



# Revisors profesjonelle skepsis ved vurdering av revisjonsbevis innehentet ved dataanalyser

*En eksperimentell studie*

**Fredrik Solbakk Andersen og Maren Louise Mørk**

**Veileder: Jonas Gaudernack**

Masteroppgave i regnskap og revisjon

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i regnskap og revisjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer innestår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

## Sammendrag

Revisjonsprofesjonen står overfor en stor teknologisk omveltning, der økt anvendelse av digitale verktøy står sentralt. Dette vil medføre en rekke fordeler, der økt treffsikkerhet og effektivitet i revisjonen er blant de viktigste (Kleive, 2018). Det teknologiske skiftet vil føre til en endring i fremgangsmåten for bevisinnhenting (KPMG, 2015; Davenport, 2016; BDO, 2016; EY, 2015). Således er det viktig å vite hvordan revisors vurderinger påvirkes av de nye metodene. Vår studie bidrar til dette forskningsområdet, og etterstreber å belyse følgende forskningsspørsmål: «Hvordan påvirker kunnskap om dataanalyser og oppgavekompleksitet revisors profesjonelle skepsis ved vurdering av revisjonsbevis innhentet ved dataanalyser?».

Tidligere forskning viser at kunnskap har betydning for profesjonell skepsis, men forskere finner imidlertid effekter i ulike retninger. Noen studier konkluderer med at kunnskap øker profesjonell skepsis, mens andre finner at kunnskap gir redusert profesjonell skepsis. Basert på dette har vi formulert en retningsnøytral hypotese, der vi forventer at økt kunnskapsnivå gir enten redusert eller økt profesjonell skepsis. Hva gjelder oppgavekompleksitet forventer vi at profesjonell skepsis reduseres når kompleksiteten øker. I tillegg forventer vi at det vil eksistere en interaksjonseffekt mellom kunnskap og kompleksitet. Denne avhenger av retningen av kunnskapseffekten, og vil resultere i enten en forsterket reduksjon av profesjonell skepsis (når økt kunnskap reduserer profesjonell skepsis) eller et moderat nivå av profesjonell skepsis (når økt kunnskap øker profesjonell skepsis).

For å besvare forskningsspørsmålet utførte vi et eksperiment på 34 deltakerne. I eksperimentet skulle deltakerne vurdere revisjonsbevis innhentet ved dataanalyser. Deltakerne ble tilfeldig inndelt i to grupper, der den ene fikk informasjon om dataanalyser og tolkning av disse, mens den andre gruppen ikke fikk slik informasjon. Alle deltakerne gjennomførte en lavkompleks og en høykompleks oppgave, og ble etter hver oppgave bedt om å foreta vurderinger knyttet til profesjonell skepsis. Resultatene våre viser en statistisk signifikant effekt i retning av at økt kunnskap reduserer profesjonell skepsis. Når det gjelder oppgavekompleksitet observerer vi en reduksjon i profesjonell skepsis som følge av økt kompleksitet, men denne effekten er ikke statistisk signifikant. For interaksjonseffekter finner vi heller ikke noe av statistisk signifikant betydning. Konklusjonen på forskningsspørsmålet er derfor at revisorer med høyt kunnskapsnivå om dataanalyser og tolkning av disse har lavere profesjonell skepsis enn revisorer som ikke har slik kunnskap, men at oppgavekompleksitet ikke er av betydning.

## **Forord**

Denne masterutredningen er skrevet som en del av masterstudiet i regnskap og revisjon (MRR) ved Norges Handelshøyskole (NHH) i Bergen. Utredningen ble gjennomført høsten 2020, og utgjør 30 studiepoeng.

Masterutredningen er skrevet innenfor revisjonsforskning, og det har vært en spennende og utfordrende prosess. Vi har blant annet lært mye av å bruke eksperimentell metode som forskningsverktøy, samt fått økt innsikt i bruken av programvarene SPSS og Tableau. Videre har vi fått et godt innblikk i bruken av digitale verktøy i revisjon og endringene bransjen står overfor. Arbeidet har vært svært relevant og nyttig for vårt fremtidige arbeid i revisjonsbransjen, da vi har fått økt forståelse for hvordan teknologi kan være med på å forme fremtidens revisjon.

Vi ønsker å takke vår veileder Jonas Gaudernack for nyttige innspill og tilbakemeldinger. Videre vil vi takke Ulf Mohrmann for god veiledning vedrørende gjennomføring og tolkning av statistiske analyser. Vi ønsker også å takke masterstudentene som gjennomførte pre-test av undersøkelsen vår. Avslutningsvis vil vi takke alle deltakere som har tatt seg tid til å gjennomføre undersøkelsen.

# Innholdsfortegnelse

SAMMENDRAG .....	II
FORORD .....	III
INNHALDSFORTEGNELSE .....	IV
FIGUROVERSIKT .....	V
TABELLOVERSIKT .....	V
FORKORTELSER .....	VI
1. INNLEDNING .....	1
2. TEORI .....	3
2.1 Dataanalyser .....	3
2.1.1 Big Data .....	3
2.1.2 Audit Data Analytics .....	3
2.1.2.1 Dataanalyser: metoder og eksempler .....	4
2.1.2.2 Fordeler ved bruk av ADA .....	4
2.1.2.3 Utdfordringer ved bruk av ADA .....	5
2.1.2.4 Status hos revisjonsselskapene i dag .....	6
2.2 Profesjonell skepsis .....	7
2.2.1 Profesjonell skepsis ved vurdering av revisjonsbevis innhentet ved Audit Data Analytics .....	9
2.3 Oppgavekompleksitet .....	11
2.4 Kunnskap og erfaring .....	13
2.5 Teori om teknologisk dominans .....	14
3. FORSKNINGSMODELL OG HYPOTESER .....	17
3.1 Valideringsrammeverk for forskningsdesignet .....	17
3.2 Hypoteseutvikling .....	18
3.2.1 Hypotese 1: Kunnskapsnivå og profesjonell skepsis .....	18
3.2.2 Hypotese 2: Oppgavekompleksitet og profesjonell skepsis .....	19
3.2.3 Hypotese 3: Interaksjon mellom kunnskapsnivå og oppgavekompleksitet .....	20
4. METODE .....	21
4.1 Forskningsdesign .....	21
4.2 Utforming av case og spørreundersøkelse .....	23
4.3 Utvalg .....	25
4.4 Pre-test .....	27
4.5 Ethiske vurderinger .....	27
4.6 Reliabilitet og validitet .....	27
5. RESULTATER .....	29
5.1 Datarensing og manipulasjonssjekk .....	29
5.2 Deskriptiv statistikk .....	30
5.2.1 Grupper .....	30
5.2.2 Demografiske variabler .....	31
5.2.2.1 Kjønn .....	31
5.2.2.2 Alder .....	31
5.2.2.3 Erfaring .....	32
5.2.2.4 Fag innen revisjon .....	33
5.2.3 Oppsummering deskriptiv statistikk .....	33
5.3 Grunnlag for analyse .....	33
5.3.1 Kunnskapstest .....	33
5.3.2 Indeks .....	34
5.4 Resultater .....	34
5.4.1 Resultater: Hypotese 1 .....	34
5.4.2 Resultater: Hypotese 2 .....	36

5.4.3 Resultater: Hypotese 3 .....	37
6. DISKUSJON OG KONKLUSJON .....	39
6.1 Diskusjon av hypoteser.....	39
6.1.1 Diskusjon: Hypotese 1 .....	39
6.1.2 Diskusjon: Hypotese 2 .....	41
6.1.3 Diskusjon: Hypotese 3 .....	42
6.1.4 Diskusjon: Andre funn.....	42
6.2 Betydning for revisjonspraksis.....	44
6.3 Begrensninger .....	45
6.4 Videre forskning.....	46
LITTERATURLISTE .....	48

## Figuroversikt

Figur 1: Teorien om teknologisk dominans.....	15
Figur 2: Valideringsrammeverk for forskningsdesignet.....	17
Figur 3: Faktorielt design .....	22
Figur 4: Deskriptiv statistikk - Grupper .....	30
Figur 5: Deskriptiv statistikk - Kjønn.....	31
Figur 6: Deskriptiv statistikk - Alder .....	31
Figur 7: Deskriptiv statistikk -Erfaring .....	32
Figur 8: Deskriptiv statistikk - Fag i revisjon.....	33
Figur 9: Profildiagram – Hypotese 1.....	34
Figur 10: Profildiagram - Hypotese 2.....	36
Figur 11: Profildiagram - Hypotese 3.....	37

## Tabelloversikt

Tabell 1: Deskriptiv statistikk og p-verdier - Hypotese 1.....	35
Tabell 2: Deskriptiv statistikk og p-verdier - Hypotese 2.....	36
Tabell 3: Deskriptiv statistikk - Hypotese 3 .....	38
Tabell 4: P-verdier - Hypotese 3 .....	38
Tabell 5: P-verdier for kontrollvariabler .....	43
Tabell 6: Gjennomsnittlig score HPSS fordelt på lavt og høyt kunnskapsnivå.....	44
Tabell 7: Gjennomsnittlig score på profesjonell skepsis fordelt på lav og høy skeptisk tilstand ..	44

## **Forkortelser**

ADA: Audit Data Analytics

AICPA: American Institute of Certified Public Accountants

ANCOVA: Kovariansanalyse

ANOVA: Variansanalyse

CAAT: Computer Assisted Audit Techniques

CRM: Customer Relationship Management

ERP: Enterprise Resource Planning

HPSS: Hurtt's Professional Skepticism Scale

IAASB: International Auditing and Assurance Standards Board

IDA: Intelligent Decision Aid

ISA: International Standards on Auditing

MRR: Masterstudiet i Regnskap og Revisjon

NHH: Norges Handelshøyskole

NSD: Norsk senter for forskningsdata

TTD: Theory of Technology Dominance

# 1. Innledning

Teknologi og dataanalyser er høyaktuelle temaer innen revisjon. Den teknologiske endringen revisjonsbransjen står overfor beskrives som revolusjonerende, og både fagtidsskrifter og forskning mener den metodiske tilnærmingen vil endres betraktelig (Kinserdal, 2017; Earley, 2015). Digitaliseringstrenden er fremtredende blant de største revisjonsselskapene i Norge, der samtlige investerer betydelige beløp i digitale verktøy for å imøtekomme det teknologiske skiftet. Av selskapenes åpenhetsrapporter fremkommer det at investeringene skal resultere i nye analyseverktøy og ny teknologi som skal innarbeides i dagens revisjonsmetodikk (PwC, 2019; EY Norge, 2020; KPMG Norge, 2019; BDO AS, 2019; Deloitte, 2020).

Det er stor entusiasme blant de største revisjonsselskapene for en mer teknologipreget revisjon, og i revisjonstidsskrifter utpekes økt effektivitet og forbedret revisjonskvalitet som noen av de viktigste fordelene digitale verktøy medbringer (Pedersen, 2016). Til tross for dette foreligger det et stort gap mellom nåværende praksis, og potensialet som ligger i avanserte, høykomplekse dataanalyser og algoritmer. Det er en rekke forhold som hindrer bransjen i å omfavne dataanalyser. Bransjens treghet til å tilpasse seg en mer teknologisk tilnærming kan blant annet skyldes strenge regulatoriske omgivelser og IT-miljøet hos kundene. Uavhengig av en slik treghet er det tydelig at den metodiske tilnærmingen til revidering av årsregnskapet på sikt vil endre seg. Derfor er det sentralt å få innsikt i hvordan grunnkonseptene i revisjon påvirkes av endret metodikk. Denne masterutredningen tar sikte på å undersøke hvordan profesjonell skepsis påvirkes når revisjonsbevis innhentes ved hjelp av dataanalyser, nærmere bestemt hvordan nivåene av kunnskap og oppgavekompleksitet virker inn. Forskningsspørsmålet vårt er således:

*Hvordan påvirker kunnskap om dataanalyser og oppgavekompleksitet revisors profesjonelle skepsis ved vurdering av revisjonsbevis innhentet ved dataanalyser?*

I fagmiljøet uttrykkes det bekymringer vedrørende revisors kunnskap og kompetanse om dataanalyser og statistikk (IDEA, 2020; Katz, 2014). For at revisor skal kunne opparbeide seg en formening om at regnskapet i det alt vesentlige er i overensstemmelse med gjeldene rammeverk for finansiell rapportering, må revisor være i stand til å vurdere revisjonsbevis på en hensiktsmessig måte. Dersom de underliggende bevisene for å støtte konklusjonen er innhentet ved avanserte dataanalyser, krever det at revisor har tilstrekkelig dyptgående

forståelse for analysemetoden (Adrain, 2017; Earley, 2015). Revisor innehar sterk faglig kompetanse innen evaluering av finansiell informasjon, men når revisjonen i større grad blir teknologipreget må revisors ferdigheter og kunnskapsområder også omfatte teknologisk og statistisk forståelse (Adrain, 2017; Earley, 2015). En mangel på sådan kan ha betydelige konsekvenser for revisjonens kvalitet, og påvirker dermed både regnskapsbrukerne og revisor selv.

Da høykomplekse algoritmer og dataanalyser for innhenting av revisjonsbevis er nye områder for revisor, er det nødvendig med forskning som belyser hvordan revisor reagerer når kompleksiteten i revisjonshandlinger øker. Tidligere forskning omhandler i stor grad hvordan kompleksitet påvirker revisors vurderingsevne<sup>1</sup>, og det er begrenset med studier som kan si noe om effekten av høykomplekse dataanalyser på revisors profesjonelle skepsis. Vår studie bidrar dermed til å belyse hvordan revisor påvirkes av nye metoder for innhenting av revisjonsbevis.

---

<sup>1</sup> Oversatt fra det engelske begrepet «judgment performance».



## **2. Teori**

Denne delen vil belyse relevant teorigrunnlag for hypoteseutvikling og forskningsspørsmål. Det vil bli redegjort for dataanalyser, profesjonell skepsis, revisjonsbevis, kompleksitet, kunnskap og erfaring, samt teknologisk dominans.

### **2.1 Dataanalyser**

#### **2.1.1 Big Data**

Big Data kan defineres som «informasjon som kjennetegnes ved høyt volum, høy hastighet og/eller høy grad av variasjon, og som krever nye metoder for prosessering og tilrettelegging for å kunne fungere som grunnlag for å forbedre beslutninger, øke innsikt og optimalisere prosesser» (PwC, 2015; Gartner, 2012). Av begrepets definisjon fremkommer det tre karakteristikk som alene eller samlet alltid foreligger ved Big Data; høyt volum, høy hastighet og stor variasjon. Volum retter seg mot vekst i mengden av data som lagres, og tilgjengeligheten av denne. Hastighet innebærer at dataene blir generert i et tempo som overstiger menneskets evne til å tolke den. Begrepets tredje karakteristikk, variasjon, retter seg mot at data uthentes fra en rekke ulike kilder, eksempelvis lokasjonsdata fra GPS eller data fra sosiale medier (PwC, 2015). Variasjonen fører således til at formateringen på dataene varierer, og det skilles her mellom strukturert og ustrukturert data. Strukturerte data kan organiseres i tabeller, og er gjerne transaksjonsdata som finnes i ERP- eller CRM-systemer. Ustrukturerte data er derimot informasjon som ikke evner å struktureres i tabellform, eksempelvis Twitter-meldinger (PwC, 2015). Formateringen i ustrukturert data vanskeliggjør forståelsen av datasettet, og således er det utfordringer knyttet til verdiskaping basert på slik informasjon (Perry, 2017).

#### **2.1.2 Audit Data Analytics**

Audit Data Analytics (dataanalyser i revisjon), heretter ADA, defineres slik av IAASB (2017): «Audit Data Analytics is the science and art of discovering and analyzing patterns, deviations and inconsistencies, and extracting other useful information in the data underlying or related to the subject matter of an audit through analysis, modelling and visualization for the purpose of planning or performing the audit». ADA må sees i sammenheng med Big Data. Virksomheter genererer store volumer med data, og ADA som revisjonsmetodikk muliggjør analyse av store datasett og forbedrer revisors evne til å forstå selskapet og dets omgivelser.

Det eksisterer et bredt spekter av ulike ADA-verktøy, og disse har stor variasjon i kompleksitet. ADA omfatter alt fra datavisualisering til høykomplekse analyser basert på kunstig intelligens og maskinlæringsalgoritmer (Bonhome et al., 2018).

### **2.1.2.1 Dataanalyser: metoder og eksempler**

Dataanalyser har et bredt anvendelsesområde, og kan benyttes i flere deler av revisjonen. I planleggingsfasen er det viktig å identifisere områder med antatt høy risiko. Et eksempel på anvendelse av ADA her er prosessutvinning<sup>2</sup>. Denne typen analyse muliggjør vurdering av forretningsprosesser slik at revisor evner å se hvilke handlinger som er utført, når de er utført og av hvem. Slik kartlegger man enhetens flyt av transaksjoner, hvilket gir revisor et tydeligere risikobilde (Rayamajhi, 2019). Clusteranalyse er en annen metode som kan brukes i både planleggingsfasen og som analytisk handling for innhenting av revisjonsbevis. Metoden baserer seg på en maskinlæringsalgoritme som grupperer observasjoner i et datasett basert på deres karakteristikk, og er således hensiktsmessig for å utforske dataen og dens struktur. Dermed kan clusteranalyse brukes for å identifisere avvik og unaturlige sammenhenger (Thiprungsri, 2011). Videre legger dataanalyser til rette for vurdering av både finansielle og ikke-finansielle data. Pedersen (2016) eksemplifiserer dette med muligheten til å knytte leverandørregisteret til korrupsjonsindeksen for å avdekke leverandører revisor bør ha ekstra fokus på. En annen form for dataanalyse med anvendelse i revisjon er datavisualisering, hvilket defineres som «seleksjon, transformasjon og presentasjon av ulike typer data i en visuell form som bistår i å fasilitere utforsking og forståelse av data» (Alawadhi, 2015). Viktige beslutninger gjøres ofte, i henhold til Kleive (2019), basert på innsikt i data, og en slik innsikt kommer ofte av datavisualisering. Videre brukes datavisualisering ofte for å presentere revisjonens hovedfunn, men har også anvendelsesområde i planleggingsfasen for å avgjøre revisjonens omfang, eller i risikovurderingsfasen for å vurdere trender eller uteliggere i dataene (Audimation, 2019).

### **2.1.2.2 Fordeler ved bruk av ADA**

Det er en rekke fordeler ved å anvende dataanalyser i revisjon. AICPA (2017) peker blant annet på to: «(1) økt forståelse for enhetens operasjoner og tilhørende risikoer, inkludert mislighetsrisiko, og (2) økt potensial for å oppdage vesentlig feilinformasjon». Videre medfører ADA økt effektivitet og mer treffsikre risikovurderinger (Kleive, 2018).

---

<sup>2</sup> Prosessutvinning er oversatt fra det engelske begrepet «process mining».

De største revisjonsselskapene mener ADA vil endre dagens revisjonsmetodikk betydelig, og nevner spesielt forbedret revisjonskvalitet som en av de største fordelene (KPMG, 2015; Davenport, 2016; BDO, 2016; EY, 2015). Ved analyse og datavisualisering av store transaksjonsmengder vil revisor få økt innsikt i klientens finansielle tall. Videre påpeker Earley (2015) i en kartleggingsstudie at tilstrekkeligheten<sup>3</sup> av revisjonsbevis vil forbedres når revisjonsmetodikken beveger seg fra utvalgsbasert testing, til analyse av 100% av populasjonen.

ADA er velegnet som planleggingsverktøy, og bidrar til en effektiv revisjon med fornuftig tidsallokering. I risikovurderinger kan eksempelvis trendanalyser og sammenlikninger med bransjetall føre til en mer treffsikker revisjon ved å synliggjøre områdene med høyest risiko (Kleive, 2018; Murphy & Tysiac, 2015). Slik kan revisor legge til rette for riktig tidsbruk. Kleive (2018) peker videre på at ny teknologi kan utføre rutinepregede oppgaver slik at revisor kan bruke den frigjorte tiden på oppgaver som krever bruk av profesjonelt skjønn. Videre mener Association of Chartered Certified Accountants (u.d.) at bruk av digitale verktøy gir effektiviseringsgevinster fordi de kan prosessere store datamengder betydelig fortere enn mennesker, og at de kan generere presise analyser som understøtter revisors konklusjon om regnskapet.

FASB (1980) slår fast at relevans og pålitelighet er det som gjør finansiell informasjon nyttig for beslutningstakere. I tidsskriftet *Revisjon og Regnskap* stiller Aurstad (2017) spørsmål vedrørende relevansen av et regnskap som fremlegges seks måneder etter årsslutt. Effektiviseringsgevinster ved bruk av ADA kan således bistå med å løse denne utfordringen, ved at regnskapet kan bli fremlagt på et tidligere tidspunkt som følge av en raskere revisjon.

### **2.1.2.3 Utfordringer ved bruk av ADA**

Det foreligger dog utfordringer tilknyttet implementering av dataanalyser i revisjon. Et selskaps ERP-system inneholder store mengder relevant informasjon, men kan være uegnet til uthenting av store datamengder siden systemene ikke alltid er kompatible med revisjonsselskapenes analyseprogrammer. Dette gjør det vanskelig for revisor å gjennomføre analyser på en effektiv måte (IDEA, 2020). Videre skriver IAASB (2017) i en utforskende rapport om bruk av teknologi i revisjon at det foreligger infrastrukturelle utfordringer

---

<sup>3</sup> ISA 500 definerer tilstrekkelighet som målet på kvantiteten av revisjonsbevis (IAASB, 2009a)

vedrørende lagring og prosessering av store datavolumer. Disse utfordringene forplanter seg videre ved kompatibilitetsproblemene mellom systemene til klienten og revisjonsselskapet, hvilket vanskeliggjør dataauthenting der systemene ikke kan kommunisere med hverandre (Adrain, 2017). I tillegg utgjør også datasikkerhet og konfidensialitet en utfordring ved implementering av ADA. For å utføre analyser må revisor lagre den mottatte informasjonen i sine systemer, hvilket er problematisk med hensyn til gjeldene lovverk (Earley, 2015).

Flere anser manglende kunnskap og kompetanse om dataanalyser og statistikk innad i revisjonsteamet som en utfordring (Earley, 2015; Adrain, 2017; IDEA, 2020). Det fremkommer av en studie utført av Turley et al. (2016) at bruk av dataanalyser i revisjon overlates til spesialister, og således er det ingen forventning til at revisorer skal utføre dette arbeidet. En utfordring med outsourcing av analysearbeid innen revisjon er at IT-spesialister har en annen tilnærming til vurdering av dataanalyser enn revisor. Forståelse av datagrunnlaget er essensielt for å kunne fatte beslutninger basert på dataanalysene. Av den grunn er det problematisk at revisor konvergerer mot å være en konsument av slike analyser (Holmstrom, 2020). Ved implementeringen av teknologi og dataanalyser i revisjonsmetodikken, må revisors kompetanseområde derfor utvides for å imøtekomme endringene bransjen står overfor (Katz, 2014). Dette er forenlig med en undersøkelse utført av Forbes og KPMG (2015) der 42% av respondentene svarte at forståelse for dataanalyser er et av de viktigste kompetanseområdene for fremtidens revisjon.

#### **2.1.2.4 Status hos revisjonsselskapene i dag**

Ifølge en studie utført av Eilifsen, Kinserdal, Messier Jr og McKee (2020) er den generelle bruken av ADA hos de fem største revisjonsselskapene i Norge på et tidlig stadium, og ingen av selskapene har obligatorisk bruk av dataanalyser. Videre finner de begrenset anvendelse av avanserte ADA som eksempelvis regresjonsanalyser, clusteranalyser og statistiske prediktive analyser. Revisjonsbevis innhentet ved ADA fungerer i hovedsak som komplementære bevis. ADA blir, ifølge studien, mest brukt på nye kunder, samt kunder som har integrerte ERP- og IT-systemer. Disse resultatene er forenlige med en rapport fra det britiske finanstilsynet (2017), som avdekket at bruk av ADA i Storbritannia i hovedsak brukes for analyse av hovedbok og test av hovedboktransaksjoner (Financial Reporting Council, 2017).

Flere mener at årsaken til begrenset omfavnelser av ADA skyldes revisjonsprofesjonens regulatoriske rammer (Kinserdal, 2017; Botez & Vasile, 2018; Eilifsen et al., 2020). I en undersøkelse utført av Forbes i samarbeid med KPMG (2015) svarer 72% av revisormedarbeidere og studenter at de anser dette som den største utfordringen ved ADA. De internasjonale revisjonsstandardene hverken oppfordrer til eller forhindrer bruk av dataanalyser i revisjon, men tilrettelegger for anvendelse av CAAT (Computer Assisted Audit Techniques). Dog er revisjonsstandardenes referanser til CAAT utarbeidet i en tid der de teknologiske hjelpemidlene ikke er sammenliknbare med dagens teknologi (IAASB, 2017). Revisjonsstandardenes manglende tilpasning til dagens teknologiske omgivelser medfører at revisjonsselskapene selv må bevise at revisjonsbevis innhentet ved ADA er like sikre som bevis innhentet med den tradisjonelle revisjonsmetoden (Kinserdal, 2017). Kinserdal (2017) skriver videre at dette ligger til grunn for at revisjonsselskapene vegrer seg for implementering av digitale verktøy. Dette støttes av rapporten fra det britiske finanstillsynet (2017), der de også fant at det strenge regelverket revisor er underlagt medfører forsiktighet i anvendelse av ADA.

## **2.2 Profesjonell skepsis**

Tillit står sentralt i revisjonsfaget. Ifølge Revisorforeningen (u.d.) gir revisjon økt tillit til regnskapstallene, og bidrar til effektive og velfungerende kapitalmarkeder. Brukerne av regnskapet må ha tillit til at regnskapet er i overensstemmelse med gjeldende rammeverk for finansiell rapportering, og ikke inneholder vesentlig feilinformasjon. Revisorloven fastslår på sin side at revisor ved utøvelse av sin virksomhet er allmenhetens tillitsperson (Revisorloven, 1999, §1-2), og Birkeland (2004) mener at en forenklet definisjon av revisjon er å skape tillit til avgitt informasjon. Videre er tillit mellom revisor og klient et viktig aspekt. For å bekrefte informasjon fra klienten, må revisor gjøre seg opp en mening om hvorvidt informasjonen er til å stole på eller om flere revisjonshandlinger skal utføres. Denne beslutningen avhenger av revisors subjektive evaluering av klientens troverdighet og betydningen informasjonen har for revisors konklusjon om regnskapet (Shaub, 1996).

Tillit og profesjonell skepsis er komplementar av hverandre, der full tillit og full profesjonell skepsis utgjør ytterpunktene (Shaub, 1996). Således er profesjonell skepsis det motsatte av tillit. Det er ifølge ISA-ene profesjonell skepsis som beskriver revisors holdning i tillitsforholdet mellom revisor og klient (Shaub, 1996; IAASB, 2009b), og i en revisjonskontekst er det av den grunn mest formålstjenlig å benytte seg av dette konseptet. Det

er forventet at revisor ved planlegging og gjennomføring av revisjonen skal opptre med profesjonell skepsis (Shaub, 1996; IAASB, 2009b). Dette skal redusere risikoen for å eksempelvis «overse uvanlige omstendigheter, overgeneralisere når det trekkes konklusjoner fra revisjonsobservasjoner og legge feilaktige forutsetninger til grunn ved fastsettelsen av typen, tidspunkt og omfanget av revisjonshandlinger og ved evaluering av resultatene av dem» (IAASB, 2009b). Profesjonell skepsis er dermed et viktig konsept innen revisjon (Nolder & Kadous, 2018). Dog finnes det ingen felles forståelse blant standardsettere, praktiserende revisorer og forskere om begrepets innhold og hvordan det kan måles. Av IAASB (2009b) defineres profesjonell skepsis som: «En holdning som innebærer at revisor stiller spørsmål og er oppmerksom på forhold som kan indikere mulig feilinformasjon som følge av feil eller misligheter, og foretar en kritisk vurdering av revisjonsbevis». Forskere har derimot andre tilnærminger til begrepets innholdsmessige betydning.

Hurt (2010) definerer profesjonell skepsis som en flerdimensjonal individuell karakteristikk. Med dette mener hun at profesjonell skepsis kan være både et personlighetstrekk og en tilstand. Et personlighetstrekk holder seg stabilt og varig, mens tilstander vil være midlertidige og kontekststøtthengige (Hurt, 2010). Det betyr at revisorer med relativt likt nivå av personlighetstrekket profesjonell skepsis vil kunne ha ulik atferd i ulike situasjoner på grunn av tilstanden profesjonell skepsis (Robinson, Robertson, & Curtis, 2018).

Nolder og Kadous (2018) har en lignende tilnærming til begrepet. De mener at profesjonell skepsis kan konseptualiseres som både et tanke sett og en holdning. Tanke sett refererer til revisors måte å tenke og prosessere informasjon på, hvilket bestemmes av blant annet kunnskap, evner, motivasjon og personlighetstrekk. Skeptisk holdning reflekterer revisors kognitive og affektive vurderinger av ledelsens påstander og revisjonsbevis. Dette påvirkes av sosiale faktorer, herunder blant annet bedriftskultur, revisjonsstandarder og press fra klienter (Nolder & Kadous, 2018).

For å måle revisorers profesjonelle skepsis, har forskere utviklet ulike måleverktøy. Hurt (2010) utviklet en 30-spørsmålsskala, kalt Hurt's Professional Skepticism Scale (HPSS), som bygger på profesjonell skepsis som et personlighetstrekk. Skalaen ble utviklet basert på de seks følgende karakteristikkene: et spørrende sinn, evne til å avvente vurdering, søken etter kunnskap, mellommenneskelig forståelse, selvtillit og autonomi. Karakteristikkene er utledet av revisjonsstandarder, psykologi, filosofi og forskning om forbrukeratferd. De tre førstnevnte

karakteristikkene er knyttet til revisors innhenting og undersøkelse av revisjonsbevis. «Mellommenneskelig forståelse» tar i betraktning det menneskelige aspektet ved vurdering av revisjonsbevis, eksempelvis individers insentiv til å begå misligheter, mens de to sistnevnte karakteristikkene ser på individets evne til å ta beslutninger basert på innhentet informasjon (Hurtt, 2010).

Robinson et al. (2018) har modifisert Hurtt's skala slik at den også skal fange opp profesjonell skepsis som tilstand. I den modifiserte skalaen er Hurtt's 30 personlighetsspørsmål omgjort til 12 kontekstuelle spørsmål. Dette er gjort ved å blant annet rette de eksisterende spørsmålene mot den spesifikke situasjonen ved å legge til frasen «gjennom denne undersøkelsen», i tillegg til at spørsmål knyttet direkte til personlighetstrekk er blitt fjernet. Modellen ble i utgangspunktet utviklet for innhenting av revisjonsbevis og review av arbeidspapir, men kan også brukes i andre sammenhenger (Robinson et al., 2018).

Nolder og Kadous (2018) har utviklet en egen modell for måling av profesjonell skepsis basert på sin todeling av konseptet, som forklart over. Skeptisk tanke sett måles ved å evaluere revisors evne til å være spørrende, våken, objektiv og mottakelig i sin vurdering av revisjonsbevis. Skeptisk holdning måles ved å vurdere revisors oppfatninger, følelser og intensjoner (Nolder & Kadous, 2018). Revisors oppfatninger måles ved hjelp av spørsmål knyttet til risikoen for vesentlig feilinformasjon og hvor rimelige ledelsens estimater er. Følelser måles derimot ved revisors følelsesmessige vurderinger, herunder grad av bekymring og hvor komfortabel en er, av de samme elementene. Videre måles intensjonene ved å evaluere revisors plan for innhenting av mer revisjonsbevis og eventuelt videre revisjonshandlinger (Nolder & Kadous, 2018).

### **2.2.1 Profesjonell skepsis ved vurdering av revisjonsbevis innhentet ved Audit Data Analytics**

I henhold til ISA 500 er revisors mål å «utforme og utføre revisjonshandlinger på en slik måte at revisor kan innhente tilstrekkelig<sup>4</sup> og hensiktsmessig<sup>5</sup> revisjonsbevis for å kunne trekke rimelige konklusjoner som grunnlag for revisors mening» (IAASB, 2009a). Profesjonell skepsis er nødvendig for en kritisk vurdering av revisjonsbevis. Dette inkluderer å stille spørsmål vedrørende påliteligheten av dokumentasjon og metoder, samt vurdere pålitelighet

---

<sup>4</sup> Tilstrekkelighet er kvantiteten av revisjonsbevis (IAASB, 2009a).

<sup>5</sup> Hensiktsmessighet er målet på kvaliteten av revisjonsbevis, herunder pålitelighet og relevans (IAASB, 2009a).

og hensiktsmessighet av revisjonsbevis (Association of Chartered Certified Accountants, u.d.). Således vil en diskusjon av profesjonell skepsis ved vurdering av revisjonsbevis rette seg mot disse faktorene.

De standardutledede kravene til revisor og til vurdering av revisjonsbevis vil gjelde uavhengig av metodisk tilnærming for bevisinnhenting. Likevel mener Brown-Liburd og Vasarhelyi (2015) at de internasjonale revisjonsstandardene ikke er tilpasset mer komplekse og avanserte teknologiske omgivelser. Videre påpeker de at karakteristikene som brukes til å definere tilstrekkelighet, relevans og pålitelighet av revisjonsbevis ikke er tilstrekkelig adekvate for det teknologiske skifte revisjonsbransjen står overfor (Brown-Liburd & Vasarhelyi, 2015). Således mener de at revisjonsstandardene bør oppdateres for å være bedre tilpasset en mer teknologisk tilnærming til revisjonen.

Ifølge Yoon, Hoogduin & Zhang (2015) vil bruk av ADA og Big Data øke tilstrekkeligheten av revisjonsbevis. Dette forklares ved at det er mulig å analysere store datavolumer med stor variasjon i datasettene. Når Big Data brukes i revisjon innhentes datagrunnlaget ofte direkte av revisor. Dette styrker påliteligheten av beviset siden informasjon innhentet direkte av revisor er mer objektivt enn eksempelvis datagrunnlag som kommer fra en ansatt hos klienten. En ulempe ved bruk av Big Data er derimot at store datavolumer kan føre til støy i datasettet. Dette kan resultere i mange falske positiver<sup>6</sup> som har evne til å redusere relevansen av beviset. Dog kan problemet løses ved å fjerne dette støyet (Yoon et al., 2015).

Hensiktsmessighet og pålitelighet av revisjonsbevis påvirkes videre av dokumentasjon av revisjonshandlinger. I henhold til revisorloven skal revisor oppbevare dokumentasjon i minst ti år (Revisorloven, 1999, §5-5). Det faktum at det i enkelte tilfeller kan være mulig å endre datasettet som ligger til grunn for revisjonsbeviset i etterkant av at revisjonshandlingen er utført, utgjør et stort problem tilknyttet pålitelighet og relevans av data (Appelbaum, Kogan, & Vasarhelyi, 2017).

Siden tillit er komplementet til profesjonell skepsis er det svært relevant å inkludere forskning som belyser hvordan teknologi påvirker revisors tillit til revisjonsbevis. Tidligere studier viser at mennesker har mindre tillit til algoritmegenerert output, enn til output generert av mennesker

---

<sup>6</sup> Falsk positiv defineres som at man forkaster en sann nullhypotese (Braut, 2014)



(Commerford, Dennis, Joe, & Wang, 2020). Eksempelvis har aksjeprediksjoner betydelig større påvirkning på individers egne vurderinger når prediksjonen er utført av et menneske fremfor en algoritmebasert modell (Onkal, Goodwin, Thomson, Gönül, & Pollock, 2009). Slik algoritmeaversjon eksisterer til tross for at algoritmebaserte prediksjoner er mer presise enn menneskeskapt prediksjoner (Commerford et al., 2020). Sett i lys av profesjonell skepsis betyr algoritmeaversjon at revisor er mer skeptisk overfor maskingenerert output, og mindre skeptisk til menneskegenerert output.

Algoritmeaversjon er problematisk i revisjon. Dette forklares ved at når revisor stoler mer på mennesker enn på kunstig intelligens, står revisor i fare for å stole mer på ledelsens subjektivt pregede estimater, enn revisjonsselskapets egne systemer, som må anses som mer objektive (Commerford et al., 2020). Videre forsterkes effekten av algoritmeaversjon når man har konkurrerende informasjonsskilder. En studie utført av Birnbaum og Stegner (1979) viste at når man har et lavere oppfattet nivå av ekspertise hos én kilde, skapes det en avveining mellom informasjonskildene. Den ene kilden blir mindre vektlagt ettersom den konkurrerende kilden blir mer vektlagt. I en revisjonskontekst betyr dette at algoritmeaversjon gjør at revisor initielt stoler mer på mennesker enn maskiner, og at effekten forsterkes når revisor står overfor de to konkurrerende informasjonsskildene menneskeskapt og algoritmebasert output (Commerford et al., 2020).

### **2.3 Oppgavekompleksitet**

Forskning og litteratur innen oppgavekompleksitet gir ingen fullstendig eller tydelig definisjon av begrepet. Akademikere har ulike oppfatninger knyttet til om oppgavekompleksitet er en funksjon av oppgavekarakteristikker, om det er psykologiske opplevelser hos individet som utfører oppgaven, eller en kombinasjon av disse to (Campbell, 1988; Bonner, 1994). Av den grunn deler en gjerne oppgavekompleksitet inn i objektiv og subjektiv oppgavekompleksitet (Campbell, 1988; Bonner, 1994; Alissa, Capkun, Jeanjean, & Suca, 2014). Objektiv oppgavekompleksitet kan påvirkes av for eksempel informasjonsmengde og informasjonstype. Derimot er det reaksjoner og egenskaper hos personen som utfører oppgaven som har innvirkning på subjektiv oppgavekompleksitet<sup>7</sup> (Campbell, 1988).

---

<sup>7</sup> Det finnes lite forskning som kun ser på oppgavekompleksitet som subjektiv, da det til en viss grad er nødvendig å hensynta objektive oppgavekarakteristikker for å kunne se en psykologisk effekt. Subjektiv oppgavekompleksitet kan således sies å representere et kunstig skille. Faktorer som kan påvirke forholdet mellom

Objektiv oppgavekompleksitet har i henhold til Wood (1986) en tredelt inndeling; komponentkompleksitet, koordinativ kompleksitet og dynamisk kompleksitet (Wood, 1986). Disse tre typene vil til sammen avgjøre oppgavens totale kompleksitet, og dermed også nødvendig kunnskap og ferdigheter for å kunne utføre oppgaven (Wood, 1986). Komponentkompleksitet er en funksjon av ulike handlinger som må utføres og mengde informasjon som må prosesseres. Her vil et økt antall handlinger og informasjon gi økt kompleksitet. Koordinativ kompleksitet refererer til koordinasjonen som kreves ved gjennomføringen av en oppgave, herunder timing, hyppighet, intensitet og plassering av handlinger (Wood, 1986). Høyere koordinativ kompleksitet vil eksempelvis være et resultat av at flere relaterte steg må gjennomføres samtidig eller at handlinger i en del av oppgaven er avhengig av tidligere utførte handlinger (Tan & Kao, 1999). Dynamisk kompleksitet er kompleksitet som følge av endringer i en oppgaves årsak-virkningssammenheng i løpet av oppgaveutførelsen (Wood, 1986).

Flere forskere har studert effekten av oppgavekompleksitet på revisors prestasjon. Bonner (1994) utviklet en modell der vurderingsevne er en funksjon av oppgavekompleksitet, samt evner og motivasjon. Modellen illustrerer at en økning i oppgavekompleksitet vil føre til redusert vurderingsevne. Det er viktig å presisere at evner og motivasjon også kan ha innvirkning på prestasjon. En studie utført av Alissa, Capkun, Jeanjean og Suca (2014) bekrefter denne sammenhengen. Studien fant i tillegg at revisors innsats forbedrer prestasjonen, men at oppgavekompleksitet demper denne effekten (Alissa et al., 2014).

Et stadig mer omfattende regelverk fører til et økt antall revisjonshandlinger, samt økt kompleksitet (Alissa et al., 2014). Dette krever et høyere kunnskaps- og innsatsnivå av revisor. Abdolmohammadi og Wright (1987) fant i en studie en positiv sammenheng mellom oppgavekompleksitet og revisors kunnskap og erfaring. Resultatene viste at oppgaver med høyere kompleksitet krever større grad av revisors dømmekraft og innsikt i alle delene av beslutningsprosessen, noe som kan oppnås ved kunnskap og erfaring. Studien illustrerer således at når kompleksiteten i revisjonsoppgaven øker, vil revisor dra nytte av relevant erfaring (Abdolmohammadi & Wright, 1987).

---

objektiv og subjektiv oppgavekompleksitet er blant annet kjennskap til oppgaven, hukommelse, konsentrasjon og tidsbegrensninger. Eksempelvis vil et individ med god kjennskap til en svært kompleks oppgave anse oppgaven som mindre kompleks enn en uten slik kjennskap (Campbell, 1988).

Graden av kompleksitet i revisjonshandlinger vil variere, og revisorer vil ha ulik kunnskap og kompetanse. For å maksimere prestasjon og en fornuftig ressursallokering, er det nødvendig at revisjonsteamet består av revisorer med ulikt kunnskapsnivå, og at revisjonshandlingene har ulik kompleksitet (Alissa et al., 2014). Dette underbygges ved at Alissa et al. (2014) fant at total prestasjon maksimeres når erfarne revisorer blir tildelt høykomplekse oppgaver, og mindre erfarne revisorer lavkomplekse oppgaver.

## **2.4 Kunnskap og erfaring**

De største revisjonsselskapene i Norge mener at digital kompetanse er avgjørende for fremtidens revisjon (KPMG, 2016; PwC, 2019; EY, 2020; Deloitte, 2020; Earley, 2015). KPMG (2016) mener videre at kunnskap og kompetanse er en avgjørende differensiator i den teknologiske utviklingen i revisjon. I henhold til Libby og Luft (1993) vil revisor tilegne seg kunnskap gjennom erfaring. De utviklet en modell som utleder sammenhengen mellom erfaring, kunnskap og prestasjonsevne. I henhold til denne modellen vil erfaring medføre økt kunnskap, og deretter forbedret prestasjon. Johari og Sanusi (2010) bekreftet denne sammenhengen. Videre viser forskning innen kognitiv psykologi at erfaring gir en fordel i beslutningstaking, og at den aktuelle fordelene fremkommer som et resultat av mer kunnskap (Libby & Luft, 1993). Basert på dette utførte Libby og Frederick (1990) en studie som fant at det foreligger fordeler ved å engasjere erfarne revisorer for å evaluere revisjonsfunn. De erfarne revisorene evnet i større grad å fremlegge en mer fullstendig vurdering av plausible årsaker til regnskapsfeil. Videre viste studien at de erfarne deltakerne kom raskere frem til en korrekt konklusjon, og således kan kunnskap og erfaring bidra til effektivitetsgevinster (Libby & Frederick, 1990).

Det foreligger sprikende forskning på om økt kunnskapsnivå fører til en økning eller reduksjon av profesjonell skepsis. Payne og Ramsay (2005) utførte en studie der de vurderer om profesjonell skepsis påvirkes av mislighetsrisikovurderinger i planleggingsfasen av revisjonen. I deres studie var senior revisormedarbeidere mindre skeptiske enn nyansatte revisormedarbeidere. De mener en plausibel forklaring på at økt erfaring gir lavere profesjonell skepsis er fordi få revisorer oppdager misligheter i løpet av karrieren sin. En annen studie utført av Robertson (2010) støtter disse resultatene. Studien vurderte hvordan studenter og erfarne revisorer ble påvirket av klienten. Resultatene anfører at studenter var mindre tilbøyelige til å

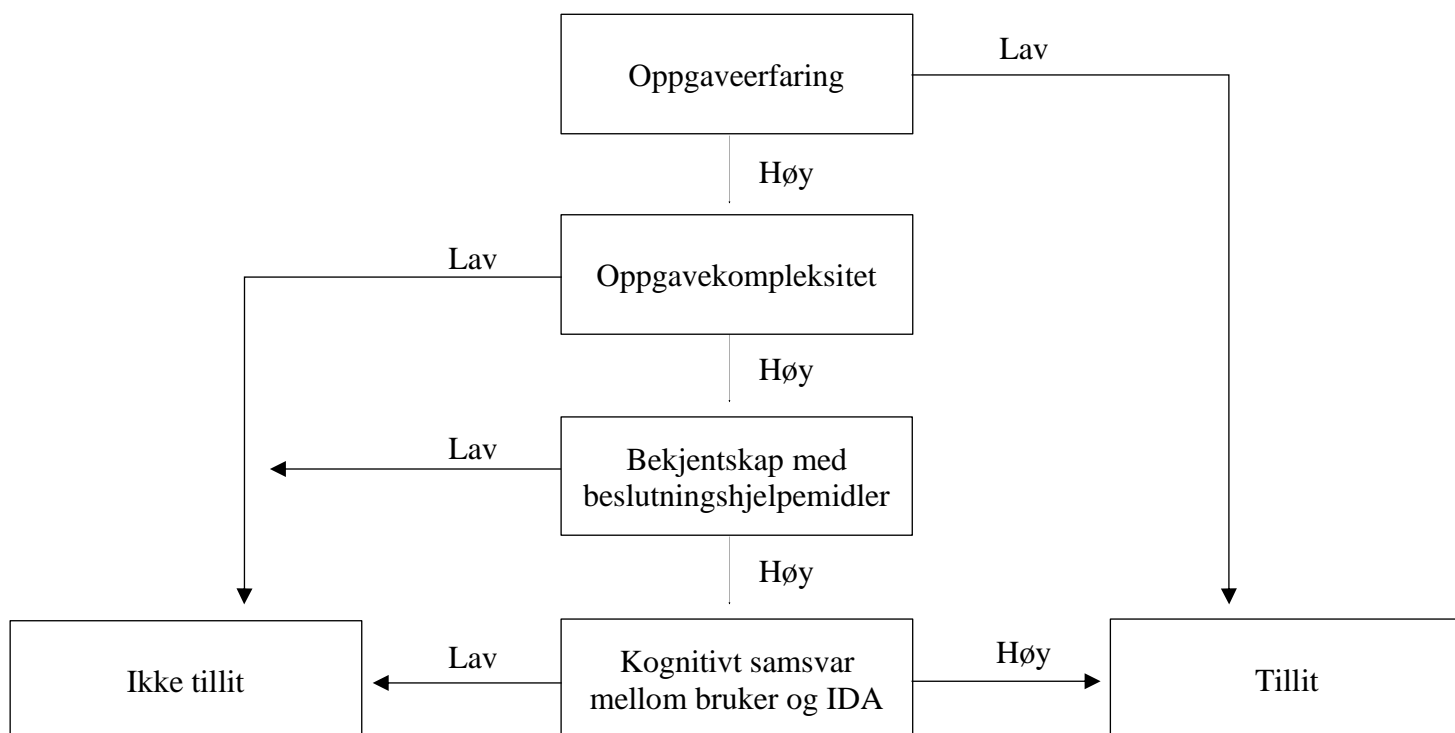
la seg påvirke av klienten enn erfarne revisorer. Således hadde studentene et høyere nivå av profesjonell skepsis enn erfarne revisorer (Robertson, 2010).

Annen forskning har funnet at økt erfaringsnivå leder til høyere profesjonell skepsis, fordi økt erfaring forbedrer revisors evne til å avdekke feil (Suryandari & Yuesti, 2017; Ratna & Anisykurlillah, 2020). I henhold til teorien om teknologisk dominans (se kapittel 2.5) er lavt kunnskapsnivå ensbetydende med lav profesjonell skepsis, og støtter således disse studiene. Holmstrom (2020) fant en tilsvarende effekt i sin studie. Hun vurderte hvordan tillit til uavhengige forventninger til regnskapsestimater innhentet ved avansert ADA og mindre avanserte metoder avhenger av oppgaveerfaring og eierskap til forventningene. Resultatene indikerte at tillitsnivå ble påvirket av oppgaveerfaring. Når forventningen genereres med avanserte ADA vil revisorer med høy grad av oppgaveerfaring stole mindre på forventningen, enn de med lav oppgaveerfaring (Holmstrom, 2020). Siden tillit er komplementet til profesjonell skepsis betyr resultatene at oppgaveerfaring øker skepsisnivået til revisor. Holmstrom (2020) mener at effekten skyldes at revisorer med høy grad av oppgaveerfaring har tilstrekkelig dyptgående forståelse til å sette spørsmålsteget ved mer avanserte metoder. På den annen side vil de med lav oppgaveerfaring sannsynligvis ikke ha tilstrekkelig forståelse for dataanalysen forventningen er basert på, hvilket leder til at de stoler blindt på output.

## **2.5 Teori om teknologisk dominans**

Teknologisk dominans defineres som «et stadium i beslutningstaking der beslutningshjelpemiddelet tar kontroll over beslutningstakingsprosessen, i stedet for brukeren» (Arnold & Sutton, 1998). Det vil si at brukeren på dette stadiet vil stole blindt på beslutningshjelpemiddelet. Arnold og Sutton (1998) utviklet en teori om teknologisk dominans (TTD), som tar sikte på å forklare hvordan ulike variabler påvirker tillit til output generert av intelligente beslutningshjelpemidler (IDA), og således når man blir utsatt for teknologisk dominans. I henhold til TTD er tillit til IDA en funksjon av oppgaveerfaring, oppgavekompleksitet, bekjentskap med IDA og kognitivt samsvar mellom bruker og IDA. Full tillit krever interaksjon mellom alle variablene, men oppgaveerfaring er eneste variabel som alene kan medføre full tillit. Arnold og Sutton (1998) presiserer at tillit er vanskelig målbart, men anfører at det er et ikke-binært mål, der tillit måles på en skala fra null til full tillit. Sett i lys av profesjonell skepsis vil full tillit bety lite profesjonell skepsis, mens null tillit vil bety høy profesjonell skepsis. Da oppgaveerfaring, oppgavekompleksitet og tillit er de eneste

variablene i TTD som er relevant for denne masterutredningen er det hovedsakelig disse som vil bli diskutert i dette kapittelet.



Figur 1: Teorien om teknologisk dominans

Hvorvidt man blir utsatt for teknologisk dominans avhenger av nivået på de ulike variablene i modellen. Når oppgaveerfaring er lav, vil revisor ha tillit til IDA, som vist i figur 1. Oppgaveerfaring er således den eneste variabelen som kan påvirke tillit alene (Hampton, 2005). Variabelen retter seg mot brukerens kunnskap om oppgaven som skal gjøres, og i hvilken grad beslutningstakeren utarbeider strategier for å fullføre eller løse oppgaven (Triki & Weisner, 2014; Hampton, 2005). I henhold til TTD vil mangel på samsvar mellom oppgaveerfaring hos en bruker og beslutningshjelpemiddelet medføre dårlig beslutningstaking (Mascha & Smedley, 2007). Dette er fordi uerfarne brukere av beslutningshjelpemiddelet vil bli utsatt for en illusjon av kontroll og således stole blindt på beslutningshjelpemiddelet (Triki & Weisner, 2014). Når dataanalyser blir en stadig større del av revisjonsmetodikken, må revisor være bevisst på teknologisk dominans og risikoene denne tilstanden medbringer.

Videre vil tillit hos de med høy oppgaveerfaring være avhengig av oppgavekompleksitet, bekjentskap med IDA og kognitivt samsvar mellom bruker og IDA (Triki & Weisner, 2014). Samtlige faktorer må befinne seg på et høyt nivå for at brukeren skal ha høy tillit til

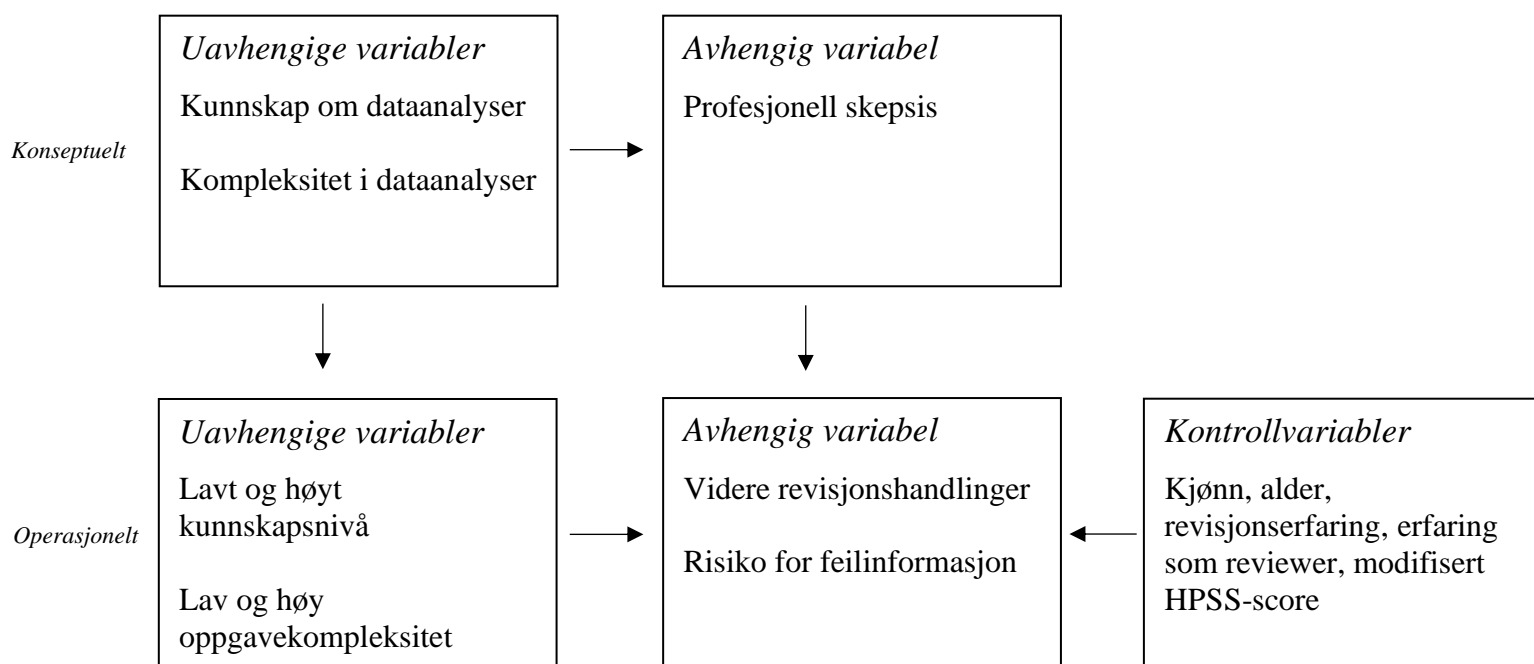
beslutningshjelpemiddelet. Oppgavekompleksitet vil i henhold til Arnold og Sutton (1998) måles etter hvilket nivå av kognitive evner som kreves av brukeren for å fullføre oppgaven, med andre ord hvilken mental innsats som kreves. Dette vil være personavhengig, men vil også bli bestemt av oppgavens karakteristikk i seg selv (Triki & Weisner, 2014). Dersom oppgavekompleksiteten er lav, vil brukeren tendere mot å ha lav tillit til beslutningshjelpemiddelet, som en kan se fra figur 1. Derimot vil brukeren tendere mot et høyere nivå dersom oppgavekompleksiteten er høy. Dette er dog avhengig av brukerens bekjentskap med IDA og kognitivt samsvar mellom bruker og IDA. Hampton (2005) slår fast at dersom oppgaveerfaring og oppgavekompleksitet begge er på et moderat til høyt nivå, vil forskjeller i brukerens tillit skyldes faktorene som nevnes over.

### 3. Forskningsmodell og hypoteser

I dette kapitlet skal vi redegjøre for forskningsmodell, samt utlede hypoteser.

#### 3.1 Valideringsrammeverk for forskningsdesignet

Basert på forskningsspørsmålet vårt har vi satt opp en modell for de konseptuelle og operasjonelle variablene som inngår i eksperimentet. Disse er presentert ved hjelp av Libby's valideringsrammeverk (Libby, 1981). Formålet med rammeverket er å rette fokus mot de viktigste faktorene for forskningsdesignets interne og eksterne validitet (Libby, Bloomfield, & Nelson, 2002). Rammeverket består av uavhengige variabler, avhengige variabler og kontrollvariabler. De uavhengige variablene skal manipuleres, mens avhengige variabler er variabler en måler variasjon i som følge av denne manipuleringen. Videre er kontrollvariabler andre variabler som potensielt kan påvirke resultatene i eksperimentet (Trotman, 2001).



Figur 2: Valideringsrammeverk for forskningsdesignet

For å kunne måle de ulike variablene, må de operasjonaliseres. Våre konseptuelle uavhengige variabler er kunnskapsnivå og kompleksitet i dataanalysene. Variablene operasjonaliseres i henholdsvis lavt og høyt kunnskapsnivå og lav og høy oppgavekompleksitet. Kunnskapsnivået manipuleres ved at halvparten av deltakerne får informasjon om dataanalyser og tolkning av disse, mens kompleksiteten manipuleres ved at deltakerne utfører både en lavkompleks og en høykompleks oppgave.

I eksperimentet skal vi måle variasjon i revisors profesjonelle skepsis som følge av manipuleringen av kunnskapsnivå og kompleksiteten i dataanalyser. Profesjonell skepsis gjøres målbart ved å operasjonalisere i to variabler; videre revisjonshandlinger og risiko for feilinformasjon. Variablene er valgt med bakgrunn i måling av skeptisk holdning fra Nolder og Kadous (2018) sitt rammeverk for profesjonell skepsis, som består av spørsmål knyttet til revisors vurderinger av revisjonsbevis. Ved å benytte dette rammeverket kan vi måle revisors skeptiske holdning til selve caseinnholdet, fremfor deres individuelle stabile skepsis.

For at eksperimentet skal ha intern validitet, må variasjon i avhengig variabel kun skyldes manipuleringen av uavhengige variabler (Trotman, 2001). For å vurdere om det eksisterer andre variabler som kan ha påvirket resultatene, hadde vi følgende kontrollvariabler: alder, kjønn, revisjonserfaring, erfaring med å gjennomføre «review» (gjennomgang av arbeidspapir utført av en annen revisormedarbeider) og modifisert HPSS-score etter Robinson et al. (2018) sitt rammeverk. Sistnevnte måler deltakernes skeptiske tilstand gjennom caset, og vil således gi et bilde på deres kontekstavhengige profesjonelle skepsis.

## **3.2 Hypoteseutvikling**

### **3.2.1 Hypotese 1: Kunnskapsnivå og profesjonell skepsis**

Det foreligger en viss diskrepans mellom uttalelser med budskap om at ADA vil revolusjonere revisjonsprofesjonen, og hva som i praksis blir gjennomført hos revisjonsselskapene. Kartleggingsstudier anfører at forståelse for dataanalyser er en av de viktigste ferdighetene for fremtidens revisor (Forbes & KPMG, 2015), men på nåværende tidspunkt utpekes manglende kompetanse om dataanalyser og statistikk som en hindring for implementering av ADA (Adrain, 2017; Earley, 2015; IDEA, 2020). Ved en slik implementering er det sentralt at revisor forstår utførelsen av dataanalysen, samt har tilstrekkelig statistisk forståelse til å analysere output (Earley, 2015). Til tross for endret fremgangsmåte for innhenting av revisjonsbevis, vil revisjonsprofesjonens grunnleggende elementer bestå. Dette betyr at profesjonell skepsis er like nødvendig uavhengig av metodisk tilnærming for bevisinnhenting.

I henhold til teorien om teknologisk dominans vil manglende kompetanse medføre at revisor ikke evner å forstå den underliggende handlingen, og således ikke foretar tilstrekkelig kritiske vurderinger (Hampton, 2005; Triki & Weisner, 2014). Dette leder til at revisor stoler blindt på intelligente beslutningshjelpemidler. I et revisjonsperspektiv er blind tillit ekvivalent med lavt



nivå av profesjonell skepsis. På bakgrunn av teori om teknologisk dominans, samt studiene til Suryandari og Yuesti (2017), Ranta og Anisykurlillah (2020; Rayamajhi, 2019) og Holmstrom (2020), vil økt kunnskap lede til økt profesjonell skepsis. Payne og Ramsay (2005), og Robertsen (2010) fant imidlertid en motsatt effekt. Deres resultater viste at økt kunnskap og erfaring reduserte profesjonell skepsis. Basert på forskningen på området ser det dermed ut til å være en sammenheng mellom kunnskap og profesjonell skepsis, men det er uvisst i hvilken retning. Derfor er det interessant å undersøke dette nærmere i kontekst av dataanalyser. Vi forventer således at økt kunnskap om dataanalyser vil påvirke profesjonell skepsis. Vår hypotese er dermed retningsnøytral, og er som følger:

*H<sub>0</sub>: Revisor har lik profesjonell skepsis uavhengig av kunnskapsnivå om dataanalyser og tolkning av disse.*

*H<sub>A</sub>: Revisorer med et høyere kunnskapsnivå om dataanalyser og tolkning av disse vil enten ha høyere eller lavere nivå av profesjonell skepsis enn de med lavt kunnskapsnivå.*

### **3.2.2 Hypotese 2: Oppgavekompleksitet og profesjonell skepsis**

Kartleggingsstudien til Eilifsen et al. (2020) fastslår at anvendelsen av avanserte dataanalyser hos de fem største revisjonsselskapene i Norge er svært begrenset. Samtidig foregår det en akselererende utvikling i teknologi som omfatter maskinlæring og kunstig intelligens, som også har et stort potensial innen revisjon (Commerford et al., 2020). Etersom den teknologiske utviklingen i revisjonsbransjen på sikt vil medføre omfavelse av mer avanserte verktøy, er det interessant å studere hvordan grunnleggende revisjonskonsepter påvirkes når kompleksiteten i revisjonshandlinger øker. Forskning innen oppgavekompleksitet viser at revisors vurderingsevne reduseres når oppgavekompleksitet øker. En slik effekt kan påvirke revisjonskvaliteten, og således ha konsekvenser for både regnskapsbrukere og revisor selv.

Basert på teorien om teknologisk dominans forventer vi at lav oppgavekompleksitet leder til økt profesjonell skepsis, og høy oppgavekompleksitet leder til redusert profesjonell skepsis. Det er dog lite eksisterende forskning på hvordan oppgavekompleksitet påvirker profesjonell skepsis, og således bidrar denne masterutredningen til å belyse denne sammenhengen. Vår hypotese er dermed som følger:

*H<sub>0</sub>: Revisor har lik profesjonell skepsis for lav- og høykompleksoppgaver*

*H<sub>A</sub>: Revisors profesjonelle skepsis reduseres når oppgavekompleksiteten øker*

### **3.2.3 Hypotese 3: Interaksjon mellom kunnskapsnivå og oppgavekompleksitet**

Som beskrevet tidligere anses revisors kunnskapsnivå relatert til dataanalyser å være lavt, samtidig som den teknologiske utviklingen indikerer økt bruk av avanserte dataanalyser i fremtiden (Commerford et al., 2020). Dermed er det hensiktsmessig å studere hvordan de to variablene kombinert påvirker profesjonell skepsis.

Som utledet over forventer vi at økt oppgavekompleksitet reduserer profesjonell skepsis. Da det imidlertid er ulike resultater fra forskning som belyser sammenhengen mellom kunnskapsnivå og profesjonell skepsis, vil hypotese 3 være todelt og avhenge av resultatene som blir funnet under hypotese 1. Dersom man legger til grunn forskning som viser at profesjonell skepsis reduseres når kunnskapsnivå øker, forventer vi at en økning i oppgavekompleksitet vil medføre en forsterket reduksjon i profesjonell skepsis, det vil si en positiv interaksjonseffekt (H<sub>A1</sub>).

Hvis man derimot går ut fra forskning som viser at økt kunnskapsnivå øker profesjonell skepsis, herunder TTD, forventer vi at effekten av økt kunnskap vil motvirke effekten av økt kompleksitet. Dette vil resultere i et moderat nivå av profesjonell skepsis, det vil si en negativ interaksjonseffekt (H<sub>A2</sub>). Det vil si at de med økt kunnskapsnivå har mer forståelse, og er derfor mer kompetente til å løse høykomplekse oppgaver. Dermed vil de stå imot effekten av at økt kompleksitet reduserer profesjonell skepsis, og ende opp med et moderat nivå av profesjonell skepsis.

Basert på dette er hypotesene våre som følger:

*H<sub>0</sub>: Revisors profesjonelle skepsis er lik uavhengig av nivå av oppgavekompleksitet og kunnskap om dataanalyser.*

*H<sub>A1</sub>: Dersom revisor har et høyt kunnskapsnivå, vil en økning i oppgavekompleksitet føre til en forsterket reduksjon i profesjonell skepsis.*

*H<sub>A2</sub>: Dersom revisor har et høyt kunnskapsnivå, vil en økning i oppgavekompleksitet føre til et moderat nivå av profesjonell skepsis.*

## 4. Metode

I dette kapittelet vil vi presentere forskningsdesignet vårt. Videre vil vi gå gjennom utformingen av caset og spørreundersøkelsen, samt redegjøre for undersøkelsens utvalg. Avslutningsvis vil vi ta for oss etiske vurderinger gjort i forbindelse med undersøkelsen, i tillegg til å vurdere eksperimentets reliabilitet og validitet.

### 4.1 Forskningsdesign

Et forskningsdesign har som hovedformål å sørge for at en får svar på angjeldende forskningsspørsmål, samt å kontrollere for feilvarians (Trotman, 2001). Det ligger mange ulike valg bak utformingen av et forskningsdesign, og vi skal her redegjøre for de valgene vi har tatt.

Det skilles gjerne mellom tre ulike typer forskningsdesign; eksplorativt, beskrivende og eksperimentelt<sup>8</sup> (Van Wyk, 2012). Et eksperimentelt design beskrives av Kerlinger (1973) som en vitenskapelig undersøkelse der uavhengige variabler manipuleres for å observere variasjon i avhengige variabler. Dermed muliggjør et slikt design manipulering av den eller de variablene som er av interesse, samtidig som at kausale sammenhenger mellom variablene kan utledes (Trotman, 2001). Med et eksperimentelt design kan vi således studere effekten kunnskapsnivå hos revisor, og oppgavekompleksitet i dataanalyser, har på revisors profesjonelle skepsis.

Videre har vi utformet et 2x2 faktorielt design, vist i figur 3. Et faktorielt design er effektivt, da en kan kombinere studier av flere variabler i ett eksperiment (Trochim & Donnelly, 2008), i vårt tilfelle både deltakernes kunnskapsnivå om dataanalyser og kompleksiteten i dataanalysen. Videre er det ved hjelp av et slikt design mulig å undersøke om det foreligger interaksjonseffekter. Dette er effekter som oppstår når variasjoner i en faktor påvirkes av nivået på en annen faktor (Trochim & Donnelly, 2008). Således forteller interaksjonseffekter hvordan to ulike faktorer kombinert påvirker den avhengige variabelen. Ved å benytte et faktorielt design, der en har deltakere som utsettes for hver av tilstandene, vil en kunne observere en potensiell interaksjon mellom faktorene.

---

<sup>8</sup> Ved et eksplorativt forskningsdesign utforsker man et forskningsområde, og målet er å identifisere variabler som kan være relevante for forskningsspørsmålet, samt begrensninger ved omgivelsene. Ved et beskrivende forskningsdesign er målet å skaffe til veie en nøyaktig og korrekt fremstilling av faktorer som er relevante for forskningsspørsmålet (Van Wyk, 2012).

		Kunnskap om dataanalyser	
		<i>Lavt</i>	<i>Høyt</i>
Kompleksitet i dataanalysen	<i>Lav</i>	Gruppe 1	Gruppe 2
	<i>Høy</i>		

*Figur 3: Faktorielt design*

Figur 3 viser eksperimentets faktorielle design. Ut fra dette kan en se at gruppe 1 og gruppe 2 kontrollerer hverandre for å måle hovedeffekten for kunnskapshypotesen, det vil si om deltakernes kunnskapsnivå påvirker deres profesjonelle skepsis. Når det gjelder hovedeffekten som dreier seg om oppgavekompleksitet, vil deltakerne kontrollere seg selv ved at de utfører både en lavkompleks- og en høykompleksoppgave. Det vil si at de fungerer både som testperson og kontrollperson. For å måle en potensiell interaksjonseffekt i hypotese 3 testes lavt nivå av både kunnskap og oppgavekompleksitet mot høyt nivå av faktorene.

For de uavhengige variablene må det bestemmes om man skal manipulere «within-subjects» eller «between-subjects». Ved bruk av et «within-subjects» design vil alle deltakere bli utsatt for alle tilstander. Derimot vil deltakerne ved et «between-subjects» design kun bli utsatt for én tilstand (Trotman, 2001). Vi har valgt å bruke et faktorielt design der vi kombinerer både «within-subjects» og «between-subjects» design. «Within-subjects» design er brukt ved at begge gruppene utfører både en lavkompleks og en høykompleks oppgave. Ved å bruke et «within-subjects» design vil deltakerne fungere som sin egen kontrollperson (Price, Jhangiani, Chiang, Leighton, & Cuttler, 2017; Trotman, 2001). Dermed vil det være behov for færre deltakere med et slikt design. Dette er fordelaktig i en masteroppgave, da det er vanskelig å få tilstrekkelig antall deltakere. En ulempe er imidlertid at det vil være enklere for deltakerne å gjette seg til hypotesen ved bruk av et «within-subjects» design. Dette er problematisk da deltakerne kan avgi svar basert på det de tror er forventet av dem. Videre vil det kunne oppstå rekkefølgeeffekter som følge av at hver deltaker får utdelt to oppgaver med ulikt kompleksitetsnivå (Price et al., 2017; Trotman, 2001). Deltakerne vil kunne lære av konklusjonen på den første oppgaven og ta med seg dette videre inn i den neste oppgaven, og på den måten få redusert profesjonell skepsis. På grunn av risikoen for rekkefølgeeffekter i eksperimentet vårt, randomiserte vi oppgaverekkefølgen, det vil si rekkefølgen på lavkompleks- og høykompleksoppgaven. Hva en oppnår med denne randomiseringen kan ses på to måter; rekkefølgeeffektene elimineres slik at total variasjon i avhengig variabel ikke vil

skyldes rekkefølgen på tilstandene, samt man kan undersøke om det eksisterer rekkefølgeeffekter (Price et al., 2017; Trotman, 2001).

«Between-subjects» design er brukt ved at kun halvparten av deltakerne (gruppe 2) får informasjon om dataanalyser og hvordan de tolkes. Bruk av et slikt design forhindrer at deltakerne klarer å gjette hypotesen som omhandler kunnskap ( $H_1$ ), og det vil heller ikke eksistere rekkefølgeeffekter. Deltakerne blir tilfeldig plassert i en av gruppene, hvilket øker intern validitet. Imidlertid medfører dette en risiko for at deltakere som allerede innehar mye kunnskap om dataanalyser havner i gruppen uten informasjon. Dette kan forstyrre effekten av manipuleringen ved at effekten av kunnskap om dataanalyser og tolkning av disse viskes ut. Likevel anså vi det som mest hensiktsmessig å ha kunnskap som en manipulert variabel fremfor en målt variabel. Dette fordi utvalget vårt i hovedsak ville bestå av studenter som går masterstudiet i regnskap og revisjon, der kurs innen digital revisjon inngår i fagplanen. Dermed kunne kunnskap som målt variabel potensielt medført en skjevfordeling av deltakere i de ulike gruppene. Da det er begrenset hvor mange deltakere en masteroppgave vil kunne få tak i, måtte jevne grupper sikres. For å vurdere deltakernes kunnskapsnivå inkluderte vi en kunnskapstest innledningsvis i spørreundersøkelsen. Vi kunne dermed måle om økt kunnskap hos gruppe 2 hadde effekt eller ikke.

## **4.2 Utforming av case og spørreundersøkelse**

Spørreundersøkelsen ble utformet i programmet Qualtrics, og bestod av tre deler; en kunnskapstest, case med tilhørende oppgaver og demografiske spørsmål. Kunnskapstesten var det første deltakerne ble utsatt for. Denne ble utformet med bakgrunn i selve caseoppgavene, og bestod av spørsmål om linjediagram, box plot og clusteranalyse. Hensikten med kunnskapstesten var å få en initiell vurdering av deltakernes kunnskap om dataanalyser, for å vurdere om kunnskapsnivået potensielt kunne påvirke effekten av manipuleringen. Dersom gruppen som ikke fikk informasjon hadde et betydelig høyere kunnskapsnivå enn gruppen som fikk informasjon, ville dette motvirket effekten av manipuleringen. Kunnskapstesten måtte derfor være med for å se om man hadde sammenliknbare grupper.

Når det kommer til utformingen av selvet caset, omhandlet dette revisjon av det fiktive vannkraftsselskapet Energi AS. Innledningsvis fikk deltakerne presentert selskapsspesifikk informasjon. Videre fikk de oppgitt at de var managere som hadde jobbet i revisjonsselskapet

i fem år, og at de skulle utføre en review av revisjonsarbeidet til en kollega som hadde jobbet i selskapet i to år. Kollegaen hadde revidert gyldigheten av inntektene fra en stor og vesentlig kunde (kalt Handel AS).

Revisjon av inntekter ble valgt hovedsakelig fordi dette er pensum i revisjonsfag på NHH og derfor noe deltakerne vil ha kjennskap til, uavhengig om de har praktisk erfaring eller ikke. Da utvalget hovedsakelig ville bestå av studenter, anså vi derfor revisjon av inntekter som hensiktsmessig. Videre inneholder inntektsposten store datavolumer, hvilket gjør det enklere å bruke dataanalyser på inntektstransaksjoner enn på andre poster.

Hver deltaker skulle utføre review av to ulike arbeidspapirer. Det ene arbeidspapiret var basert på revisjonshandlinger utført med lavkomplekse dataanalyser (linjediagram og box plot), og det andre var utført med en høykompleks dataanalyse (clusteranalyse). Kompleksiteten i dataanalysene er basert på objektiv oppgavekompleksitet, herunder komponentkompleksitet og koordinativ kompleksitet<sup>9</sup> (Wood, 1986). For den lavkomplekse oppgaven er det færre handlinger som må gjennomføres og mindre informasjon å prosessere enn for den høykomplekse oppgaven. Ved sistnevnte økes oppgavekompleksiteten ved at deltakeren i tillegg til å vurdere selve spredningsplotet og clustere, også må ta stilling til deskriptiv statistikk og om modellen er god eller dårlig ved å se på p- og F-verdi. Når det gjelder koordinativ kompleksitet foreligger det ved den lavkomplekse dataanalysen ingen avhengig rekkefølge og timing for de utførte handlingene, i motsetning til ved den høykomplekse dataanalysen. Ved clusteranalysen er en nødt til å prosessere de ulike delene i riktig rekkefølge, samt se de i sammenheng, for at det skal gi mening, og dette gjør derfor at oppgaven blir mer kompleks. De to oppgavene ble vist i randomisert rekkefølge. Etter hvert arbeidspapir fikk deltakerne tre spørsmål vedrørende innhenting av ekstra revisjonsbevis, risiko for feilinformasjon og om de var i stand til å konkludere på gyldigheten av inntektene, etter Nolder og Kadous (2018) sitt rammeverk om skeptisk holdning. Som svaralternativer ble Likert-skala med syv nivåer brukt, der 1 tilsvarte «veldig lite» og 7 var «veldig mye/stor». Tillit og profesjonell skepsis er ikke-binært, og det er derfor naturlig å ha en skala som gjenspeiler at det finnes flere nivåer av tillit og profesjonell skepsis. Ved å ha syv nivåer tvinges deltakerne til å i større grad gi en nyansert vurdering.

---

<sup>9</sup> Etter Wood (1986) deles objektiv oppgavekompleksitet i tillegg inn i dynamisk kompleksitet. Dynamisk kompleksitet er imidlertid ikke aktuelt ved reviewing av allerede utførte revisjonshandlinger, da det ikke foreligger endringer i oppgavens årsaks-virkningssammenheng i løpet av reviewingen.

I tilknytning til oppgavene fikk halvparten av deltakerne (gruppe 2) informasjon om dataanalyser, mens den resterende halvparten (gruppe 1) ikke fikk dette. Informasjonen besto av hvordan man skal tolke de ulike dataanalysene, samt deskriptiv statistikk og hvordan man skal vurdere denne. Deltakerne hadde tilgang til informasjonen under gjennomføringen av oppgavene. Dette var for å forhindre at informasjonen bare ble raskt gjennomlest. For at manipulasjonen skal ha effekt er det nødvendig at deltakerne leser informasjonen nøye, noe vi ønsket å tilrettelegge for ved å ha den tilgjengelig under gjennomføringen.

Etter at deltakerne hadde gjennomført caset, var det inkludert en manipulasjonssjekk. Manipulasjonssjekken har som hensikt å undersøke om manipuleringen av de uavhengige variablene har fungert som tiltenkt (Libby, Bloomfield, & Nelson, 2002). Vi spurte her deltakerne om de fikk oppgitt spesifikk informasjon om de ulike dataanalysene og hvordan disse skal tolkes. Det ble her brukt nominalt målenivå<sup>10</sup>. Deltakere som ikke besto manipulasjonssjekken ble eliminert fra datasettet.

Avslutningsvis samlet vi inn demografiske data, herunder data om alder, kjønn, erfaring og kurs, samt svar på spørsmål i modifisert HPSS. Modifisert HPSS måler profesjonell skepsis som en tilstand, og vil dermed bidra til å observere deltakernes kontekstavhengige skepsis i tilknytning til vår undersøkelse. For de demografiske spørsmålene ble det brukt nominal måling, mens ved modifisert HPSS benyttet vi en Likert-skala med syv nivåer.

Fullstendig spørreundersøkelse kan ses i vedlegg 1.

### **4.3 Utvalg**

En populasjon er den gruppen av mennesker et statistisk utvalg trekkes fra (Kenton, 2020). Populasjonen vi ønsket å se på i vårt eksperiment var revisorer. Etter å ha tatt kontakt med NHH, fikk vi imidlertid beskjed om at vi ikke fikk lov til å kontakte fagavdelingene i revisjonsselskapene for videre utsendelse til revisorer. Det ble imidlertid gitt tillatelse til bruk av eget nettverk. I tillegg til å henvende oss til bekjente, var vi nødt til å benytte oss av studenter for å øke antallet respondenter. Da vi ikke har tilgang til hele populasjonen, men benytter oss

---

<sup>10</sup> Nominalt målenivå brukes der en har to eller flere kategorier, og det ikke er rangering mellom kategoriene fra eksempelvis høyest til lavest. Dette kan eksempelvis være kjønn (UCLA Statistical Consulting Group, 2020a).

av de respondentene vi kan få tak i, blir utvalget vårt et ikke-sannsynlighetsutvalg, og et såkalt bekvemmelighetsutvalg (Rooney & Evans, 2018). Dette er raskere og enklere å gjennomføre enn et sannsynlighetsutvalg, og er derfor hensiktsmessig å benytte ved masterutredninger der en har begrenset med tid og ressurser. Ved bruk av et ikke-sannsynlighetsutvalg konstaterer imidlertid Rooney og Evans (2018) at det foreligger en mulighet for at utvalget ikke er representativt for populasjonen.

Når det gjelder bruken av studenter som deltakere i eksperiment, er forskere splittet i sine synspunkter. Sears (1986) argumenterer for at studenter har andre tanker og holdninger enn praktiserende revisorer, samt at de er vant til å løse kognitive oppgaver og dermed har lettere for å skjønne hypoteser og dermed endre atferd. Peecher og Solomon (2001) mener imidlertid at studenter bør være standarddeltakeren. Dette begrunnes med at de fleste karakteristika ved et eksperiment kan manipuleres slik at det er tilpasset bruk av studenter som deltakere, og at den eksterne validiteten dermed ikke vil svekkes. Videre argumenterer de for at bruk av praktiserende revisorer er en mer kostbar og begrenset ressurs sammenliknet med studenter (Peecher & Solomon, 2001).

Med bakgrunn i elementene over satte vi som krav at deltakerne måtte ha arbeidserfaring innen revisjon eller at de måtte ha tatt eller påbegynt grunnleggende fag innen revisjon. Dermed ville også 4.årsstudenter som nylig har påbegynt en master i regnskap og revisjon (MRR) på NHH kunne delta. Dette begrunnes med at studentene kun hadde én forelesning igjen i faget Revisjon 1 på tidspunktet undersøkelsen ble sendt ut, og de har således gjennomgått tilnærmet alt pensum som inngår i faget. Vi inkluderte spørsmål om fag/kurs og arbeidserfaring avslutningsvis i undersøkelsen for å kontrollere at deltakerne oppfylte kravene.

Undersøkelsen ble sendt ut 20.10.2020. Den ble lagt ut i Facebook-grupper for MRR-studenter på NHH, kalt «MRR NHH kull 2019 og 2020» og «MRR NHH 2019-2021». Videre sendte vi melding til bekjente i revisjonsselskapene, samt oppfølgingsmelding til bekjente på MRR. Vi endte opp med å få 34 svar, hvorav 31 var gyldige.



#### **4.4 Pre-test**

Vi gjennomførte pre-test på to sisteårsstudenter på masterstudiet i regnskap og revisjon, samt på veilederen vår. Formålet med testene var å få innspill om undersøkelsens lengde, om caset, instruksjoner og spørsmål var forståelige, og om de klarte å forstå hva hypotesene var. Vi fikk tilbakemelding om at undersøkelsen var i overkant tidkrevende. Derfor ble tekst og informasjon vi anså som mindre relevant og uten påvirkning for gjennomføringen av eksperimentet fjernet. I tillegg tok vi bort noen spørsmål i tilknytning til caset, og spisset de resterende spørsmålene slik at målingen av vår avhengige variabel ble foretatt på en hensiktsmessig måte. Videre endret vi på enkelte formuleringer der testdeltakerne oppga at det var uklarerheter.

#### **4.5 Etiske vurderinger**

Det var frivillig å delta i undersøkelsen, og deltakerne kunne trekke seg når som helst i løpet av gjennomføringen. For å tilrettelegge for ærlig avgivelse av svar, ønsket vi at undersøkelsen skulle være anonym. Ved utforming av undersøkelser i Qualtrics har man muligheten til å skru av lagring av IP-adresse, noe vi gjorde for å sikre denne anonymiteten. Videre stilte vi ingen personidentifiserende spørsmål da dette ikke var nødvendig. For å forsikre oss om at vi fulgte gjeldende retningslinjer for behandling av personopplysninger og for å undersøke om meldeplikten var utløst, tok vi kontakt med NSD (Norsk senter for forskningsdata). De bekreftet at meldeplikten ikke var utløst.

#### **4.6 Reliabilitet og validitet**

Valgene vi har tatt ved utformingen av vårt eksperiment har påvirket eksperimentets validitet og reliabilitet. Validitet, herunder intern og ekstern validitet, er viktig for å vurdere et eksperimentelt designs kvalitet, nøyaktighet og nytte (McDermott, 2011). Målet er å maksimere både intern og ekstern validitet, men det må ofte foretas kompromisser mellom disse. Ekstern validitet angir i hvilken grad resultatene av et eksperiment kan anvendes på tvers av andre populasjoner og situasjoner, det vil si generaliseringen av resultatene i eksperimentet (McDermott, 2011). En masteroppgave har begrenset med tilgjengelig tid og ressurser, og det er derfor vanskelig å få en stor utvalgsstørrelse. Videre benyttet vi som nevnt de deltakerne vi kunne få tak i, og fikk dermed ikke brukt et sannsynlighetsutvalg. Disse faktorene svekker den eksterne validiteten til eksperimentet vårt, og gjør det vanskelig å generalisere resultatet til hele populasjonen.

Den interne validiteten angir i hvilken grad variasjoner i avhengig variabel skyldes endringer i uavhengige variabler (Peecher & Solomon, 2001). For å øke den interne validiteten implementerte vi randomisert inndeling i de ulike gruppene i eksperimentet. Qualtrics allokerer hver deltaker tilfeldig til en gruppe med eller uten informasjon om dataanalyser, samt lavkompleks eller høykompleks oppgave først. Videre inkluderte vi en manipulasjonssjekk som skulle teste om deltakerne fikk med seg manipuleringen de ble utsatt for. Deltakere som feilet på denne sjekken ble eliminert fra datasettet. Intern validitet blir videre styrket dersom forskeren eller hans assistenter er til stede når deltakerne utfører eksperimentet (Trotman, 2001). På grunn av koronasituasjonen så ble dette vanskelig for oss, og vi var derfor nødt til å benytte oss av kun digital gjennomføring.

Reliabiliteten i et eksperiment refererer til påliteligheten i målingene, det vil si at resultatene skal være konsekvente (Price et al., 2017). For å styrke reliabiliteten gjennomførte vi en pre-test. Dette bidro til å oppdage feil og uklarheter som ble rettet opp før undersøkelsen ble sendt ut. I tillegg fjernet vi enkelte spørsmål knyttet til caset og spisset de resterende, da vi fikk tilbakemeldinger på at undersøkelsen var for omfattende. Spørsmålene ble formulert nøytralt for å unngå å lede deltakerne i en retning. Videre var undersøkelsen anonym, noe som gir deltakerne større insentiv til å svare så ærlig som mulig.

## 5. Resultater

Dette kapitlet vil presentere datarensing og manipulasjonssjekk, deskriptiv statistikk for analysens datagrunnlag og resultater fra analysene. Alle testene som er omfattet av dette kapitlet har benyttet et signifikansnivå på 5%. Det betyr at dersom p-verdien er mindre enn 0,05 forkaster vi nullhypotesen.

### 5.1 Datarensing og manipulasjonssjekk

Undersøkelsen hadde initielt 34 deltakere. Før vi begynte å analysere, foretok vi imidlertid en rensing av datasettet, der tre av respondentene ble eliminert. En deltaker ble fjernet fordi vedkommende ikke hadde besvart spørsmålene tilknyttet vurderingen av profesjonell skepsis for høykompleksoppgaven. De andre to ble eliminert grunnet feil svar på manipulasjonssjekk. Sjekken gikk ut på at deltakerne skulle svare på om de hadde fått informasjon om dataanalyser og tolkningen av disse, og sjekket således om manipulasjonen av kunnskapsnivå hadde fungert som tiltenkt.

Undersøkelsen ble utformet på bakgrunn av objektiv oppgavekompleksitet. Imidlertid vil subjektiv oppgavekompleksitet også spille en rolle i den enkeltes vurdering av hvilken av oppgavene som er mest kompleks. Vi inkluderte derfor et spørsmål i undersøkelsen for å undersøke om manipulasjonen av oppgavekompleksitet fungerte. 58 % av deltakerne svarte at clusteranalysen var mest kompleks, 13 % svarte at linjediagram og box plot var mest kompleks, mens resterende 29 % svarte at de var like komplekse. Manipulasjonen av kompleksitet fungerte dermed bare delvis som tiltenkt. Vi fjernet imidlertid ikke respondenter fra datasettet, da graden av kompleksitet ikke var av betydning for hypotese 1 om kunnskapsnivå, i tillegg til at dette ville ført til en svært liten utvalgsstørrelse. Vi tok allikevel hensyn til problemet i hypotese 2 om kompleksitet, ved å gjennomføre analysen både med og uten de respondentene som ikke svarte at clusteranalyse var mest kompleks.

For å kontrollere for rekkefølgeeffekter, ble rekkefølgen på lav- og høykompleksoppgaven randomisert slik at noen av deltakerne fikk lavkompleksoppgaven først, mens andre fikk høykompleksoppgaven først. Vi undersøkte derfor om det eksisterte en rekkefølgeeffekt, og således om randomiseringen hadde eliminert denne effekten. Deltakerne som fikk lavkompleksoppgaven først hadde en gjennomsnittlig score på profesjonell skepsis for lavkompleksoppgaven på 3,85, mens scoren for samme oppgave for de som fikk

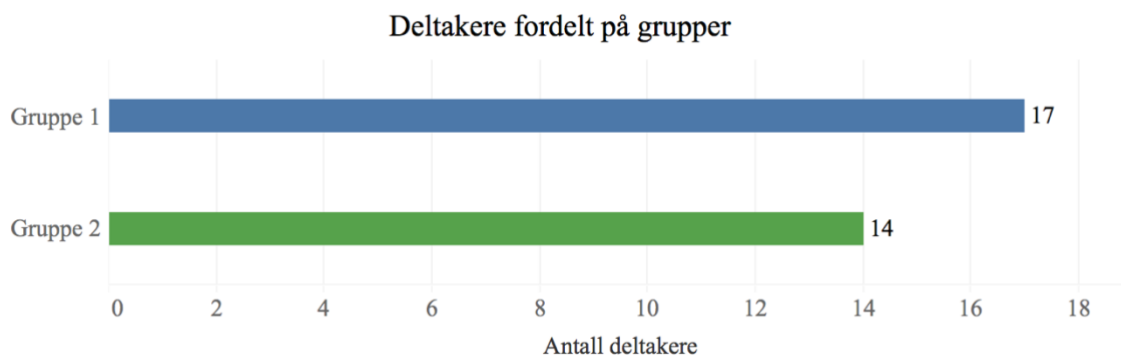
høykompleksoppgaven først var på 3,81. For høykompleksoppgaven var gjennomsnittene henholdsvis 4,23 og 4,20. Ingen av testene viste statistisk signifikans. Det er dermed ikke statistisk grunnlag for å si at det eksisterer rekkefølgeeffekter i eksperimentet.

## 5.2 Deskriptiv statistikk

Deskriptiv statistikk er fremlegging av informasjon som brukes til å beskrive datagrunnlaget analysen bygger på (Trochim & Donnelly, 2008). I dette kapittelet vil vi presentere gruppefordeling og demografiske variabler som belyser gruppenes karakteristikk.

### 5.2.1 Grupper

Etter eliminering av ufullstendige svar, samt svar som feilet manipulasjonssjekken, bestod datagrunnlaget av 31 anvendbare svar. Qualtrics allokerer deltakere tilfeldig til hver gruppe. Gruppe 1 ble kun presentert for revisjonsbevis innhentet ved dataanalyser og uten informasjon om de aktuelle analysene. Gruppe 2 fikk derimot utfyllende informasjon om dataanalyser og statistikk, samt hvordan de skulle tolke resultatene. Av figur 4 ser vi at gruppe 1 hadde 17 deltakere, og gruppe 2 hadde 14 deltakere. Til tross for en liten forskjell i antall deltakere i gruppene anså vi fordelingen til å være tilfredsstillende for analyseformål. Det er viktig å presisere at dette er en liten utvalgsstørrelse.

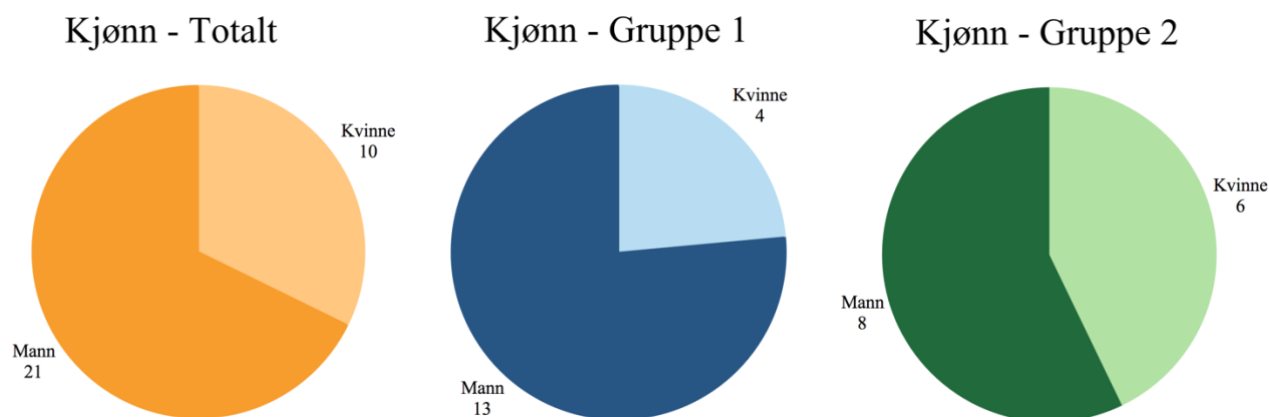


Figur 4: Deskriptiv statistikk - Grupper

## 5.2.2 Demografiske variabler

### 5.2.2.1 Kjønn

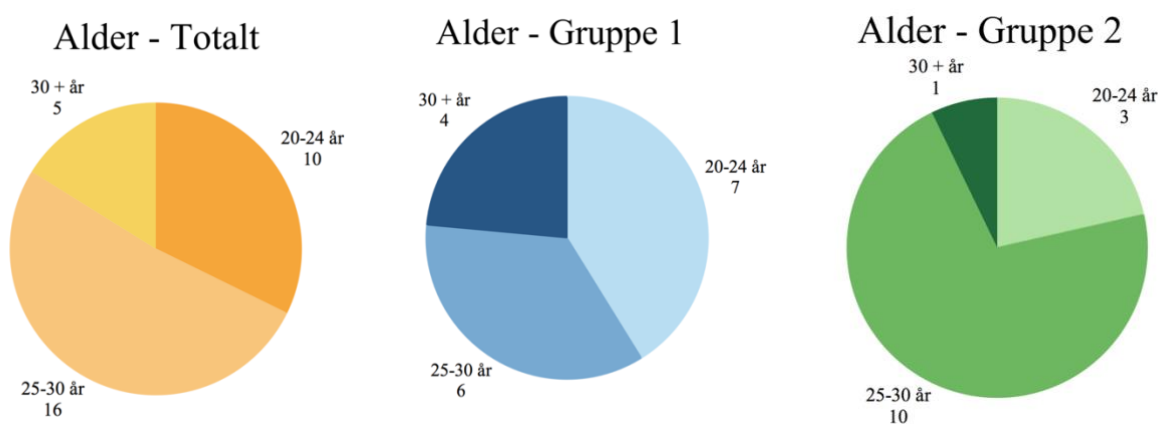
Av totalt 31 deltakere var 21 menn og 10 kvinner. Dette er noe ujevnt, hvilket gjenspeiles i gruppene. Av figur 5 fremkommer det at det er flere kvinner i gruppe 2 enn i gruppe 1. Den prosentvise fordelingen er 23,5% og 42,9% kvinner for henholdsvis gruppe 1 og gruppe 2.



Figur 5: Deskriptiv statistikk - Kjønn

### 5.2.2.2 Alder

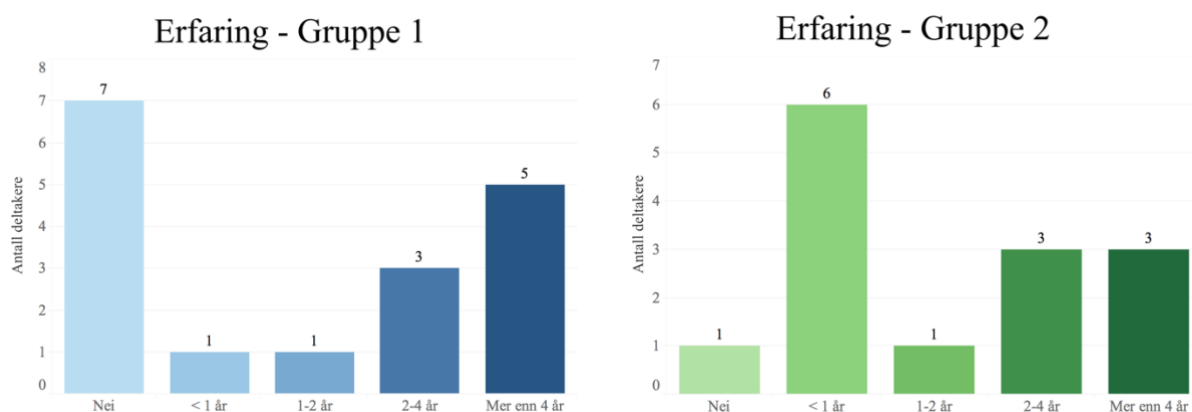
Alder anses å være en sentral demografisk variabel. Dette er fordi deltakernes verdier og holdninger kan avhenge av alder, og kan således ha betydning for resultatene våre. Majoriteten av deltakerne befinner seg i aldersgruppen 20-30 år, og kun 5 deltakere er over 30 år. Gruppene er noe skjevfordelt med hensyn til alder. Gruppe 1 har flere deltakere i både den øvre aldersgruppen (30+ år) og i den nedre aldersgruppen (20-24 år). Videre har gruppe 2 mange deltakere i aldersgruppen 25-30 år.



Figur 6: Deskriptiv statistikk - Alder

### 5.2.2.3 Erfaring

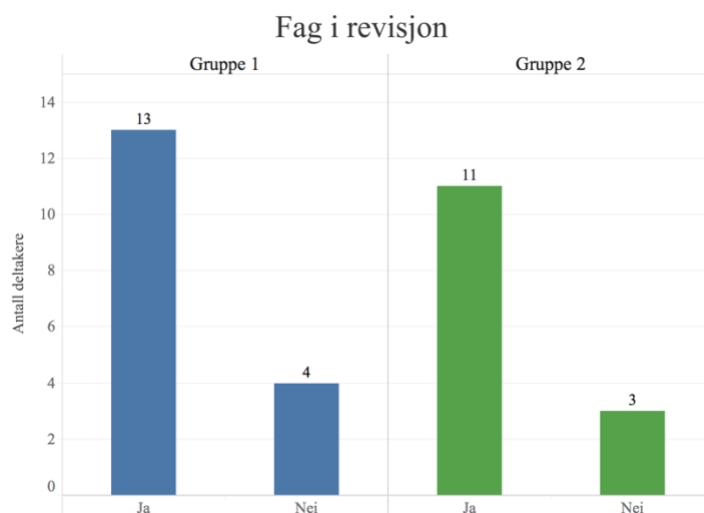
Erfaring som demografisk variabel retter seg mot antall år deltakerne har jobbet med revisjon. Dette er en viktig variabel da erfaringsnivå har evne til å påvirke nivå av profesjonell skepsis. Eksempelvis kan studenter ha en annen tilnærming til vurdering av revisjonsbevis enn deltakerne med flere år med relevant arbeidserfaring. Gruppene er svært ulike med hensyn til erfaringsnivå. Gruppe 1 har syv studenter uten arbeidserfaring, mens gruppe 2 kun har én. Videre er det vesentlig forskjell i antall deltakere med mindre enn ett års erfaring innen revisjon. Det er en ulempe at gruppene er erfaringsmessig ulike for disse erfaringsnivåene. Det er viktig å presisere at majoriteten av deltakerne med mindre enn ett års arbeidserfaring har tiltrådt i stilling som revisjonsmedarbeider denne høsten, og således kun jobbet et par måneder, eller at de er studenter som har hatt internship. Det er uvisst hvor stor forskjellen faktisk er mellom MRR-studenter og deltakere med kun noen måneder med revisjonserfaring. Gruppene er derimot relativt like med hensyn til antall deltakere med mer enn to års erfaring. Den prosentvise størrelsen for gruppe 1 og 2 er henholdsvis 47% og 42,9%.



Figur 7: Deskriptiv statistikk -Erfaring

#### 5.2.2.4 Fag innen revisjon

Denne demografiske variabelen retter seg mot oppnådde studiepoeng innen revisjonsfaget. Dette er i utgangspunktet et kontrollspørsmål for å se at kriteriene for å delta i undersøkelsen er tilfredsstillende, men det er også hensiktsmessig å se hvor mange det er som har hatt revisjonsfag. Det er totalt 24 deltakere som har hatt minst ett revisjonsfag. Videre er fordelingen mellom gruppene jevn, hvilket er en fordel.



Figur 8: Deskriptiv statistikk - Fag i revisjon

#### 5.2.3 Oppsummering deskriptiv statistikk

Totalt sett er gruppene skjevfordelt med hensyn til kjønn, alder og erfaring. Vi anser at skjevfordelingen av deltakerens alder og erfaring har størst betydning for resultatene våre. Skjevheten medfører en svekkelse av sammenliknbarheten mellom gruppene. Dette er fordi en mulig avvikende tilnærming til vurdering av revisjonsbevisene kan forekomme basert på disse demografiske variablene.

### 5.3 Grunnlag for analyse

#### 5.3.1 Kunnskapstest

Deltakernes initiale kunnskap om dataanalyser ble målt i en kunnskapstest. Vi anså det som hensiktsmessig å måle denne variabelen innledningsvis for å få en objektiv vurdering av deltakernes forkunnskaper. For å vurdere sammenliknbarheten mellom gruppene basert på kunnskapsnivå gjorde vi en t-test for å vurdere gjennomsnittlig score på kunnskapstesten. Gruppe 1 hadde et gjennomsnitt på 3,71 riktige svar, og gruppe 2 hadde et gjennomsnitt på 3,79 riktige svar. Resultatene fra t-testen gir ikke statistisk grunnlag for å påstå at gruppene er

kunnskapsmessig ulike. Dette gir dermed grunnlag for å måle effekten av kunnskapsmanipuleringen.

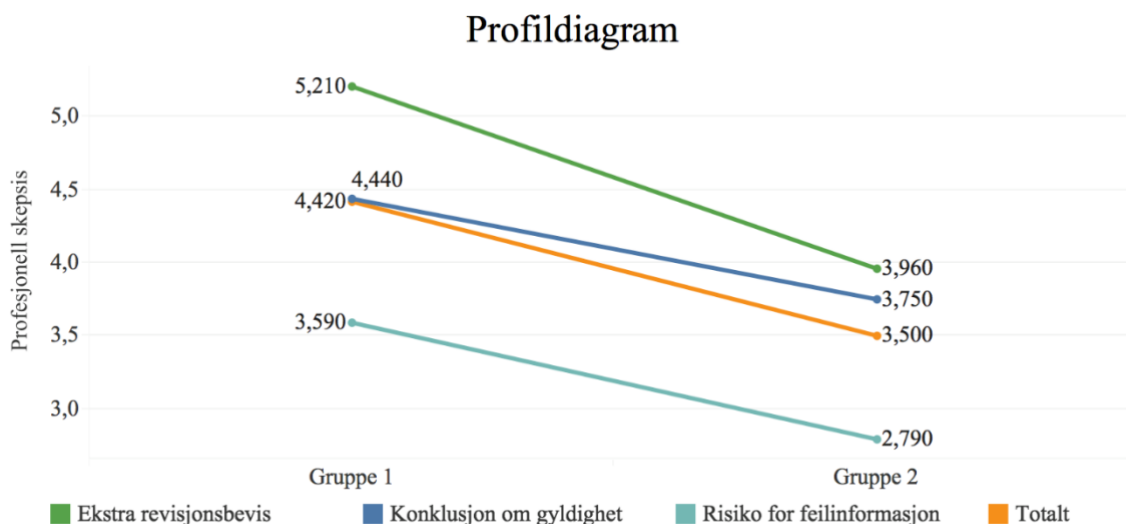
### 5.3.2 Indeks

Det ble laget en indeks for å vurdere samlet profesjonell skepsis for de ulike spørsmålene. Indeksen er satt sammen av de tre spørsmålene om risiko for feilinformasjon, konklusjon om gyldighet og behov for ekstra revisjonsbevis. Det vil således bli beregnet en gjennomsnittlig score for profesjonell skepsis for hver deltaker basert på de tre spørsmålene. Det er mulig å lage en slik indeks dersom det er tilstrekkelig intern konsistens i svarene. Intern konsistens beregnes ved Cronbach's alpha<sup>11</sup>. Indeksen for profesjonell skepsis har en Cronbach's alpha på 0,75, hvilket anses som tilstrekkelig for å lage en indeks.

## 5.4 Resultater

### 5.4.1 Resultater: Hypotese 1

Hypotese 1 tester hvorvidt kunnskapsnivå har påvirkning på profesjonell skepsis. Spørsmålene tilknyttet deltakernes vurdering av profesjonell skepsis analyseres enkeltvis og samlet (indeks).



Figur 9: Profildiagram – Hypotese 1

<sup>11</sup> Cronbach's alpha brukes for å vurdere reliabiliteten til flere spørsmål på en Likert-skala, og er således et mål på intern reliabilitet, og forteller oss om spørsmålene latent måler den samme skjulte, eller ikke-observerbare variabelen. Videre sier Cronbach's alpha i hvor stor grad spørsmålene er korrelert. Jo nærmere Cronbach's alpha er 1, jo mer korrelert er spørsmålene. Når Cronbach's alpha er høyere enn 0,7 anses intern reliabilitet og korrelasjon mellom spørsmålene til å være tilstrekkelig god, og man kan slå sammen spørsmålene til en indeks (Stephanie, 2014; UCLA Statistical Consulting Group, 2020b)



Figur 9 viser et profildiagram for hypotese 1. Deltakerne i gruppe 2 har i gjennomsnitt lavere profesjonell skepsis enn gruppe 1. Dette gjelder for alle spørsmålene enkeltvis, og totalt sett (indeksen).

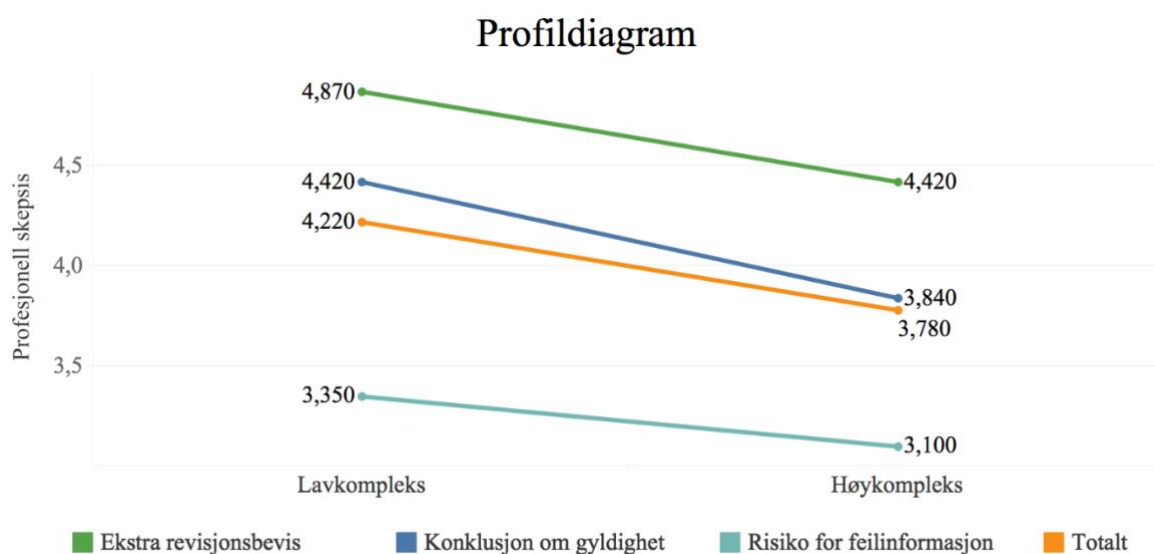
		Risiko	Konklusjon om gyldighet	Ekstra revisjonsbevis	Samlet
<b>Gruppe 1</b> Lavt kunnskapsnivå	Gjennomsnitt	3,59	4,44	5,21	4,42
	Standardavvik	1,395	1,481	1,591	1,289
<b>Gruppe 2</b> Høyt kunnskapsnivå	Gjennomsnitt	2,79	3,75	3,96	3,5
	Standardavvik	1,228	1,713	1,815	1,167
<b>Samlet</b>	P-verdi	0,021	0,094	0,006	0,005

*Tabell 1: Deskriptiv statistikk og p-verdier - Hypotese 1*

Av deskriptiv statistikk fremkommer det at gruppe 1 har et gjennomsnittlig nivå av profesjonell skepsis på 4,42 og at gruppe 2 har et gjennomsnittlig nivå på 3,5. Gruppen som ikke fikk informasjon om dataanalyser og tolkning av disse (gruppe 1) er således mer skeptisk til revisjonsbevisene enn gruppen som fikk slik informasjon (gruppe 2). For å vurdere om gjennomsnittene til gruppene er signifikant forskjellige fra hverandre foretas det en enveis ANOVA, samt en Wilcoxon test. ANOVAen viste at spørsmålene om risiko for feilinformasjon, konklusjon om gyldighet og tilstrekkelighet av revisjonsbevis har en p-verdi på henholdsvis 0,021, 0,094 og 0,006. Således er det signifikant forskjell mellom gruppene for vurderingen av risiko for feilinformasjon og tilstrekkelighet av revisjonsbevis. På 5% signifikansnivå er det ikke statistisk grunnlag for å vurdere gjennomsnittet mellom gruppene som ulike for spørsmålet om konklusjon om gyldighet. For spørsmålene samlet sett gir ANOVAen et signifikansnivå på 0,005. Gruppe 1, som ikke fikk informasjon om dataanalyser, hadde dermed signifikant høyere profesjonell skepsis enn gruppe 2, som fikk informasjon om dataanalyser. Det er også foretatt en fordelingsfri test siden det er brudd på forutsetning om normalitet for ANOVAen. Resultatene fra en Wilcoxon test gir en p-verdi på 0,022, noe som støtter resultatene fra ANOVAen. Konklusjonen for hypotese 1 er at vi kan forkaste  $H_0$ , og det er statistisk grunnlag for å påstå at økt kunnskapsnivå reduserer profesjonell skepsis.

## 5.4.2 Resultater: Hypotese 2

Hypotese 2 tester hvorvidt kompleksitet i dataanalysen påvirker profesjonell skepsis.



Figur 10: Profildiagram - Hypotese 2

Figur 10 viser et profildiagram for hypotese 2. Deltakerne hadde i gjennomsnitt lavere nivå av profesjonell skepsis for høykompleksitetsoppgaven, enn for lavkompleksitetsoppgaven. Dette gjelder både for spørsmålene enkeltvis og totalt sett (indeksen).

		Risiko	Konklusjon om gyldighet	Ekstra revisjonsbevis	Samlet
<b>Lavkompleks</b> (Grafer, box plot)	Gjennomsnitt	3,35	4,42	4,87	4,22
	Standardavvik	1,279	1,689	1,765	1,354
<b>Høykompleks</b> (Clusteranalyse)	Gjennomsnitt	3,10	3,84	4,42	3,78
	Standardavvik	1,469	1,508	1,822	1,246
<b>Samlet</b>	P-verdi	0,464	0,158	0,326	0,198

Tabell 2: Deskriptiv statistikk og p-verdier - Hypotese 2

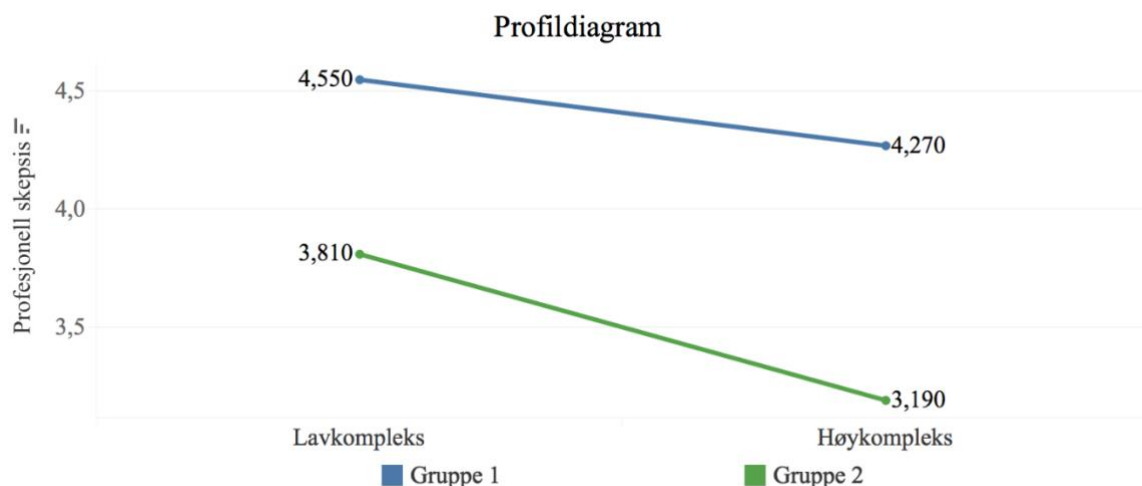
Gjennomsnittlig nivå av profesjonell skepsis for lavkompleksitetsoppgaven og høykompleksitetsoppgaven er henholdsvis 4,22 og 3,78. Dette betyr at nivå av profesjonell skepsis for høykompleksitetsoppgaven er lavere enn nivået for lavkompleksitetsoppgaven. Det gjøres en ANOVA for å vurdere om gruppene er signifikant forskjellige fra hverandre. For

spørsmålene enkeltvis gir testen p-verdier på 0,464, 0,158 og 0,326 for henholdsvis risiko for feilinformasjon, konklusjon om gyldighet og tilstrekkelighet av revisjonsbevis. For spørsmålene som helhet gir testen en p-verdi på 0,198. Det foretas også en fordelingsfri test da det er brudd på forutsetning om normalitet for ANOVAen. Resultatene fra en Wilcoxon test gir en p-verdi på 0,052, noe som støtter resultatene fra ANOVAen. Konklusjon for hypotese 2 er at vi beholder  $H_0$ , som betyr at revisors profesjonelle skepsis er lik for både lav- og høykompleksoppgaver.

Det er også utført en Wilcoxon test kun for de deltakerne som har svart at clusteranalysen var mest kompleks<sup>12</sup>. Gjennomsnittlig profesjonell skepsis er 4,05 og 3,72 for henholdsvis lavkompleksoppgaven og høykompleksoppgaven. Testen gir en p-verdi på 0,196, og det er ikke statistisk grunnlag for å forkaste  $H_0$ . Dette støtter resultatene fra hovedanalysen.

### 5.4.3 Resultater: Hypotese 3

Hypotese 3 tester eventuelle interaksjonseffekter mellom variablene kunnskapsnivå og oppgavekompleksitet. For å teste dette ble det utført en toveis ANOVA.



Figur 11: Profildiagram - Hypotese 3

Figur 11 viser et profildiagram tilhørende hypotese 3. Av diagrammet fremkommer det at gruppe 1 har høyere profesjonell skepsis enn gruppe 2. Dette gjelder for både lav- og høykompleksitetsoppgaven. Begge gruppene har høyere profesjonell skepsis for

<sup>12</sup> Det gjøres en Wilcoxon test på grunn av at mange svar er eliminert, og således er det svært få observasjoner. Dette medfører brudd på normalitet som gjør en fordelingsfri test mer egnet.

lavkompleksitetsoppgaven enn for høykompleksitetsoppgaven. Dette er det samme som vi ser av analysene fra hypotese 1 og hypotese 2.

		Gjennomsnitt	Standardavvik
<b>Gruppe 1</b> Lavt kunnskapsnivå	<b>Lavkompleks</b>	4,55	1,359
	<b>Høykompleks</b>	4,27	1,243
<b>Gruppe 2</b> Høyt kunnskapsnivå	<b>Lavkompleks</b>	3,81	1,279
	<b>Høykompleks</b>	3,19	0,993

*Tabell 3: Deskriptiv statistikk - Hypotese 3*

	P-verdi
Gruppe	0,005
Kompleksitet	0,161
Gruppe*kompleksitet (interaksjonseffekt)	0,587

*Tabell 4: P-verdier - Hypotese 3*

Av tabellene kan en se at gjennomsnittlig nivå for profesjonell skepsis for lavkompleksoppgaven og høykompleksoppgaven i gruppe 1 er henholdsvis 4,55 og 4,27. Videre er nivået for lavkompleksoppgaven og høykompleksoppgaven i gruppe 2 henholdsvis 3,81 og 3,19. Resultatene viser en p-verdi for interaksjonseffekt på 0,587. Vi beholder derfor  $H_0$ , hvilket vil si at det er ikke statistisk grunnlag for å si at det eksisterer interaksjonseffekter.

## **6. Diskusjon og konklusjon**

I dette kapittelet vil resultatene fra analysen bli diskutert. Videre vil vi belyse studiens betydning for revisjonspraksis, begrensninger og forslag til videre forskning.

### **6.1 Diskusjon av hypoteser**

#### **6.1.1 Diskusjon: Hypotese 1**

Resultatene fra hypotese 1 gir statistisk grunnlag for å påstå at kunnskap om dataanalyser og tolkning av disse har betydning for profesjonell skepsis. Resultatene viser at økt kunnskap reduserer profesjonell skepsis, og effekten er signifikant på 5% nivå. Resultatene våre sammenfaller med forskningen til Payne og Ramsay (2005) og Robertson (2010), men er motstridende med teorien om teknologisk dominans og Holmstroms (2020) studie. For spørsmålene enkeltvis observeres det en statistisk signifikant effekt for ønsket om å innhente ekstra revisjonsbevis og/eller forklaring fra klient og risiko for at det foreligger vesentlig feilinformasjon. Derimot finner en ikke tilsvarende statistisk signifikant effekt for konklusjon om gyldighet.

En årsak til at gruppen med lavt kunnskapsnivå har høyere grad av profesjonell skepsis kan være at de har manglende forståelse for dataanalyser, noe som gjør at de ikke vet hvordan de skal vurdere revisjonsbeviset. Hvis de ikke forstår prosessen fra input til output kan dette medføre økt profesjonell skepsis ved at sammenhengen mellom analysen og påstanden som blir testet er uklar. Således stoler ikke deltakerne blindt på beslutningshjelpemiddelet når de ikke har tilstrekkelig forståelse for hvordan det fungerer, slik TTD anfører. I revisjonsbeviset som har anvendt clusteranalyse blir deltakerne presentert for deskriptiv statistikk som oppsummerer analysen. P-verdien og F-verdien forteller deltakerne at det er en god statistisk modell som egner seg for det underliggende datasettet. Derimot vil denne informasjonen være verdiløs for deltakerne dersom de ikke vet hva parameterne betyr. Effekten av lav kunnskap vil også gjelde for lavkompleksoppgaven. Dersom deltakerne ikke vet hvordan analyseverktøyene fungerer, og ikke får informasjon om dette, vil de ha økt skepsisnivå siden de ikke ser sammenhengen mellom den foretatte analysen og konklusjonen i revisjonsbeviset.

Videre vil gruppen med høyt kunnskapsnivå ha større tiltro til bruken av dataanalyser i revisjonen enn de med lavt kunnskapsnivå, da de ser hvordan verktøyene faktisk kan brukes i en revisjonssammenheng. Dette tydeliggjør anvendelsesområdet i analysen, slik at deltakerne

kan se sammenhengen mellom dataanalysen og revisjonsbeviset, og således er bedre rustet til å konkludere. Større tiltro til bruken av dataanalyser vil kunne føre til redusert profesjonell skepsis.

Når det gjelder spørsmålet om risiko for feilinformasjon kan en mulig forklaring på signifikante resultater være at økt kunnskap gjør deltakerne i stand til å evaluere risikobildet tydeligere. Dersom de har forståelse for hvordan analysen er utført, og av den grunn har større tiltro til output, er det naturlig at de også mener at det er lavere risiko for at det foreligger feilinformasjon. Da de i tillegg får oppgitt en positiv konklusjon, og at det ikke foreligger feilinformasjon basert på innhentet bevis, vil informasjonen om dataanalyser og tolkning av disse fungere som en dobbel bekreftelse. Dette er fordi de kan sjekke resultatene av analysen opp mot den vedlagte informasjonen. En slik dobbelt bekreftelse kan bidra til å redusere profesjonell skepsis. Gruppen som ikke får informasjon om dataanalyser vil derimot ikke ha en slik bekreftende effekt, siden de ikke kan sjekke resultatene av revisjonshandlingen mot vedlagt informasjon.

Spørsmålet vedrørende nødvendighet for ekstra revisjonsbevis har også et statistisk signifikant resultat. Dette kan ses i sammenheng med at manglende forståelse gjør at revisor vil innhente mer revisjonsbevis for å i større grad kunne underbygge konklusjonen sin. Det kan dog også forklares med at deltakerne mener, uavhengig av om de stoler på revisjonshandlingene som allerede er utført eller ikke, at det er behov for ytterligere bevis for å understøtte konklusjonen. Dermed vil det fremstå som at deltakeren er mer skeptisk til revisjonsbevisene, til tross for at vedkommende egentlig stoler på det underliggende beviset. Dette gjelder uavhengig av om de har fått økt kunnskap eller ikke. Effekten kan forklares ved at det dreier seg om revisjon av inntekter, der revisor skal ta utgangspunkt i at det foreligger feilinformasjon og således skal være mer oppmerksom.

Profesjonell skepsis kan, som nevnt i teoridelen, både være et stabilt personlighetstrekk og en kontekstavhengig tilstand (Hurtt, 2010). Det vil si at avhengig av konteksten kan et individ ha lite profesjonell skepsis når det er grunn til det, og høy profesjonell skepsis når det er grunn til det. Informasjonen gitt til gruppe 2 i undersøkelsen vil være med på å øke deltakernes forståelse for dataanalysene som er brukt. Dette kan medføre en dobbel bekreftelse på at konklusjonen er korrekt, og dermed redusere profesjonell skepsis. Konteksten vil dermed ikke tilsi at det er

grunn for å ha høy profesjonell skepsis. Dette kan også være grunnen til at forskning finner avvikende resultater på sammenhengen mellom kunnskap og profesjonell skepsis.

Videre har vi evaluert tidsbruken til deltakerne i begge grupper. I gjennomsnitt brukte gruppe 1 og gruppe 2 henholdsvis 16,47 minutter og 18,52 minutter. En enveis ANOVA gir ikke statistisk grunnlag for å si at gjennomsnittlig tidsbruk mellom gruppene er ulikt (uteliggere ble fjernet fra analysen<sup>13</sup>). Dette kan ha betydning for resultatene i vårt eksperiment, da informasjon om dataanalyser kun vil ha effekt dersom deltakerne leser informasjonen nøye. Ut fra gjennomsnittlig tidsbruk har deltakerne som fikk informasjon om dataanalyser kun brukt to minutter på å lese gjennom denne informasjonen, noe som ikke anses å være tilstrekkelig.

### **6.1.2 Diskusjon: Hypotese 2**

Hypotese 2 anfører at økt oppgavekompleksitet fører til redusert profesjonell skepsis. Resultatene fra analysen viste i samsvar med dette at gjennomsnittlig nivå av profesjonell skepsis er lavere for høykompleksitetsoppgaven enn for lavkompleksitetsoppgaven. Testresultatene gir imidlertid ikke grunnlag for å forkaste nullhypotesen om likt gjennomsnitt mellom gruppene, og det er dermed ikke statistisk grunnlag for å si at økt kompleksitet reduserer profesjonell skepsis. Det kan være flere grunner til at det ikke er statistisk signifikante resultater. Først og fremst kan det være at økt kompleksitet faktisk ikke fører til en endring i profesjonell skepsis. Dette er dog ikke i samsvar med teorien om teknologisk dominans som sier at økt kompleksitet leder til redusert skepsis.

Videre kan for lav utvalgsstørrelse være en grunn til at en ikke får et resultat som er statistisk signifikant. Vi har en utvalgsstørrelse på 31, noe som er lite. Dersom vi hadde hatt flere deltakere i eksperimentet er det en mulighet for at vi hadde fått et statistisk signifikant resultat. Dette underbygges ved en betydelig lavere p-verdi (0,052) for en fordelingsfri test (Wilcoxon test), noe som kan tyde på at brudd på normalitet er av betydning for resultatene av en ANOVA.

En tredje grunn kan være at manipulasjonen av kompleksitet ikke har vært tydelig nok. Oppgavene i eksperimentet ble utformet kun basert på objektiv oppgavekompleksitet, da det er vanskelig å utforme oppgaver som også hensyntar subjektiv oppgavekompleksitet. Dette skyldes at subjektiv oppgavekompleksitet påvirkes av deltakernes egenskaper, erfaring og

---

<sup>13</sup> Observasjonene som oversteg øvre grense av 95% konfidensnivå mht. tidsbruk ble fjernet fra datasettet.

reaksjoner til oppgaven. Vi hensynte dog subjektiv oppgavekompleksitet til en viss grad ved å inkludere spørsmål om hvilken av oppgavene deltakerne syntes var mest kompleks. Dersom deltakere syntes lavkompleksoppgaven var mest kompleks, altså det motsatte av tiltenkt manipulasjon, vil dette skyldes subjektiv oppgavekompleksitet. Dette kan medføre lavere skeptisk holdning til lavkompleksoppgaven enn til høykompleksoppgaven, med andre ord det motsatte av vår forventning. Således har dette evne til å påvirke resultatene. Derfor gjennomførte vi analysen på nytt kun med deltakerne som mente at clusteranalysen var mest kompleks. For høykompleksoppgaven var nivå av profesjonell skepsis tilnærmet likt (3,78 for hele utvalget og 3,72 for de som mente clusteranalyse var mest kompleks), mens det ble observert en større forskjell for lavkompleksoppgaven (4,22 for hele utvalget, og 4,05 for de som mente clusteranalyse var mest kompleks). P-verdien var 0,196, altså tilnærmet likt som for ANOVAen (p-verdi på 0,198). Disse resultatene er sammenfallende med analysen for utvalget som helhet, og subjektiv oppgavekompleksitet anses således ikke å ha påvirket resultatene.

### **6.1.3 Diskusjon: Hypotese 3**

Da funnene fra hypotese 1 antyder at kunnskap reduserer profesjonell skepsis, vil  $H_{A1}$  være gjeldende. I  $H_{A1}$  forventer vi en positiv interaksjonseffekt, altså at dersom revisor har høyt kunnskapsnivå vil en økning i kompleksiteten medføre en forsterket reduksjon i profesjonell skepsis. Ut fra profildiagrammene og resultatene fra analysene, observerer vi en slik positiv interaksjonseffekt. Effekten er imidlertid ikke signifikant. En årsak til dette kan være at kunnskap har signifikant effekt på profesjonell skepsis, men at kompleksitet derimot ikke er signifikant. Vi har således en signifikant og en ikke-signifikant variabel, hvilket vanskeliggjør konklusjon om statistisk signifikante interaksjonseffekter. Elementene fra diskusjonen av hypotese 2 vedrørende liten utvalgsstørrelse og at manipulasjonen av oppgavekompleksitet ikke fungerte helt som tiltenkt vil også gjelde her.

### **6.1.4 Diskusjon: Andre funn**

Vi gjennomførte en ANCOVA<sup>14</sup> for å undersøke om kontrollvariablene hadde innvirkning på variasjonen i profesjonell skepsis, der resultatet kan ses av tabell 5. Resultatet viste at deltakernes alder var av signifikant betydning, med en p-verdi på 0,031. Det vil si at noe av

---

<sup>14</sup> En ANCOVA er en ANOVA som også tar hensyn til kontrollvariabler. En ANCOVA ser etter forskjeller i justerte gjennomsnitt og gir dermed også mulighet for å statistisk kontrollere for andre variabler enn de uavhengige variablene som en tror kan påvirke resultatet (Lærd Statistics, 2018a).



variasjonen i profesjonell skepsis blant deltakerne skyldes alderen deres. Vi undersøkte dette funnet videre ved å se isolert på variasjon i profesjonell skepsis som følge av alder. Deltakerne ble delt inn i tre grupper, der gruppe 1 var i aldersgruppen 20-24 år, gruppe 2 var i aldersgruppen 25-30 år og gruppe 3 var deltakere som var over 30 år gamle. Av gjennomsnittene, på henholdsvis 3,85, 4,06 og 4,1 kan en se at de eldste deltakerne i gjennomsnitt hadde marginalt høyere profesjonell skepsis enn resten. Det var imidlertid ingen signifikante funn i analysen. Siden vi i ANCOVAen tester mange kontrollvariabler, vil variablene alltid kunne forklare noe av variasjonen, og dette kan således være grunnen til at vi fikk et signifikant funn for alder.

Kontrollvariabel	P-verdi
Erfaring	0,464
Alder	0,031
Kjønn	0,509
Scroe HPSS	0,871
Fag innen revisjon	0,701
Erfaring med revisjon av inntekter	0,559
Erfaring med review	0,095

*Tabell 5: P-verdier for kontrollvariabler*

Av den deskriptive statistikken (se figur 7) kunne en se at erfaringsnivået til deltakerne er svært skjevfordelt. Vi undersøkte derfor om erfaringsnivå var av betydning for respondentenes profesjonelle skepsis. Deltakerne ble delt inn i tre grupper, der gruppe 1 ikke hadde erfaring, gruppe 2 hadde litt erfaring (fra 1 måned til 2 år) og gruppe 3 hadde mer erfaring (over 2 år). Analysen viste at de mest erfarne hadde høyere profesjonell skepsis enn de andre, dog var det kun statistisk signifikant forskjell mellom gruppe 2 og gruppe 3. Forskning viser at økt erfaring medfører økt kunnskap (Libby & Luft, 1993), men i vårt eksperiment vil erfaring være rettet mot revisjon og ikke spesifikk kunnskap om dataanalyser. De mer erfarne revisorene er dermed generelt mer skeptiske, uten at dette nødvendigvis er av betydning for skepsis i sammenheng med dataanalyser.

Vi kontrollerte også for, ved hjelp av Robinson et al. (2018) sin modifiserte HPSS, deltakernes skeptiske tilstand. Basert på deres 12 kontekststilhengende spørsmål ble det lagd en indeks.

Indeksen har en Cronbach's alfa på 0,785, hvilket anses som tilstrekkelig for å lage en indeks. Vi gjennomførte en ANOVA for å se om det var forskjeller i skeptisk tilstand mellom gruppene med høy og lav kunnskap<sup>15</sup>. Analysen viste gjennomsnitt for de to gruppene på henholdsvis 52,88 og 51,71, med en p-verdi på 0,739. Forskjellen mellom de to gruppene er dermed ikke statistisk signifikant, og analysen viser således at det ikke er grunn til å si at det ikke er likt nivå av skeptisk tilstand mellom gruppene.

Gruppe	Gjennomsnittlig score HPSS
Lavt kunnskapsnivå	51,71
Høyt kunnskapsnivå	52,88

Tabell 6: Gjennomsnittlig score HPSS fordelt på lavt og høyt kunnskapsnivå

Videre ble deltakerne delt inn i grupper for høy skeptisk tilstand og lav skeptisk tilstand basert på om de hadde skeptisk tilstand over eller under gjennomsnittet. Det ble deretter utført en ANOVA med samlet profesjonell skepsis for revisjonshandlingene som avhengig variabel. Denne viste at gruppen med høy skeptisk tilstand i gjennomsnitt hadde høyere profesjonell skepsis til revisjonshandlingene enn gruppen med lav skeptisk tilstand. Dette viser således at deltakere som hadde en høy skeptisk holdning etter Nolder og Kadous (2018) sitt rammeverk, også hadde høy skeptisk tilstand etter den modifiserte HPSS. Effekten var imidlertid ikke signifikant.

Gruppe	Gjennomsnittlig score profesjonell skepsis (Nolder og Kadous)
Lav skeptisk tilstand	3,78
Høy skeptisk tilstand	4,23

Tabell 7: Gjennomsnittlig score på profesjonell skepsis fordelt på lav og høy skeptisk tilstand

## 6.2 Betydning for revisjonspraksis

Vår studie ilegger oppmerksomhet til de teknologiske endringene revisjonspraksisen står overfor. Resultatene våre indikerer at økt oppgavekompleksitet reduserer profesjonell skepsis hos revisor, til tross for at denne effekten ikke er statistisk signifikant. Dette er allikevel en

<sup>15</sup> Tilsvarende analyse ble ikke gjennomført for lav- og høykompleksitetsoppgaven da deltakerne her er sin egen kontrollperson.

viktig observasjon, da høykomplekse dataanalyser enda ikke er utbredt hos revisjonsselskapene i Norge. Vår studie bidrar således til bevisstgjøring av en sentral sammenheng, noe som kan ha stor betydning når komplekse og avanserte metoder blir en stadig større del av arbeidshverdagen til fremtidens revisor. En slik bevisstgjøring kan bidra til å opprettholde et høyt skepsisnivå til revisjonsbevis innhentet ved høykomplekse teknikker.

Videre indikerer funnene våre at økt kunnskap reduserer revisors profesjonelle skepsis. Dette belyser en viktig sammenheng i revisjon. Økt kunnskap gjør revisor i stand til å evaluere revisjonsbevis mer korrekt, hvilket kan bidra til at revisjonsselskapene i større grad tør å ta i bruk dataanalyser. Imidlertid viser tidligere studier motstridende resultater på området, og det er derfor nødvendig med mer forskning som belyser hvordan revisors profesjonelle skepsis påvirkes ved bruk av dataanalyser i revisjon.

### **6.3 Begrensninger**

Denne studien har begrensninger det er nødvendig å belyse. Et eksperimentelt forskningsdesign har en iboende begrensning ved at deltakerne er i en ikke-reell situasjon. Dette medfører at deltakerne ikke har eierskap til revisjonsbeviset de blir presentert for, og det er naturlig at de ikke har den samme ansvarsfølelsen som ved utførelse av faktiske arbeidsoppgaver. En slik effekt gjør seg spesielt gjeldende i revisjonsstudier, da det foreligger betydelige negative konsekvenser dersom man kommer frem til feil konklusjon basert på innhentet bevis. Effekten av konsekvenstenkning vil ikke eksistere på likt nivå i en eksperimentell undersøkelse som på et reelt revisjonsoppdrag, og har således evne til å påvirke resultatene.

Studien blir videre begrenset av eksperimentets omfang. I eksperimentelle undersøkelser vil deltakerne miste konsentrasjon og vilje dersom undersøkelsen tar lang tid, og det er således ikke mulig å utforme et eksperiment som er like omfattende som i en reell arbeidssituasjon. På et reelt revisjonsoppdrag ville revisor gjort flere revisjonshandlinger for å understøtte sin konklusjon. Deltakerne tar stilling til et revisjonsbevis som omhandler inntekter, og det er naturlig at for en så stor og vesentlig regnskapspost ville revisor gjort et mer omfattende arbeid. Et annet aspekt ved eksperimentets omfang er tilgjengeligheten av relevant informasjon. Kunnskap om den reviderte enhetens omgivelser er sentralt ved utførelse av revisjonshandlinger, og deltakerne har svært begrenset informasjon vedrørende dette i

eksperimentet. Dette kan påvirke deltakernes oppfatning av bransjespesifikke risikoer som er relevante for den aktuelle regnskapsposten.

Revisjonsselskapene har utviklet egne programvarer til analyseformål. Dataanalysene som benyttes i eksperimentet er utformet ved bruk av en annen programvare, og deltakerne med arbeidserfaring vil dermed ikke være kjent med output fra denne programvaren. Dette fører til et gap mellom eksperimentsituasjonen og en arbeidssituasjon. Videre er dataanalysene i eksperimentet valgt på grunnlag av relevante studier som kartlegger hvilke former for dataanalyser som det er mulig å anvende i revisjon. Basert på Eilifsen et al. (2020) sin studie er avanserte dataanalyser lite brukt. Dette kan medføre at deltakerne ikke ser relevansen av clusteranalyser i revisjon, og at graden av profesjonell skepsis kan ha blitt påvirket av dette.

Vår studie baserer seg på svar fra 31 respondenter. Dette anses som en liten utvalgsstørrelse, og er dermed lite egnet til å trekke konklusjoner for populasjonen som helhet. Få respondenter medfører videre statistiske begrensninger. En forutsetning for ANOVA er normalfordelte observasjoner. Spørsmålene om risiko for feilinformasjon, konklusjon om gyldighet og innhenting av ekstra revisjonsbevis tilfredsstillende ikke forutsetningen om normalitet. Dermed må resultatene tolkes med varsomhet. Når man legger sammen spørsmålene til en samlet indeks er dataene normalfordelt, men bare så vidt<sup>16</sup> (p-verdi på Shapiro-Wilk test på 0,165). Problemet med få respondenter forplanter seg videre til demografiske forskjeller mellom gruppene. Det foreligger en skjevfordeling av respondenter med hensyn til kjønn, alder og erfaring innen revisjon. Av den deskriptive statistikken i kapittel 5.2.2.3 fremkommer det eksempelvis at gruppe 2 består av deltakere med generelt høyere erfaringsnivå. Siden deltakere med ulikt erfaringsnivå vil ha ulike tilnærminger til oppgavene vil også resultatene påvirkes av dette. Det er derfor vanskelig å vurdere gruppene som sammenliknbare, hvilket utgjør en stor begrensning. Dersom undersøkelsen hadde fått flere respondenter ville gruppene sannsynligvis blitt likere rent demografisk sett.

## 6.4 Videre forskning

Til vår undersøkelse fikk vi deltakere nok til en utvalgsstørrelse på 31, noe som er lavt. Dette medførte at fordelingen av deltakere til de to gruppene ble ujevn, spesielt når det gjelder

---

<sup>16</sup> Shapiro-Wilk test anvendes for å teste forutsetningen om normalitet for en ANOVA. Denne testen ble valgt over andre tester for normalfordeling da den er velegnet for liten utvalgsstørrelse. Dersom p-verdien for testen overstiger 0,05 er det ikke statistisk grunnlag for å påstå at dataene avviker betydelig fra normalfordelingen (Lærd Statistics, 2018b).

deltakernes erfaringsnivå og alder. Det kunne derfor vært nyttig å gjennomføre et lignende eksperiment, men med en større utvalgsstørrelse. Ved å ha større utvalg er det i tillegg mulig at en ser en signifikant effekt når det kommer til oppgavekompleksitet.

Videre undersøkte vi i vårt eksperiment kun to av komponentene i teorien om teknologisk dominans. Det kunne derfor vært interessant å se hvordan de to andre komponentene, bekjentskap med beslutningshjelpemidler og kognitivt samsvar mellom bruker og IDA, virker inn på revisors profesjonelle skepsis. Det vil også være interessant å undersøke mer rundt kunnskapskomponenten og dens betydning for profesjonell skepsis, da vi fikk resultater som ikke var i overensstemmelse med teorien om teknologisk dominans.

Det er en økende bruk av ADA i revisjon, og det vil derfor være interessant å forske videre på implikasjoner knyttet til denne bruken. Det kunne eksempelvis vært hensiktsmessig å se på effektivitetsgevinster ved bruk av ADA, og herunder hvordan revisors kunnskapsnivå påvirker effektiviteten. Videre kunne det vært nyttig å se isolert på høykomplekse dataanalyser, da disse kan utføre grundige analyser og analysere større datavolumer, noe som vil kreve et høyere kunnskapsnivå hos revisor.

## Litteraturliste

- Abdolmohammadi, M., & Wright, A. (1987, Januar). An Examination of the Effects of Experience and Task Complexity on Audit Judgments. *The Accounting Review*, 62, ss. 1-13.
- Adrain, A. (2017, Mai 29). *Audit data analytics: Rising to the challenge*. Hentet fra ICAS: <https://www.icas.com/professional-resources/audit-and-assurance/audit-data-analytics-rising-to-the-challenge>
- AICPA. (2017, Desember 5). *Audit Data Analytics (ADAs) Can Transform Audits; New AICPA Guide Will Help Auditors Apply ADA Techniques*. Hentet fra AICPA: <https://www.aicpa.org/press/pressreleases/2017/audit-data-analytics-new-aicpa-guide-will-help-auditors-apply-ada-techniques.html>
- Alawadhi, A. (2015). *The application of data visualization in auditing (Doktoravhandling)*. New Jersey: Graduate School- Newark Rutgers.
- Alissa, W., Capkun, V., Jeanjean, T., & Suca, N. (2014). An empirical investigation of the impact of audit and auditor characteristics on auditor performance. *Accounting, Organization and Society*, 39, ss. 495-510.
- Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017, Februar 1). Big Data and Analytics in the Modern Audit Engagement: Research Needs. *Auditing A Journal of Practice & Theory*, 36(4), ss. 1-27.
- Arnold, V., & Sutton, S. (1998). The theory of technology dominance: understanding the impact of intelligent decision aids on decision makers' judgments. *Advances in accounting behavioral research* 1(3), ss. 175-194.
- Association of Chartered Certified Accountants. (u.d.). *Data analytics and the auditor*. Hentet fra ACCA Global <https://www.accaglobal.com/gb/en/student/exam-support-resources/professional-exams-study-resources/p7/technical-articles/data-analytics.html>
- Audimation. (2019, September 11). *Planning & Scoping Your Audit with Data Visualization*. Hentet fra Audimation: <https://www.audimation.com/planning-scoping-your-audit-with-data-visualization/>
- Aurstad, T. (2017). Revisjonsbevis i en digital hverdag. *Revisjon og Regnskap*, 7, ss. 26-27.
- BDO (2019). *Åpenhetsrapport 2019*. Hentet fra BDO: <https://www.bdo.no/getattachment/Om-BDO/Apenhetsrapport/2019/BDO-Apenhetsrapport-2019.pdf.aspx?lang=nb-NO>
- BDO. (2016, Juni). *Data analytics & enhancing audit quality*. Hentet fra BDO Insight: <https://www.bdo.com/insights/assurance/corporate-governance/data-analytics-enhancing-audit-quality>
- Birkeland, L. (2004). Revisors arbeid og signering. *Revisjon og Regnskap*, 6.
- Birnbaum, M. H & Stegner, S. E. (1979, Januar). Source Credibility in Social Judgment: Bias, Expertise, and the Judge's Point of View. *Journal of Personality and Social Psychology* 37(1), ss.48-74.
- Bonhome, O. D., Gjymshana, E., Jans, M., Kroes, D., Marissen, M., Simpelaere, F., Verachtert, S. (2018, Januar 15). *Data Analytics: The Future of Audit*. Hentet fra Institut des Réviseurs d'Entreprises: <https://www.ibr-ire.be/fr/actualites/news-detail/data-analytics-the-future-of-audit>
- Bonner, S. E. (1994). A Model of the Effects of Audit Task Complexity. *Accounting, Organizations and Society*, 19, ss. 213-234.
- Botez, D., & Vasile, A. (2018). Recent Challenge for Auditors: Using Data Analytics in the Audit of the Financial Statements. *BRAIN. Broad Research In Artificial Intelligence And Neuroscience*, 9(4), ss. 61-71.
- Braut, G. S. (2014, Oktober 27). *Store Norske Leksikon*. Hentet fra [https://snl.no/type\\_I-feil](https://snl.no/type_I-feil)

- Brown-Liburud, H., & Vasarhelyi, M. A. (2015, Desember 1). Big Data and Audit Evidence. *Journal of Emerging Technologies in Accounting* 12(1), ss. 1-16.
- Campbell, D. J. (1988, Oktober). Task Complexity: A Review and Analysis. *The Academy of Management Review*, 13, ss. 40-52.
- Commerford, B. P., Dennis, S. A., Joe, J., & Wang, J. (2020, Februar 24). *Man versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence*. Hentet fra SSRN: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3422591](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3422591)
- Davenport, T. H. (2016). *The power of advanced audit analytics*. Hentet fra Deloitte: <https://www2.deloitte.com/us/en/pages/deloitte-analytics/articles/us-the-power-of-advanced-audit-analytics.html>
- Deloitte. (2020, 31 August). *2020 Åpenhetsrapport*. Hentet fra Deloitte: [https://info.deloitte.no/rs/777-LHW-455/images/%C3%85penhetsrapporten\\_2020.pdf](https://info.deloitte.no/rs/777-LHW-455/images/%C3%85penhetsrapporten_2020.pdf)
- Earley, C. (2015, September). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, 58(5), ss. 493-500.
- Eilifsen, A., Kinserdal, F., Messier Jr, W., & McKee, T. (2020, Juni 12). An Exploratory Study into the Use of Audit Data Analytics on Audit Engagements. *Accounting Horizons* 19(121).
- EY. (2015, April 1). *How big data and analytics are transforming the audit*. Hentet fra EY: [https://www.ey.com/en\\_gl/assurance/how-big-data-and-analytics-are-transforming-the-audit](https://www.ey.com/en_gl/assurance/how-big-data-and-analytics-are-transforming-the-audit)
- EY Norge. (2020, September 24). *Åpenhetsrapport 2020*. Hentet fra EY: [https://assets.ey.com/content/dam/ey-sites/ey-com/no\\_no/home-index/ey-aapenhetsrapport-2020.pdf](https://assets.ey.com/content/dam/ey-sites/ey-com/no_no/home-index/ey-aapenhetsrapport-2020.pdf)
- FASB. (1980, Mai). *Qualitative Characteristics of Accounting Information*. Hentet fra Fasb: [https://www.fasb.org/jsp/FASB/Document\\_C/DocumentPage?cid=1218220132599&acceptedDisclaimer=true](https://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1218220132599&acceptedDisclaimer=true)
- Financial Reporting Council. (2017). *Audit Quality Thematic Review: The use of data analytics in the audit of financial statements*. Hentet fra FRC: [https://www.frc.org.uk/getattachment/4fd19a18-1beb-4959-8737-ae2dca80af67/AQTR\\_Audit-Data-Analytics-Jan-2017.pdf](https://www.frc.org.uk/getattachment/4fd19a18-1beb-4959-8737-ae2dca80af67/AQTR_Audit-Data-Analytics-Jan-2017.pdf)
- Forbes & KPMG. (2015, Juni). *Audit 2020: A Focus on Change*. Hentet fra Forbes Insight: <https://images.forbes.com/forbesinsights/StudyPDFs/KPMG-AFocusOnChange-REPORT.pdf>
- Gartner. (2012). *Big Data*. Hentet fra Gartner: <https://www.gartner.com/en/information-technology/glossary/big-data>
- Hampton, C. (2005, Desember). Determinants of reliance: An empirical test of the theory of technology dominance. *International Journal of Accounting Information Systems* 6(4), ss. 217-240.
- Holmstrom, K. (2020, Juni 5). *The Effect of Opaque Audit Methods and Auditor Ownership on Reliance on Independent Expectations*. Hentet fra [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3596478](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3596478)
- Hurt, R. K. (2010, Mai). Development of a Scale to Measure Professional Skepticism. *Auditing: A Journal of Practice & Theory*, 29(1), ss. 149-171.
- IAASB. (2009a). *ISA 500 Revisjonsbevis*. New York, NY: International Federation of Accountants; Revisorforeningen
- IAASB. (2009b). *ISA 200 Overordnede mål for den uavhengige revisor og gjennomføringen av en revisjon i samsvar med de internasjonale revisjonsstandardene*. New York, NY: IFAC: International Federation of Accountants; Revisorforeningen.
- IAASB. (2017). *Exploring the Growing Use of Technology in the Audit, with a Focus on Data Analytics*. New York, NY: International Federation of Accountants.

- IDEA. (2020, Juli 4). *IAASB discusses challenges with the use and adoption of data analytics*. Hentet fra IDEA: <https://idea.caseware.com/blog/iaasb-discusses-challenges-with-the-use-and-adoption-of-data-analytics>
- Johari, R. J., & Sanusi, Z. M. (2010). *The effect of knowledge, effort and ethical orientation on audit judgement performance*. Hentet fra [https://www.researchgate.net/profile/Zuraidah\\_Mohd-Sanusi/publication/266268556\\_THE\\_EFFECT\\_OF\\_KNOWLEDGE\\_EFFORT\\_AND\\_ETHICAL\\_ORIENTATION\\_ON\\_AUDIT\\_JUDGMENT\\_PERFORMANCE/links/54bce4cd0cf24e50e940b533/THE-EFFECT-OF-KNOWLEDGE-EFFORT-AND-ETHICAL-ORIENTATION-ON-AUDIT-JUDGMENT-PERFORMANCE.pdf](https://www.researchgate.net/profile/Zuraidah_Mohd-Sanusi/publication/266268556_THE_EFFECT_OF_KNOWLEDGE_EFFORT_AND_ETHICAL_ORIENTATION_ON_AUDIT_JUDGMENT_PERFORMANCE/links/54bce4cd0cf24e50e940b533/THE-EFFECT-OF-KNOWLEDGE-EFFORT-AND-ETHICAL-ORIENTATION-ON-AUDIT-JUDGMENT-PERFORMANCE.pdf)
- Katz, D. (2014, April 15). *Regulators Fear Big Data Threatens Audit Quality*. Hentet fra CFO: <https://www.cfo.com/auditing/2014/04/regulators-fear-big-data-threatens-audit-quality/>
- Kenton, W. (2020, Mars 19). *Population Definition*. Hentet fra Investopedia: <https://www.investopedia.com/terms/p/population.asp>
- Kerlinger, F. N. (1973). *Foundations of Behavioral Research, 2nd edition*. New York: Holt, Rinehart & Winston.
- Kinserdal, F. (2017). NHH skal forske på digitalisering i revisjonsbransjen. *Magma*, 6, ss. 79-86.
- Kleive, K. (2018). Bedre revisjon med dataanalyser. *Revisjon og Regnskap* 8, ss. 16-18.
- Kleive, K. (2019). Datavisualisering. *Revisjon og regnskap*, 5, ss. 30-33.
- KPMG. (2015, Februar 25). *Audit Data & Analytics: Unlocking the value of audit*. Hentet fra KPMG: <https://home.kpmg/xx/en/home/insights/2015/02/audit-data-analytics-unlocking-value-of-audit.html>
- KPMG. (2016, April 8). *Fremtidens revisorer må ha digital kompetanse*. Hentet fra KPMG: <https://home.kpmg/no/nb/home/nyheter-og-innsikt/2016/04/fremtidens-revisorer-ma-ha-digital-kompetanse.html>
- KPMG Norge. (2019, Desember). *Åpenhetsrapport 2019*. Hentet fra KPMG: [https://home.kpmg/content/dam/kpmg/no/pdf/2019/12/Aapenhetsrapport\\_KPMG\\_Norge\\_2019.pdf](https://home.kpmg/content/dam/kpmg/no/pdf/2019/12/Aapenhetsrapport_KPMG_Norge_2019.pdf)
- Lærd Statistics. (2018a). *One-way ANCOVA in SPSS statistics*. Hentet fra Lærd Statistics: <https://statistics.laerd.com/spss-tutorials/ancova-using-spss-statistics.php>
- Lærd Statistics. (2018b). *Testing for Normality using SPSS Statistics*. Hentet fra Lærd Statistics: <https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php>
- Libby, R. (1981). *Accounting and Human Information Processing: Theory and Applications*. Englewood Cliffs, N.J: Prentice-Hall.
- Libby, R., & Frederick, D. M. (1990). Experience and the Ability to Explain Audit Findings. *Journal of Accounting Research*, 28(2), ss. 348-367.
- Libby, R., & Luft, J. (1993, Juli). Determinants of Judgment Performance in Accounting Settings: Ability, Knowledge, Motivation and Environment. *Accounting, Organizations and Society*, 18(5), ss. 425-450.
- Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental Research in Financial Accounting. *Accounting, Organization and Society*, 27, ss. 775-810.
- Mascha, M. F., & Smedley, G. (2007, Juni 7). Can computerized decision aids do “damage”? A case for tailoring feedback and task complexity based on task experience. *International Journal of Accounting Information Systems*, 8(2), ss. 73-91.
- McDermott, R. (2011). Chapter 3: Internal and External Validity. I R. McDermott, N. J. Druckman, D. P. Greene, J. H. Kuklinski, & A. Lupia, *Cambridge Handbook of Experimental Political Science* (ss. 27-40). New York: Cambridge University Press.



- Murphy, M. L., & Tysiac, K. (2015, April 13). *Data analytics helps auditors gain deep insight*. Hentet fra Journal of Accountancy: <https://www.journalofaccountancy.com/issues/2015/apr/data-analytics-for-auditors.html>
- Nolder, C. J., & Kadous, K. (2018, Mai). Grounding the Professional Skepticism Construct in Mindset and Attitude Theory: A Way Forward. *Accounting, Organization and Society*, 67, ss. 1-14.
- Onkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009, Oktober). The Relative Influence of Advice From Human Experts and Statistical Methods on Forecast Adjustments. *Journal of Behavioral Decision Making*, 22(4), ss. 390 - 409.
- Payne, E. A., & Ramsay, R. J. (2005, April). Fraud risk assessments and auditors' professional skepticism. *Managerial Auditing Journal*, 20(3), ss. 321-330.
- Pedersen, J. S. (2016). Dataanalyse i revisjon. *Revisjon og regnskap*, 7, ss. 30-31.
- Peecher, M. E., & Solomon, I. (2001). Theory and Experimentation in Studies of Audit Judgments and Decisions: Avoiding Common Research Traps. *International Journal of Auditing*, 5 (3), ss. 193-203.
- Perry, S. (2017, Mai 22). *What is Big Data? More than volume, velocity and variety*. Hentet fra IBM Developer: <https://developer.ibm.com/technologies/analytics/blogs/what-is-big-data-more-than-volume-velocity-and-variety/>
- Price, P. C., Jhangiani, R., Chiang, I.-C. A., Leighton, D. C., & Cuttler, C. (2017). *Research Methods in Pshychology*. Pressbooks. Hentet fra <https://opentextbc.ca/researchmethods/>
- PwC. (2015, Mars 4). *Hva er Big Data, og hva betyr Big Data for deg?* Hentet fra PwC: <https://www.pwc.no/no/publikasjoner/information-management/big-data.pdf>
- PwC. (2019). *Åpenhetsrapport 2020 Innsyn i vår kvalitetssikring*. Hentet fra PwC: <https://www.pwc.no/no/om-oss/aapenhetsrapport-2019.pdf>
- Ratna, T. D., & Anisykurlillah, I. (2020, September 27). The Effect of Experience, Independence, and Gender on Auditor Professional Scepticism with Professional Ethics as Moderating. *Accounting Analysis Journal*, 9(2), ss. 138-145.
- Rayamajhi, D. B. (2019). Dataanalyser i revisjon – metodikken. *Revisjon og regnskap*, 1, ss. 15-17.
- Revisorforeningen. (u.d.). *Revisjon gir tillit*. Hentet fra Revisorforeningen: <https://revisorforeningen.no/om-revisjon/Revisjon-gir-tillit/>
- Revisorloven. (1999). *Lov om revisjon og revisorer (LOV-1999-01-15-2)*. Hentet fra <https://lovdata.no/dokument/NL/lov/1999-01-15-2>
- Robertson, J. C. (2010). The Effects of Ingratiation and Client Incentive on Auditor Judgment. *Behavioral Research in Accounting*, 2, ss. 69–86.
- Robinson, S. N., Robertson, J. C., & Curtis, M. B. (2018, Februar). Disentangling the Trait and State Components of Professional Skepticism: Specifying a Process for State Scale Development. *Auiting: A Journal of Practice & Theory*, 37 (1), ss. 215-235.
- Rooney, B. J., & Evans, A. N. (2018). *Methods in Pshycological Research, 3rd edition*. New York City: SAGE Publications.
- Sears, D. (1986). College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature. *Journal of Personality and Social Psychology*, 51(3), ss. 515-530.
- Shaub, M. K. (1996). Trust and Suspicion: The Effects of Situational and Dispositional Factors on Auditors' Trust of Clients. *Behavioral Research in Accounting*, 8, ss. 154-174.
- Stephanie. (2014, Desember 8). *Cronbach's alpha: Simple definition, use and interpretation*. Hentet fra Statistics how to: <https://www.statisticshowto.com/cronbachs-alpha-spss/>

- Suryandari, N. N., & Yuesti, A. (2017, September). Professional Scepticism and auditors ability to detect fraud based on workload and characteristics of auditors. *Scientific Research Journal*, 5(4), ss. 109-115.
- Tan, H.-T., & Kao, A. (1999, Vår). Accountability Effects on Auditors' Performance: The Influence of Knowledge, Problem-Solving Ability, and Task Complexity. *Journal of Accounting Research*, 37, ss. 209-223.
- Thiprungsi, S. (2011, Januar). Cluster analysis for anomaly detection in accounting. *The international journal of digital accounting research*, 11(17), ss. 69-84.
- Triki, A., & Weisner, M. M. (2014). Lessons from the Literature on the Theory of Technology Dominance: Possibilities for an Extended Research Framework. *Journal of Emerging Technologies in Accounting* 11(1), ss. 41-69.
- Trochim, W. M., & Donnelly, J. P. (2008). *Research Methods Knowledge Base*. Ohio: Atomic Dog/Cengage Learning.
- Trotman, K. T. (2001). Design Issues in Audit JDM Experiments. *International Journal of Accounting* 5, ss. 181-192.
- Turley, S., Humphrey, C., Samsonova-Taddei, A., Siddiqui, J., Woods, M., Basoudis, I., & Richard, C. (2016). *Skills, competencies and the sustainability of the modern audit*. Institute of Chartered Accountants Scotland and UK Financial Reporting Council. Hentet fra <https://www.icas.com/thought-leadership/research/skills,-competencies-and-the-sustainability-of-the-modern-audit>
- UCLA Statistical Consulting Group (2020a). *What is the difference between categorical, ordinal and numerical variables?* Hentet fra <https://stats.idre.ucla.edu/other/mult-pkg/whatstat/what-is-the-difference-between-categorical-ordinal-and-numerical-variables/>
- UCLA Statistical Consulting Group. (2020b). *What does Cronbach's alpha mean?* Hentet fra Statistical Consulting: <https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/>
- Van Wyk, B. (2012). *Research design and methods, Part 1*. University of the Western Cape. Hentet fra [https://d1wqtxts1xzle7.cloudfront.net/35022489/Research\\_and\\_Design\\_I.pdf?1412649196=&response-content-disposition=inline%3B+filename%3DResearch\\_and\\_Design\\_I.pdf&Expires=1608309275&Signature=ZSbfOw1vHNEXIy9QKKrfCcCe~-aoT5DGLVwXUoGN~TQDCfbbbw7YR~uD0bMJUxXqLYz1snzZLhMErQsAEGlv t2pql2u~zqENnYgjBxoirwhB3IQlenQjvV8XfdrbUU2FpSMBkY3q6Ym1DjQylFUte onqxJkJCo~2pq6sB8IwOTSIJhVnxcMgWNDYM053ZhBmFU3sXqCH4WBcdTfHz E8VQvVtj1Qc6AWiFNjH50dprjErYEtimPOGoIfF~jiFHKvjeQyzlOnVCpXeZuFv5RBHLuChhCOZxhMzsqJqsZFCkV~8oJcTOyYj40LBXQ4INrQ~-LbEavxI4g~CTx33sJTnQ\\_\\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](https://d1wqtxts1xzle7.cloudfront.net/35022489/Research_and_Design_I.pdf?1412649196=&response-content-disposition=inline%3B+filename%3DResearch_and_Design_I.pdf&Expires=1608309275&Signature=ZSbfOw1vHNEXIy9QKKrfCcCe~-aoT5DGLVwXUoGN~TQDCfbbbw7YR~uD0bMJUxXqLYz1snzZLhMErQsAEGlv t2pql2u~zqENnYgjBxoirwhB3IQlenQjvV8XfdrbUU2FpSMBkY3q6Ym1DjQylFUte onqxJkJCo~2pq6sB8IwOTSIJhVnxcMgWNDYM053ZhBmFU3sXqCH4WBcdTfHz E8VQvVtj1Qc6AWiFNjH50dprjErYEtimPOGoIfF~jiFHKvjeQyzlOnVCpXeZuFv5RBHLuChhCOZxhMzsqJqsZFCkV~8oJcTOyYj40LBXQ4INrQ~-LbEavxI4g~CTx33sJTnQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA)
- Wood, R. E. (1986). Task Complexity: Definition of the Construct. *Organizational Behavior and Human Decision Processes* 37, ss. 60-82.
- Yoon, K., Hoogduin, L., & Zhang, L. (2015, Juni 1). Big Data as Complementary Audit Evidence. *Accounting Horizons*, 29(2), ss. 431-438.

## Vedlegg 1: Spørreundersøkelse

I forbindelse med vår masteroppgave i regnskap og revisjon har vi utarbeidet denne spørreundersøkelsen. Svarene du gir er anonyme, og dataene behandles med full konfidensialitet. Det er viktig at du leser informasjonen nøye, da det ikke er mulig å gå tilbake til tidligere spørsmål. Ved gjennomføring av undersøkelsen skal du ikke diskutere oppgavene med andre. Undersøkelsen vil ta ca. 20 minutter.

Vi setter stor pris på at du tar deg tid til å gjennomføre undersøkelsen. På forhånd, tusen takk.

**Du får nå se to dashboard. Svar etter beste evne på de etterfølgende spørsmålene.**

Dashboard 1 viser visualisering av data om antall turister, skatteinntekter fra turistnæringen og utviklingen i BNP. Basert på dette, svar på spørsmålene under.

Presisering: Det er ikke nødvendig med utregninger for å komme frem til riktig svar.



### Spørsmål 1:

Hva var den gjennomsnittlige økningen i skatteinntekter fra uke til uke?

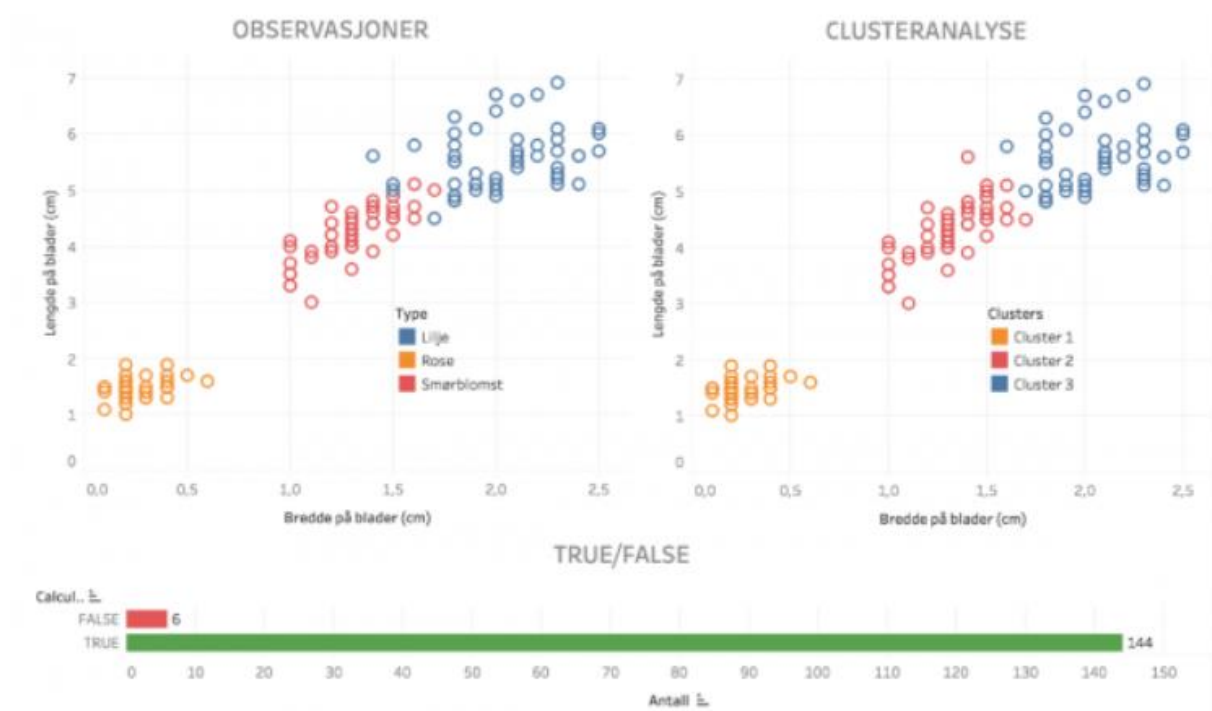
- 150,97
- 1508,01
- 413,93
- Vet ikke

## Spørsmål 2:

Basert på variasjonen i datasettet, for hvilke kvartaler er det en unaturlig sammenheng mellom antall turister og skatteinntekter fra turistnæringen?

- Q1 og Q3
- Q4 og Q1
- Q2 og Q4
- Vet ikke

Dashboard 2 er basert på et datasett som inneholder informasjon om lengden og bredden på bladene til tre ulike typer blomster; roser, smørblomst og liljer. Det er gjort en clusteranalyse basert på denne dataen. Svar på spørsmålene under.



### Summary Diagnostics

Number of Clusters: 3  
 Number of Points: 150  
 Between-group Sum of Squares: 26.686  
 Within-group Sum of Squares: 1.7051  
 Total Sum of Squares: 28.392

### Centers

Clusters	Number of Items	Sum of Lengde på blader	Sum of Brekke på blader
Cluster 1	50	1.464	0.244
Cluster 2	52	4.2962	1.325
Cluster 3	48	5.5667	2.0562
Not Clustered	0		

### Analysis of Variance:

Variable	F-statistic	p-value	Model		Error	
			Sum of Squares	DF	Sum of Squares	DF
Sum of Brekke på blader	69.2	0.0	14.18	2	15.07	147
Sum of Lengde på blader	68.96	0.0	12.5	2	13.33	147

**Spørsmål 1:**

For hvilke typer blomster er det mest sannsynlig at en observasjon er gruppert til feil blomstertype i clusteranalysen?

- Liljer og roser
- Roser og smørblomst
- Smørblomst og liljer
- Vet ikke

**Spørsmål 2:**

Hvor mange observasjoner er gruppert feil i clusteranalysen?

- 5
- 6
- 12
- 144
- 3
- Vet ikke

**Spørsmål 3:**

Vurder informasjonen i «Summary Diagnostics» og «Analysis of Variance». Er clusteranalysen en god modell?

- Ja
- Nei
- Vet ikke

I resten av undersøkelsen skal du ta stilling til utførte revisjonshandlinger, og svare på spørsmål vedrørende disse. Det er viktig at du leser den oppgitte informasjonen nøye.

***Case: Energi AS*****Bakgrunnsinformasjon**

Du skal i dette caset tenke deg at du er med på et revisjonsteam og skal revidere selskapet Energi AS. Du har stilling som manager, og har jobbet i selskapet i fem år. Revisjonsselskapet du jobber for har hatt Energi AS som kunde i fem år, og hittil har dere kun gode erfaringer med selskapet. Det er ikke avdekket vesentlig feilinformasjon tidligere år. Energi AS produserer strøm i et vannkraftverk. Bransjen Energi AS operer i kjennetegnes ved at produksjon avhenger av nedbørmengde og vannmengde i magasinene, og konsum avhenger av årstid; henholdsvis vinter, vår, sommer og høst. Alle kundene til Energi AS betaler strøm hver uke.

En på revisjonsteamet ditt, som har to års erfaring med revisjon, har revidert inntektene til Energi AS. Kollegaen din har testet påstanden gyldighet for inntektene fra en stor, vesentlig kunde, Handel AS. Det er gjennomført en analytisk handling for å vurdere inntektstransaksjonene. Du skal nå gjøre en review av kollegaen din sitt arbeid, det vil si å gjennomgå og vurdere jobben som er gjort.

For å vurdere inntektstransaksjonene har kollegaen din innhentet data om Energi AS sine totale salgstall, og salgstall fra kunden Handel AS. Dataene er hentet fra salgssystemet til Energi AS.

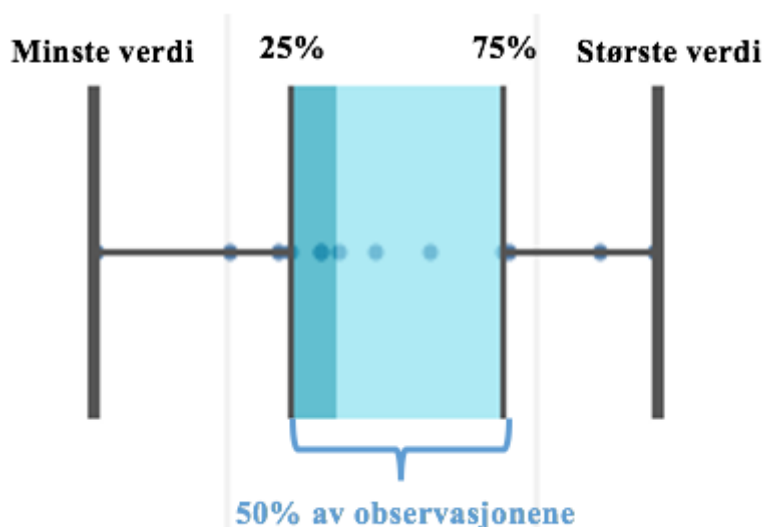
Revisor vurderer selskapets interne kontroller som gode, og salgssystemet anses således som pålitelig. Det er brukt et analyse- og visualiseringsverktøy som er vanlig i revisjonsbransjen, og verktøyet anbefales av Den norske Revisorforeningen.

**Du får nå to oppgaver. Det er viktig at du vurderer disse separat.**

## Informasjon om dataanalyser

### Box Plot

Et box plot er nyttig for å sammenlikne variasjon i ulike grupper. Det markerte området angir starten på 25% persentilen og slutten av 75% av persentilen. Det betyr at 50% av observasjonene ligger innenfor dette området. Medianen fremkommer av skillet mellom mørk/lys farge i box plotet. De ytterste strekene angir den laveste og den høyeste verdien i datasettet ditt.



Måten man tolker et box plot på er å vurdere størrelsen på de ytterste linjene, og det markerte fargelagte området. Dersom den fargelagte boksen er veldig liten, betyr det at det er liten variasjon i observasjonene dine. Likevel kan de ytterste strekene være langt unna, noe som indikerer at du har ekstremverdier som avviker fra resten av datasettet. Dersom det er betydelig ulik variasjon mellom to variabler du forventer er korrelerte, indikerer dette at forventningen din ikke holder.

### Linjediagram

Et linjediagram viser utviklingen i ulike variabler. Slike diagrammer er spesielt hensiktsmessig når man ønsker å sammenlikne ulike variabler, eller ønsker å vurdere mønstre i dataene. Slike mønstre kan være sesongvariasjoner eller trender. Dette gjør deg i stand til å identifisere motstridende mønstre, eller trender som avviker fra forventningen din. Basert på dette kan du vurdere hvilke områder som bør undersøkes nærmere. Videre er diagrammet nyttig for å identifisere enkeltobservasjoner med ekstremverdier som man burde sjekke opp.

## Oppgave 1

Du gjør en review av et arbeidspapir for å sjekke gyldighet av inntekter fra kunden Handel AS. Det er en på revisjonsteamet som har utformet arbeidspapiret ved hjelp av selskapets analyse- og visualiseringsverktøy. Ta stilling til de følgende to revisjonshandlingene som er utført, og svar på spørsmålene.

### Revisjonshandling 1

<b>Formål:</b>	Kontrollere om det er en plausibel sammenheng mellom inntekter fra kunden og volum kjøpt for de ulike sesongene.															
<b>Revisjons-handling:</b>	Vurdere variasjon innad i sesonger for å identifisere unaturlige sammenhenger.															
<b>Analyse:</b>	<p style="text-align: center;"><b>Variasjon i inntekter og strømforbruk</b></p> <table border="1"><thead><tr><th>Sesong</th><th>Inntekter fra Handel AS (K)</th><th>Volum kjøpt av Handel AS (K)</th></tr></thead><tbody><tr><td>Vinter</td><td>30 - 35</td><td>45 - 50</td></tr><tr><td>Vår</td><td>15 - 20</td><td>40 - 45</td></tr><tr><td>Sommer</td><td>10 - 15</td><td>30 - 40</td></tr><tr><td>Høst</td><td>25 - 26</td><td>45 - 50</td></tr></tbody></table>	Sesong	Inntekter fra Handel AS (K)	Volum kjøpt av Handel AS (K)	Vinter	30 - 35	45 - 50	Vår	15 - 20	40 - 45	Sommer	10 - 15	30 - 40	Høst	25 - 26	45 - 50
Sesong	Inntekter fra Handel AS (K)	Volum kjøpt av Handel AS (K)														
Vinter	30 - 35	45 - 50														
Vår	15 - 20	40 - 45														
Sommer	10 - 15	30 - 40														
Høst	25 - 26	45 - 50														
<b>Resultat:</b>	OK. Identifisert to unaturlige sammenhenger. Disse forklares ved at prisendringer utjevner effekten av volumendringer. Det vil si lav pris med stort volumkjøp og omvendt.															

## Revisjonshandling 2

<b>Formål:</b>	Kontrollere at inntektene avhenger av sesong.
<b>Revisjons-handling:</b>	Sammenlikne inntekter for 2019 med inntekter i 2017 og 2018.
<b>Analyse:</b>	
<b>Resultat:</b>	<p>OK.</p> <p>En unormal overgang mellom sommer og høst i 2019. Dette skyldes en kraftig økning i strømprisene i sammenheng med økt volum solgt til kunden. Det er også funnet høyere nivå på inntektene for våren 2019 sammenliknet med samme periode i 2017. Dette forklares med unormalt store nedbørsmengder i 2017 som drev prisene kraftig ned.</p>

**Konklusjon:** Inntektene er i det alt vesentlige gyldige. Alle unaturlige sammenhenger/avvik er sjekket og kan forklares.

Svar på følgende spørsmål, der 1 er veldig lite, og 7 er veldig stor:

	Veldig lite			Middels			Veldig stor	
	1	2	3	4	5	6	7	
Basert på innhentet revisjonsbevis, i hvilken grad mener du det at det er risiko for at inntektene fra kunden inneholder vesentlig feilinformasjon?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
I hvilken grad føler du at du er i stand til å konkludere med at inntektene er gyldige?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
Hvor sannsynlig er det at du vil innhente ekstra revisjonsbevis og/eller forklaring fra klient?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Du skal nå se bort fra oppgaven du nettopp gjorde.



## Informasjon om dataanalyser

### Clusteranalyse

Clusteranalyse er en analyseform der man bruker en maskinlæringsalgoritme for å gruppere observasjoner basert på karakteristikken til observasjonene. Hver gruppe (cluster) vil således inneholde observasjoner som likner på hverandre og som er sterkt korrelert. Grupperingen foregår slik at algoritmen først finner et antall optimale grupper basert på spredningen av observasjoner i datasettet. Deretter tilordnes hver observasjon en gruppe ved å minimere avstanden mellom observasjonen og sentrum i gruppen.



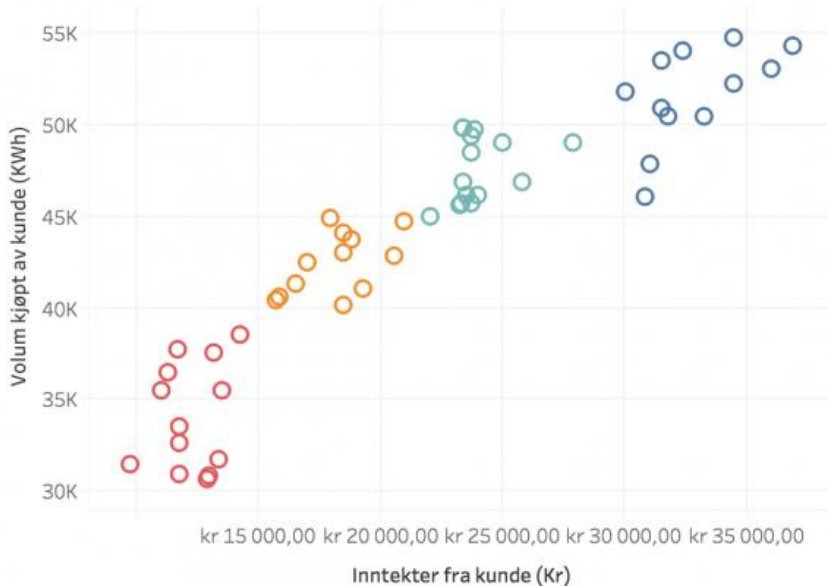
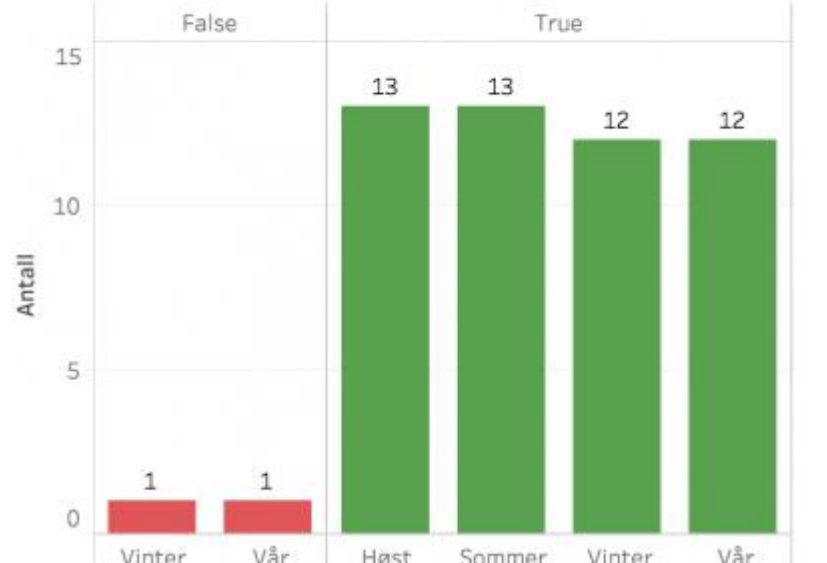
Programvaren bestemmer selv hvor mange grupper som blir laget. Dersom observasjoner ikke passer inn i noen grupper, det vil si at avstanden mellom observasjonen og sentrum i en gruppe er veldig stor, vil ikke algoritmen tilordne en gruppe for denne observasjonen. Dersom vi har mange observasjoner som ikke grupperes, vil dette være en svakhet i modellen.

En clusteranalyse baserer seg på statistikk, og man kan derfor vurdere styrken til analysen. Programvaren bruker en «Analysis of Variance» for å vurdere om gjennomsnittet mellom gruppene er signifikant ulike fra hverandre. Dersom gjennomsnittet for gruppene er like, vil man på generell basis få en F-statistic på 1. Det vil si at jo høyere verdien er for en F-statistic utover 1, jo større er variasjonen mellom gruppene. Med andre ord betyr en høy F-statistic en god statistisk modell. Variasjonen i observasjonene innad i en gruppe er gitt ved «Within-group Sum of Squares». Siden de ulike gruppene er basert på observasjonenes karakteristikker forventer vi en lav variasjon. Dermed kan man si at man har en god modell dersom «Within-group Sum of Squares» er lav. Dette må dog ses i sammenheng med p-verdien. Et signifikansnivå på  $p < 0,05$  anses som en statistisk valid modell.

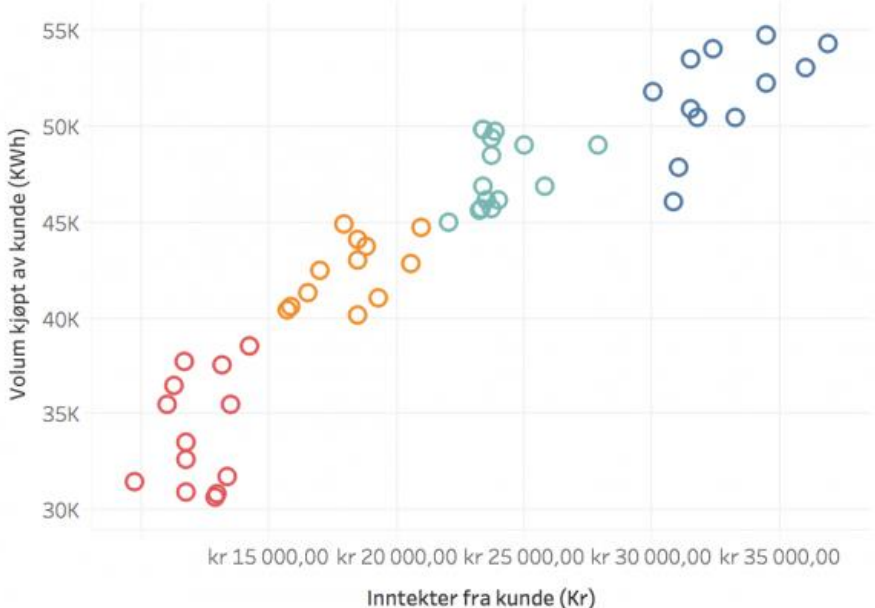
### Oppgave 2

Du gjør nå en review av et arbeidspapir for å sjekke gyldighet av inntekter fra kunden Handel AS. Det er en på revisjonsteamet som har utformet arbeidspapiret ved hjelp av selskapets analyse- og visualiseringsverktøy. Ta stilling til de følgende to revisjonshandlingene som er utført. Svar deretter på spørsmålene.

## Revisjonshandling 1

<b>Formål:</b>	Kontrollere at inntektene og volum kjøpt avhenger av sesong.
<b>Revisjonshandling:</b>	Sammenlikne inntekter og volum kjøpt av Handel AS for hver sesong for å vurdere sesongvariasjoner.
<b>Analyse:</b>	<p style="text-align: center;"><b>Clusteranalyse</b></p>  <p style="text-align: center;"><b>TRUE/FALSE fordelt på sesong</b></p> 
<b>Resultat:</b>	OK. Avvik på to observasjoner. Sjekket mot faktura, og avviket forklares med stor prisendring for disse ukene. Konklusjonen er at det foreligger sesongvariasjoner.

## Revisjonshandling 2

<b>Formål:</b>	Kontrollere at inntektene avhenger av volum som er kjøpt.																																															
<b>Revisjonshandling:</b>	Vurdere om det foreligger en plausibel sammenheng mellom inntektene fra kunden og volum som er kjøpt.																																															
<b>Analyse:</b>	<p style="text-align: center;"><b>Clusteranalyse</b></p>  <p style="text-align: center;">Volum kjøpt av kunde (kWh)</p> <p style="text-align: center;">Inntekter fra kunde (Kr)</p> <p><b>Summary Diagnostics</b></p> <p>Number of Clusters: 4  Number of Points: 52  Between-group Sum of Squares: 7.667  Within-group Sum of Squares: 0.61372  Total Sum of Squares: 8.2808</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th rowspan="2">Clusters</th> <th rowspan="2">Number of Items</th> <th colspan="2">Centers</th> </tr> <tr> <th>Sum of Inntekter fra kunde</th> <th>Sum of Volum kjøpt av Handel AS</th> </tr> </thead> <tbody> <tr> <td>Cluster 1</td> <td>12</td> <td>32859.0</td> <td>51577.0</td> </tr> <tr> <td>Cluster 2</td> <td>15</td> <td>23997.0</td> <td>47519.0</td> </tr> <tr> <td>Cluster 3</td> <td>12</td> <td>18176.0</td> <td>42418.0</td> </tr> <tr> <td>Cluster 4</td> <td>13</td> <td>12232.0</td> <td>34078.0</td> </tr> <tr> <td>Not Clustered</td> <td>0</td> <td></td> <td></td> </tr> </tbody> </table> <p><b>Analysis of Variance:</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th rowspan="2">Variable</th> <th rowspan="2">F-statistic</th> <th rowspan="2">p-value</th> <th colspan="2">Model</th> <th>Error</th> </tr> <tr> <th>Sum of Squares</th> <th>DF</th> <th>Sum of Square</th> </tr> </thead> <tbody> <tr> <td>Sum of Inntekter fra kunde</td> <td>15.32</td> <td>3.985e-07</td> <td>3.915</td> <td>3</td> <td>4.088</td> </tr> <tr> <td>Sum of Volum kjøpt av Handel AS</td> <td>14.32</td> <td>8.597e-07</td> <td>3.752</td> <td>3</td> <td>4.193</td> </tr> </tbody> </table>	Clusters	Number of Items	Centers		Sum of Inntekter fra kunde	Sum of Volum kjøpt av Handel AS	Cluster 1	12	32859.0	51577.0	Cluster 2	15	23997.0	47519.0	Cluster 3	12	18176.0	42418.0	Cluster 4	13	12232.0	34078.0	Not Clustered	0			Variable	F-statistic	p-value	Model		Error	Sum of Squares	DF	Sum of Square	Sum of Inntekter fra kunde	15.32	3.985e-07	3.915	3	4.088	Sum of Volum kjøpt av Handel AS	14.32	8.597e-07	3.752	3	4.193
Clusters	Number of Items			Centers																																												
		Sum of Inntekter fra kunde	Sum of Volum kjøpt av Handel AS																																													
Cluster 1	12	32859.0	51577.0																																													
Cluster 2	15	23997.0	47519.0																																													
Cluster 3	12	18176.0	42418.0																																													
Cluster 4	13	12232.0	34078.0																																													
Not Clustered	0																																															
Variable	F-statistic	p-value	Model		Error																																											
			Sum of Squares	DF	Sum of Square																																											
Sum of Inntekter fra kunde	15.32	3.985e-07	3.915	3	4.088																																											
Sum of Volum kjøpt av Handel AS	14.32	8.597e-07	3.752	3	4.193																																											
<b>Resultat:</b>	OK. Statistisk signifikant modell. Konkluderer med at det foreligger sammenheng mellom variablene inntekter og volum som er kjøpt.																																															

**Konklusjon:** Inntektene er i det alt vesentlige gyldige. Alle unaturlige sammenhenger/avvik er sjekket og kan forklares.

Svar på følgende spørsmål, der 1 er veldig lite, og 7 er veldig stor:

	Veldig lite			Middels			Veldig stor	
	1	2	3	4	5	6	7	
<b>Basert på innhentet revisjonsbevis, i hvilken grad mener du det at det er risiko for at inntektene fra kunden inneholder vesentlig feilinformasjon?</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>I hvilken grad føler du at du er i stand til å konkludere med at inntektene er gyldige?</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Hvor sannsynlig er det at du vil innhente ekstra revisjonsbevis og/eller forklaring fra klient?</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Avslutningsvis skal du besvare noen generelle spørsmål.

Hvilken av metodene synes du var mest kompleks?

- Grafer (linjediagram og box plot)
- Clusteranalyse (spredningsplot og tilhørende deskriptiv statistikk)
- Like komplekse

Fikk du spesifikk informasjon om de ulike metodene (linjediagram/box plot/clusteranalyse) og hvordan disse skal tolkes?

- Ja
- Nei

Hvilket kjønn er du?

- Mann
- Kvinne
- Annet

Hva er din alder?

- 20-24 år
- 25-30 år
- 30 + år

Har du påbegynt eller har du tidligere tatt fag innen revisjon, eksempelvis MRR411/BUS426 (Revisjon 1)? Påbegynt fordrer at du startet med faget dette semesteret

- Ja
- Nei

Har du arbeidserfaring innen revisjon?

- Nei
- Ja, mindre enn ett år (eksempelvis internship eller nylig begynt i jobb)
- Ja, 1-2 år
- Ja, 2-4 år
- Ja, mer enn fire år

På en skala fra 1-7, der 1 er veldig lite og 7 er veldig mye, hvor mye erfaring har du med revisjon av inntekter?

1	2	3	4	5	6	7
---	---	---	---	---	---	---

På en skala fra 1-7, der 1 er veldig lite og 7 er veldig mye, hvor mye erfaring har du med å være reviewer?

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Diskuterte du eksperimentet med andre under gjennomføringen?

- Ja
- Nei

	Veldig lite		Middels			Veldig mye	
	1	2	3	4	5	6	7
På generell basis hadde jeg en tendens til å stille spørsmål ved informasjonen gitt i dette caset.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ved gjennomføringen av caset stilte jeg spørsmål ved elementer jeg så eller leste.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg hadde en tendens til å ikke godta informasjon gitt i dette caset, med mindre jeg hadde bevis på at det var sant.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ved gjennomføring av caset tok jeg med god tid til å ta beslutninger.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Gjennom dette eksperimentet likte jeg ikke å måtte ta beslutninger uten å få sett på all tilgjengelig informasjon.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg likte ikke å måtte ta raske beslutninger mens jeg jobbet med dette caset.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mens jeg jobbet med dette caset forsøkte jeg å forsikre meg om at jeg hadde vurdert all tilgjengelig informasjon før jeg fattet beslutninger.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ved gjennomføring av eksperimentet ventet jeg med å ta beslutninger frem til jeg fant informasjon jeg trengte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg føler at det ville økt mine muligheter til å ta riktige beslutninger i caset, dersom jeg hadde hatt muligheten til å se underliggende revisjonsbevis.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg forsøkte å lete etter mer revisjonsbevis for å styrke mine muligheter til å ta korrekte beslutninger i caset.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg søkte aktivt etter all informasjon som var tilgjengelig ved gjennomføringen av caset.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg benyttet meg av alle ressurser som var tilgjengelig for å få mest mulig informasjon ved gjennomføringen av caset.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

We thank you for your time spent taking this survey.  
Your response has been recorded.

## Vedlegg 2: Nolder & Kadous' rammeverk for skeptisk holdning

**Table 2**  
Attitude and intention measures.

Question(s)	Target	Scales Anchors	
		Good	Bad
		<b>Attitude Component: Beliefs (B)</b>	
B1	What is the risk that [revenue] is materially misstated?	Management's Assertions	Extremely risky (probable)
B2	How reasonable is management's estimate of [the warranty reserve]?	Management's Assertion	Unreasonable
B3	How appropriate is [procedure] for testing management's assertion that [accounts payable is complete]?	Evidence	Inappropriate
B4	When testing management's [valuation assertion regarding inventory], would [procedure] provide sufficient evidence regarding the fair presentation of management's assertion?	Evidence	Insufficient
B5	To what extent are you prepared to conclude that the balance is fairly stated?	Evidence	Not at all prepared to conclude (Insufficient)
		<b>Attitude Component: Feelings (F)</b>	
F1	How worried are you that [revenue] is materially misstated? How fearful are you that there is an undiscovered misstatement in the financial statements?	Management's Assertions	Very worried/fearful
F2	How comfortable are you with [management's estimate of uncollectibles]? How satisfied are you that management's [revenue recognition policies] are acceptable?	Management's Assertions	Not at all comfortable/satisfied
F3	How worried are you that the evidence supporting management's assertions is not sufficient to conclude?	Evidence	Very worried
F4	How satisfied are you with the quantity and quality of evidence you have collected to support your conclusion? How comfortable are you that the evidence is sufficient?	Evidence	Not at all satisfied/comfortable
		<b>Intentions (I)</b>	
I1	Which procedures should be conducted?	N/A	<sup>a</sup>
I2	How will you allocate time to the various procedures?	N/A	<sup>a</sup>
I3	How much additional testing would you recommend? How many hours of work are needed?	N/A	Limited testing
I4	How likely is it that you would seek additional evidence and/or explanation from the client?	N/A	Extremely likely
I5	What percentage of the population do you want to sample?	N/A	100%

<sup>a</sup> While these measures do not include a bimodal continuous scale anchored with some degree of favor or disfavor, they do proxy for the auditors' intentions and/or behaviors in performing the audit. After collecting these responses, investigators can transform responses to a continuous scale that reflects how persuasive the evidence is in supporting auditors' opinion regarding the fair presentation of the financial performance of the company (i.e., financial performance of the company (i.e., management's assertions)).

## Vedlegg 3: Innledende tester<sup>17</sup>

Cronbach's alfa			
<b>Case Processing Summary</b>			
		N	%
Cases	Valid	62	96.9
	Excluded <sup>a</sup>	2	3.1
	Total	64	100.0
a. Listwise deletion based on all variables in the procedure.			
<b>Reliability Statistics</b>			
	Cronbach's Alpha	N of Items	
	.750	3	

T-test for score på kunnskapstesten										
<b>Group Statistics</b>										
		Gruppe	N	Mean	Std. Deviation	Std. Error Mean				
Riktige svar	1		34	3.71	1.088	.187				
	2		28	3.79	.876	.166				
<b>Independent Samples Test</b>										
		Levene's Test for Equality of Variances			t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Riktige svar	Equal variances assumed	.374	.543	-.313	60	.755	-.080	.255	-.589	.430
	Equal variances not assumed			-.320	59.978	.750	-.080	.249	-.579	.419

<sup>17</sup> Utvalgsstørrelsen er 62 for mange av testene (dette gjelder også for testene videre i vedleggene). Dette er fordi hver deltaker besvarte spørsmål tilknyttet to ulike arbeidspapir (ett utført med lavkomplekse dataanalyser, og ett utført med høykomplekse dataanalyser). Hver deltaker har derfor to observasjoner tilknyttet sitt avgitte svar på undersøkelsen.



## Vedlegg 4: Forutsetninger for ANOVA

Shapiro-Wilk test for normalitet																										
Risiko for feilinformasjon	<p style="text-align: center;"><b>Tests of Normality</b></p> <table border="1"> <thead> <tr> <th rowspan="2"></th> <th colspan="3">Kolmogorov-Smirnov<sup>a</sup></th> <th colspan="3">Shapiro-Wilk</th> </tr> <tr> <th>Statistic</th> <th>df</th> <th>Sig.</th> <th>Statistic</th> <th>df</th> <th>Sig.</th> </tr> </thead> <tbody> <tr> <td>Risiko</td> <td>.178</td> <td>62</td> <td>.000</td> <td>.926</td> <td>62</td> <td>.001</td> </tr> </tbody> </table> <p>a. Lilliefors Significance Correction</p>							Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			Statistic	df	Sig.	Statistic	df	Sig.	Risiko	.178	62	.000	.926	62	.001
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk																						
	Statistic	df	Sig.	Statistic	df	Sig.																				
Risiko	.178	62	.000	.926	62	.001																				
Konklusjon om gyldighet	<p style="text-align: center;"><b>Tests of Normality</b></p> <table border="1"> <thead> <tr> <th rowspan="2"></th> <th colspan="3">Kolmogorov-Smirnov<sup>a</sup></th> <th colspan="3">Shapiro-Wilk</th> </tr> <tr> <th>Statistic</th> <th>df</th> <th>Sig.</th> <th>Statistic</th> <th>df</th> <th>Sig.</th> </tr> </thead> <tbody> <tr> <td>Gyldighet</td> <td>.145</td> <td>62</td> <td>.002</td> <td>.935</td> <td>62</td> <td>.003</td> </tr> </tbody> </table> <p>a. Lilliefors Significance Correction</p>							Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			Statistic	df	Sig.	Statistic	df	Sig.	Gyldighet	.145	62	.002	.935	62	.003
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk																						
	Statistic	df	Sig.	Statistic	df	Sig.																				
Gyldighet	.145	62	.002	.935	62	.003																				
Ekstra revisjonsbevis/ forklaring fra klient	<p style="text-align: center;"><b>Tests of Normality</b></p> <table border="1"> <thead> <tr> <th rowspan="2"></th> <th colspan="3">Kolmogorov-Smirnov<sup>a</sup></th> <th colspan="3">Shapiro-Wilk</th> </tr> <tr> <th>Statistic</th> <th>df</th> <th>Sig.</th> <th>Statistic</th> <th>df</th> <th>Sig.</th> </tr> </thead> <tbody> <tr> <td>EkstraRevisjonsbevis</td> <td>.194</td> <td>62</td> <td>.000</td> <td>.887</td> <td>62</td> <td>.000</td> </tr> </tbody> </table> <p>a. Lilliefors Significance Correction</p>							Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			Statistic	df	Sig.	Statistic	df	Sig.	EkstraRevisjonsbevis	.194	62	.000	.887	62	.000
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk																						
	Statistic	df	Sig.	Statistic	df	Sig.																				
EkstraRevisjonsbevis	.194	62	.000	.887	62	.000																				
Indeks	<p style="text-align: center;"><b>Tests of Normality</b></p> <table border="1"> <thead> <tr> <th rowspan="2"></th> <th colspan="3">Kolmogorov-Smirnov<sup>a</sup></th> <th colspan="3">Shapiro-Wilk</th> </tr> <tr> <th>Statistic</th> <th>df</th> <th>Sig.</th> <th>Statistic</th> <th>df</th> <th>Sig.</th> </tr> </thead> <tbody> <tr> <td>INDEX</td> <td>.104</td> <td>62</td> <td>.093</td> <td>.972</td> <td>62</td> <td>.165</td> </tr> </tbody> </table> <p>a. Lilliefors Significance Correction</p>							Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			Statistic	df	Sig.	Statistic	df	Sig.	INDEX	.104	62	.093	.972	62	.165
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk																						
	Statistic	df	Sig.	Statistic	df	Sig.																				
INDEX	.104	62	.093	.972	62	.165																				

### Levene-test for homogen varians

Risiko for feilinformasjon

#### Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Risiko	Based on Mean	1.012	1	60	.318
	Based on Median	.726	1	60	.398
	Based on Median and with adjusted df	.726	1	59.235	.398
	Based on trimmed mean	1.001	1	60	.321

Konklusjon om gyldighet

#### Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Gyldighet	Based on Mean	1.023	1	60	.316
	Based on Median	.459	1	60	.501
	Based on Median and with adjusted df	.459	1	56.211	.501
	Based on trimmed mean	.989	1	60	.324

Ekstra revisjonsbevis/  
forklaring fra klient

#### Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
EkstraRevisjonsbevis	Based on Mean	1.321	1	60	.255
	Based on Median	1.053	1	60	.309
	Based on Median and with adjusted df	1.053	1	59.053	.309
	Based on trimmed mean	1.310	1	60	.257

Indeks

#### Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
INDEX	Based on Mean	1.149	1	60	.288
	Based on Median	.944	1	60	.335
	Based on Median and with adjusted df	.944	1	59.894	.335
	Based on trimmed mean	1.181	1	60	.281

## Vedlegg 5: Hypoteser

### Wilcoxon test for rekkefølgeeffekter - Lavkompleksoppgave

Indeks

#### Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
L først (LK)	16	3.85416667	1.43485062	1.33333333	6.00000000
H først (LK)	14	3.80952381	1.29900873	2.00000000	6.66666667

#### Wilcoxon Signed Ranks Test

##### Ranks

		N	Mean Rank	Sum of Ranks
H først (LK) - L først (LK)	Negative Ranks	6 <sup>a</sup>	8.08	48.50
	Positive Ranks	6 <sup>b</sup>	4.92	29.50
	Ties	2 <sup>c</sup>		
	Total	14		

#### Test Statistics<sup>a</sup>

H først (LK) - L først (LK)	
Z	-.748 <sup>b</sup>
Asymp. Sig. (2-tailed)	.454

## Wilcoxon test for rekkefølgeeffekter – Høykompleksoppgave

Indeks

### Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
L først (HK)	16	4.22916667	1.41273966	1.33333333	6.33333333
H først (HK)	15	4.20000000	1.08963369	2.00000000	6.00000000

### Wilcoxon Signed Ranks Test

#### Ranks

		N	Mean Rank	Sum of Ranks
H først (HK) – L først (HK)	Negative Ranks	8 <sup>a</sup>	7.44	59.50
	Positive Ranks	7 <sup>b</sup>	8.64	60.50
	Ties	0 <sup>c</sup>		
	Total	15		

#### Test Statistics<sup>a</sup>

		H først (HK) – L først (HK)
Z		-.029 <sup>b</sup>
Asymp. Sig. (2-tailed)		.977

## ANOVA – Hypotese 1

Risiko for feilinformasjon

**Descriptives**

Risiko

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	34	3.59	1.395	.239	3.10	4.08	1	7
2	28	2.79	1.228	.232	2.31	3.26	1	6
Total	62	3.23	1.372	.174	2.88	3.57	1	7

**ANOVA**

Risiko

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9.889	1	9.889	5.654	.021
Within Groups	104.950	60	1.749		
Total	114.839	61			

Konklusjon om gyldighet

**Descriptives**

Gyldighet

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	34	4.44	1.481	.254	3.92	4.96	2	7
2	28	3.75	1.713	.324	3.09	4.41	1	7
Total	62	4.13	1.614	.205	3.72	4.54	1	7

**ANOVA**

Gyldighet

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7.335	1	7.335	2.903	.094
Within Groups	151.632	60	2.527		
Total	158.968	61			

Ekstra revisjonsbevis/  
forklaring fra klient

**Descriptives**

EkstraRevisjonsbevis

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	34	5.21	1.591	.273	4.65	5.76	2	7
2	28	3.96	1.815	.343	3.26	4.67	2	7
Total	62	4.65	1.793	.228	4.19	5.10	2	7

**ANOVA**

EkstraRevisjonsbevis

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	23.670	1	23.670	8.232	.006
Within Groups	172.523	60	2.875		
Total	196.194	61			

Indeks		Descriptives							
INDEX						95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum	
1	34	4.4118	1.28984	.22121	3.9617	4.8618	1.67	6.67	
2	28	3.5000	1.16711	.22056	3.0474	3.9526	1.33	6.00	
Total	62	4.0000	1.30851	.16618	3.6677	4.3323	1.33	6.67	
INDEX		ANOVA							
		Sum of Squares	df	Mean Square	F	Sig.			
Between Groups		12.765	1	12.765	8.354	.005			
Within Groups		91.680	60	1.528					
Total		104.444	61						

Indeks		Wilcoxon-Test – Hypotese 1				
		Descriptive Statistics				
	N	Mean	Std. Deviation	Minimum	Maximum	
Index G1	34	4.41176471	1.28984332	1.66666667	6.66666667	
Index G2	28	3.50000000	1.16710750	1.33333333	6.00000000	
INDEX		Wilcoxon Signed Ranks Test				
		Ranks				
		N	Mean Rank	Sum of Ranks		
Index G2 – Index G1	Negative Ranks	20 <sup>a</sup>	14.20	284.00		
	Positive Ranks	7 <sup>b</sup>	13.43	94.00		
	Ties	1 <sup>c</sup>				
	Total	28				
		Test Statistics <sup>a</sup>				
			Index G2 – Index G1			
Z			-2.286 <sup>b</sup>			
Asymp. Sig. (2-tailed)			.022			
a. Wilcoxon Signed Ranks Test						
b. Based on positive ranks.						

## ANOVA – Hypotese 2

Risiko for feilinformasjon

<b>Descriptives</b>									
Risiko									
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	
					Lower Bound	Upper Bound			
1	31	3.35	1.279	.230	2.89	3.82	1	6	
2	31	3.10	1.469	.264	2.56	3.64	1	7	
Total	62	3.23	1.372	.174	2.88	3.57	1	7	

<b>ANOVA</b>					
Risiko					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.032	1	1.032	.544	.464
Within Groups	113.806	60	1.897		
Total	114.839	61			

Konklusjon om gyldighet

<b>Descriptives</b>									
Gyldighet									
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	
					Lower Bound	Upper Bound			
1	31	4.42	1.689	.303	3.80	5.04	1	7	
2	31	3.84	1.508	.271	3.29	4.39	2	7	
Total	62	4.13	1.614	.205	3.72	4.54	1	7	

<b>ANOVA</b>					
Gyldighet					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.226	1	5.226	2.039	.158
Within Groups	153.742	60	2.562		
Total	158.968	61			

Ekstra revisjonsbevis/  
forklaring fra klient

<b>Descriptives</b>									
EkstraRevisjonsbevis									
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	
					Lower Bound	Upper Bound			
1	31	4.87	1.765	.317	4.22	5.52	2	7	
2	31	4.42	1.822	.327	3.75	5.09	2	7	
Total	62	4.65	1.793	.228	4.19	5.10	2	7	

<b>ANOVA</b>					
EkstraRevisjonsbevis					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.161	1	3.161	.983	.326
Within Groups	193.032	60	3.217		
Total	196.194	61			

Indeks

**Descriptives**

INDEX

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	31	4.2151	1.35414	.24321	3.7184	4.7118	1.33	6.67
2	31	3.7849	1.24588	.22377	3.3280	4.2419	1.67	6.67
Total	62	4.0000	1.30851	.16618	3.6677	4.3323	1.33	6.67

**ANOVA**

INDEX

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.867	1	2.867	1.694	.198
Within Groups	101.577	60	1.693		
Total	104.444	61			

**Wilcoxon-Test – Hypotese 2**

Indeks

**Descriptive Statistics**

	N	Mean	Std. Deviation	Minimum	Maximum
Index_LK	31	4.21505376	1.35413875	1.33333333	6.66666667
Index_HK	31	3.78494624	1.24587731	1.66666667	6.66666667

**Wilcoxon Signed Ranks Test**

**Ranks**

	N	Mean Rank	Sum of Ranks
Index_HK - Index_LK	Negative Ranks	16 <sup>a</sup>	251.50
	Positive Ranks	10 <sup>b</sup>	99.50
	Ties	5 <sup>c</sup>	
	Total	31	

**Test Statistics<sup>a</sup>**

	Index_HK - Index_LK
Z	-1.945 <sup>b</sup>
Asymp. Sig. (2-tailed)	.052

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.



## Wilcoxon-Test – Hypotese 2 (kun deltakere som har svart at clusteranalyse var mest kompleks)

Indeks

### Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
LK_1	18	4.05555556	1.30484266	1.33333333	5.66666667
HK_1	18	3.72222222	1.24852855	1.66666667	5.66666667

### Wilcoxon Signed Ranks Test

#### Ranks

		N	Mean Rank	Sum of Ranks
HK_1 - LK_1	Negative Ranks	9 <sup>a</sup>	9.17	82.50
	Positive Ranks	6 <sup>b</sup>	6.25	37.50
	Ties	3 <sup>c</sup>		
	Total	18		

#### Test Statistics<sup>a</sup>

	HK_1 - LK_1
Z	-1.294 <sup>b</sup>
Asymp. Sig. (2-tailed)	.196

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

## ANOVA – Hypotese 3

Indeks

### Descriptive Statistics

Dependent Variable: INDEX

Gruppe	H/L tall	Mean	Std. Deviation	N
1	1	4.5490	1.35882	17
	2	4.2745	1.24295	17
	Total	4.4118	1.28984	34
2	1	3.8095	1.27912	14
	2	3.1905	.99326	14
	Total	3.5000	1.16711	28
Total	1	4.2151	1.35414	31
	2	3.7849	1.24588	31
	Total	4.0000	1.30851	62

### Tests of Between-Subjects Effects

Dependent Variable: INDEX

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	16.088 <sup>a</sup>	3	5.363	3.520	.020
Intercept	961.152	1	961.152	630.929	.000
Gruppe	12.765	1	12.765	8.379	.005
HLtall	3.065	1	3.065	2.012	.161
Gruppe * HLtall	.456	1	.456	.299	.587
Error	88.357	58	1.523		
Total	1096.444	62			
Corrected Total	104.444	61			

a. R Squared = ,154 (Adjusted R Squared = ,110)

## Vedlegg 6: Andre funn

Cronbach's alfa HPSS			
HPSS	<b>Case Processing Summary</b>		
		N	%
	Cases	Valid	31 93.9
		Excluded <sup>a</sup>	2 6.1
		Total	33 100.0
	a. Listwise deletion based on all variables in the procedure.		
	<b>Reliability Statistics</b>		
	Cronbach's Alpha	N of Items	
	.785	12	

ANOVA – HPSS score fordelt på gruppe 1 og gruppe 2									
HPSS Indeks	<b>Descriptives</b>								
	HPSS								
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
	1	34	52.8824	11.01449	1.88897	49.0392	56.7255	37.00	73.00
	2	28	51.7143	7.09124	1.34012	48.9646	54.4640	40.00	63.00
	Total	62	52.3548	9.39323	1.19294	49.9694	54.7403	37.00	73.00
	<b>ANOVA</b>								
	HPSS								
			Sum of Squares	df	Mean Square	F	Sig.		
	Between Groups	20.950	1	20.950	.234	.630			
	Within Groups	5361.244	60	89.354					
	Total	5382.194	61						

## ANOVA – HPSS score som gruppefordeling (lavskepsis gruppe og høyskepsis gruppe)

Indeks

**Descriptives**

INDEX	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	32	3.7813	1.40559	.24848	3.2745	4.2880	1.33	6.67
2	30	4.2333	1.17493	.21451	3.7946	4.6721	1.67	6.00
Total	62	4.0000	1.30851	.16618	3.6677	4.3323	1.33	6.67

**ANOVA**

INDEX	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.165	1	3.165	1.875	.176
Within Groups	101.280	60	1.688		
Total	104.444	61			

## ANCOVA

Indeks

**Tests of Between-Subjects Effects**

Dependent Variable: INDEX

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39.215 <sup>a</sup>	10	3.921	3.015	.005
Intercept	23.016	1	23.016	17.694	.000
Erfaring_tall	.709	1	.709	.545	.464
Kjønn_tall	.575	1	.575	.442	.509
INDEX_HPSS_DENNE	.034	1	.034	.026	.871
Alder_tall	6.420	1	6.420	4.936	.031
Fag_tall	.193	1	.193	.149	.701
Erfaringmedinntekter	.451	1	.451	.347	.559
Erfaringmedreview	3.780	1	3.780	2.906	.095
HLtall	3.087	1	3.087	2.373	.130
Gruppe	1.661	1	1.661	1.277	.264
HLtall * Gruppe	.495	1	.495	.380	.540
Error	63.739	49	1.301		
Total	1076.333	60			
Corrected Total	102.954	59			

a. R Squared = .381 (Adjusted R Squared = .255)

## ANOVA – Erfaring som faktor

Indeks

### Descriptives

INDEX	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	16	3.8750	1.01013	.25253	3.3367	4.4133	1.67	5.33
2	18	3.3333	1.18818	.28006	2.7425	3.9242	1.33	6.00
3	28	4.5000	1.35931	.25689	3.9729	5.0271	1.67	6.67
Total	62	4.0000	1.30851	.16618	3.6677	4.3323	1.33	6.67

### ANOVA

INDEX	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15.250	2	7.625	5.044	.010
Within Groups	89.194	59	1.512		
Total	104.444	61			

### Multiple Comparisons

Dependent Variable: INDEX  
Tukey HSD

(I) Erfaring_tall	(J) Erfaring_tall	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	.54167	.42246	.411	-.4740	1.5574
	3	-.62500	.38533	.244	-1.5514	.3014
2	1	-.54167	.42246	.411	-1.5574	.4740
	3	-1.16667*	.37146	.007	-2.0597	-.2736
3	1	.62500	.38533	.244	-.3014	1.5514
	2	1.16667*	.37146	.007	.2736	2.0597

\*. The mean difference is significant at the 0.05 level.

## ANOVA – Alder som faktor

Indeks

### Descriptives

INDEX	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	20	3.8500	1.08942	.24360	3.3401	4.3599	2.00	6.00
2	32	4.0625	1.34521	.23780	3.5775	4.5475	1.33	6.00
3	10	4.1000	1.67074	.52833	2.9048	5.2952	2.33	6.67
Total	62	4.0000	1.30851	.16618	3.6677	4.3323	1.33	6.67

### ANOVA

INDEX	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.675	2	.337	.192	.826
Within Groups	103.769	59	1.759		
Total	104.444	61			

### Multiple Comparisons

Dependent Variable: INDEX

Tukey HSD

(I) Alder_tall	(J) Alder_tall	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.21250	.37802	.841	-1.1214	.6964
	3	-.25000	.51363	.878	-1.4849	.9849
2	1	.21250	.37802	.841	-.6964	1.1214
	3	-.03750	.48046	.997	-1.1926	1.1176
3	1	.25000	.51363	.878	-.9849	1.4849
	2	.03750	.48046	.997	-1.1176	1.1926