

Ministeren for videnskab, teknologi og udvikling

Udvalget for Videnskab og Teknologi
Folketinget
Christiansborg
1240 København K

Hermed fremsendes i 5 eksemplarer svar på spørgsmål nr. 39, 40, 41 og 42 (Alm. del) stillet af Udvalget for Videnskab og Teknologi den 27. februar 2006.

Med venlig hilsen

Helge Sander

**Ministeriet for Videnskab,
Teknologi og Udvikling**

Bredgade 43
1260 København K
Telefon 3392 9700
Telefax 3332 3501
E-post vt@vtu.dk
Netsted www.vtu.dk
CVR-nr. 1680 5408

Spørgsmål nr. 39, 40, 41 og 42 stillet af Udvalget for Videnskab og Teknologi den 2. marts 2006 til ministeren for videnskab, teknologi og udvikling (Alm. del)

Spørgsmål 39

Vil ministeren oversende Videnskabsministeriets notat af 24. februar 2005 om undersøgelse af forskellige prøveformers indflydelse på gennemsnitskaraktererne på Aarhus Universitet (sagsnr. 61587, dok-id 405190)?

Svar:

./ Vedlagt oversendes Videnskabsministeriets notat af 24. februar 2005 om undersøgelse af forskellige prøveformers indflydelse på gennemsnitskaraktererne på Aarhus Universitet (sagsnr. 61587, dok-id 405190).

Spørgsmål 40

Hvor er der i notatet belæg for at konkludere, at den enkelte studerendes præstation ved gruppeeksamener bedømmes mere lemfældigt end ved individuelle eksamener?

Svar:

Notatet indeholder ikke udsagn om, at bedømmelsen ved gruppeeksamen er lemfældig.

Spørgsmål 41

Finder ministeren det rimeligt at basere sine konklusioner på et notat om et universitet, der kun i meget ringe omfang benytter og har erfaringer med gruppeeksamener?

Svar:

Jeg vil gerne understrege, at regeringens beslutning om at afskaffe gruppeeksamen ikke baserer sig på en konkret undersøgelse.

Regeringen har det klare standpunkt, at der bag enhver karakter på eksamensbeviset skal ligge en individuel eksamination.

Regeringen er af den opfattelse, at det ved gruppeprøver sløres - og ikke afsløres - hvad det enkelte gruppemedlem kan præstere. Vi mener, at karakteren ved gruppeprøver ikke er så præcist et udtryk for den enkeltes standpunkt, som den kunne være, hvis der var eksamineret individuelt. Vi lægger vægt på, at eksamensbeviset tilhører den enkelte og derfor også skal afspejle den enkeltes viden og kompetence. Derfor afskaffer vi gruppeprøver.

Spørgsmål 42

Hvad er ministerens kommentar til følgende passus i notatet: "Der er ingen tvivl om, at gruppeprøver generelt bliver højere og 'smallere' bedømt. Men hvad årsagen er er svært at sige. Om dette er fordi bedømmelsen af den enkelte er ringere, eller at folk faktisk kan mere, viser modellen intet om. Blot at de bedømmes højere. Det er muligt, at der ikke sker en ordentlig bedømmelse af den enkelte, fordi denne forsvinder i mængden, men der er også mulighed for, at de

studerende har lært mere, får højere karakter, kan nogenlunde det samme og får nogenlunde den samme karakter, fordi de *netop* er i gruppe. Det er ikke muligt på baggrund af de forhåndenværende data at se, hvor de ellers ville have landet, hvis det havde været individuelle eksaminer.”

Svar:

Jeg henviser til mit svar på spørgsmål 41.

**Ministeriet for Videnskab,
Teknologi og Udvikling**

Bilag

Undersøgelse af forskellige prøveformers indflydelse på gennemsnitskaraktererne på Aarhus universitet.

Indledning

I relation til bedømmelsen af de studerende til eksamen foregår der til stadighed en diskussion om, hvad forskellige prøveforhold betyder. I den forbindelse har regeringen iværksat en undersøgelse, der skal se på bedømmelsen af den studerende i forbindelse med en redegørelse om modernisering af prøver, eksamener og karakterer.

Det følgende er en generaliseret lineær regressionsanalyse af den indflydelse, forskellige forhold omkring prøver på universitetet har på gennemsnitskarakteren for studerende, der gennemførte en eksamen på Aarhus Universitet i 2003 og 2004. Herunder forhold som censur, fakultet, eksamenstype og prøveform.

Ministeriet for Videnskab,
Teknologi og Udvikling

Før resultaterne præsenteres, præsenteres de data, som undersøgelsen baserer sig på, samt nogle af de forbehold man er nødt til at tage.

Præsentation

Samplet indeholder 117.969 opgivne prøver. Ikke alle disse indgår i regressionen, da nogle af dem er bedømt bestået/ikke-bestået og derfor ikke har en angivet karakter. Dette betyder, at gennemsnittet, der bliver omtalt i det følgende, vil være gennemsnittet af de, der har fået tildelt en karakter og bestod deres eksamen. Dette trækker gennemsnittet op og gør det en smule højere end gennemsnittet over alle elever. Dertil kommer, at 13-skalaen ikke indeholder alle punkter, men bryder den fuldstændig lineære struktur i springet fra 11 til 13. Dette er også med til at løfte gennemsnittet en smule, om end ikke meget.

Dette vil dog ikke influere på retningen konklusionen tager. Den er stadig valid.

Det skal også bemærkes, at de valgte variable ikke er dækkende for alle de forhold, der bestemmer den studerendes karakter, men meget vel skulle være dækkende for, hvad man kunne kalde de institutionelle omstændigheder ved eksamen.

Nedenstående tabel viser en oversigt over de variable, der er medtaget i regressionen i forhold til hvilke omstændigheder omkring eksamen, der kan tænkes at have en påvirkning.

Tabel 1. Variabeloversigt.

Variabel.	Mulige udfald.
KARAKTER (Den afhængige)	06-13

Variabel.	Mulige udfald.
1.1 CENSUR	Ekstern
	Intern
EKSAMENSTYPE	Rene mundtlige prøver
	Rene skriftlige prøver
	Skriftlige opgaver med mundtligt forsvar
FAKULTET	Humaniora
	Naturvidenskab
	Samfundsvidenskab
	Sundhedsvidenskab
	Teologi
FORM	Gruppe
	Individuel

Ministeriet for Videnskab,
Teknologi og Udvikling

Kolonnen variabel angiver navnet på variablen og anden kolonne de omstændigheder, som eksamen har været afholdt under. Dvs. om den har været bedømt internt eller eksternt, hvilken type prøve det har været, på hvilket fakultet den har været afholdt under og til sidst, om den har været afholdt som en gruppeeksamen eller som individuel eksamen.

Metode

Det præsenterede datasæt er herefter behandlet med statistikprogrammet SAS. Til analysen er der anvendt 72.276 besvarelser, idet de resterende ikke har angivet en karakter. Modellen og resultaterne er præsenteret i følgende tabel, som vil blive gennemgået. Først skal det dog lige forklares, hvordan modellen kan bruges. Dette er præsenteret i følgende formel, der viser gennemsnitskarakteren for en gruppe givet de estimerede værdier af karakteristika og efterfulgt af et regneeksempel.

Gennemsnitskarakter =
 Intercept + (Censur(intern/ekstern)) + (Eksamenstype(R.Skr/R.mundt/Skr. M.
 Mundt.)) +
 (Fakultet(Hum/Nat/Sam/Sun/Teo)) + Form(Gruppe/Individuel))

Gennemsnitskarakteren for en internt bedømt, rent skriftlig gruppeeksamen på det humanistiske fakultet er altså:

Gennemsnitskarakter =

$$\text{Intercept } (8,9756) + \text{Censur } (0,00) + \text{Eksamenstype } (-0,4014) + \text{Fakultet } (0,1198) + \text{Form } (0,4560) = 8,9756 + 0,00 - 0,4014 + 0,1198 + 0,4560 = 9,15$$

I modellen er nogle af de estimerede værdier 0,00. Disse er det, man kalder referenceværdier og er dem, som de andre værdier refererer til. Hvordan dette foregår vil blive uddybet senere.

Spredningsmålene.

Til at uddybe analysen af de forskellige forhold omkring eksamen er der i analysen medtaget andre mål end estimationen. Dette er standardafvigelsen og variansen.

Standardafvigelse

Standardafvigelsen fortæller, hvor spredt data er. Ca. 65% falder inden for st. afvigelsen +/- gennemsnittet. Dvs. Jo mindre st. afvigelse, jo mindre spredning.

Varians

Den gennemsnitlige afvigelse fra gennemsnittet og et andet mål for spredningen. Det beskriver den gennemsnitlige afvigelse fra gennemsnittet blandt besvarelsene. Dvs. jo større varians, jo større spredning.

Ministeriet for Videnskab,
Teknologi og Udvikling

Resultaterne

Nedenstående tabel viser de estimerede bidrag til gennemsnitskarakteren, som er den variabel, som alle de nedenstående forholder sig til. Variablerne vil blive gennemgået umiddelbart efter tabellen. Kolonnerne viser variabelens navn, de udfald som variabelen har, 'tilskuddet' til gennemsnitsværdien i forhold til referencen og niveauet for signifikans. Signifikansniveauet ligger for næsten alle variable på under 1/10 af en promille. Dvs. der er under 1/10 promille chance for, at estimationen ikke er korrekt. Humaniora ligger på ½ promille, hvilket dog ikke rykker ved, at estimationerne generelt er højsignifikante og dermed gældende.

Tabel 2. Regressionsmodel over prøveforhold på Aarhus Universitet.

Variabel.	Mulige udfald	Estimeret værdi	Signifikans niveau
Intercept		8.9756	<.0001
Censur	Ekstern	0.0559	<.0001
	Intern	0.0000*	.
Eksamenstype	Rene mundtlige prøver	0.2118	<.0001
	Rene skriftlige prøver	-0.4014	<.0001
	Skriftlige opgaver med mundtligt forsvar	0.0000*	.
Fakultet	HUM	0.1198	0.0005

Variabel.	Mulige udfald	Estimeret værdi	Signifikans niveau
	NAT	0.1875	<.0001
	SAM	-0.3446	<.0001
	SUN	-0.3136	<.0001
	TEO	0.0000*	.
Form	Gruppe	0.4560	<.0001
	Individuel	0.0000*	.

*Referencekategori

Ministeriet for Videnskab,
Teknologi og Udvikling

I det følgende gennemgås variablenes betydning i modellen én for én, og der knyttes yderligere kommentarer til dem på baggrund af andre målemetoder.

Analyse

Censur

I variabelen 'Censur' er de eksaminer, der er internt bedømt, valgt som reference. Derfor bidrager de ikke til at ændre gennemsnitskarakteren positivt eller negativt. Derimod viser det sig, at de eksaminer, der er bedømt af eksterne censorer, gennemsnitligt ligger 0,0559 karakter over de internt bedømte.

Nedenstående tabel viser, at standardafvigelsen og variansen er stort set den samme for bedømmelser under intern og ekstern censur, og dermed er der ikke er nogen nævneværdi forskel på den måde, de bedømmer på.

Tabel 3. Variabeloversigt: Censur

CENSUR	N	Manglende	Fordeling	Gennemsnit	St. Afvigelse	Varians	Median
Ekstern	56248	7733	78%	8,71	1,57	2,46	9
Intern	16028	35622	22%	8,74	1,58	2,50	9

Det betyder, at der stort set ikke er forskel på interne og eksterne bedømmere i forhold til fordelingen af bedømmelserne på karakterskalaen.

Fordelingen mellem antallet af eksterne og interne prøver er i dette tilfælde ikke den faktiske fordeling af internt og eksternt bedømte, da en stor del af de internt bedømte bliver bedømt bestået/ikke bestået, uden at der gives karakter. Tæller man disse prøver med, bliver fordelingen 55% eksternt bedømt og 45% internt bedømte.

Fortolkning:

Ekstern bedømmelse har altså et lille positivt bidrag til karakteren i forhold til den interne bedømmelse, og der bliver bedømt på stort set den samme måde i forhold til fordelingen af karaktererne på karakterskalaen. En del af forklaringen på dette kan ligge i, at en del af de eksternt bedømte eksaminer er de store og

vigtige eksamener som fx bacheloropgave eller speciale, hvor de studerende dels har vejledning og dels sætter sig ekstra op til eksamen. Dette vil trække gennemsnittet op for denne gruppe.

Eksamenstype

Variablen 'Eksamenstype' angiver typen af prøve, og her er skriftlige prøver med mundtlig forsvar referencen. Det viser sig, at de rene skriftlige prøver gennemsnitlig giver 0,40 karakter lavere end de, der er suppleret med et mundtligt forsvar. Det viser sig også, at de mundtlige prøver giver 0,21 karakter højere end de skriftlige med mundtligt forsvar (og dermed 0,61 karakter højere end de rene skriftlige prøver¹). Rene mundtlige prøver bidrager altså positivt til karakteren i forhold til skriftlige opgaver med mundtligt forsvar, og rene skriftlige bedømmes gennemsnitlig lavere end tilsvarende med tilhørende mundtligt forsvar.

Tabel 4. Variabeloversigt: Eksamenstype

Ministeriet for Videnskab,
Teknologi og Udvikling

Eksamenstype	N	Manglende	Fordeling	Gennemsnit	St. Afvigelse	Varians	Median
Rene mundtlige prøver	19008	25513	26%	9,11	1,53	2,35	9
Rene skriftlige prøver	49477	18653	68%	8,53	1,56	2,43	9
Skriftlige opgaver med mundtligt forsvar	4278	1040	6%	9,16	1,48	2,18	9

Spredningen mellem de forskellige eksamenstyper er ikke stor, men viser, at bedømmelserne fordeler sig nogenlunde ens ud over karakterskalaen. Dog er der en svag tendens til, at de skriftlige opgaver med mundtligt forsvar samler sig en smule mere omkring gennemsnittet, der ligger en smule lavere for de rene skriftlige opgaver, som den ovenstående regressionsmodel forudsagde.

Også i denne variabel mangler der mange besvarelser af karakter, hvilket igen kan tilskrives, at en stor del af de observerede eksamener og bedømt bestået/ikke bestået. Tages disse med i fordelingen mellem de forskellige eksamenstyper, bliver fordelingen 38% rent mundtlige prøver, 58% rent skriftlige prøver og 5% skriftlige prøver med mundtligt forsvar.

Fortolkning:

Den mundtlige eksamination kræver, at et ekstensivt stof præsenteres i meget komprimeret form i løbet af kort tid, hvor den skriftlige stiller eleven mere til ansvar for, hvad der præsenteres, idet alt, hvad der påstås, er dokumenteret i opgaven. Dermed kan det være nemmere for den studerende at skjule de mangler, der er i hans kendskab til stoffet i en situation med mundtlig eksamen. At de skriftlige med mundtligt forsvar ligger imellem de to rent karaktermæssigt, kan skyldes, at disse opgaver i mange tilfælde er store, og nogle har vejledning

¹ Den numeriske forskel mellem estimerne af rent skriftlige prøver og rent mundtlige prøver. $(0,2118 - (-0,4014)) = 0,2118 + 0,4014 = 0,6132$

gennem forløbet og dermed er bedre gennemarbejdet. Denne type eksamen udgøre kun 6% af dem, der er medtaget i analysen.

Fakultet

Analysen viser, at naturvidenskab generelt bliver bedømt højest efterfulgt af humaniora. Samfundsvidenskab bliver hårdest bedømt, efterfulgt af sundhedsvidenskab. Spredningssmålet viser, at naturvidenskab har størst spredning i deres bedømmelser og sundhedsvidenskab mindst. Mellem de to er der tale om en forskel på 1 karakter i spredningen for ca. 65% af eksaminerne. Dog kan man ikke alene på baggrund af dette konkludere, at fx. samfundsvidenskabeligt uddannede har lavere gennemsnitskarakterer, men skal ses i forhold til de andre variable. Fx kunne man forestille sig, at der var flere mundtlige prøver på samfundsvidenskab end fx. naturvidenskab, hvilket vil kunne løfte karaktergennemsnittet på samfundsvidenskab over naturvidenskab, selvom man generelt bedømmer højere der, når man alene ser på fakultetet.

Ministeriet for Videnskab,
Teknologi og Udvikling

Tabel 5. Variabeloversigt: Fakultet

FAKULTET	N	Manglende	Fordeling	Gennemsnit	St. Afvigelse	Varians	Median
HUM	18499	11212	25%	8,95	1,48	2,20	9
NAT	14912	6544	20%	8,99	1,75	3,07	9
SAM	24013	5465	33%	8,44	1,53	2,33	8
SUN	13042	20084	18%	8,52	1,45	2,09	9
TEO	2297	1901	3%	9,01	1,50	2,24	9

Spredningen i karaktererne på de forskellige fakulteter ligger også nogenlunde jævnt, dog med en svag tendens til at naturvidenskab spreder sig en smule mere på karakterskalaen. Samfundsvidenskab er det fag, der har det laveste karaktergennemsnit, som også regressionsmodellen forudsagde, mens de andre gennemsnit ligger nogenlunde ens. Noget tyder altså på, at samfundsvidenskab bedømmes lavere uanset andre omkringværende forhold omkring eksamen. Til gengæld er samfundsvidenskab, forudsat at de manglende karaktergivninger er eksaminer, der er bedømt bestået/ikke bestået, faget, hvor flest prøver resulterer i en karakter. Hele 81% er bedømt ved karaktergivning i modsætning til fx sundhedsvidenskab, hvor 61% er bedømt bestået/ikke bestået.

Fortolkning:

Det er en smule svært at sige noget om, hvorfor der er forskel på de forskellige fakulteter. For samfundsvidenskab kan det dog have noget at gøre med, at der er mange flere prøver, der bliver bedømt med en karakter end fx naturvidenskab. Det er muligt, at det forholder sig sådan, at man sætter sig mere op til de eksaminer, der bliver bedømt med karakter, hvis denne slags eksaminer optræder sjældnere. Analysen slår dog fast, at naturvidenskabelige uddannelser generelt bedømmes højest og med mest bredde i brug af karakterskalaen og samfundsvidenskab lavest. Sundhedsvidenskab bliver generelt bedømt mest omkring midten af karakterskalaen.

Form

Når man ser på formen, hvorunder eksamen har været udført, viser modellen, at gruppeeksamener gennemsnitligt bliver bedømt 0,45 karakter højere end de individuelle eksamener. Gruppeeksamener yder dermed et betydeligt bidrag til gennemsnitskarakteren. Analysen af spredningen af karaktererne for de individuelle eksamener over for gruppeeksamenerne viser, at der er forskel på de to grupper ud over bidraget til gennemsnitskarakteren. Analysen viser, at gruppeprøver har en mindre standardafvigelse og større varians end de individuelle.

Tabel 6. Variabeloversigt: Form

FORM	N	Manglende	Fordeling	Gennemsnit	St. Afvigelse	Varians	Median
Gruppe	2236	1352	3%	9,33	1,27	1,61	9
Individuel	70527	43854	97%	8,70	1,58	2,48	9

Ministeriet for Videnskab,
Teknologi og Udvikling

Dette betyder, at eksaminander, der har været til gruppeeksamen, er blevet bedømt mere samlet end de, der har gået til individuel eksamen. Gruppeeksamener har altså en tendens til at gruppere sig mere omkring gennemsnittet end de individuelle eksemper, der viser en større spredning.

Fortolkning:

Der er ingen tvivl om, at gruppeprøver generelt bliver højere og 'smallere' bedømt. Men hvad årsagen er er svært at sige. Om dette er fordi bedømmelsen af den enkelte er ringere, eller at folk faktisk kan mere, viser modellen intet om. Blot at de bedømmes højere. Det er muligt, at der ikke sker en ordentlig bedømmelse af den enkelte, fordi denne forsvinder i mængden, men der er også mulighed for, at de studerende har lært mere, får højere karakter, kan nogenlunde det samme og får nogenlunde den samme karakter, fordi de *netop* er i gruppe. Det er ikke muligt på baggrund af de forhåndenværende data at se, hvor de ellers ville have landet, hvis det havde været individuelle eksaminer.

Mulige elaborerende analyser

For at lave en bedre analyse af gruppeprøver ville det nok være muligt at sammenligne de enkelte elever under de bestemte prøveformer over tid. Dvs. se på eleverne en for en og se om deres gennemsnitlige gruppekarakter adskiller sig fra de karakterer, den studerende normalt får i de individuelt bedømte eksaminer. Derved ville man få et fingerpeg om, hvorvidt gruppeprøver 'trækker' nogen i en retning, hvor de ikke bør være i forhold til de individuelt bedømte eksaminer. Denne analyse vil dog kræve lidt samtale med Aarhus Universitet om, hvordan de har anonymiseret deres studerende, men skulle kunne lade sig gøre.

**Ministeriet for Videnskab,
Teknologi og Udvikling**