

Studiebarometeret: Rapport 11–2014

Hvor pålitelige er resultatene fra Studiebarometeret 2014?

Estimerte gjennomsnittsverdier og tilhørende konfidensintervaller

Hilde Karlsen



NOKUT 

Innhold

Innledning.....	3
Noen relevante statistiske konsepter	3
Utvalg og populasjon, estimat og parameter.....	3
Gjennomsnittsverdier med tilhørende konfidensintervaller Studieprogrammene fra pilotundersøkelsen.....	5
Konfidensintervallene til pilotprogrammene og tilhørende utdanningstyper.....	7
Appendix.....	9
Litt mer teknisk om konfidensintervallet, bruk av fordelinger, samt beregning av utvalgsstørrelse og konfidensintervaller når populasjonen er liten.....	9

Innledning

I spørreundersøkelsen Studiebarometeret 2014 stilles 2. års-studenter ved norske høyskoler og universiteter spørsmål som handler om hvordan de opplever kvaliteten ved studieprogrammet de går på. Ved hjelp av studentenes svar (data på individnivå) på en rekke spørsmål, kan vi beregne gjennomsnittsverdier for de forskjellige spørsmålene (variablene) for hvert studieprogram. Imidlertid er det slik at vi ikke har svar fra samtlige studenter ved hvert studieprogram, derfor hefter det en viss usikkerhet ved gjennomsnittsverdien vi har funnet. Usikkerheten knytter seg til om gjennomsnittet vi har funnet i utvalget vårt, er en god tilnærming til det sanne gjennomsnittet i populasjonen (dvs. forventningen). Dette er ikke spesielt for Studiebarometeret; det er ekstremt sjelden at undersøkelser når frem til absolutt alle vi ønsker informasjon om – med andre ord kjenner vi så å si aldri til det sanne gjennomsnittet i populasjonen. Det vi kan gjøre, er imidlertid å undersøke om gjennomsnittestimatet i utvalget vårt (eksempelvis studenter ved et bestemt studieprogram som besvarte undersøkelsen), er et mer eller mindre pålitelig estimat på det sanne gjennomsnittet i populasjonen¹ som utvalget er trukket fra (samtlige studenter ved det aktuelle studieprogrammet). Dette kan gjøres ved å beregne et konfidensintervall, det vil si feilmarginer, rundt gjennomsnittsverdien vi har funnet. Konfidensintervall lar oss altså anslå usikkerheten rundt et gjennomsnittsestimat, noe som forteller oss hvilken tillit (eng. confidence, herav konfidensintervall) vi kan ha til at gjennomsnittet vi har funnet basert på utvalget vårt, er et mer eller mindre pålitelig estimat på det sanne gjennomsnittet i populasjonen.

Noen relevante statistiske konsepter

Utvalg og populasjon, estimat og parameter

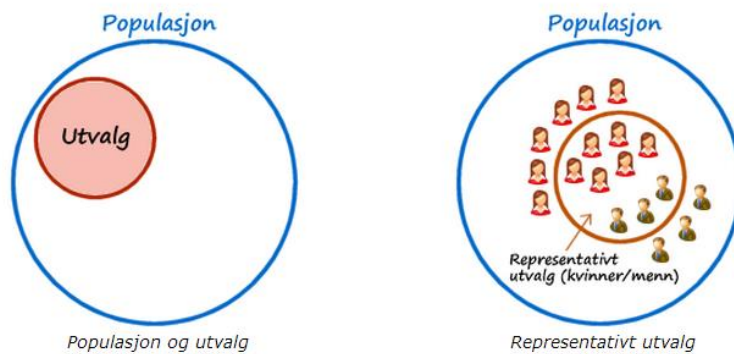
Når vi skal trekke slutninger fra *utvalg* til *populasjon*, får vi en usikkerhet som knytter seg til ulike egenskaper ved utvalget vårt. Dette skriver seg som nevnt fra at vi tross alt ikke har intervjuet alle studentene, og utvalget vårt er således bare ett av flere mulige utvalg fra den aktuelle populasjonen. Vi kan av ulike grunner ha fått en seleksjonsskjevhet i utvalget vårt (for mange kvinner/for mange eldre studenter/for mange spesielt fornøyde/misfornøyde studenter) og vi kan ha lav deltakelse i undersøkelsen ved noen studieprogrammer. Slike situasjoner kan føre til et misforhold mellom variasjonen i utvalget vårt og den faktiske variasjonen i populasjonen, fordi utvalget vårt ikke er tilstrekkelig representativt for den gruppen av studenter vi ønsker si noe om.

Hvor mange studenter vi har fra populasjonen vi ønsker si noe om, er av betydning for hvor trygge vi kan være på at utvalgsgjennomsnittet fra undersøkelsen vår er et godt estimat på det sanne populasjonsparameteret. Et utvalg bestående av mange studenter gir tross alt mer informasjon om studentpopulasjonen enn det et lite utvalg av studenter gjør.

For å få et grep om disse størrelsene og hvordan de påvirker spredningen og konfidensintervallet i Studiebarometer-data, skal vi i denne artikkelen ta som eksempel gjennomsnittsverdier fra et

¹ Studiebarometeret orienterer seg mot studieprogram-nivået, så slik sett kan populasjonen her sies å være hvert enkelt studieprogram. Dersom vi ville undersøke om eksempelvis Bachelorstudenter i sykepleie ved Haraldsplass diakonale høyskole skiller seg fra studentmassen ved alle andre (heltids) bachelorprogram i Sykepleie som vi har i Norge, ville populasjonen imidlertid vært samtlige (heltids) bachelorutdanninger i Sykepleie i Norge. Hva populasjonen er, vil altså i noen grad variere med analysens formål.

spørsmålsbatteri som angår læringsmiljøet ved det aktuelle utdanningsprogrammet. For enkelhets skyld skal vi avgrense oss til å se på hvor smale/brede konfidensintervallene blir når vi har å gjøre med små og større utvalg og populasjoner. Som vi skal se, vil små utvalg med lite variasjon gi brede konfidensintervaller, fordi den statistiske usikkerheten blir større når vi har lite informasjon å trekke på i slutningene våre fra utvalg til populasjon.



Det vi har er et utvalg av studenter som inngår i studentpopulasjonen (eksempelvis 66 av 74 sykepleierstudenter ved Haraldsplass diakonale høgskole og 37 av 64 barnevernspedagogstudenter ved HiOA). Det vi gjør er å benytte disse tilgjengelige målingene til å anslå hvilket resultat vi ville fått dersom vi hadde samlet inn svar på spørreundersøkelsen fra absolutt alle studenter ved det enkelte studieprogram, og deretter beregnet gjennomsnittet. Det vi beregner er en feilmargin som tilkjenner usikkerheten rundt gjennomsnittet, ved gjentatte uttrekk av utvalg fra en og samme populasjon. Ved normalfordelte variabler (konfidensintervaller avledes fra en variabels antatte fordeling, ofte normalfordelingen; se appendix for mer informasjon) er formelen vi bruker for å beregne et 95 % konfidensintervall som følger:

$$CI = [m - 1,96 \cdot s; m + 1,96 \cdot s]$$

Figur 2: Formel for et 95% konfidensintervall rundt gjennomsnittet

Her er CI den engelske forkortelsen for konfidensintervallet, m er utvalgets gjennomsnittsverdi og s er utvalgsgjennomsnittets standardavvik.

La oss tenke oss et eksempel til å begynne med: Dersom gjennomsnittsverdien (på en skala fra 1 til 5) i utvalget er 3,3 og vi ved hjelp av formelen i Figur 2 beregner et 95 prosent konfidensintervall med nedre grensegjennomsnitt lik 3,0 og øvre lik 3,6, er det dette intervallet vi har tillit (konfidens) til. Med andre ord: hadde vi trukket et nytt utvalg ville gjennomsnittsverdien sikkert ikke vært nøyaktig 3,3 – men med 95 prosent sannsynlighet ville vi funnet et tall mellom 3,0 og 3,6. Et konfidensintervall viser oss altså med hvilken presisjon et utvalgsgjennomsnitt treffer. Konfidensintervall kan være enklere å forstå om vi tenker prosent. La oss si at en undersøkelse viste en 20 prosent reduksjon i frafall fra sykepleierutdanningen blant mannlige studenter. Dersom denne reduksjonen på 20 prosent hadde et smalt konfidensintervall – la oss si fra 18 til 23 prosent, er prosentandelen vi avdekket «mer til å stole på» (we can be more *confident* in it) enn om prosentandelen var forbundet men en stor feilmargin, for eksempel 5 til 40 prosent. Da kunne vi tenkt: Ok, vi får 20 prosent i utvalget vårt, men vi må trekke flere utvalg og gjøre flere undersøkelser, for estimatet vårt er forbundet med veldig stor usikkerhet.

For å være trygge på at konfidensintervallet inneholder den sanne populasjonsverdien, må vi velge et bredt konfidensintervall – eksempelvis et 99 prosent konfidensintervall. Et slik intervall vil i 99 prosent av tilfellene inneholde den sanne populasjonsverdien. Brede konfidensintervaller har imidlertid den ulempen at de er lite informative, ettersom den sanne populasjonsverdien da anslås å kunne ligge så å si hvor som helst på en veldig bred skala. Vi kan være mer sikre på at intervallet inneholder den sanne populasjonsverdien (enn om vi velger et 90% konfidensintervall), men vi må tåle at graden av nøyaktighet i gjennomsnittet vi har funnet, blir mindre jo bredere intervall vi beregner. Verdier som ligger utenfor intervallet vi beregner, sies å avvike signifikant fra forventningen.

Gjennomsnittsverdier med tilhørende konfidensintervaller

Studieprogrammene fra pilotundersøkelsen

Utvalget av studieprogrammer for denne analysen er for enkelhets skyld basert på utvalget fra pilotundersøkelsen til Studiebarometeret, som ble gjennomført våren 2013, med 4 frivillig deltagende læresteder og totalt 6 studieprogrammer: Bachelor i sykepleie ved Haraldsplass diakonale høyskole, Bachelor i barnevern og Master i produktdesign ved HiOA, Master i fornybar energi og Master i økonomi og administrasjon ved UMB, samt Bygg og miljøteknikk ved NTNU.

Spørsmålsbatteriet om Studie- og læringsmiljø som ble transformert til én indeks, består av i alt 6 spørsmål med en tilfredshetsskala fra 1 til 5, der 1 er «Ikke tilfreds» og 5 er «Tilfreds». Spørsmålene som ble stilt var «Hvor tilfreds er du med ...» a) Det sosiale miljøet blant studentene på studieprogrammet, b) Det faglige miljøet blant studentene på studieprogrammet, c) Lokaler for undervisning og øvrig studiearbeid, d) Tilgjengelig utstyr og hjelpemidler i undervisningen, e) Bibliotek og bibliotekstjenester, f) IKT-tjenester (f.eks. læringsplattformer, internett og pc-tilgang). Cronbachs alpha for spørsmålsbatteriet var 0,8.

Tabell 1 viser populasjonen og resulterende utvalg for pilotundersøkelsens studieprogram, hvor stor andel utvalget gjør av populasjonen (svarprosent), hva den estimerte gjennomsnittsverdien på indeksen «Læringsmiljø» ble, standardfeilen og standardavviket til gjennomsnittet, samt øvre og nedre grenseverdi for et 95% konfidensintervall.

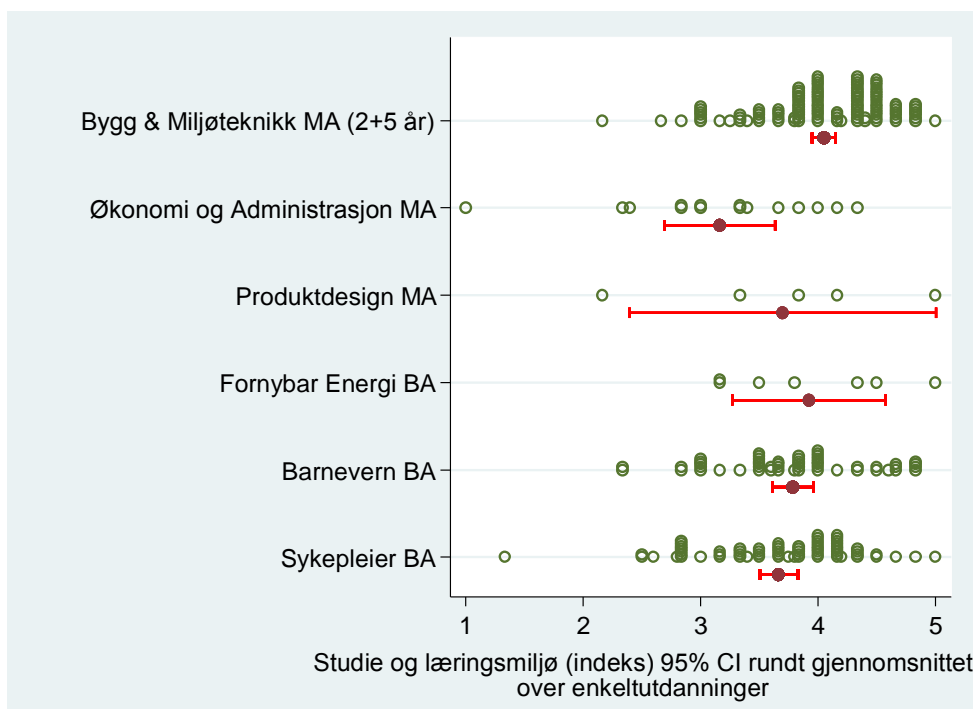
Tabell 1: Konfidensintervaller for indeks om læringsmiljø for pilot-studieprogrammene i Studiebarometer 2014-data

Institusjon	Studieprogram	Populasjon	Utvalg (N)	Svarprosent	Læringsmiljø (gj.snitt)	SE	SD	CI-95%	CI+95%
Haraldsplass	Sykepleie BA	74	66	0,89	3,67	0,08	0,65	3,51	3,83
HiOA	Barnevern BA	64	37	0,58	3,87	0,1	0,63	3,66	4,08
NTNU	Bygg&miljø MA2&5	272	112	0,41	4,05	0,05	0,54	3,95	4,15
UMB	Fornybar energi BA	18	7	0,39	3,92	0,27	0,71	3,27	4,58
HiOA	Produktdesign MA	16	5	0,31	3,70	0,47	1,05	2,40	5,00
UMB	Øk&Adm MA	54	15	0,28	3,16	0,22	0,85	2,69	3,64

Som vi ser varierer utvalgsstørrelsen betydelig, fra N = 112 til N = 5, og det samme gjør utvalgets andel av populasjonen (antall studenter ved det aktuelle studieprogrammet).

I Figur 3 ser vi hvordan konfidensintervallenes øvre og nedre grenseverdi strekker seg ut rundt de estimerte gjennomsnittsverdiene for de seks studieprogrammene. Mens feilmarginene er små rundt gjennomsnittet for sykepleierstudentene ved Haraldsplass, er det svært bredt rundt det estimerte

gjennomsnittet for produktdesign. Dette er ikke overraskende med tanke på at utvalget for Master i produktdesign bare består av 5 personer, noe som dessuten tilsvarer under en tredjedel av populasjonen (av studentmassen ved det aktuelle studieprogrammet). Vi ser av de grønne punktene, som viser hvordan studentene sprer seg på indeksens skala, at de 5 studentene på Produktdesign har gitt ganske forskjellige svar. Hvordan studentene sprer seg, betyr altså noe for det estimerte gjennomsnittet. Dersom det er få studenter og stor spredning, slik vi har i dette tilfellet, får vi høye feilmarginer rundt gjennomsnittet, ettersom usikkerheten omkring studentmassens erfaringer, og da altså det estimerte gjennomsnittet, er stor. Bygg og miljøteknikk har de smaleste feilmarginene, og som vi ser av Tabell 1 er dette også det studieprogrammet som har mest informasjon å trekke på – i alt 112 studenter. De presise grenseverdiene for intervallet kan gjenfinnes i de to kolonnene lengst til høyre i Tabell 1.



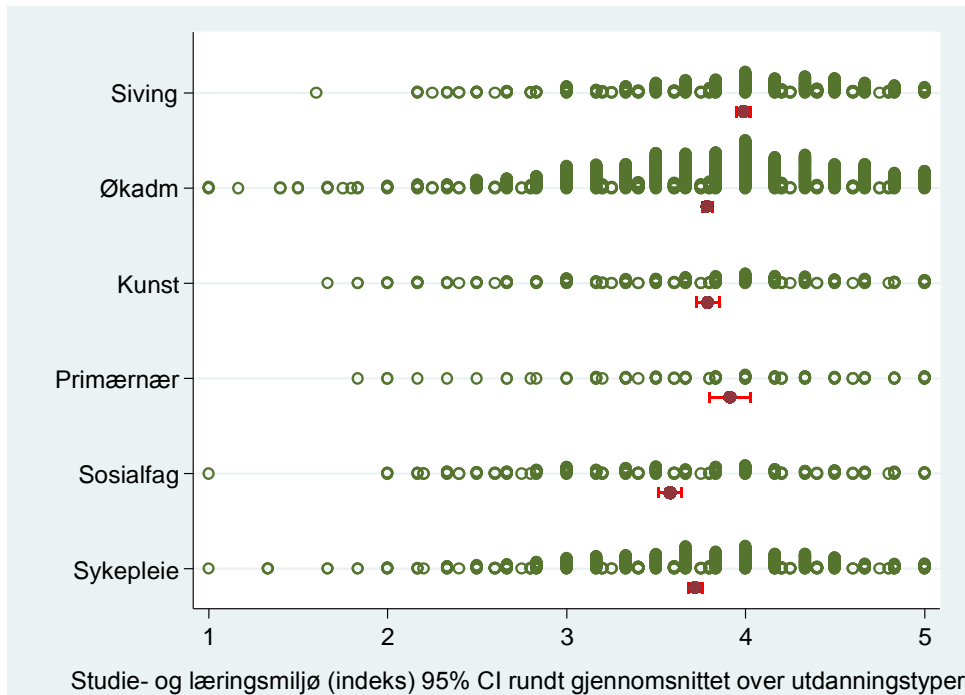
Figur 3: Konfidensintervaller for gjennomsnittene av indeksen for læringsmiljø over studieprogrammene i pilot-utdanningene. Alle er heltidsstudier på campus. (Studiebarometer-2014-data)

Tabell 2 viser utvalgsstørrelsen til utdanningstypene som de 6 studieprogrammene tilhører (data fra den ordinære undersøkelsen), den estimerte gjennomsnittsverdien på indeksen «Læringsmiljø», standardfeilen og standardavviket til gjennomsnittet, samt øvre og nedre grenseverdi for et 95% konfidensintervall. De 6 enkeltutdanningene er ikke inkludert i tabellen.

Tabell 2: Konfidensintervaller for indeks om læringsmiljø for utdanningsgruppene tilsvarende pilotutdanningene (uten enkeltutdanningene fra piloten; altså minus de 250 individene fra pilot-studieprogrammene) i Studiebarometeret 2014-data

Utdanningstype	Studieprogramnavn	Utvalg	Gjennomsnitt	SE	SD	CI-95%	CI+95%
Sykepleie	Sykepleie BA	1277	3,73	0,02	0,65	3,70	3,77
Sosialfag	Barnevern BA	457	3,59	0,03	0,68	3,52	3,65
Siving	Bygg&miljø MA2&5	872	3,98	0,02	0,59	3,94	4,02
Primærnær	Fornybar energi BA	132	3,91	0,06	0,69	3,79	4,03
Kunst	Produktdesign MA	443	3,80	0,03	0,69	3,73	3,86
Økadm	Øk&Adm MA	2815	3,79	0,01	0,68	3,77	3,82

Figur 4 viser hvordan konfidensintervallene til gruppene av studieprogrammer som de 6 enkeltprogrammene kan sies å tilhøre. Her ser vi at konfidensintervallet for utdanningsgruppen «Kunst» er svært smalt sammenliknet med konfidensintervallet for Produktdesign, som faller inn under utdanningsgruppen «Kunst». Faktisk er det slik at samtlige konfidensintervaller er temmelig smale, sammenliknet med konfidensintervallene for enkeltprogrammer i Figur 3.



Figur 4: Konfidensintervaller for gjennomsnittene av indeksen for læringsmiljø over utdanningstyper. Enkeltprogrammene fra Figur 3 inngår i de ulike utdanningstypene slik: Bygg og miljøteknikk=Siving, Økonomi og administrasjon=Økadm; Produktdesign=Kunst, Fornybar energi=Primærnær, Sosionom=Sosialfag og Sykepleier=Sykepleie. Alle er heltidsstudier på campus. (Studiebarometer-2014-data)

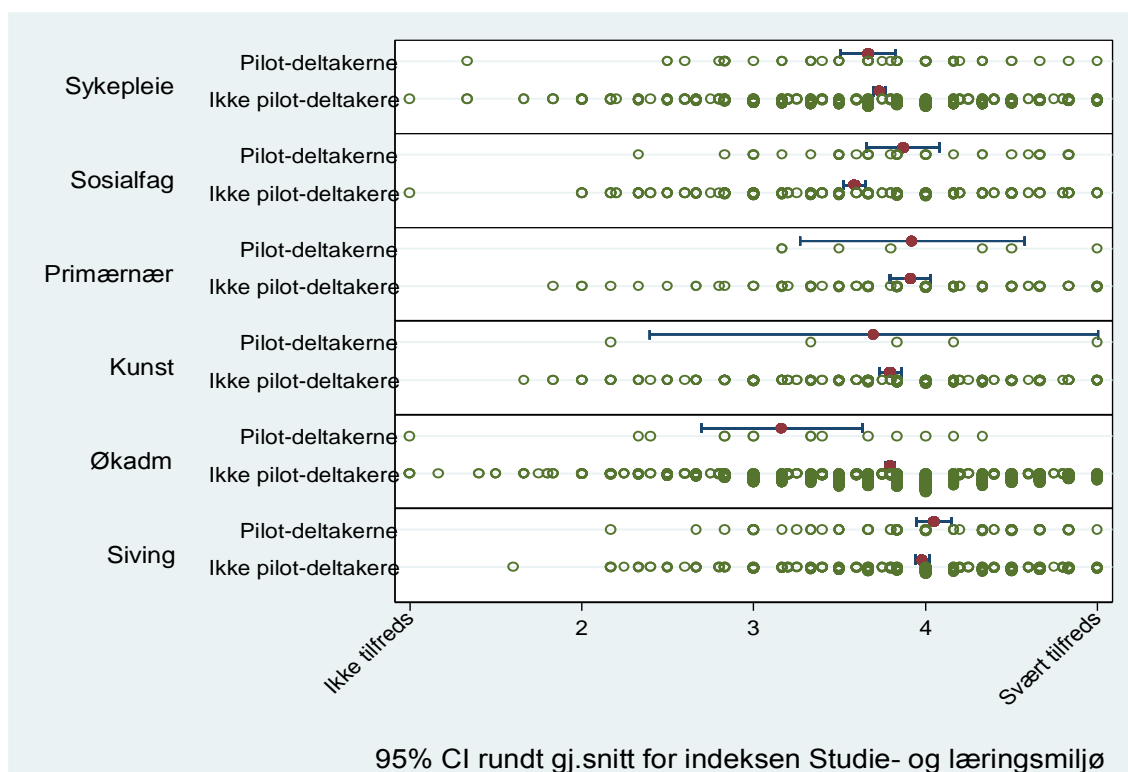
Når det er få observasjoner, får vi altså mindre informasjon å basere beregningen av populasjonsgjennomsnittet på, noe som gir oss et usikkert estimat på den sanne verdien i populasjonen, og dermed bredere feilmarginer for gjennomsnittsverdien vi har funnet. Motsatt gir mange observasjoner oss et sikrere estimat på den sanne verdien i populasjonen.

Konfidensintervallene til pilotprogrammene og tilhørende utdanningstyper

Figur 5 viser konfidensintervaller rundt de estimerte gjennomsnittsverdiene separat for programmene som deltok i piloten, samt tilsvarende studieprogrammer som ikke deltok i piloten. (Det vil eksempelvis si at i gruppen «Sykepleie, Pilotdeltakerne» finner vi de 66 sykepleierne fra Haraldsplass, mens i «Sykepleie, Ikke pilot-deltakere» finner vi samtlige andre bachelorprogrammer i sykepleie (men altså ikke Haraldsplass-sykepleierne). Figuren viser de to gruppene: pilot-deltakerne og ikke-pilot-deltakerne². Intervallene for enkeltprogrammene faller stort sett inn under intervallene for utdanningsgruppen de tilhører, hvilket betyr at det med all sannsynlighet ikke er noen signifikant

² I teorien kunne vi utført en t-test for å undersøke hvorvidt gjennomsnittet på indeksen «Studie- og læringsmiljø» blant for eksempel sykepleierstudentene ved Haraldsplass, skiller seg fra gjennomsnittet til alle de andre sykepleierstudentene på tilsvarende utdanningsnivå (heltids- bachelorprogrammer på campus, - altså ikke deltidstudier eller nett-/fjernstudier o.l.). Av plasshensyn faller ikke dette innenfor rammen av denne Aktuelle-Analyser artikkelen.

forskjell mellom gjennomsnittsverdiene i disse gruppene. Unntaket er Økadm og Master i Økonomi og administrasjon, samt (så vidt) Sosialfag og Bachelor i sosionom, hvor intervallene ikke overlapper.



Figur 5: Konfidensintervaller for gjennomsnittene av indeksen for læringsmiljø over utdanningsgrupper, separat for studenter som deltok i piloten (Pilot-deltakere) og studenter ved samme type utdanninger som ikke deltok i piloten (Ikke pilot-deltakere) (Studiebarometer-2014-data)

Appendix

Litt mer teknisk om konfidensintervallet, bruk av fordelinger, samt beregning av utvalgsstørrelse og konfidensintervaller når populasjonen er liten

Konfidensintervallet handler dypst sett om hvordan en populasjon (for eksempel alle studenter ved Høgskolen i Oslo og Akershus) *fordeler seg* rundt gjennomsnittet på en variabel.

Normalfordelingskurven er i denne sammenheng vesentlig fordi mange naturlige fenomener, slik som høyde og vekt, fordeler seg normalfordelt – altså klokkeformet (se Figur 1). Mange statistiske prosedyrer der vi tester forskjeller mellom gjennomsnitt, antar at gjennomsnittene omtrentlig følger normalfordelingen.

Som følge av sentralgrenseteoremet³, vil slike statistiske tester fungere selv om den opprinnelige fordelingen av gjennomsnitt bare er tilnærmet normalfordelt.

For å forstå hva statistiske tester og konfidensintervaller dreier seg om, må vi vite hvilke egenskaper som kjennetegner normalfordelingen og forstå hvorfor vi forholder oss til disse egenskapene når vi trekker slutninger fra utvalget vårt til populasjonen. Hvorfor er det for eksempel vanlig å snakke om 95 % konfidensintervaller, og hvorfor ser vi i denne forbindelse gjerne nevnt en såkalt «kritisk verdi» 1,96 (denne så vi første gang i formelen for konfidensintervallet i innledningen)? Er det noe spesielt med tallene 95 %, og 1,96?

Ser vi på Figur 1, forstår vi at ja, det er noe spesielt med disse tallene. Under normalfordelingen er det nemlig slik at 95% prosent av populasjonen vil befinne seg 1.96 (altså om lag 2) standardavvik fra gjennomsnittet. Med andre ord, dersom vi vet at gjennomsnittlig høyde blant norske rekrutter som er inne til sesjon⁴ er 180 cm, og vi vet at ett standardavvik er 5 cm, betyr det 95 prosent av alle norske rekrutter vil befinne seg i intervallet ca 170 cm til 190 cm høyde (helt presist: $\pm 5 \cdot 1,96$ standardavvik på hver side av 180 cm = 170,2 cm og 189,8 cm).

Om vi trekker et utvalg på 30 rekrutter, bør altså 95% av dette utvalget ligge innenfor 1.96 standardavvik – altså være mellom ca 170 og 190 cm høye – dersom vi skal kunne si at utvalget følger normalfordelingen (forventningen).

Sentrale egenskaper ved normalfordelinger er følgende:

- De er symmetriske rundt gjennomsnittet
- Gjennomsnittet, medianen og modus i en normalfordeling er identiske
- Området under normalfordelingskurven (Gauss-kurven) er lik 1 (dvs. 100%)
- De er tettere (høyere) på midten og mindre høy langs de to halene
- De er definert av to parametere: gjennomsnittet (μ) og standardavviket (σ)

³ Sentralgrenseteoremet sier at dersom vi gjentatte ganger trekker ut gjennomsnitt fra en populasjon, vil den resulterende utvalgsfordelingen/samplingsfordelingen (altså, fordelingen av gjennomsnittene vi trakk fra populasjonen), begynne å likne på en normalfordelingskurve. Det er altså primært når utvalgene våre er små (ved studieprogrammer med veldig få respondenter) at vi vil støte på problemer med beregningen. Utvalgsstørrelsen (n) inngår nemlig i beregningen av standardavviket, og standardavviket inngår videre i beregningen av konfidensintervallet, - derfor er det ikke likegyldig hvor stor utvalgsstørrelsen vår er.

⁴ Eksempel hentet fra <http://www.matematikk.org/artikkel.html?tid=68746>

- f) 68 % av området under normalfordelingskurven ligger 1 standardavvik fra gjennomsnittet (altså 34 % under og 34 % over gjennomsnittet – symmetrisk på hver side av gjennomsnittet)
- g) 95 % av området under normalfordelingskurven ligger innenfor 1.96 standardavvik fra gjennomsnittet (altså 47,5% under og 47,5% over gjennomsnittet- igjen symmetrisk)

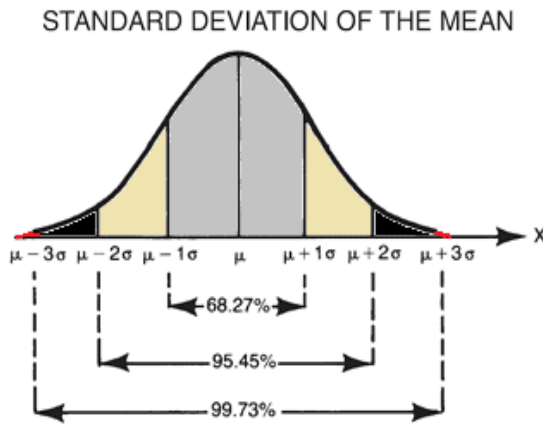


Figure 2

Percent	99.73%	99%	95.45%	95%	90%	80%	68.27%
No. of $\pm \sigma$'s	3.00	2.58	2.00	1.96	1.645	1.28	1.00

Figur 1: Normalfordelingskurven med gjennomsnitt μ og standardavvik σ .

Selv om selve gjennomsnittet og standardavviket i en normalfordeling vil variere (altså, i et tilfelle kan μ være 4 og σ være 0,5, mens i et annet tilfelle kan μ være 8 og σ være 3), vil fordelingen av gjennomsnitt alltid være lik. At fordelingen alltid er lik, betyr mer konkret at i en normalfordeling vil 68,3 % av observasjonene ligge innenfor ett standardavvik fra gjennomsnittet (grå farge); 95,5 % er innenfor to standardavvik fra gjennomsnittet (gul farge); 99,7 % innenfor tre standardavvik (sort farge) og 99,994 % er innenfor fire standardavvik fra gjennomsnittet (rød farge).

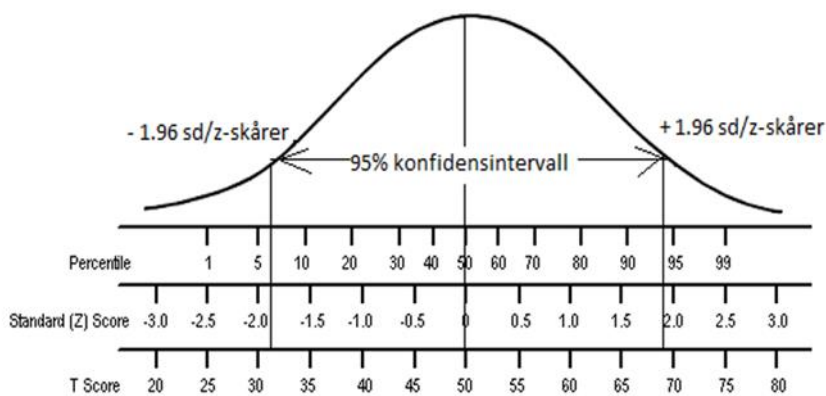
En normalfordeling med et gjennomsnitt lik 0 og standardavvik lik 1 kalles en standard normalfordeling. Standard normalfordelingens skårer kalles z-skårer.

Til forskjell fra eksempelet med de norske rekruttene, er det imidlertid uvanlig at vi kjenner til det sanne gjennomsnittet og standardavviket, for det er, som tidligere nevnt, uvanlig vi har tilgang til hele populasjonen. Av denne grunn forholder vi oss til enda en fordeling, nemlig Student's t-fordeling⁵.

$$^5 \text{ Normalfordelt Z - verdi} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}, \text{ der } \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$\text{T - fordelt t - verdi} = \frac{\bar{X} - \mu}{s / \sqrt{n}}, \text{ der } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Den første formelen viser hvordan z-verdien beregnes. I denne formelen inngår standardavviket til populasjonen, derfor viser formel nummer to hvordan standardavviket til populasjonen (σ) beregnes. Den tredje formelen viser hvordan t-verdien beregnes; hvilket er nøyaktig som vi beregner z-verdien, med unntak av at vi benytter standardavviket til utvalget (s), siden vi ikke kjenner til det sanne populasjons-standardavviket. Den fjerde formelen viser på samme vis som for normalfordelingen hvordan standardavviket for utvalget beregnes. Den eneste, men viktige forskjellen er at vi trekker fra en frihetsgrad fra antallet observasjoner, n.



Figur 2: Den klokkeformede kurven går igjen i alle tre fordelinger, både standard normalfordelingen (gjennomsnitt = 0), Student's t-fordeling (gjennomsnitt lik 50) og persentilfordelingen (gjennomsnitt = 50). Når N er høy (ved store utvalg) nærmer t-fordelingen seg normalfordelingen, og vi kan benytte normalfordelingens kritiske verdier 1.96, fremfor t-fordelingens verdier ved ulike frihetsgrader.

Prosentandel under α/s	95%
Normalfordelingen, kritisk verdi	1.96
Students t-fordeling, kritisk verdi	
Kritisk verdi ved 1 df	6.31
Kritisk verdi ved 3 df	2.35
Kritisk verdi ved 5 df	2.02
Kritisk verdi ved 10 df	1.81
Kritisk verdi ved 20 df	1.73
Kritisk verdi ved 40 df	1.68
Kritisk verdi ved 60 df	1.67
Kritisk verdi ved 100 df	1.66

Tabell 1: Utdrag av t-fordelingen med ulike frihetsgrader (df), under et 95 % konfidensintervall.

Navnet «Student's t-fordeling» kommer av at mannen som fant frem til denne fordelingen kalte seg «Student» da han la frem sine kalkyler for fordelingen. Hans egentlige navn var William S. Gossett (1876-1937), og han arbeidet med å produsere øl ved Guinness bryggeri i Dublin i Irland. William hadde å gjøre med svært små utvalg av humle og bygg i ølbryggerarbeidet sitt, og når det gjaldt bryggeripresisjon fungerte det simpelthen ikke godt nok å benytte standardavviket til utvalget som et estimat på det ukjente standardavviket i populasjonen. Med andre ord innså William at han ikke kunne bruke normalfordelingen, ettersom den faktiske fordelingen avhang av utvalgsstørrelsen (n).

Denne innsikten ledet ham til å oppdage t-fordelingen. Mens standard normalfordelingen har z-skårer som forteller oss hvor mange standardavvik fra gjennomsnittet ulike prosentandeler av populasjonen befinner seg, forteller t-skårer det samme, men tar høyde for at vi kun har et utvalg av populasjonen, og ikke kjenner til standardavviket til populasjonen. Ved svært små utvalg har t-fordelingen noe tykkere haler og en noe mindre «høy» kurve enn normalfordelingen. Følgen av dette (at fordelingen spres mer utover ved bruk av variansen) er at mindre enn 95 prosent av observasjonene vil befinne seg 1,96 standardavvik fra gjennomsnittet. Ved store utvalg – som gir mange frihetsgrader – oppfører t-fordelingen seg svært likt normalfordelingen, slik at forskjellene mellom dem blir uvesentlige. Frihetsgrader kan beskrives som det antall skårer som kan variere fritt⁶.

Det er også en egen formel som gjelder for å beregne konfidensintervallet til et gjennomsnitt når populasjonen er begrenset/kjent⁷. Denne formelen inkorporerer en finite population correction (fpc) factor, som legges til standardformelen for beregning av konfidensintervaller rundt gjennomsnitt.

⁶ Eksempel: La oss si at du kastet tre terninger, og den totale poengsummen ble 15. Hvis du fikk 4 på den første terningen og 6 på den andre, så vet du at den tredje terningen må være en 5 (ellers, vil det ikke adderes til 15). I dette eksemplet er to terningskårer fri til å variere, mens det tredje ikke er det. Derfor har vi i dette tilfellet to frihetsgrader. I mange situasjoner er frihetsgradene lik antallet av observasjoner minus 1. Dersom utvalgsstørrelsen er 66, vil det være 66 observasjoner, og frihetsgrader er $66-1$, altså 65 frihetsgrader.

⁷ For at denne tilleggsformelen skal ha noen reell innvirkning på konfidensintervallets bredde, er det nødvendig med en utvalgsstørrelse n som er minst 5% av populasjonsstørrelsen N . Kilder med eksempler på utregning: http://wps.pearsoned.co.uk/wps/media/objects/10721/10978811/Ch_08/levine-smume6_topic_08-07.pdf og <https://onlinecourses.science.psu.edu/stat414/node/264>

Formelen når N er stor, populasjonsstørrelsen N ikke er kjent eller SD til populasjonen er kjent, er:

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}}$$

Finite population correction-faktoren er: $\sqrt{\frac{N-n}{N-1}}$

Dette gir følgende formel for å finne øvre og nedre grenseverdi i konfidensintervallet til et gjennomsnitt når populasjonen N er liten/N er en kjent størrelse/SD til populasjonen er ukjent:

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Fpc-faktoren kan også brukes for å anslå hvor stort utvalg som trengs når populasjonen er liten og kjent⁸. For å gjøre dette må vi først beregne standardfeilen til det estimerte utvalgsgjennomsnittet, som er:

$$e = Z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

For å beregne utvalgsstørrelsen som trengs for å estimere gjennomsnittet, brukes følgende formel:

$$n_0 = \frac{Z^2 \sigma^2}{e^2}$$

Her er n_0 utvalgsstørrelsen uten justering for fpc-faktoren.

For å gå videre til å beregne utvalgsstørrelsen justert for små/kjente utvalg, blir formelen som følger:

$$n = \frac{n_0 N}{n_0 + (N - 1)}$$

På tross av til tider svært små populasjoner og tilhørende små utvalg, har vi i denne analysen lagt normalfordelingen til grunn for vår beregning av konfidensintervallene. Mer presise metoder som tar høyde for små populasjoner er ikke-parametriske metoder for beregning av konfidensintervaller, som vil gi noe mer nøyaktige resultater. En tommelfingerregel hva gjelder sentralgrenseteoremet, er at det trengs et utvalg på minimum 30. Dette er imidlertid bare en av flere oppfatninger i fagmiljøet; så mye som 100 har blitt påpekt å være et minimum. Når vi har å gjøre ikke bare med små utvalg, men også med små populasjoner, kan vi imidlertid ikke lene oss på sentralgrenseteoremet. En mulighet er som nevnt å beregne konfidensintervaller manuelt, ved bruk av fpc-faktoren beskrevet over.

Alternative, automatiserte ikke-parametriske beregningsmåter for konfidensintervallet er såkalt «Bootstrap resampling», som er en statistisk rutine der det fra utvalget trekkes mange nye utvalg for hvilke det beregnes separate gjennomsnitt som danner grunnlag for et estimert populasjonsgjennomsnitt. Slike gjentatte utvalg (re-sampling) med bootstrap legger ikke til grunn antagelser om at distribusjonen følger en bestemt fordeling, slik som normalfordelingen, hvilket er en vanlig (og ofte ikke uttalt) antakelse i standardberegninger av konfidensintervaller. En annen ikke-parametrisk prosedyre er såkalt «Jackknife», som også baserer seg på å trekke utvalg fra utvalget (resampling) for å estimere et populasjonsgjennomsnitt. Til slutt har vi også Chebysev's ulikhetsmetode, men denne metoden har den ulempen at den antar at populasjonsvariansen eller

⁸ For utregning og eksempler, se:

http://wps.pearsoned.co.uk/wps/media/objects/10721/10978811/Ch_08/levine-smume6_topic_08-07.pdf

standardavviket er kjent, noe som sjelden er tilfelle, og som heller ikke er tilfellet i Studiebarometer-data.

Når vi ikke har anvendt slike ikke-parametriske rutiner i denne analysen, skyldes det i første omgang at NOKUTs mandat og rolle i utviklingen av Studiebarometeret ikke er å bedrive signifikanstesting eller andre typer mer dyptgripende statistiske analyser. Formålet med denne analysen har vært å vise hvordan konfidensintervaller blir bredere og mer usikre når utvalgene er små, uansett om populasjonen utvalget er trukket fra er stor eller liten. Dette funnet ville vært det samme om vi hadde anvendt en ikke-parametrisk rutine for beregning av intervallets grenseverdier. Det er nettopp fordi NOKUT er bevisst på utfordringer knyttet til utvalgsstørrelse når det gjelder dataenes robusthet, at vi i mange tilfeller ikke viser detaljerte tall for enkelte studieprogrammer i nettportalen.

Forskningsmiljøer, som har større kapasitet og ressurser til å bedrive statistiske analyser, vil kunne gjøre mer dyptgripende undersøkelser av sammenhenger mellom utvalgsstørrelse og Studiebarometer-dataenes robusthet, enn det som i øyeblikket er mulig for NOKUT.