmane.

Dei største begrensingane ved filmane var at dei var av ei vesentleg lengde, og i utgangspunktet ikkje utarbeidd for sensorane nytta i forsøket. Felles for alle gjennomgatte studiar der biometriske sensorar freistar å måle reaksjon på eit stimuli, er at dei måler eit og eit stimuli av gongen, og varigheita er gjerne nokre sekund, med pauseskjerm mellom kvar stimuli for «nullstilling» av respondenten. Då blir det gjerne målt høgaste GSR-amplitude (mikrosiemens) eller kva kjensle som har høgast intensitet innanfor tidsramma til stimuliet (Danner et al., 2014; Lei et al., 2017; Leitch et al., 2015; Zhi et al., 2018). Sidan filmane er eit langt stimuli, medfører det at ein frå ferdigprosesserte resultatfiler kun har mogelegheit til å sjå kor mange reaksjonar som inntreff for kvar film. Men det er ikkje likegyldig når i filmane

respondentane reagera, og ved å dele filmene inn i kortare epokar som inneheld eit stimuli kvar, kunne ein tenkjast å fange opp variasjonane ikkje berre innad i responsar og respondentar, men og innad i kvar film.

Om sjølve innhaldet i filmene er det innhaldet som ville vere best egna til å forklare grunnleggande statistikk til forskjellige personlegdomstypar er uvisst. Sjølv om Kathrine og Hilde har forskjellig tilnærming til undervisning, og har dels forskjellige personlegdomstypar, er det ikkje utenkjeleg at filmene ikkje var forskjellige nok. Riktig nok synte dei innleiande undersøkingane av spørjeundersøkinga ein samanheng mellom korleis ein lærer best sjølve og kva ein trur andre vil lære av. Men er det ikkje nødvendigvis slik at forskjellane i personlegdomstypene til Hilde og Kathrine vil kome til uttrykk i filmene på ein slik måte at filmene vil appellere meir til nokon personlegdomstypar enn andre. Filmene kunne med fordel vore utarbeidd med hensyn til kva personlegdomstypene ville føretrekke. Ein må og ta høgde for at undervisningsfilmar om tema frå grunnleggande statistikk mogelegvis ikkje er det som set kjenslene mest i kok hjå respondentane.

Undervisningsfilmene er heller ikkje laga med blikksporingsforsøk i tankane. I blikksporingsforsøk er vanlege forsøksdesign basert på merka Area Of Interest (AOI) på eit stimuli. AOI kan til dømes vere relevante opplysningar blant irrelevante opplysningar, eit relevant bilete assosiert med tekst, ein knapp eller symbol. Ofte vil ein telje kor lang tid det tar før respondenten fiksera på AOIet, kor mange fikseringar på AOIet, om respondenten fiksera på to AOI som er i samanheng (Alemdag & Cagiltay, 2018). Filmene si oppbygning gjer det vanskeleg å setje slike AOI i etterkant, sidan filmene ikkje inneheld irrelevant informasjon. Det er heller ikkje scener der bilde og tekst er knytt saman. Kva dynamiske visuelle variablar som er nytta til formidling av informasjon er og forskjellig for dei to filmene. Begge nyttar skriving på smartskjerm, men i den eine filmen er det nytta animasjon med fleire raude felt i beveging, noko som er sterkt merksemdsvekkjande. Skal ein på nokon måte kunne nyttegjere seg av blikksporing, må stimuliet i utgangspunktet vere utarbeidd for formålet.

At filmene si lengde medfører rådata av eit omfang på vel 252 000 målingar av 179 variablar for kvar einskild respondent, burde i seg sjølve vere motivasjon til å heller ha fleire og kortare stimuli. Men det store omfanget av rådata belysar og den svært viktige kjennsgjerning at ein student sine bevisste og ubevisste responsar på ein undervisningsfilm i seg sjølve kan vere så omfattande og komplekse at det er vanskeleg, om ikkje umogeleg å favne om dei. For å tilføre ytterleg med vanskar, vil det vere ei rekkje faktorar som vil påverke responsane frå kvar einskild respondent. Sjølv om ein freistar å utføre alle dei einskilte forsøka under så like

forhold som råd, er det ei rekkje faktorar som ikkje vert att utanfor døra til biometrilaben. Dette kan vere respondentane si oppseding, deira tidlegare erfaringar med statistikk og matematiske formlar, om dei har ein dårleg hårdag, om dei er trøtte eller svoltne, eller om dei gler seg til å reise heim for å ha fredagstaco med familien sin for første gong på lenge. For å freiste å skilje ut kva som er ein respons på sjølv undervisningsmetoda og kva som er ein respons på undervisningsmetoda i samanheng med personlegdomstypen bør framtidige prosjekt førebuast på ein heilt anna måte. I alle høve må fleire forsøk til for å ikkje berre kaste lys over STAT100-studentar sine opplevingar av forskjellige undervisningsmetodar, men for å finne høvelege måtar å måle, analysere og tolke desse opplevingane i lys av personlegdomstypar.

#### **4.1 VIDARE ARBEID**

Det er først når ein skal drøfte vidare arbeid at dette prosjektet verkeleg kjem til sin rett. Som nybrottsarbeid flest, låg ikkje verdien av prosjektet i resultata, men i grundig utforsking av responsar og at svakheiter i forsøksdesign vart avdekt. Komande arbeid kan då starte ut med kjennskap til korleis dei ferdigprosesserte resultatfilene er oppbygd og kva dei held av informasjon. Komande prosjekt har høve til å trekkje mykje lærdom av biometriske sensorar sine styrkar og svakheiter, og korleis forsøk kan førebuast med hensyn til desse. Først når desse hensyna er ivaretekne vil det vere mogelege å gje eit påliteleg svar på om personlegdomstypane frå utdanningstesten har samanheng med opplevingar målt ved biometriske sensorar. Svært mykje av vanskane med denne type problemstillingar er at samhengane gjerne er svært spreidde, mogelegvis og ikkje-lineære og komplekse. For å nøste ut av slike fenomen treng ein mest sannsynleg store data, større enn alle STAT100-studentar kvart år i mange år. Difor vil anbefalingar for vidare arbeid fokusere på korleis ein kan nytte små datasett eller svært enkel datafangst til å utvinne små brikkar i eit stort puslespel.

Anbefalingar for vidare arbeid har fleire forskjellige opprinningar. Dei kjem dels frå utfordringar undervegs i prosjektet, dels frå inspirerande samtalar med respondentar, veiledarar og vener og dels frå korleis ein hadde gått fram om ein hadde hatt høve til å gjere prosjektet på nytt. Og nokon har si opprinning frå dei mange spanande sidespora ein gjerne skulle forfølgt, men som ikkje passa inn i rammene for oppgåva.

Noko som ein lett kan gjere endringar på er spørjeundersøkinga. Det kunna vore gjort mykje betre nytte av spørjeundersøkinga enn kva som har vore gjort i dette forsøket. Undersøkinga burde innehaldd spørsmål om filmane som var meir retta mot responsar frå

ansiktsgjennkjenningsalgoritma og mot personlegdomstypene. Ei mogelegheit kan vere å be studentane rangere frå 0 til 100 kor mykje dei sjølve meiner dei følte av kvar kjensle når dei vart eksponert for stimuliet, for samanlikning mot AFFDEX statistics. Ei anna mogelegheit er å be respondentane velje kva kjensle dei sjølve meiner dei følte mest av når dei vart eksponert for stimuliet. I begge høva har ein mogelegheit til å nytte responsen frå spørjeundersøkinga og responsen frå AFFDEX statistics både kvar for seg, samla og knytt saman med personlegdomstypar. Ein kunne og spurt om kva rekkefølge dei ville sett filmene i, om dei fekk sjå begge filmene ein gong for å sjå om føretrekt rekkefølge hadde samanheng med personlegdomstypene. Dette kunne gje nyttig innsikt, særskild sidan fleire av respondentane ga uttrykk for at dei tykte filmene utfylte kvarandre.

Spørjeundersøkinga si eintydige oppbygning natur er ofte ein fordel, men kan og vere eit hinder når ein freistar å kartleggje noko så komplekst som opplevinga av undervisningsfilmar. Etter kvart som tekniske løysingar har utvikla seg kan mindre rigide variantar forsøkast. Til dømes kunne respondentane blitt bedd om å velje nokre ord frå mange mogelege som beskriv kva dei tykte om filmene, kva kjensler dei opplevde mest av og kva dei likte minst og mest med filmene. Fordelen med å kunne avdekkje fleire nyansar av opplevinga må målast mot ulempa med ein meir komplisert modell. Dette forsøket kan og nyttast utan biometriske sensorar, og utført på e-læringsplattform kunne ein fått mange fleire respondentar og mogelegvis kunne avdekt spanande samanhengar mellom undervisningsfilmene og personlegdomstypene.

For å avdekkje korleis forskjellige respondentar opplev ei undervisningsmetode, vil det vere svært viktig å vite korleis respondenten opplev noko som ikkje er ei undervisningsmetode. Om ein respondent til dømes syner mykje av ansiktsuttrykk tolka av AFFDEX som glede medan vedkomande er eksponert for ei undervisningsmetode, er det uråd å vite om denne respondenten er svært begeistra for undervisningsmetodar eller rett og slett var glad til sinns den dagen forsøket vart utført. Ved å eksponere respondentane for bilete eller filmar som kan forventast å vekke forskjellige kjensler, kan ein lage ei basislinje for kvar einskild respondent og samanlikne reaksjonar som relative forskjellar heller enn absoluttverdiar. Å danne basislinjer for dei einskilte respondentane vil og vere svært nyttig når ein skal analysere resultat frå GSR, då ein kan freiste å skilje til dømes ein verkeleg reaksjon frå klamme hender.

Eit anna mogleg forsøk kan vere å syne respondentane nøytrale bilete og bilete frå både undervisningsfilmar (figurar, formalar) og andre undervisningssituasjonar, som til dømes

forelesingssal, eigenstudie og gruppearbeid. Ein kan be respondentane syne ansiktsuttrykk som syner kva kjensler som blir vekkt av bileta og i tillegg gje kvart bilete en score frå 0 (svært negativt) til 100 (svært positivt). Desse resultatane kan nyttast for å sjå om einsskilte personlegdomstypar er meir positive til nokre undervisningsmetodar enn andre. Ein har og høve til å sjå om bilete av forskjellige studiesituasjonar vekker andre kjensler enn nøytrale bilete i tillegg til å dra nytte av (Recio & Sommer, 2018) sine funn om tidsforskjell mellom ubevisst og bevisst respons. Sjølv om ein gjerne vil favne respondentane sine ubevisste kjensler, kan forsøk i e-læringsplattform og gje nyttig lærdom om studentane sine kjensler for forskjellige undervisningsmetodar sidan ein kan nå fleire respondentar. Ein kan og gjenta forsøket fleire gongar gjennom STAT100 for å spore om studentane sine kjensler endrar seg. Særskilt spanande kan det vere å undersøkje om respondentane sine scorar på bilete som syner kollokviearbeid vert betre eller dårlegare etter kvart som respondentane får meir erfaring med kollokviearbeid gjennom kurset. Dette kan vere eit mogeleg supplement til upubliserte analyser frå omlegginga til omvendt klasserom.

Ein anna type meir studentaktivt forsøk som kan gjerast i e-læringsplattform og kan kaste lys over korleis studentar i STAT100 føretrekk å lære, skriv seg frå inspirerende samtalar med respondentane etter forsøka var gjennomført. Fleire av respondentane har uttrykt at filmene utfyller kvarandre. Ein spennande tanke kan vere å spele inn korte filmar som kan sjåast enkeltståande og som tek for seg, i uordna rekkefølge og med ei kort beskriving av innhald til dømes kvifor temaet er viktig, konseptuell teori, matematisk teori, praktisk konseptuelt eksempel og praktisk rekneeksempel. Ved å la respondentane sjølve velje kva dei vil sjå, og i kva rekkefølge kan ein avdekkje ny kunnskap om korleis studentar i STAT100 føretrekk å bli undervist. Ein kan og sjå om forskjellige personlegdomstypar føretrekk same rekkefølge. Forsøket kan gjennomførast fleire gongar gjennom STAT100 for å sjå om føretrekte filmar og rekkefølge endrar seg gjennom kurset.

## **4.2 UTFORDRINGAR SOM FØLGJE AV COVID-19**

Etter å ha diskutert resultat og vidare arbeid kan det arbeidet som skulle vore gjort, men som ikkje var råd å utføre grunna covid-19, trekkjast fram. Det er ikkje til å kome vekk frå at å skrive masteroppgåve under ein verdsomspennande pandemi kan by på heilt unike utfordringar. I dette einsskilte høve har den største utfordringa vore at tilgangen til boimetrialaben på Vitenparken, med ei kraftfull datamaskin og iMoitions programvare, vart

brått avstengd. Planar om utforsking av bruk av forskjellige terskelverdiar til etterprosessering av AFFDEX statistics og grundige undersøkingar av rådata grunnstøytte brått grunna avstengd tilgang til programvare og tilstrekkeleg prosessorkraft. Vidare vert det presentert nokre av dei spørsmåla som i dette høvet må stå ubesvarte, og korleis dei ville vore forsøkt løyst om ikkje covid-19 hadde inntruffe.

Dei ferdigprosesserte resultatfilene frå iMotions kun gjer eit lite glimt av kva den einskilde respondent har gjett oss av kjensler og opplevingar, og det fortsett er ei rekkje ubesvarte spørsmål om kva desse filene faktisk representerar. Desse datasetta er og av eit svært beskjedent omfang, og det er store begrensningar i kva statistiske metodar som kan nyttast. Det er ikkje urimeleg å anta at utforsking av rådata, både i seg sjølve og sett i samanheng med både ferdigprosesserte resultatfiler og personlegdomstypar, kunne gjett grobotn for spanande anvendingar av statistiske metodar og auka innsikta i samhengane mellom personlegdomstypar og opplevinga av forskjellige undervisningsmetodar. Dette var og eit av delmåla for prosjektet, som ein hadde håpa kunne tilføre ny innsikt i forskingsspørsmålet og samstundes letta arbeidet for komande prosjekt.

Det var planlagt å nytte rådata til utforsking av tersklinga AFFDEX statistics si påverking, både som heilheit og for den einskilde respondent. Kjenslene vart registrert på ein skala frå 0 (ikkje til stades) til 100 (fullt til stades), som lett kan reknast om til heiltal mellom 0 og 255 og representerast som eit gråtonebilete. Ved å generere eit undersett som består av den høgaste intensiteten målt i kvar kjensle i kvart sekund i filmane, kan kjenslene representerast ved eit gråtonebilete på 12x600 pixlar. Då er styrken på kjensla ivareteken utan terskling, samstundes som ein har tidfesta når i filmen kjensla inntreff. Dette kan opne for bruk av bilete-analyseteknikkar for å talfeste kjenslene, eller variasjon i kjensler. For terskling kan ein og undersøke og samanlikne forskjellige tersklingsalgoritmer, som kan gje ei terskling tilpassa den einskilde respondent, både for kjenslebiletet som heilheit og for kvar einskild kjensle. Sidan intensiteten i kjenslene og kan sjåast på som z-verdiar i eit koordinatsystem bestående av kjensle (y) og tid i sekund (x) kan desse kjenslebiletta med glatting og interpolering av flater og representerast som kjenslelandskap. Dette kan opne for teknikkar frå både geografiske analyser og form-analyser. Eit døme kan vere same prinsipp som vert nytta i vatnskilte-analyser, at ein tenkjer seg at ein heller vatn i kjenslelandskapa til dei mest tydelege toppane kjem fram, kunne gjett ny innsikt i korleis respondentane opplevde filmane.

Ved utforsking av rådata frå GSR kunne det tenkjast at ein kunne fått svar på om kalibrerte GSR-data er kalibrert etter kvar einskild respondent slik at terskelen på 800 faktisk definerar



ein topp for den einskilde respondenten, eller for ein tenkt gjennomsnittsrespondent. Rådata var og planlagt nytta til ei anna mogleg samanknytning av variablar, ved å nytte rådata frå GSR til vekting av kjenslelandskapa. Då kunne kjensletoppar der GSR var låg for den einskilde respondenten bli flata ut, men kjensletoppar der respondenten synte ein for vedkomande høg GSR blitt meir framtrudande.

Dette er nokre dømer på nye anvendingar av metodar frå andre fagfelt som ein hadde håpa skulle syne seg nyttige. Dette satte covid-19 satte ein brå stoppar for, særskilt når siste respondent gjennomførte sitt forsøk 20. Februar, og NMBU stengde dørene 12.Mars. Hadde ein på forhånd visst kva konsekvensar som ville følgje av covid-19, hadde mykje vore planlagt og gjort annaleis. Ein kan kun vonast at å dele desse døma kan motivere komande arbeid, både til å forsetje arbeidet med å undersøke opplevingar av forskjellige undervisningsmetodar, men og til å freiste å finne inspirasjon frå andre fagfelt.

## 5 OPPSUMMERING

---

Etter at all litteratur er lest og lest oppatt etter at analysane vart utført, alle testar er gjort og alle resultat er tolka og drøfta er tida komen for å sjå heile prosjektet under eitt.

Det er formodentleg sjeldan ein kjem i berøring med prosjekt der alle prosjektet sine svakheiter kan syne seg å vere prosjektet sin største styrke. Å utføre eit nybrottsarbeid frå start til slutt byr på mykje lærdom, ikkje berre om statistiske testar, men og om kor mykje arbeid som må gjerast før ein kan utføre den aller første analysa. Ein må ikkje berre lære å meistre det tekniske utstyret og programvara. Utforsking av eksportmogelegheiter og innhald i resultatfilene eit møysomeleg arbeid, og variablane er mange. Alle desse variablane held eit stort spenn i korleis dei er strukturert, kva datatype dei er og kva fordeling dei stammar frå. Utan å kunne stø seg til nokon som hadde utført liknande forsøk før, var arbeidet med å gjere seg kjend med datasetta ved litteraturgjennomgang og utforsking, krevjande. Det kunne, for å nytte litt uakademiske vendingar, samanliknast med å leggje eit 40 000 brikker stort puslespel der halva av motivet er blå himmel, og den andre halva er gras.

Med dette prosjektet er det ikkje resultatata i seg sjølve som har den største betydninga, men ein kan vonast at kandidatar som skal forsetje arbeidet kan ta mykje lærdom av alt som ikkje har verka etter hensikta og dermed unngå fleire av fallgruvene dette prosjektet har avdekt. Det vonast og at manualar, samt varsam utforsking og grundig beskriving av datasetta kan gje komande arbeid eit betre utgangspunkt og ein lettare start. Eller for å forsetje den noko uakademiske puslespelmetaforen, at komande arbeid kan forsetje å fylle inn brikkar der dette prosjektet har fått på plass eit hjørne og litt av kantane i puslespelet.

Det vetle hjørnet i puslespelet avdekt av dette prosjektet skriv seg ikkje frå resultatata, men frå lærdom om viktigheita av heilheitsvurderingar av resultat. At ein i nybrottsarbeid må freiste å finne mange forskjellige vegar inn under huden på eit datasett, og ha eit mykje større fokus på å kontrollere resultatata enn på resultatata i seg sjølve. Ein må lære å støtte seg til meir enn berre p-verdiar og  $R^2$ . Ein lærer å ikkje la seg skremme frå å nytte framande metodar, og vel så viktig lærer ein å ikkje la seg forlede til å tru at noko som framstår som nytt og avansert ikkje nødvendigvis må vere betre enn det enkle og kjende. Ein lærer mykje om kva som kan vere av betydning, og kva som ikkje treng å vere det. Og om ein skulle trekkje fram ein observasjon som er av betydning, må det vere den store spreinga i respondentane sine kjensler og meiningar.

Studentar er svært, svært forskjellige. Dette er noko som nær sagt ikkje kan formidlast ofte nok til studentane, men som eit faktum, ei trøyst og ei oppmuntring. Absolutt ikkje som ei unnskyldning til å ikkje gjere sin beste innsats, vilkårleg av kva undervisningsmetode ein føretrekk. Om ein skal trekkje fram eit einaste einskild resultat, vil det vere frå den innleiande undersøking av spørjeundersøkinga. Enn så forskjellige som studentar er, så syner undersøkinga at det er ein korrelasjon på 0.9 mellom kva respondentane lærte mest av, og kva dei trudde studentar flest ville lære mest av. Ei mogleg antaking kan vere at ein ynskjer det beste for andre, og difor kan tenkast å ville dele av det som er best for seg sjølve, noko som og vart gjenspegla i Hilde og Kathrine sine undervisningsfilmar. Ein kan vonast at ein, både som student, forelesar og øvingslærer, kan drage lærdom av at enn så forskjellige ein er, så har ein lett for å tru at det som er best for ein sjølv kan vere det beste for andre.

Det er heller ikkje slik at ein stoggar å tileigne seg kunnskapar og ferdigheiter i den augneblunken ein går ut døra på universitetet for aller siste gong. Det er svært sannsynleg at ein i arbeidslivet må klare å tileigne seg lærdom. Det er ikkje sikkert at den ein då skal lære av, er ein dedikert forelesar som vil gjere sitt aller beste for å nå inn til flest mogeleg. Så om ein kan tenkje seg å ikkje berre best mogeleg tilretteleggje for einskilte grupper av studentar, men å fostre kandidatar som har trening i å tileigne seg kunnskap og ferdigheiter på mange forskjellige måtar. Ved å variere undervisningsmetodar og arbeidsformer, kan ein betre legge til rette for at studentar lærer seg å lære på mange forskjellige måtar, og betre auke forståinga for kva ein sjølve føretrekk og kor forskjellige ein er.

Det er i alle høve tydeleg at avdekking av STAT100-studentar sine svært forskjellige opplevingar av undervisningsmetode, både i seg sjølve og saman med personlegdomstypar er ei utømmelig kjelde til viktige prosjekt. Det vonast at dette prosjektet både kan brøyte veg for, og inspirere til komande arbeid. Sjølv om prosjektet har sine feil og manglar, vil det vere å ansjå som vellukka om det kan anspore til arbeid som i ytterste konsekvens bidreg til at ein STAT100-student kan lene seg tilbake i stolen og utbryte: -Åh, var det berre så enkelt??



## Referanseliste

- Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education, 126*, 334-345.  
doi:10.1016/j.compedu.2018.07.021
- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education, 125*, 413-428.  
doi:10.1016/j.compedu.2018.06.023
- Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science, 15*(4), 225-231.
- Andres, J. M. L., Rodrigo, M. M. T., Baker, R. S., Paquette, L., Shute, V. J., & Ventura, M. (2015). *Analyzing Student Action Sequences and Affect While Playing Physics Playground*. Paper presented at the EDM (Workshops).
- Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics, 17*(3), 166-173. doi:10.1002/cem.785
- Boucsein, W. (2012). *Electrodermal activity*: Springer Science & Business Media.
- Brovold, H. (2014). *Invariants drøftet i et nevropsykologisk perspektiv med spesiell referanse til realfaglig kognisjon. 'Fire veier inn i matematikken'*. (PhD). NTNU-trykk, Trondheim.
- Coffield, F., Moseley, D., Hall, E., Ecclestone, K., Coffield, F., Moseley, D., . . . Ecclestone, K. (2004). Learning styles and pedagogy in post-16 learning: A systematic and critical review.
- Dandy, K. L., & Bendersky, K. (2014). Student and faculty beliefs about learning in higher education: implications for teaching. *International Journal of Teaching and Learning in Higher Education, 26*(3), 358-380.
- Danner, L., Sidorkina, L., Joechl, M., & Duerrschmid, K. (2014). Make a face! Implicit and explicit measurement of facial expressions elicited by orange juices using face reading technology. *Food Quality and Preference, 32*, 167-172.
- Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology, 3*, 429.
- Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models*: CRC press.
- Druggeri, K., Dempster, M., Hanna, D., & Cleary, C. (2008). Experiences and Expectations: The Real Reason Nobody Likes Stats. *Psychology Teaching Review, 14*(2), 75-83.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*: Springer Science & Business Media.
- Hayes, N. (2000). *Foundations of psychology*: Cengage Learning EMEA.
- Hogg, R. V., McKean, J., & Craig, A. T. (2005). *Introduction to mathematical statistics*: Pearson Education.
- Horsley, M., Eliot, M., Knight, B. A., & Reilly, R. (2013). *Current trends in eye tracking research*: Springer.
- Huey, E. B. (1908). *The psychology and pedagogy of reading*: The Macmillan Company.
- imotions. The iMotions Platform. Retrieved from <https://imotions.com/platform/>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review, 87*(4), 329.
- Kenney, T., Huang, T., & Gu, H. (2019). Poisson PCA: Poisson Measurement Error corrected PCA, with Application to Microbiome Data. *arXiv preprint arXiv:1904.11745*.

- Kloke, J. D., & McKean, J. W. (2012). Rfit: Rank-based estimation for linear models. *The R Journal*, 4(2), 57-64.
- Kvam, M. A. (2020). Greie og ugreie fag ved NMBU. *Tuntreet*, 2020(14. April), 30-31. Retrieved from <https://issuu.com/tuntreet/docs/tt01>
- Lei, J., Sala, J., & Jasra, S. K. (2017). Identifying correlation between facial expression and heart rate and skin conductance with iMotions biometric platform. *Journal of Emerging Forensic Sciences Research*, 2(2), 53-83.
- Leitch, K., Duncan, S., O'keefe, S., Rudd, R., & Gallagher, D. (2015). Characterizing consumer emotional response to sweeteners using an emotion terminology questionnaire and facial expression analysis. *Food Research International*, 76, 283-292.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Magdin, M., Benko, L., & Koprda, Š. (2019). A case study of facial emotion classification using affdex. *Sensors*, 19(9), 2140.
- Mardia, K. V., Kent, J. T., & Bibby, J. (1979). *Multivariate analysis*. London: Academic Press.
- Matté, J. A. (2000). *Examination and cross-examination of experts in forensic psychophysiology using the polygraph*: JAM Publications.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153-157.
- Mendez, K. M., Reinke, S. N., & Broadhurst, D. I. (2019). A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. *Metabolomics*, 15(12), 150.
- Myers, I. B. (1980). *Gifts Differing: Understanding Personality Type*.
- Nasjonalt senter for realfagsrekruttering. (2017a). Om utdanningstesten. Retrieved from <https://velgriktig.no/for-radgivere/om-utdanningstesten-laerveiledning/>
- Nasjonalt senter for realfagsrekruttering. (2017b). Utdanningstesten. Retrieved from <https://velgriktig.no/utdanningstesten/>
- Newton, P. M., & Miah, M. (2017). Evidence-Based Higher Education – Is the Learning Styles ‘Myth’ Important? *Frontiers in Psychology*, 8(444). doi:10.3389/fpsyg.2017.00444
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning Styles: Concepts and Evidence.(Report). *Psychological Science in the Public Interest*, 9(3), 105. doi:10.1111/j.1539-6053.2009.01038.x
- Pham, T. V., & Jimenez, C. R. (2012). An accurate paired sample test for count data. *Bioinformatics*, 28(18), i596-i602. doi:10.1093/bioinformatics/bts394
- Recio, G., & Sommer, W. (2018). Copycat of dynamic facial expressions: Superior volitional motor control for expressions of disgust. *Neuropsychologia*, 119, 512-523. doi:10.1016/j.neuropsychologia.2018.08.027
- Rencher, A. C. (1995). *Methods of Multivariate Analysis*. New York: JOHN WILEY & SONS, INC. .
- Rohrer, D., & Pashler, H. (2012). Learning styles: where’s the evidence? *Medical Education*, 46(7), 634-635. doi:10.1111/j.1365-2923.2012.04273.x
- Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, 50(4), 1446-1460. doi:10.3758/s13428-017-0996-1
- Sæbø, S., Almøy, T., & Brovold, H. (2015). Does academia disfavor contextual and extraverted students. *Uniped*, 38(4).

- Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and instruction, 54*, 93-103.
- Taub, M., Azevedo, R., & Mudrick, N. V. (2018). How Do Different Levels of AU4 Impact Metacognitive Monitoring During Learning with Intelligent Tutoring Systems? In *Intelligent Tutoring Systems* (pp. 223-232).
- Wade, N. J. (2010). Pioneers of Eye Movement Research. *i-Perception, 1*(2), 33-68.  
doi:10.1068/i0389
- Was, C., Sansosti, F., & Morris, B. (2016). *Eye-tracking technology applications in educational research*: IGI Global.
- Yeh, Y.-c., Lai, S. C., & Lin, C.-W. (2016). The dynamic influence of emotions on game-based creativity: An integrated analysis of emotional valence, activation strength, and regulation focus. *Computers in Human Behavior, 55*, 817-825.
- Zhi, R., Wan, J., Zhang, D., & Li, W. (2018). Correlation between hedonic liking and facial expression measurement using dynamic affective response representation. *Food Research International, 108*, 237-245.

# VEDLEGG

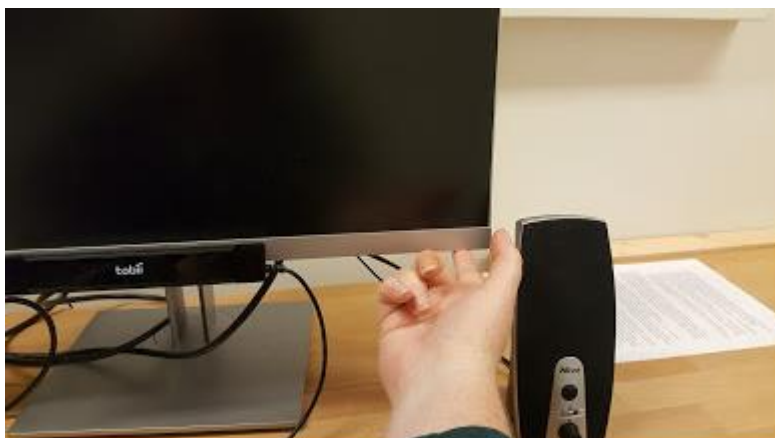
---

## VEDLEGG 1. MANUAL FOR OPPSTART AV BIOMETRILAB OG FORSØK



Du vil typisk finne 1) høytalerledningane fråkobla, 2) shimmeren avskrudd og 3) eye-tracker frakoblet. Begge skjermene og datamaskinen vil være avskrudd. Dette er utgangspunktet for oppstarten.

1) Skru på den lille skjermen. Knappen er på undersida av skjermen, mot høyre. Det vil kome på ett lite blått lys.



2) Ta ut den lille USB-stikken som tilhører shimmeren. Og pass for all del godt på den!





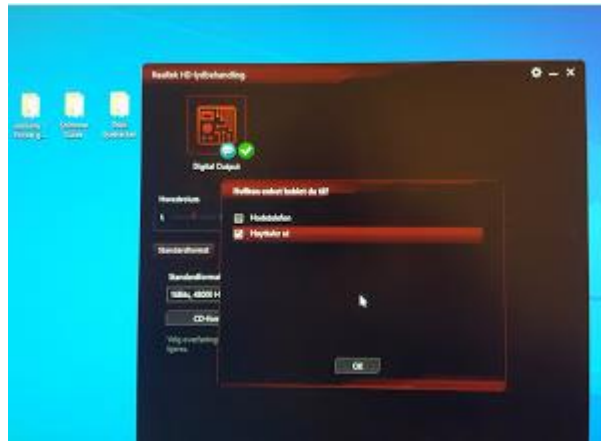
3) Skru på selve datamaskinen.



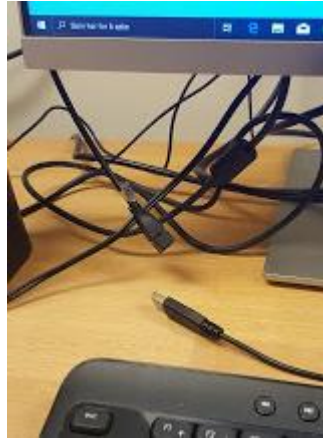
4) Du blir møtt av ett fint landskapsbilde. Du må taste inn pin-koda for å kome inn på PC-en.



5) Sett i kablene til høgtaleren, pass på at du huker av for «høytaler ut» i vindauget som kjem opp på skjermen.



6) Koble på blikksporaren. Det vil ikke kome lys i denne før du skrur på imotions, så ingen grunn til panikk. Ennå.



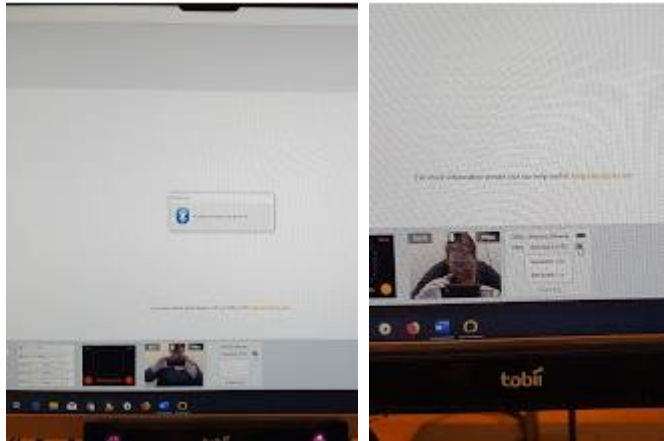
7) Start opp imotions 8. Det vil no etterkvart kome lys i blikksporaren



8) Sett i USB-stikken til shimmeren, og skru på denne.

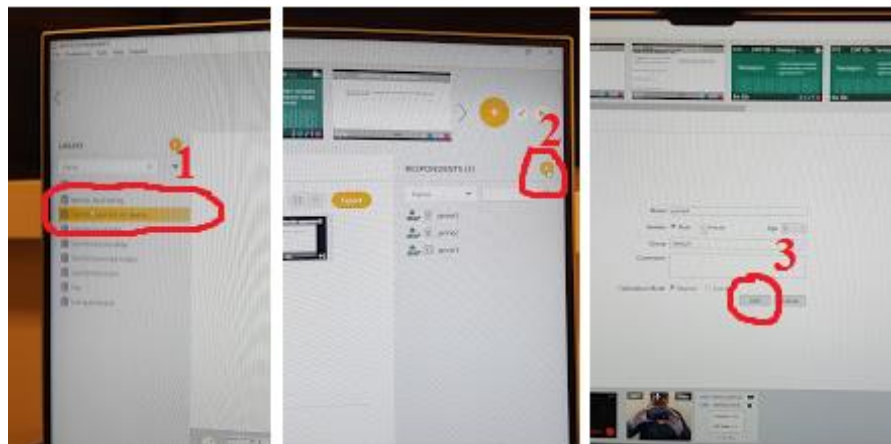


9) Vent på at dialogboks for Blue Tooth (BT) kommer opp. Det kan ta ett lite minutt. Du trykker så på den lille play-pila i statusruta til shimmeren. Det kan hende du må trykke på denne for å få BT-dialogboksen, det er litt tilfeldig fra gong til gong.

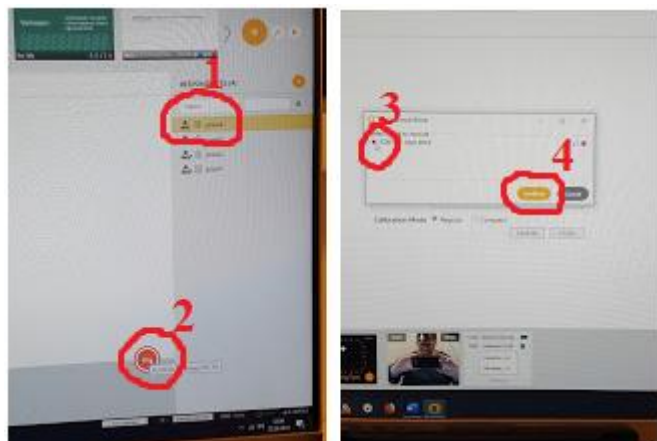


10) 1) velg riktig forsøk 2) legg til respondent 3) fyll inn info og trykk «add»

Husk å koble shimmeren på respondenten!



11) 1) velg en respondent 2) trykk «record» 3) huk av for forsøket 4) trykk «continue».



Etter dette vil kalibreringa starte automatisk. Denne beregnast, og etter hvert får ein opp kvalitet. Den er oftest excellent eller good. Om man er fornøyd, trykk en «continue», og forsøket er i gang.

Blikksporaren har en tendens til å dovne av om den blir ståande uvirksom. Det anbefalast derfor å skru av og på imotions rett før neste respondentens.

## **VEDLEGG 2. PROTOKOLL FOR UTFØRING AV FORSØK**

Forsøket foregår på biometrilaben i Vitenparken. Rommet må vere i fritt for bileter, oppslag og anna som kan trekkje merksemda vekk frå det som foregår på skjermen. Rommet må ha god belysning, dog vere skjerma for naturleg lys med den hensikt å gje like forhold under forsøket. Rommet må og ha ei beliggenheit som skjermar mot støy, trafikk i gongar og liknande. Døra inn til rommet må merkast med eit skilt med teksta «forsøk pågår».

### **I forkant av forsøket.**

Studentane har på forhånd blitt rekruttert til forsøket, mellom anna i forelesingar. Dei har og har fått noko informasjon om hensikta med forsøket, og om korleis forsøket foregår.

### **Mottaking av studenten.**

Vi må freiste å gje alle studentar om lag same mottaking. Studentane må og kjenne seg godt ivaretekne, utan å gå for mykje på bekostning av tida vi har til rådighet i ei hektisk datafangstperiode. Ei kort helsing, ein takk for at dei stiller opp, og kanskje eit par spørjesmål om kvar dei kjem frå og kva dei studerar burde vere høveleg.

Respondenten blir vist inn i forsøksrommet, og blir bedt om å sete seg i den innerste stolen. Der ligg samtykkeskjema og penn klart. Respondentenles gjennom denne i ro og mak medan forsøket blir satt opp, og får dermed høve til å bli tilvent lysforhold og temperatur. Så blir respondenten flytta til stolen framfor den minste skjermen, der syner vi studentane dei forskjellige sensorane. Forklaringane til dei forskjellige sensorane må haldast korte og enkle. Det er kjekt disom dei let seg engasjere av teknologien, men vi må freiste å halde fokus på forsøket.

### **Tilkobling av sensorar og kalibrering.**

Først finn ein ein god posisjon av stolen. Den må stå midt framfor blikksporaren, og respondenten må sete behageleg. Best avstand er mellom 55cm og 65cm. Best resultat får ein om respondenten set om lag 62-63cm frå sensoren, då dei gjerne sig i hop og lenar seg framover når forsøket startar.

Vidare skal den galvaniske hudleiingssensoren setjast på. Respondentane er gjerne litt spente, så ein liten fleip om «løgndetektorutstyret» får dei til å slappe av. Då kan det skje at dei synk ein del saman i stolen, og denne kan etterjusterast.

Når respondenten har kome på plass i stolen med alle sensorane er klare, understreker vi igjen at dette er frivillig, og at dei kva tid som helst kan forlate forsøket. Vi forsikrar respondenten om at vi er like ute i gangen, og at dei berre kan komme ut om noko skulle skje. Deretter gjennomførast kalibrering. Når denne er gjort spør vi igjen om respondenten har spørsmål. Etter eventuelle spørsmål er besvart startar forsøket. Vi må også passe på at ingen forstyrrer forsøkspersonen undervegs.

### **Etter forsøket.**

Når forsøket er ferdig, blir respondenten spurd det gjekk vel, og om det hende noko som kan ha betydning for forsøket som svikt i utstyr, PC-trøbbel eller liknande. Respondenten blir og spurt om kva film som vart synt først.

Det er viktig å bruke litt tid på respondentane etter forsøket. Spørre korleis det gjekk og korleis det var å vere med på dette. Eventuelle innspel frå respondenten om forsøket noterast til seinare høve.

Respondentane blir oppfordra til å ikkje diskutere filmar og spørsmål med andre, for å unngå systematiske trendar som følge av forutinntattheit.

Nokre studentar er interessert i å få sjå på sine egne resultat frå forsøket. Desse blir bedd om å avtale tid etter datafangstperioden er over, og vil bli synt korleis deira data ser ut.

Vi kan kanskje oppleve at dei har lyst å spørre om faglege ting dei lurar på i stat100, og vi må passe oss for å bli øvingslærarar.

### VEDLEGG 3. TABELLAR OG FIGURAR FRÅ UNDERSØKING AV SPØRJEUNDERSØKINGA

Tabell 3-A Tabell som syner spørsmåla nytta i spørjeundersøkinga med tilhøyrande nøkkeltal.

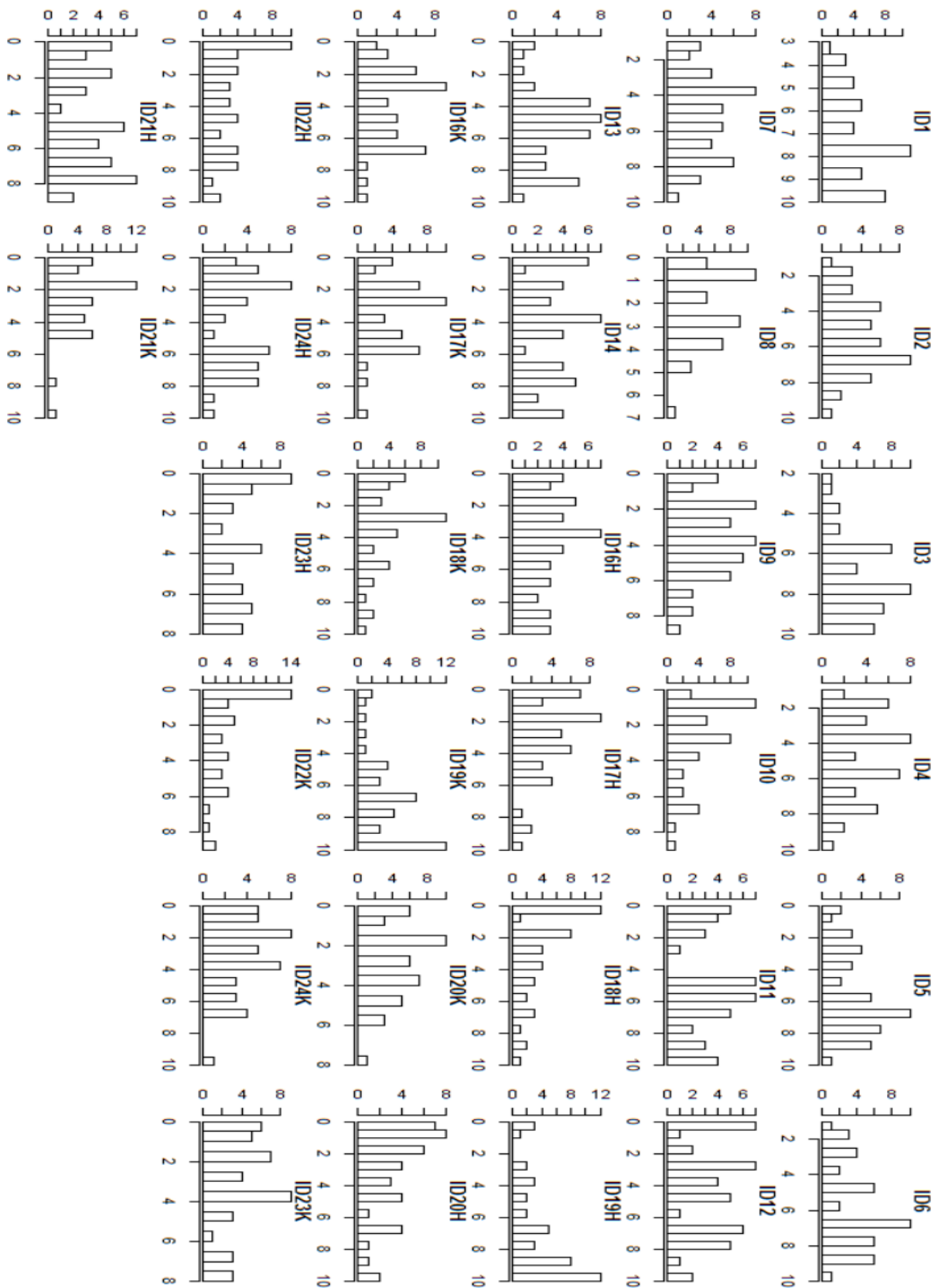
Spørsmål	Min	Max	Median	Middel	Standard avvik
ID1 Statistikk vil vere nyttig i fremtidige studier	3	10	8	7,5	2
ID2 Statistikk vil vere nyttig i dagliglivet	1	10	6	5,7	2,2
ID3 Statistikk vil vere nyttig i arbeidslivet	2	10	8	7,4	2
ID4 Utrekningene går bra, men det er vanskelig å forstå konseptene.	1	10	5	5	2,4
ID5 Konseptene er lette å forstå, men utregningene er vanskelige.	0	10	7	5,8	2,7
ID6 Både konseptene og utregningene er enkle.	1	10	7	6,1	2,4
ID7 Både konseptene og utregningene er vanskelige.	1	10	5	5,3	2,4
ID8 Jobber mye med faget.	0	7	2	2,4	1,7
ID9 Utbytte av Kathrine sin film	0	9	4	3,9	2,3
ID10 Utbytte av Hilde sin film.	0	9	3	3,1	2,4
ID11 Hvilken av videoene følte du du lærte mest av. *	0	10	6	5,1	3,2
ID12 Hvilken av videoene tror du studenter flest får best utbytte av i stat100. *	0	10	4	4,5	3
ID13 Hvilken av videoene var lengst. *	0	10	5	5,6	2,5
ID14 Jeg hadde lest om/hørt om temaet før jeg kom hit	0	10	4	4,8	3,2
ID16H Filmen engasjerte meg -Hilde	0	10	4	4,5	3
ID16K Filmen engasjerte meg -Kathrine	0	10	4	4,2	2,5
ID17H Jeg forsto det som ble presentert -Hilde	0	10	3	3,3	2,6
ID17K Jeg forsto det som ble presentert -Kathrine	0	10	3	3,7	2,3
ID18H Det var greit å følge med på den tekniske notasjonen -Hilde	0	10	2	3,1	2,9

ID18K Det var greit å følge med på den tekniske notasjonen -Kathrine	0	10	3	3,6	2,7
ID19H Filmen var frustrerende -Hilde	0	10	8	7,1	3,1
ID19K Filmen var frustrerende -Kathrine	0	10	7	7,1	2,9
ID20H Jeg likte oppbygningen av filmen -Hilde	0	10	2	3,3	3
ID20K Jeg likte oppbygningen av filmen -Kathrine	0	8	3	2,9	2
ID21H Eksempelet var interessant -Hilde	0	9	5	4,6	2,9
ID21K Eksempelet var interessant -Kathrine	0	10	2	2,8	2,1
ID22H Jeg forsto hvorfor variasjonen til gjennomsnittet var mindre enn variasjonen til enkeltverdiene -Hilde	0	10	3	3,8	3,2
ID22K Jeg forsto hvorfor variasjonen til gjennomsnittet var mindre enn variasjonen til enkeltverdiene -Kathrine	0	9	2	2,7	2,8
ID23H Temaet følte relevant for studiene mine -Hilde	0	8	4	3,6	2,8
ID23K Temaet følte relevant for studiene mine -Kathrine	0	8	3	3,2	2,4
ID24H Filmen fikk meg til å forstå den praktiske relevansen til dette temaet -Hilde	0	10	4	4,3	2,9
ID24K Filmen fikk meg til å forstå den praktiske relevansen til dette temaet -Kathrine	0	10	3	3,3	2,4

\*Svara blir gjedd på ein skala frå Hilde (0) til Kathrine (10)

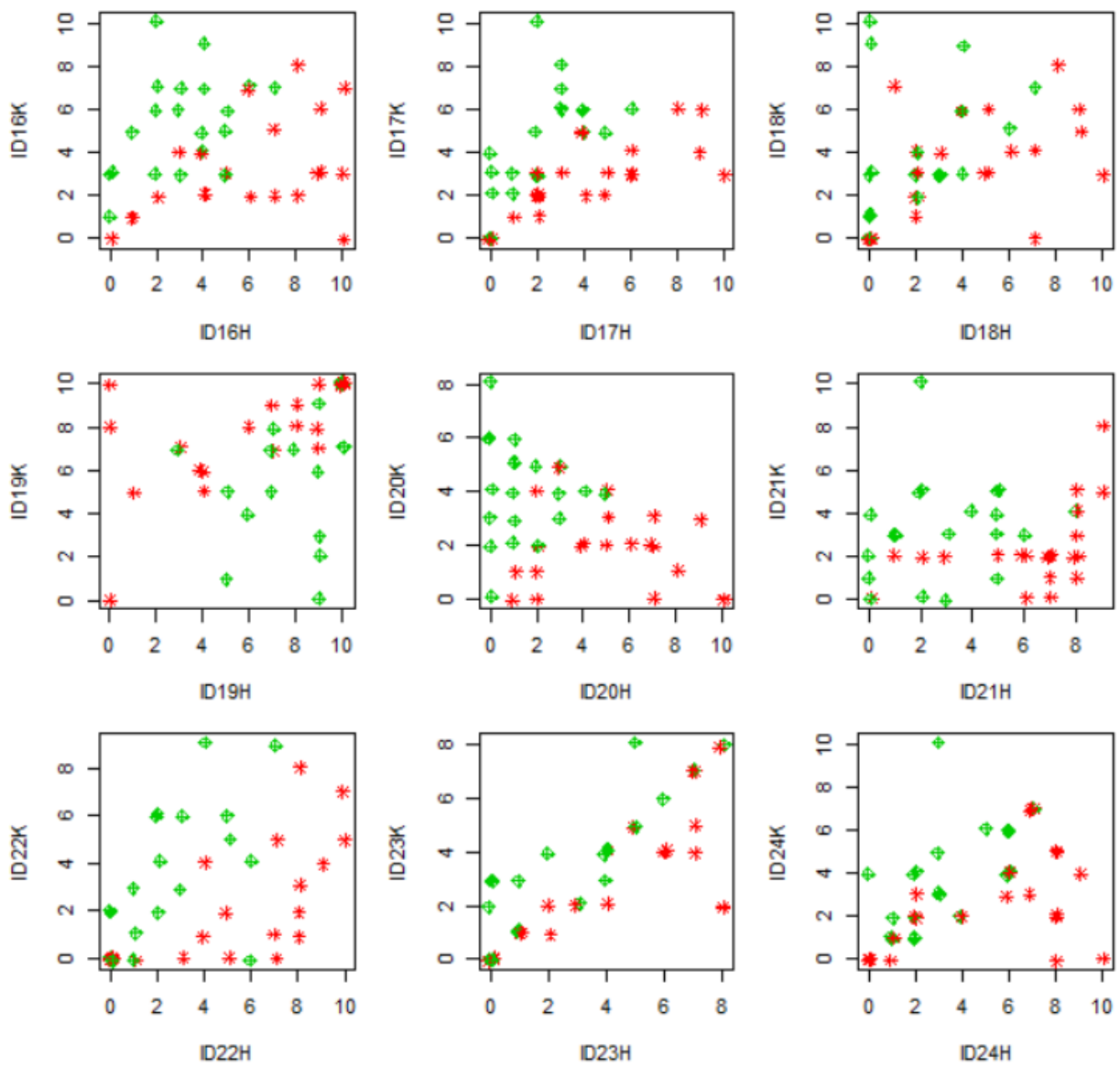
Tabell 3-B Tabell som syner fordelinga av respondentar etter det eine binære spørsmålet, ID15.

Spørsmål	Hilde	Kathrine
ID15 Om du bare fikk se en av filmene i stat100, hvilken hadde du valgt	20	21



Figur 3-A Frekvensdiagram som syner antal respondentar etter avgjett svar på spørjeundersøkinga.





Figur 3-B Spredningsdiagram som syner respondentane representert ved dei parvise spørsmåla i spørjeundersøkinga. Respondentane er fargelagt etter ID15, kva film dei ville valt om dei kun fekk sjå ein film i STAT100. Grøne respondentar ville valt Hilde sin film, Raude respondentar ville valt Hilde sin film.

## VEDLEGG 4. TABELLAR OG FIGURAR FRÅ UNDERSØKING AV AFFDEX STATISTICS

Tabell 4-A Resultat frå Hilde sin film. For kvar respondent er det gjeitt kor stor prosent av tida filmen varte kvar variabel oversteig terskelen. Kvitt bakgrunn markerar observasjonar lågare enn første tertil, lys grå bakgrunn markerar observasjonar mellom første og andre tertil, mørk grå bakgrunn markerar observasjonar høgare enn andre tertil.

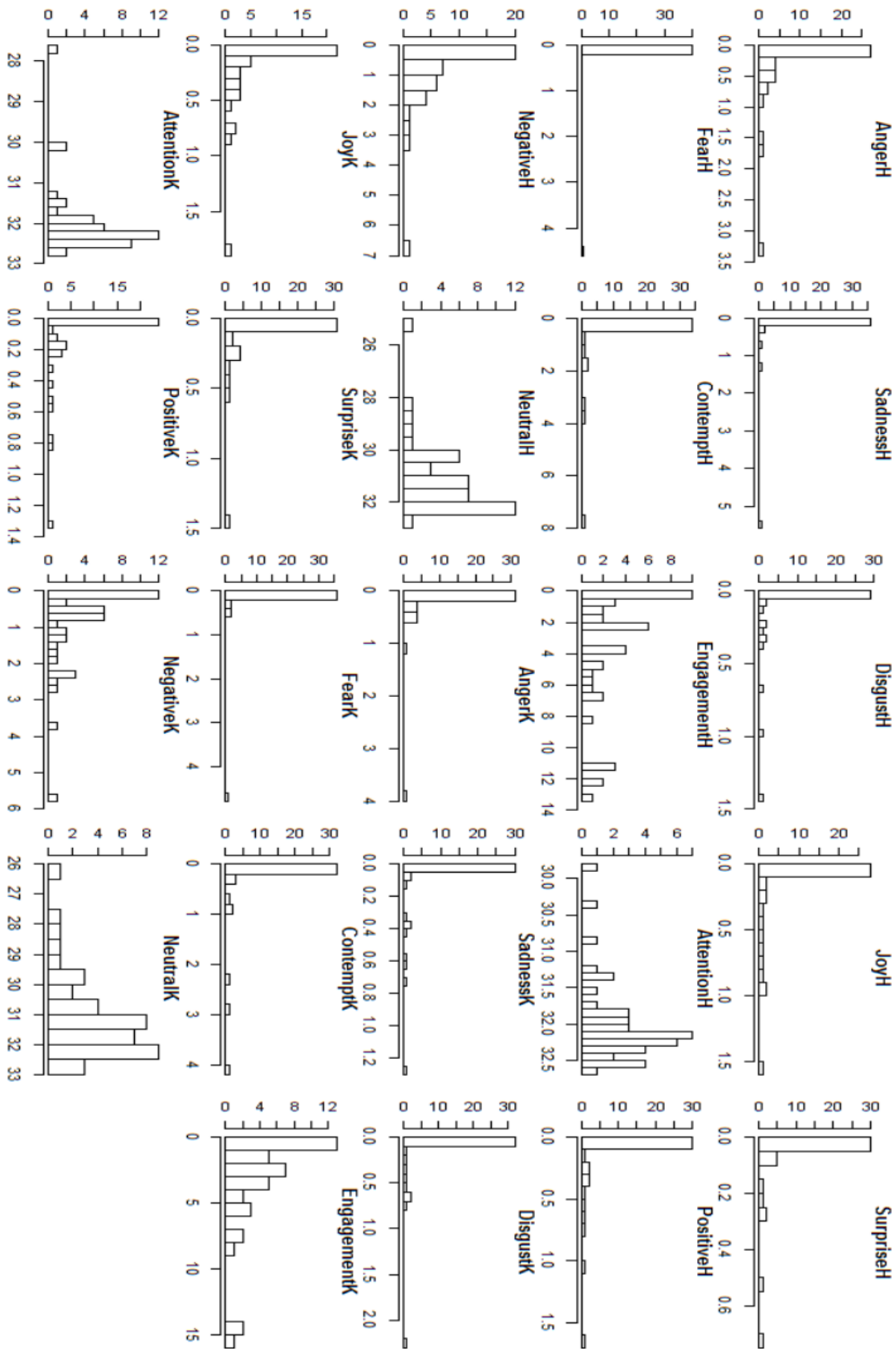
Respo ndent	anger	sad ness	dis gust	joy	surp rise	fear	cont empt	engag ement	atten tion	posi tive	nega tive	neut ral
1	0,15	0,06	0,03	0	0,18	0,1	0,01	0,65	32,02	0	2,19	30,05
2	0,01	0	0,25	0,01	0,09	0	0	3,58	30,82	0	0,68	30,38
3	0,59	0,36	0,02	0	0	0	0,04	11,03	32,4	0	1,26	31,15
4	0	0	0,02	0,79	0	0	0,03	1,53	32,26	0,72	0,51	31,11
5	0	0	0	0,21	0	0	0	0,24	32,24	0,2	0	32,04
6	0	0,02	0	0	0	0	0	0	32,47	0	0	32,47
7	1,59	5,45	0	0	0	0	0,03	11,3	32,16	0	6,86	25,31
8	0	0	0	0,05	0	0,01	0	0,11	32,54	0,05	0	32,5
9	0,15	0	0	0	0,05	0	0	0,84	32,34	0	0	32,39
10	0,48	0,05	0,98	1,54	0,29	0,01	0,23	6,62	31,4	1,65	1,23	28,8
11	0	0,01	0,01	0,82	0	0,01	0	2,29	31,99	0,38	0,62	31,2
12	0,83	0,08	0,1	0,33	0,23	0	0,02	5,26	32,55	0,21	1,95	30,42
13	0,01	0,19	0,27	0,6	0	0	3,45	3,85	32,36	0,42	1,07	30,94
14	0,8	0,07	0	0	0	0,14	0,6	1,86	32,13	0	1,21	31,14
15	0	0	0	0	0	0	0	0	32,61	0	0	32,61
16	0	0	0,01	0,93	0	0	0,08	2,19	32,55	0,67	0	31,88
17	0	0	0,01	0,2	0	0	0,03	0,51	32,6	0,1	0,27	32,23
18	0	0	0	0	0	0	0	0,1	32,26	0	0	32,26
19	0	0	0	0	0	0	0	0	32,28	0	0	32,28
20	0,01	0	0	0,25	0,02	0	0	5,95	32,2	0,25	0,47	31,49
21	1,8	0,68	0	0,1	0,02	0	7,58	13,39	32,43	0,06	1,9	30,64
22	0,6	0,09	0,15	0,12	0,54	0	0	3,83	32,25	0,08	0,29	32,06
23	0	0	0	0	0	0	0,26	0,5	32,16	0	0,18	32,09
24	0	0	0,02	0	0	0	0,08	0,32	31,89	0	0,61	31,48
25	0,18	1,24	0,02	0,1	0,1	0,01	0,12	12,37	32,02	0,1	0,21	31,75

26	0,1	0,03	0,37	0,97	0,05	0	0,25	4,93	31,9	1,05	1,1	29,78
27	0,25	0,07	0,32	0,05	0,06	0	0,05	2,12	31,6	0,06	0,39	31,76
28	0,17	0,02	0,02	0,47	0,03	0	0,01	6,22	32,23	0,33	1,4	30,5
29	0	0	0	0	0	0	0	0	32,36	0	0	32,36
30	0,1	0,19	0,33	0,06	0,07	0	0,18	3,65	30,38	0	0,99	31,1
31	0,27	0	0,66	0,01	0	0	0,05	1,2	32,12	0	0,56	31,56
32	3,37	0	0	0	0,26	4,49	0,2	12,33	32,1	0	1,86	30,41
33	0	0	0	0,62	0,08	0	1,07	2,07	31,3	0,57	0,62	30,15
34	0,02	0	0,06	0	0	0	0,06	1,17	31,76	0	0,29	31,89
35	0,52	0	0,01	0	0	0,11	0	4,74	31,86	0	0	31,88
36	0	0	0,04	0	0	0,02	3,9	2,17	31,98	0	1,65	30,62
37	0	0	0	0	0	0	0	2,42	32,14	0	0	32,14
38	0,29	0,39	0	0,03	0	0	0	11,04	31,97	0	0,1	31,9
39	0,67	0	1,41	0,02	0	0	1,52	6,65	31,38	0,03	2,65	29,35
40	0	0	0	0	0	0	0	0,21	32,19	0	0	32,19
41	0,28	0	0,21	0	0,75	0,06	1,56	8,23	29,88	0	3,38	28,34

Tabell 4-B Resultat frå Kathrine sin film. For kvar respondent er det gjett kor stor prosent av tida filmen varte kvar variabel oversteig terskelen. Kvitt bakgrunn markerar observasjonar lågare enn første tertil, lys grå bakgrunn markerar observasjonar mellom første og andre tertil, mørk grå bakgrunn markerar observasjonar høgare enn andre tertil.

respo ndent	anger	sad ness	dis gust	joy	surp rise	fear	cont empt	engag ement	atten tion	posi tive	nega tive	neut ral
1	0	0,39	0	0,58	0,14	0,2	0,02	1,99	32,03	0,51	2,25	29,48
2	0,07	0	0,64	0,35	0,02	0	0,06	2,72	32,31	0,16	0,43	31,74
3	0	0,36	0,09	0	0	0	0	3,44	32,48	0	0,45	32,18
4	0	0	0,03	0,2	0	0,45	0	0,73	32,41	0	0,6	31,83
5	0	0	0	0,79	0	0	0	0,83	32,2	0,82	0	31,38
6	0	0	0	0	0	0	0	0,07	32,41	0	0,17	32,24
7	0	0,72	0	0	0	0	0	1,94	32,52	0	3,67	28,85
8	0	0	0	0,02	0	0	0,16	0,08	32,35	0	0	32,35
9	0	0	0	0	0	0	0,02	0,07	32,53	0	0,01	32,59
10	0,12	0	0,75	0,74	1,49	0	0,27	8,88	31,47	0,56	1,4	29,85

11	0,13	0	0,05	0,5	0,01	0,56	0,06	1,43	32,06	0,03	0,65	31,52
12	0,49	0,57	0	0	0,09	0	0,06	3,93	32,48	0	0,81	31,68
13	0,02	0,07	0,47	0,43	0	0	2,99	2,5	32,31	0,41	1,65	30,41
14	0,24	0,02	0	0	0	0,32	0,24	2,86	32,26	0	1,09	31,26
15	0	0	0	0	0	0,05	0	0,47	32,64	0	0	32,64
16	0	0	0	0,18	0	0	0,03	0,7	32,59	0,23	0,12	32,24
17	0	0	0	0	0	0	0,06	0,24	32,77	0,01	0	32,76
18	0	0	0	0	0	0	0	0,1	32,33	0	0	32,33
19	0	0	0	0,42	0	0	0	0,6	32,4	0,14	0	32,26
20	0,41	0	0	0,06	0,21	0	0,1	3,58	32,2	0	0,79	31,41
21	1,2	0,42	0	0,33	0	0	0	14,82	32,6	0,24	0,58	31,78
22	0,18	0,11	0,17	0,38	0,25	0	0,09	4,27	32,27	0,17	0,61	31,54
23	0,09	0	0,02	0,3	0,02	0	0,09	2,91	32,29	0,17	1,48	30,63
24	0	0	0	0	0	0	0,03	0,24	32,44	0	0,1	32,34
25	0,34	1,29	0,23	0,17	0,01	0	0,1	15,53	31,48	0,21	2,28	29,55
26	0,2	0,03	0,05	0,29	0,02	0,01	0,93	5,01	31,82	0,31	1,03	30,54
27	0,43	0,1	0,61	0,02	0,08	0	0,08	2,88	31,81	0	0,78	31,28
28	0	0	0	0,16	0,03	0	0,08	2,61	32,24	0,2	0,76	31,37
29	0	0,31	0	0	0,06	0,01	0	0,69	32,09	0,02	0	32,19
30	0,06	0	0,39	0,07	0,38	0	0,22	7,41	30,15	0,06	1,29	30,86
31	0,4	0	0,52	0	0	0	0	1,31	31,93	0	0,49	31,44
32	3,94	0,04	0,03	0	0,57	4,67	0	14,28	32,25	0	2,25	30,02
33	0	0	0	1,89	0,21	0	0,79	5,3	31,36	1,32	1,95	28,48
34	0	0	0	0,82	0,27	0	0	3,11	32,28	0,78	0,3	31,38
35	0,01	0,02	0,01	0,13	0,05	0,06	0,01	1,02	32,11	0,14	0,21	31,78
36	0	0	0,05	0,22	0,03	0,24	4,02	5,25	30,11	0	2,74	27,51
37	0	0,02	0	0	0,16	0	0,02	2,06	31,89	0	0,52	31,41
38	0,34	0,61	0	0	0	0	0	4,48	31,74	0	0,78	30,96
39	0,53	0	2,26	0,01	0	0	0,98	7,56	27,62	0,02	2,49	29,82
40	0	0	0	0	0	0	0	0,05	32,35	0	0	32,35
41	0	0	0	0	0,46	0,01	2,37	3,92	31,84	0	5,73	26,47



Figur 4-A Frekvensdiagram frå variablane i AFFDEX statistics.

## VEDLEGG 5. OVERSIKT OVER VARIABELNAMN MED SUFFIKS OG PREFIKS

Responstype	Namn	Kort beskriving
Spørjeundersøking	PCA_ES	Prinsipalkomponentanalyse av enkeltspørsmåla. Prinsipalkomponentane vart benevnt som PC1_ES, PC2_ES og vidare.
	PCA_PS	Prinsipalkomponentanalyse av dei parvise spørsmåla. Prinsipalkomponentane vart benevnt som PC1_PS, PC2_PS og vidare.
	PCA_SU	Prinsipalkomponentanalyse av den fulle spørjeundersøkinga unnateke ID11 og ID15. Prinsipalkomponentane vart benevnt som PC1_SU, PC2_SU og vidare.
AFFDEX statistics	kjensleH	Kjenslene frå AFFDEX statistics som kjem frå Hilde sin film får suffiks H, som angerH, joyH og vidare.
	kjensleK	Kjenslene frå AFFDEX statistics som kjem frå Kathrine sin film får suffiks K, som angerK, joyK og vidare.
	diffkjensle	Skilnad mellom kjensler frå dei to filmene. Framkjem ved å trekkje verdiane for kjensla i Kathrine sin film frå verdiane for kjensla i Hilde sin film, til dømes angerH – angerK. Kjensledifferandane vert benevnt som diffanger, diffjoy og vidare.
	PCA_H	Prinsipalkomponentanalyse av kjenslene unnateke attention og neutral for Hilde sin film.

		Prinsipalkomponentane vart benevnt som PC1_H, PC2_H og vidare.
	pPCA_H	Poissonkorrigert prinsipalkomponentanalyse av kjenslene unnateke attention og neutral for Hilde sin film. Prinsipalkomponentane vart benevnt somp PC1_H, pPC2_H og vidare.
	PCA_K	Prinsipalkomponentanalyse av kjenslene unnateke attention og neutral for Kathrine sin film. Prinsipalkomponentane vart benevnt som PC1_K, PC2_K og vidare.
	pPCA_K	Poissonkorrigert prinsipalkomponentanalyse av kjenslene unnateke attention og neutral for Kathrine sin film. Prinsipalkomponentane vart benevnt som pPC1_K, pPC2_K og vidare.
	PCA_AS	Prinsipalkomponentanalyse av kjenslene unnateke attention og neutral for begge filmene. Prinsipalkomponentane vart benevnt som PC1_AS, PC2_AS og vidare.
	pPCA_AS	Poissonkorrigert prinsipalkomponentanalyse av kjenslene unnateke attention og neutral for begge filmene. Prinsipalkomponentane vart benevnt som pPC1_AS, pPC2_AS og vidare.
GSR summary scores	TPM_H	Toppar per minutt for Hilde sin film.
	TPM_K	Toppar per minutt for Kathrine sin film.

	diffTPM	Skilnad mellom TPM frå dei to filmene. Framkjem ved å trekkje verdiane for TPM i Kathrine sin film frå verdiane for TPM i Hilde sin film, $TPM_H - TPM_K$ .
GSR binned peak detection	SBP_H	Sum antal intervall med minst ein topp for Hilde sin film.
	SBP_K	Sum antal intervall med minst ein topp for Kathrine sin film.
	diffSBP	Skilnad mellom SBP frå dei to filmene. Framkjem ved å trekkje verdiane for SBP i Kathrine sin film frå verdiane for SBP i Hilde sin film, $SBP_H - SBP_K$ .





**Norges miljø- og biovitenskapelige universitet**  
Noregs miljø- og biovitenskapelige universitet  
Norwegian University of Life Sciences

Postboks 5003  
NO-1432 Ås  
Norway