

# Myter og sannheter om årsaker til studenters prestasjoner i programmering

Tom Heine Nätt, Christian F. Heide og Monica Kristiansen  
Avdeling for informasjonsteknologi  
Høgskolen i Østfold

## Sammendrag

I denne artikkelen studerer vi studentenes karakterer gitt i et introduksjonsemne i programmering de åtte siste årene innenfor bachelorstudiene ingeniørfag - data, informatikk, informasjonssystemer, digital medieproduksjon og årsstudiet i informasjonsteknologi ved Høgskolen i Østfold. Spesielt undersøker vi hvorvidt studentenes karakterer i dette emnet korrelerer med karakterer i andre emner som også krever abstrakt tenkning (matematikk og objektorientert programmering), samt studentenes gjennomsnittskarakter i alle andre emner. I tillegg undersøker vi om mannlige studenter generelt presterer bedre enn kvinnelige studenter, og om dette i noen grad varierer mellom de ulike studieretningene. Resultatene fra undersøkelsene viser bl.a. at kvinnelige studenter gjør det 0.7 karaktertrinn dårligere enn sine mannlige medstudenter i introduksjonsemnet Innføring i programmering, men at dette langt ifra er signifikant for alle studieprogrammene. Videre viser resultatene korrelasjoner i størrelsesorden 0.5-0.6 mellom karakterene gitt i introduksjonsemnet og karakterene gitt i matematikk, objektorientert programmering og studentenes gjennomsnittskarakter.

## 1 Introduksjon

Når man avholder et introduksjonsemne i programmering, tenker man ofte at eksamenresultatene skiller seg fra slik man kjenner i andre kurs. Vår erfaring er at karakterene ofte får en "kamelfordeling" i stedet for en normalfordeling, der en del av kullet gjør det svært godt, en annen del gjør det svært dårlig, mens relativt få gis middelkarakteren C.

Vi oppfatter at det verserer mange ideer om årsakene til denne spredningen, og en del av påstandene som blir fremsatt kan for oss virke lite fundamentert i faktiske data. Er mannlige studenter generelt flinkere enn kvinnelige studenter i programmering? Må en student være god i matematikk for å bli god i programmering? Er det slik at studenter i et studieprogram er dårligere enn de i et annet når de tar et slikt kurs?

I denne artikkelen gjør vi en analyse av hvordan studentenes karakterer i emnet Innføring i programmering ved Høgskolen i Østfold varierer med studieprogram og kjønn. Spesielt undersøker vi hvorvidt studentenes karakterer i dette emnet korrelerer med karakterer i andre emner som også krever abstrakt tenkning. Blant annet gjør vi en sammenligning med emnet Objektorientert programmering (OOP), da vi ønsker å se om

---

*Denne artikkelen ble presentert på konferansen NIK-2013; se <http://www.nik.no/>.*

resultatene studentene får i det første programmeringssemnet er direkte overførbare til det neste. Vi gjør også en sammenligning med studentenes matematikkresultater, da svært mange hevder at ferdigheter i matematikk er nødvendige for å beherske programmering.

Basert på dette utgangspunktet, ønsker vi å finne svar på følgende forsknings spørsmål:

- Varierer studentens prestasjon i introduksjonsemnet i programmering mellom de ulike studieprogrammene?
- Finnes det noen korrelasjon mellom en students eksamensresultat i introduksjonsemnet i programmering og studentens:
  - gjennomsnittskarakter i alle matematikkemner.
  - karakter i emnet Objektorientert programmering (OOP).
  - gjennomsnittskarakter i alle emner (unntatt innføring i programmering).

Varierer disse korrelasjonene m.h.p. kjønn og studieprogram?

- Gjør mannlige studenter det generelt bedre i introduksjonsemnet i programmering enn sine kvinnelige medstudenter, og i hvilken grad varierer dette mellom de ulike studieprogrammene?

Data for den statistiske undersøkelsen er basert på eksamensresultater for studenter ved avdeling for informasjonsteknologi ved Høgskolen i Østfold i perioden 2005-2012.

## 2 Relevant litteratur

Flere studier har funnet en sammenheng mellom hvor godt studentene gjør det i grunnleggende IT-emner og deres evner innen matematiske fag [2, 3, 4, 13, 15].

Det har også vært gjennomført studier hvor en har funnet sammenheng mellom resultater i programmering og tidligere programmerings erfaring og også med tidligere erfaring med datamaskiner som ikke omfattet programmering [4, 6, 8, 14, 15].

Evne til problemløsning og andre kognitive faktorer har også blitt identifisert som faktorer som har korrelasjon med oppnådde resultater innen programmering [5, 9].

Det har videre blitt påvist en sammenheng mellom studenters mentale modeller for programmering og tiltro til egen mestringsevne [14]. Det er også funnet at karakteren en student forventer å få i et introduksjonsemne var den viktigste indikatoren for studentens prestasjoner [12].

Bergin og Reilly [1] har gjort en undersøkelse av hvordan 15 ulike faktorer influerer på hvordan studentene gjør det i et førsteårsemne i objektorientert programmering. Disse faktorene var fordelt på fire kategorier: hvorvidt studentene hadde erfaring fra tidligere innen akademia og/eller datamaskiner, studentens kognitive ferdigheter målt ved en test, personlig info som kjønn og alder, samt studentens vurdering av hvor godt han/hun selv gjorde det i emnet.

De fant relativt sterke korrelasjoner (omkring 0.5 - 0.75) mellom resultater i programmering og karakterer fra videregående skole i fag som matematikk, fysikk og biologi. Ved fordeling av disse korrelasjonene på kjønn, fant de at korrelasjonene var svake for gutter og sterke for jenter. De fant imidlertid ingen korrelasjon med kjemikarakteren fra videregående skole.

Videre fant de en korrelasjon på 0.52 mellom resultatet i programmering og det de kalte "comfort level", som var et mål de brukte på hvor komfortabel en student føler seg med emnets innhold. Et tilsvarende resultat kan sees av studien til Wilson og Shrock [15].

Den sterkeste korrelasjonen Bergin og Reilly fant var på 0.76, som var med studentens vurdering av sin egen forståelse for innholdet i emnet [1]. Dette målet baserte seg på et gradert svar på spørsmålet: ”How do you rate your level of understanding of the programming module?”

Våre data skiller seg imidlertid fra mange av disse undersøkelsene, da vi også har med tverrfaglige studier som informasjonssystemer og digital medieproduksjon. Vi mener også at det er interessant at vi her ser på data fra det norske utdanningssystemet.

### 3 Datagrunnlag

Datamaterialet som benyttes i denne artikkelen er hentet ut fra FS (Felles Studentsystem). FS er et studieadministrativt system utviklet for universiteter og høyskoler. Her finnes blant annet informasjon om når en student startet sin utdanning, hvilket studieprogram studenten tilhører, hvilke emner studenten har tatt, hvilken vurderingsform som er benyttet i hvert emne og karakterene studenten har fått.

Alle studenter i datamaterialet har vært eller er studenter ved avdeling for informasjonsteknologi ved Høgskolen i Østfold, og hadde sitt oppstartsår i perioden 2005 til 2012. I denne perioden har avdelingen hatt fem studieprogram<sup>1</sup> som har hatt innføringsemnet i programmering:

- Bachelorstudium i ingeniørfag - data (DAT)<sup>2</sup>
- Bachelorstudium i informatikk - design og utvikling av IT-systemer (ITBDES)
- Bachelorstudium i informasjonssystemer (ITBINF)
- Bachelorstudium i digital medieproduksjon (ITDM)
- Informasjonsteknologi, årsstudium (ITEKAA)

Under perioden som denne artikkelen benytter som datagrunnlag, har programmeringsspråket i innføringsemnet variert. Emnet har hatt følgende navn og emnekoder:

- ITF10005 - Innføring i programmering (Java, 2005-2006)
- ITM20606 - Webprogrammering med PHP (PHP, 2006-2007)
- ITF10007 - Innføring i programmering (Java, 2007)
- ITF10208 - Webprogrammering 1 (PHP, 2008-2011)
- ITF10212 - Innføring i programmering (PHP, 2012)

For enkelhets skyld vil vi gjennom hele denne artikkelen omtale disse ulike emneko-  
dene/navnene som Innføring i programmering, forkortet InnfPro.

Alle emner har blitt vurdert med en 4-timers skriftlig eksamen, unntatt ITF10005 (40 % prosjekt, 60 % eksamen) og ITF10007 (50 % prosjekt og 50 % eksamen). Som vist i bl.a. Heide et. al. [7] kan det at vurderingsformen er en mappe, og da også trolig at den består delvis av en mappe, påvirke eksamenskarakterene i positiv retning. Dette bør man derfor ta hensyn til under analysen av resultatene.

Alle variantene av emnet har blitt undervist og gjennomført på hva som må kalles en tradisjonell måte for et innføringskurs i programmering. Alle studenter har emnet i høstsemesteret første år, foruten digital medieproduksjon der det i perioder har blitt gitt i

År	ITBINF	DAT	ITBDES	ITDM
2012	PHP	PHP	PHP	PHP (2. år)
2011	PHP	PHP	PHP	PHP (2. år)
2010	PHP	PHP	PHP	PHP (2. år)
2009	PHP	PHP	PHP	-
2008	PHP	PHP	PHP	PHP
2007	PHP	Java	Java	PHP (1. og 2. år)
2006	PHP	Java	Java	PHP (2. år)
2005	Java	Java	Java	-

Tabell 1: Programmeringsspråk i de ulike studieretningene hvert år.

høstsemesteret andre år. Hvilke studieretninger som har hatt hvilket programmeringsspråk vises i tabell 1.

I utgangspunktet bestod vårt utplukk fra FS av 1169 studenter. Av disse er det 753 studenter som har gjennomført eksamen i emnet Innføring i programmering, og dermed er med i vårt datamateriale. Årsaken til at såpass mange ikke har tatt emnet Innføring i programmering er hovedsakelig at studenten har sluttet før eksamen i dette emnet ble avlagt eller at studenten kun tar andre enkeltemner.

For å kunne lage en sammenlignbar statistikk, har vi utelatt emner som vurderes med bestått / ikke bestått. Videre har vi, i motsetning til hva som blir benyttet på et vitnemål, valgt å benytte studentens første karakter i et emne der hvor studenten har gjort flere forsøk. Vår oppfatning er at den initielle karakteren bedre representerer det generelle nivået ved en slik sammenligning som vi her gjør. Disse 753 studentene har med disse begrensningene til sammen tatt 119 ulike emner og fått 7289 karakterer.

Ved å sammenstille karakteren i introduksjonsemnet i programmering med karakteren i objektorientert programmering, snittet av karakterene i alle matematikkemnene og snittet totalt, får vi datasettet som det vises et utdrag fra i tabell 2. For hver enkelt student er følgende informasjon inkludert<sup>3</sup>:

- En unik ID for studenten. Denne anonymiserer studentens navn og studentnummer gjennom enveisalgoritmen MD5 [11].
- Karakteren studenten har fått i emnet Innføring i programmering.
- Året som studenten fikk sin karakter i emnet Innføring i programmering (første forsøk).
- Kjønn til studenten.
- Studieprogrammet studenten tilhører.
- Gjennomsnittskarakter for alle matematikkemner som studenten har tatt.
- Karakter i emnet Objektorientert programmering.
- Gjennomsnittskarakter i alle emner studenten har tatt, der karakter i emnet innføring i programmering er ekskludert.

<sup>1</sup>Enkelte studieprogrammer har byttet navn underveis. Vi velger å benytte det nåværende navnet.

<sup>2</sup>Her er også Bachelorstudium i ingeniørfag - data, Y-veien inkludert.

<sup>3</sup>P.g.a. en svakhet i måten data kan hentes ut fra FS, vil flere forsøk på samme emnet i de tilfellene der emnekoden har endret seg mellom hvert forsøk, bli regnet som ulike emner med hver sin karakter. Dette er manuelt rettet i tallmaterialet for Innføring i programmering, objektorientert programmering og matematikkemnene. Avviket vil ikke ha noen nevneverdig betydning for resultatene.

Student-id	InnfPro	InnfPro_år	Kjønn	Studieprogram	Gj_matte	OOP	Gj_andre
d6e7fea73	2	2012	K	ITBINF	<i>NULL</i>	0.00	1.67
dd369f1ef	5	2011	M	ITBDES	3.00	0.00	4.00
77b21d3bb	4	2011	M	DAT	4.50	5.00	4.57
6d7d37d80	5	2011	M	ITBDES	2.50	1.00	3.86

Tabell 2: Tilfeldig utdrag av datasettet. Student-id består av 38 tegn, men er her forkortet til 9.

Alle karakterer er kodet med F = 0, E = 1, D = 2 etc.

Informatikkstudiet har et 10 studiepoengs matematikkemne kalt Matematikk for IT som i hovedsak består av diskret matematikk. Dataingeniørstudiet har i tillegg til dette et 10 studiepoengs matematikkemne kalt Matematikk 1, som er et typisk innføringsemne i matematikk på universitetsnivå med hovedvekt på reell analyse og lineær algebra.

## 4 Metodebeskrivelse

For å beskrive hvordan studentenes prestasjoner i emnet Innføring i programmering har endret seg de senere årene i de ulike studieprogrammene, har vi benyttet deskriptiv statistikk og histogrammer

For å undersøke om det finnes noen korrelasjon mellom en students eksamensresultat i emnet Innføring i programmering og studentens:

- eksamenskarakter i emnet Objektorientert programmering
- gjennomsnittskaracter i alle matematikkemner
- gjennomsnittskaracter i alle andre emner enn Innføring i programmering

benytter vi Pearsons korrelasjonskoeffisient ( $r$ ) og prinsipalkomponentanalyse (PCA).

Pearsons korrelasjonskoeffisient måler grad av lineær avhengighet mellom to variable. Korrelasjonskoeffisienten gir en verdi mellom  $-1$  og  $1$ . Dess nærmere absoluttverdien til koeffisienten er  $1$ , jo mer lineært avhengige er variablene. En korrelasjonskoeffisient lik null indikerer at det ikke er noen lineær avhengighet mellom variablene. Dersom en variabel øker når den andre avtar blir korrelasjonskoeffisienten negativ. På samme måte, dersom begge variablene øker eller avtar simultant blir korrelasjonskoeffisienten positiv.

For å teste om det er signifikant lineær avhengighet mellom to variable, gjennomfører vi en tosidig test på korrelasjonskoeffisienten. Nullhypotesen og den alternative hypotesen blir da som følger:

$$H_0: r = 0$$

$$H_1: r \neq 0$$

Det er viktig å påpeke at lineær korrelasjon ikke medfører et årsak-virkning-forhold mellom variablene. For å undersøke hvor godt modellen faktisk tilpasser data, kan  $R^2$  beregnes.  $R^2$  gir et mål på hvor mye av variasjonen i data som er forklart ved modellen og kan bestemmes ved å kvadrere  $r$  og deretter multiplisere resultatet med  $100\%$ . For eksempel vil en  $r = 0.5$  kun forklare  $25\%$  av variasjonen i data.

Videre er det viktig å huske på at korrelasjonskoeffisienten er svært sensitiv til ekstremverdier (utliggere). Faktisk kan en enkelt ekstremverdi påvirke koeffisienten i stor grad.

I PCA blir kovariansstrukturen eller korrelasjonsstrukturen til et sett av variabler forklart gjennom noen få ukorrelerte lineære kombinasjoner av disse variablene [10]. Disse lineære kombinasjonene kalles prinsipalkomponenter. Hovedmålsettingen med en PCA er ofte datareduksjon. Selv om det kreves  $m$  variabler til å forklare all variasjonen i data, kan ofte mye av variasjonen forklares ved et lite antall  $k$  ukorrelerte prinsipalkomponenter ( $k \leq m$ ). Hvis det er slik, kan disse  $k$  prinsipalkomponentene erstatte de  $m$  variablene og datasettet reduseres.

Et av resultatene fra en PCA er en grafisk fremstilling av data. Skåringsplott og ladningsplott er begge to grafiske fremstillinger som bør studeres nøye. Skåringsplott uttrykker grafisk variasjonen i data, mens ladningsplott uttrykker de originale variablenes bidrag til å beskrive denne variasjonen. Spesielt er to punkter som faller nær hverandre i ladningsplottet av spesiell interesse. Dette indikerer at variablene er høyt korrelert og forklarer samme type variasjon i data.

For å undersøke om mannlige studenter generelt presterer bedre enn kvinnelige studenter i emnet Innføring i programmering, har vi gjennomført hypotesetester på differansen mellom to forventningsverdier  $\mu_m$  og  $\mu_k$ . Her er  $\mu_m$  mannlige studenters forventede gjennomsnittskaraktter i emnet Innføring i programmering, og  $\mu_k$  er kvinnelige studenters forventede gjennomsnittskaraktter i det samme emnet. Nullhypotesen og den alternative hypotesen blir som følger:

$H_0$ : Mannlige studenter presterer like godt eller dårligere enn kvinnelige studenter i emnet Innføring i programmering ( $\mu_m \leq \mu_k$ )

$H_1$ : Mannlige studenter presterer bedre enn kvinnelige studenter i emnet Innføring i programmering ( $\mu_m > \mu_k$ )

Denne testen er gjennomført for alle studentene i hvert enkelt bachelorstudium (ITBINF, DAT, ITBDES og ITDM), årsstudiet, samt for alle studentene samlet. For å illustrere testene, har vi i tillegg benyttet boksploott.

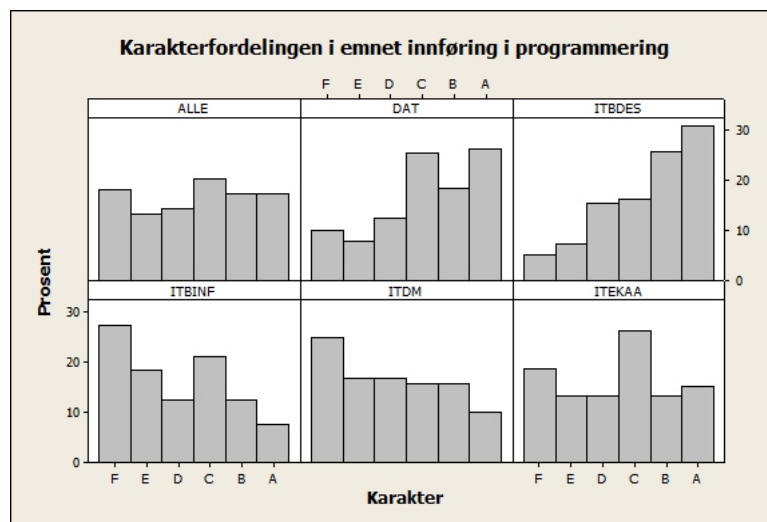
For å gjennomføre disse analysene har vi benyttet statistikkprogrammet Minitab, samt Microsoft Excel.

## 5 Resultater

I figur 1 vises studentenes karakterfordeling i prosent for emnet Innføring i programmering for hvert av de fem studieprogrammene, samt for alle studentene samlet. Vi ser at studentene i dataingeniørstudiet og informatikkstudiet har en langt større andel gode karakterer enn de tre andre studieprogrammene.

Dette framgår også av tabell 3, som viser gjennomsnittskaraktteren for Innføring i programmering fordelt på år. Av tabellen ser vi videre at gjennomsnittskaraktterene for de ulike studieprogrammene varierer en god del fra år til år, men at informatikk- og dataingeniørstudentene stort sett ligger høyere enn studentene i de øvrige studieprogrammene. Differensen er i de fleste tilfeller i størrelsesorden ett karaktertrinn, med en topp i 2011 på omkring to hele karaktertrinn. Differensen mellom informatikk og digital medieproduksjon var dette året på over 2.4 karaktertrinn.

I tabell 4 finner vi gjennomsnittskaraktteren for Innføring i programmering, Objektorientert programmering, matematikk og gjennomsnittet av karakterene i andre emner enn Innføring i programmering. Vi ser at studenter knyttet til informatikkstudiet har en gjennomsnittskaraktter på 3.4 i emnet Innføring i programmering, altså omtrent



Figur 1: Histogram over studentenes karakterfordeling for InnfPro i de ulike studieprogrammene, samt for alle studentene samlet, aggregert over perioden 2005-2012.

Gruppe	2005	2006	2007	2008	2009	2010	2011	2012
DAT	3.16	3.09	2.16	3.22	3.36	3.62	3.67	2.63
ITBDES	3.50	2.56	2.94	3.44	3.85	3.65	4.11	3.16
ITBINF	2.00	1.94	2.72	2.67	1.77	1.77	1.83	1.52
ITDM	-	3.19	2.37	1.68	-	2.08	1.67	1.64
ITEKAA	2.00	3.00	2.50	2.43	5.00	4.33	0.67	2.30
Kvinner	2.57	1.95	2.78	2.14	2.25	1.82	0.85	1.96
Menn	2.90	3.06	2.43	2.65	3.13	2.78	2.87	2.17
Alle	2.85	2.80	2.49	2.50	3.08	2.58	2.61	2.12

Tabell 3: Gjennomsnittskarakterer i InnfPro fordelt på studieprogram, eksamensår og kjønn. Digital medieproduksjon hadde ikke eksamen i dette emnet i 2005 og 2009.

Gruppe	InnfPro	OOP	Matematikk	Andre
DAT	3.12 (169)	2.31 (111)	2.27 (127)	2.63 (135)
ITBDES	3.42 (137)	2.52 (104)	2.58 (91)	2.95 (137)
ITBINF	1.96 (183)	1.70 (37)	1.75 (8)	2.46 (179)
ITDM	2.11 (209)	2.07 (14)	3.00 (1)	3.00 (208)
ITEKAA	2.47 (55)	2.53 (15)	1.00 (1)	2.44 (53)
Kvinner	2.01 (141)	2.22 (32)	3.04 (23)	2.87 (136)
Menn	2.70 (612)	2.32 (249)	2.30 (205)	2.71 (577)
Alle	2.57 (753)	2.31 (281)	2.38 (228)	2.74 (713)

Tabell 4: Gjennomsnittskarakter i InnfPro, OOP, matematikkemner og alle emner unntatt InnfPro, aggregert over perioden 2005-2012, fordelt på studieprogram og kjønn. Tallene i parentes angir hvor mange studenter som er med i gjennomsnittet. Dataingeniørene har kun InnfPro i høstsemesteret som gir bokstavkarakter. Etersom vi avsluttet datainnsamlingen i januar 2013 har 34 studenter i første år dataingeniør ikke noen andre karakterer i dette datamaterialet. Videre har én student på digital medieproduksjon, fire på informasjonssystemer og to på årsstudium kun tatt eksamen i InnfPro før de har sluttet.

Gruppe	$r$	P-verdi
DAT	0.509	$< 10^{-3}$
ITBDES	0.334	0.001
Kvinner	0.465	0.026
Menn	0.495	$< 10^{-3}$
Alle	0.471	$< 10^{-3}$

Tabell 5: Korrelasjon mellom studentenes karakterer i InnfPro og gjennomsnittskarakterer i alle matematikkemner.

midt mellom C og B, mens studenter på studiet i informasjonssystemer har en gjennomsnittskarakter rett under 2, altså litt dårligere enn D.

Som vi nevnte i introduksjonen, blir det av enkelte hevdet at det er en entydig sammenheng mellom studenters prestasjoner i matematikk og programmering. I tabell 5 ser vi korrelasjonen ( $r$ ) mellom karakterene studentene har fått i emnet Innføring i programmering og gjennomsnittskarakterene i alle matematikkemnene som inngår i studiet. P-verdien i tabellen angir sannsynligheten for å anta lineær avhengighet mellom variablene gitt at de er lineært uavhengige. Denne sannsynligheten ønsker vi skal være så lav som mulig.

Vi ser av tabellen at alt i alt har vi en korrelasjon på 0.47 mellom karakterene studentene har fått i emnet Innføring i programmering og gjennomsnittskarakterene i matematikkemnene. En korrelasjon i denne størrelsesorden indikerer at det er en viss samvariasjon, men at det langt ifra er noen entydig sammenheng mellom disse størrelsene. Det er interessant å merke seg at denne korrelasjonen stemmer godt overens med korrelasjonen på 0.46 som Bergin og Reilly [1] fant mellom karakteren i et innføringsemne i programmering og matematikkresultatene fra videregående skole. Vi ser videre at korrelasjonen er noe høyere for menn enn for kvinner, og vesentlig større for dataingeniørstudiet enn informatikkstudiet. Alle korrelasjonene er signifikante.

Gruppe	$r$	P-verdi
DAT	0.541	$< 10^{-3}$
ITBDES	0.629	0.001
ITBINF	0.637	$< 10^{-3}$
Kvinner	0.508	0.003
Menn	0.619	$< 10^{-3}$
Alle	0.610	$< 10^{-3}$

Tabell 6: Korrelasjon mellom studentenes karakterer i InnfPro og OOP.

I tabell 6 ser vi korrelasjonen mellom karakterene studentene har fått i høstemnet Innføring i programmering og karakterene de har fått i det påfølgende våremnet Objektorientert programmering. Her ville vi intuitivt forventet en meget høy korrelasjon. Når vi ser på alle studentene samlet er korrelasjonen på 0.61, altså høyere enn korrelasjonen med matematikkresultatene, men lavere enn det vi hadde forventet. Vårt materiale gir ingen forklaring på dette forholdet, så dette må vi bare forsøke å finne plausible forklaringer på ut fra sunn fornuft. En mulig forklaring kan være at emnet Objektorientert programmering er mer krevende intellektuelt sett enn Innføring i programmering, og man kan tenke seg at studenter som har gjort det bra i Innføring i programmering på grunn av god innsats, faller igjennom i Objektorientert programmering



fordi deres evnenivå gjør at de ikke mestrer dette selv med god innsats. Videre kan det tenkes at noen studenter i løpet av studieåret går lei og ikke yter tilstrekkelig innsats i vårsemesteret. Det omvendte vil også kunne inntreffe: studenter som gjør det dårlig til juleeksamen får ved det en vekker og forstår at de må yte mer dersom de skal kunne gjennomføre studiet med brukbare karakterer, og yter derfor atskillig større innsats i vårsemesteret, med de bedre resultater det medfører.

Også her ser vi at korrelasjonen er større for menn (0.62) enn for kvinner (0.51). Videre ser vi at korrelasjonen er noe større for studenter på informatikkstudiet (0.63) enn for studenter på dataingeniørstudiet (0.54). Alle korrelasjonene er signifikante.

Gruppe	$r$	P-verdi
DAT	0.614	$< 10^{-3}$
ITBDES	0.643	$< 10^{-3}$
ITBINF	0.555	$< 10^{-3}$
ITDM	0.551	$< 10^{-3}$
ITEKAA	0.707	$< 10^{-3}$
Kvinner	0.378	$< 10^{-3}$
Menn	0.597	$< 10^{-3}$
Alle	0.543	$< 10^{-3}$

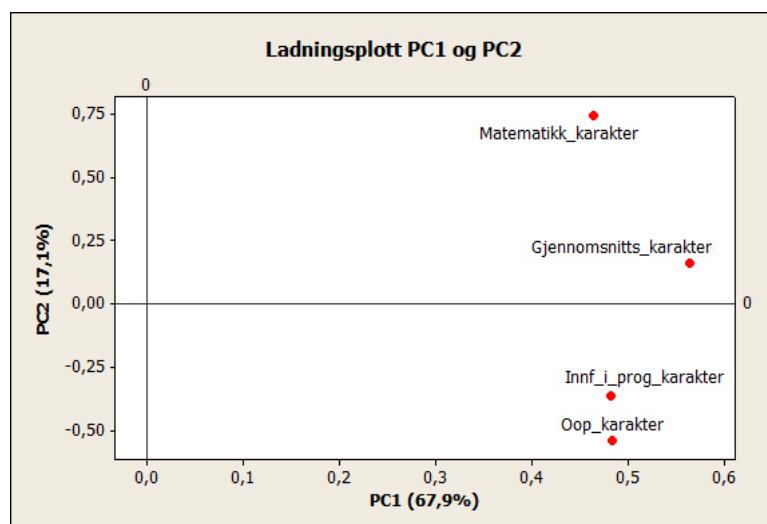
Tabell 7: Korrelasjon mellom karakterene i InnfPro og gjennomsnittskarakterene i alle andre kurs.

I tabell 7 ser vi korrelasjonen mellom karakterene studentene har fått i emnet Innføring i programmering og gjennomsnittskarakterene i studentenes øvrige emner. Når vi ser på alle studentene samlet er korrelasjonen på 0.54. Gjennomsnittet av alle karakterene i et studium, burde gi en indikasjon på studentens evnenivå og arbeidsinnsats, så det ville være rimelig å anta at denne korrelasjonen var stor. Imidlertid vet vi at en del studenter kan være gode i noen fag, men ikke så gode i programmering. I motsatt fall kan det tenkes at studenter er svært gode i programmering og har stor interesse for dette, men får dårlige resultater i støttfag på grunn av manglende interesse. Slike forhold kan være med på å forklare at korrelasjonen ikke er enda større.

Også her ser vi at korrelasjonen er større for menn (0.60) enn for kvinner (0.38). Videre ser vi at korrelasjonen er størst for studenter på årsstudiet (0.71) og minst for studenter i digital medieproduksjon (0.55) og informasjonssystemer (0.56). Etter vår erfaring, er studentene ved digital medieproduksjon og informasjonssystemer de som sterkest uttrykker manglende interesse og motivasjon for programmeringsemnet. Dette kan forklare noe av denne korrelasjonen ved at de legger ned lite innsats og dermed gjør det dårligere i programmering enn i andre emner. Støtte for dette kan vi finne i tabell 4, der disse to studentgruppene er de som gjør det dårligere i programmering enn sine andre emner. Alle korrelasjonene er signifikante.

For å utføre en multivariat analyse av korrelasjonen mellom studentenes karakterer i emnene Innføring i programmering og objektorientert programmering, samt studentens gjennomsnittskarakterer i alle matematikkemner og i de øvrige emner, har vi gjennomført en prinsipalkomponentanalyse. Analysen viser at en PCA-modell med tre komponenter forklarer 96 % av variasjonen i data. Videre viser analysen at en signifikant andel av variasjonen i data, 85 %, er forklart med kun to komponenter.

Figur 2 viser ladningsplottet til PC1 og PC2. Av figuren ser vi at alle variablene lader nokså likt til PC1. En mulig forklaring på dette er at PC1 hovedsakelig forklarer



Figur 2: Ladningsplott for PC1 og PC2.

variasjonen i data som skyldes studentenes varierende grad av grunnkompetanse og evnenivå. Videre ser vi at programmeringsemnene og matematikk lader motsatt til PC2. En mulig forklaring på dette er at PC2 forklarer variasjonen i data som skyldes forskjellen mellom såkalte “kreative” emner (programmering) og mer “oppskriftbaserte” emner (matematikk).

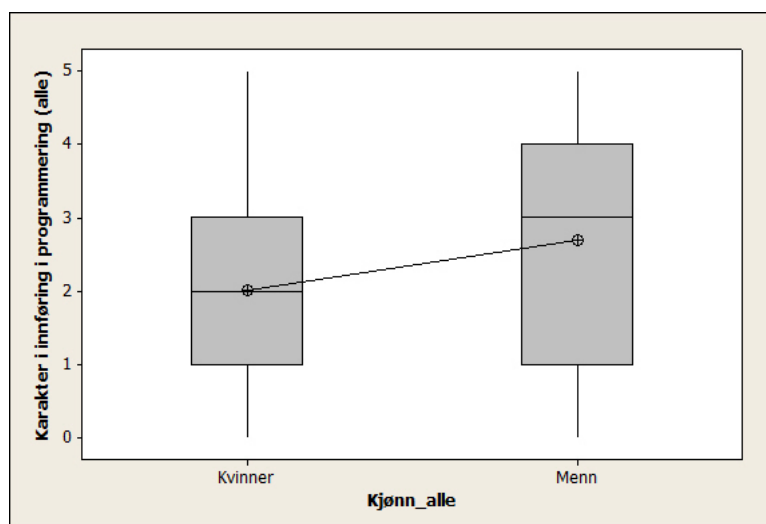
Gruppe	Antall menn	Antall kvinner	Snitt menn	Snitt kvinner	P-verdi
DAT	155	14	3.14	3.00	0.377
ITBDES	122	16	3.47	2.94	0.052
ITBINF	152	30	1.99	1.80	0.267
ITDM	144	65	2.34	1.60	0.001
ITEKAA	39	16	2.62	2.07	0.122
Alle	612	141	2.70	2.01	$< 10^{-3}$

Tabell 8: Studentenes gjennomsnittskarakterer i InnfPro fordelt på studieprogram og kjønn, aggregert over perioden 2005-2012.

Figur 3 og tabell 8 viser studentenes gjennomsnittskarakterer fordelt på kjønn for emnet Innføring i programmering. Når vi ser på alle studentene samlet er gjennomsnittskaracteren for mannlige studenter på 2.70 (noe under C) mens den for kvinnelige studenter er på 2.01 (dvs. D). Med en differanse på 0.7 karaktertrinn og en P-verdi nærmest lik null, kan det altså synes som om mannlige studenter generelt gjør det bedre i dette emnet enn kvinnelige studenter. Undersøker vi derimot de ulike studieprogrammene hver for seg, ser vi imidlertid at differansen mellom mannlige og kvinnelige studenter ikke er signifikant for studieprogrammene dataingeniør og informatikk. Det kan også være interessant å se at i matematikk har kvinnene i gjennomsnitt fått 0.7 karaktertrinn bedre enn mennene.

## 6 Konklusjon

Resultatene fra våre undersøkelser viser at studenter i studieprogrammene digital medieproduksjon og informasjonssystemer de senere årene har gjort det vesentlig



Figur 3: Studentenes gjennomsnittskarakterer i InnfPro fordelt på kjønn.

dårligere enn studenter i studieprogrammene informatikk og dataingeniør. Forskjellen har de siste årene ligget i området ett til to karaktertrinn.

Videre har vi funnet at korrelasjonen mellom karakterene i Innføring i programmering og matematikk er 0.47, korrelasjonen mellom karakterene i Innføring i programmering og Objektorientert programmering er 0.61, mens korrelasjonene mellom karakterene i Innføring i programmering og gjennomsnittet av alle andre emner er 0.54. Disse korrelasjonene indikerer en viss samvariasjon, men viser også at det på ingen måte er slik at en student som gjør det veldig bra i emnet Innføring i programmering uten videre vil gjøre det bra i resten av studiet.

Resultatene våre viser også at kvinnelige studenter gjør det 0.7 karaktertrinn dårligere enn sine mannlige medstudenter i introduksjonsemnet Innføring i programmering. Differansen mellom mannlige og kvinnelige studenter er imidlertid ikke signifikant for studieprogrammene dataingeniør og informatikk.

Det bør bemerkes at karakterne studentene får kan være noe avhengig av f.eks. vurderingsform, foreleser, sensor osv. Resultatene kan derfor i en viss grad være påvirket av dette.

## Referanser

- [1] Susan Bergin og Ronan Reilly. "programming: factors that influence success". I *Proceedings of the 36th SIGCSE technical symposium on Computer science education*, SIGCSE '05, side 411–415, New York, NY, USA, 2005. ACM. ISBN 1-58113-997-7.
- [2] Pat Byrne og Gerry Lyons. "The effect of student attributes on success in programming". I *Proceedings of the 6th annual conference on Innovation and technology in computer science education*, ITiCSE '01, side 49–52, New York, NY, USA, 2001. ACM. ISBN 1-58113-330-8.
- [3] Patricia F. Campbell og George P. McCabe. "Predicting the success of freshmen in a computer science major". *Commun. ACM*, 27(11):1108–1113, november 1984. ISSN 0001-0782.

- [4] Gerald E. Evans og Mark G. Simkin. What best predicts computer proficiency? *Communications of the ACM*, 32(11):1322–1327, 1989.
- [5] David C. Gibbs. The effect of a constructivist learning environment for field-dependent/independent students on achievement in introductory computer programming. I *Proceedings of the thirty-first SIGCSE technical symposium on Computer science education*, SIGCSE '00, side 207–211, New York, NY, USA, 2000. ACM. ISBN 1-58113-213-1.
- [6] Dianne Hagan og Selby Markham. Does it help to have some programming experience before beginning a computing degree program? I *Proceedings of the 5th annual SIGCSE/SIGCUE ITiCSE conference on Innovation and technology in computer science education*, ITiCSE '00, side 25–28, New York, NY, USA, 2000. ACM. ISBN 1-58113-207-7.
- [7] C. F. Heide, M. Kristiansen og T. H. Nätt. Vurderingsform og karakterbruk. *Norsk informatikkonferanse (NIK)*, 2012.
- [8] Edward Holden og Elissa Weeden. The impact of prior experience in an information technology programming course sequence. I *Proceedings of the 4th conference on Information technology curriculum*, CITC4 '03, side 41–46, New York, NY, USA, 2003. ACM. ISBN 1-58113-770-2.
- [9] Terry R. Hostetler. Predicting student success in an introductory programming course. *SIGCSE Bull.*, 15(3):40–43, september 1983. ISSN 0097-8418.
- [10] Richard Arnold Johnson. *Applied multivariate statistical analysis*. Prentice Hall, Upper Saddle River, NJ [u.a.], 4. ed utgave, 1998. ISBN 013834194X.
- [11] Ronald L. Rivest. The MD5 Message-Digest Algorithm. Internet RFC 1321, April 1992.
- [12] Nathan Rountree, Janet Rountree og Anthony Robins. Predictors of success and failure in a cs1 course. *SIGCSE Bull.*, 34(4):121–124, desember 2002. ISSN 0097-8418.
- [13] Laurie Honour Werth. Predicting student performance in a beginning computer science class. I *Proceedings of the seventeenth SIGCSE technical symposium on Computer science education*, SIGCSE '86, side 138–143, New York, NY, USA, 1986. ACM. ISBN 0-89791-178-4.
- [14] Susan Wiedenbeck, Deborah Labelle og Vennila N. R. Kain. Factors affecting course outcomes in introductory programming. I *In 16th Annual Workshop of the Psychology of Programming Interest Group*, side 97–109, 2004.
- [15] Brenda Cantwell Wilson og Sharon Shrock. Contributing to success in an introductory computer science course: a study of twelve factors. I *Proceedings of the thirty-second SIGCSE technical symposium on Computer Science Education*, SIGCSE '01, side 184–188, New York, NY, USA, 2001. ACM. ISBN 1-58113-329-4.