

Sumarverkefni

Er hægt að mæla framfarir en ekki bara
utanaðbókarlærdóm?

Útdráttur

Inngangur: Spurningar í tutor-web eru stundum búnar til þannig að kennari semur fyrst einn haus og svo tvo hauga af fullyrðingum, einn með réttum svörum og einn með röngum svörum. Svo eru spurningarnar búnar til með því að velja slembið rétt svar og slurk af röngum svörum (kannski með allt/ekkert ofangreint er rétt). Ef nemandi lærir ekki bara utanað, heldur er að auka skilning á verkefnið, þá ættu að sjást smátt og smátt framfarir í einkunn líka fyrir nýjar fullyrðingar. Svo markmið rannsóknarinnar er að sjá hvort hægt er að skoða framfarir nemenda, en ekki bara utanbókarlærdóm. **Aðferð:** Í þessari rannsókn var byrjað á því að setja upp myndir tengt meðaltali réttra svara eftir fjölda spurninga svaraðra, til að sjá hvort það sæist framfarir eftir því sem fleiri spurningum er svarað. Eftir það voru stillt upp þrjú logistík líkön til að skoða hvort marktækni fjölda spurninga svaraðra er enn sterk með áhrif utanbókarlærdóms í sama líkani. Það var gert með því að hafa eitt líkann byggt í kringum fjölda spurninga svarað, annað byggt í kringum hlutfall rangra svarmöguleika sem hafa sést áður og í lokinn líkann sem er byggt á báðum breytunum. **Niðurstöður:** Frá meðaltals myndunum þá sást að það gæti verið hægt að skoða framfarir, en óvissa hvort framfarirnar sem sást voru aðallega áhrif utanbókarlærdómar. Þegar skoðað var líköninn, þá sást að fjöldi spurninga svarað fram að þessu var ennþá marktækt, með áhrifum utanbókarlærdóms innihaldinn í líkönunum, sem gæti verið að benda til þess að hægt sé að skoða þá framfarir í gegnum breytu tengt fjölda spurninga svarað. **Umræða:** Það var set up bootstrap fyrir gögninn til að fá gott áhorf á AUC og standardized brier skorinu fyrir líkönin þrjú, þar sást ekki mikill munur milli líkananna. Það gæti verið að segja að utanbókarlærdóms áhrif hlutfall rangra svara sem hafa sést áður er líkt styrkleikanum á að skoða fjölda spurninga svarað, sem sýnir stóru áhrif utanbókarlærdóms á líkur nemanda að svara næstu spurningu.

Efnisyfirlit

Útdráttur	2
1 Inngangur	4
2 Aðferð	4
2.1 Þáttakendur	4
2.2 Tölfræðileg úrvinnsla	4
2.2.1 Hugbúnaður	4
2.2.2 Gögn	4
2.2.3 Gagnavinnsla	5
2.2.4 Breytur	5
2.2.5 Aðferðarfræði við líkanasmið	6
3 Niðurstöður	7
3.1 lýsandi tölfræði	7
3.1.1 öll gögn	7
3.1.2 Stytt gögninn	8
3.1.3 Smá samantekt fyrir myndirnar	16
3.2 líkanasmið	16
3.2.1 breytur	16
3.2.2 val á líkani	17
3.2.3 Matsgæði líkana	17
3.2.4 Mat stika lokalíkana	21
4 Umræða	21
4.1 fyrsta líkanið	26
4.2 Annað líkanið	26
4.3 Þriðja líkanið	27
4.4 Smá loka niðurstöður	27
4.5 Smá upphát hugsanir um hugmyndir til að skoða í framtíðinni	27

1 Inngangur



Hugsað upp í kringum 2000 af honum Gunnari Stefánssyni, tutor-web er lærdóms umhverfi á netinu með fyrirlestrum, dæmi og krossaspurningar. Með einbeitingu á stærðfræði og tölfraði. Spurningar sem eru svaraðar eru svo geymdar til rannsóknar sem gæti komið fram seinna. (Þetta er góð byrjun, veit ekki hvort þörf er á meira fyrir þennan hóp)

Spurningar í tutor-web eru stundum búnar til þannig að kennari semur fyrst einn haus og svo tvo hauga af fullyrðingum, einn með réttum svörum og einn með röngum svörum. Svo eru spurningarnar búnar til með því að velja slembið rétt svar og slurk af röngum svörum (kannski með allt/ekkert ofangreint er rétt). Ef nemandi lærir ekki bara utanað, heldur er að auka skilning á verkefnið, þá ættu að sjást smátt og smátt framfarir í einkunn líka fyrir nýjar fullyrðingar. (Frá Gunnari þegar kynnt var verkefnið. Þarf að endur orða)



Markmið rannsóknarinnar fór því yfir hvort hægt væri að sjá framfarir nemenda, en ekki bara utanbókarlærdóm. Með því að skoða hvernig líkurnar á að svara spurningu breytist eftir því sem fleiri spurningum er svarað og skoðað fyrir sitthvor tilvik eftir því hvað af svarmöguleikunum hafa sést áður.



2 Aðferð

2.1 Þáttakendur

Úrtak rannsóknarinnar voru nemendur í líkindareikningur og tölfraði á vormisserinu 2020. Þar voru 293 nemendur sem svöruðu samtals 108.017 spurningum í heild.

2.2 Tölfræðileg úrvinnsla

2.2.1 Hugbúnaður

Öll tölfræðileg úrvinnsla, meðhöndlun gagna og líkanasmíð fór fram á forritunarmálinu R v4.0.0 í viðmóttinu RStudio v1.3.959



2.2.2 Gögn

Gögninn sem notast var við, voru annaðhvort tekin beint úr SQL gagnagrunni eða afhent sem .txt skrá. Frá SQL gagnagrunninum var fengið gagnasett fyrir öll svör nemanda, gagnasett fyrir allar spurningar og svo gagnasett fyrir stillingu nemenda innan við hvern fyrirlestur. Svo texta skrár með hash fyrir alla réttu svarmöguleikana og ranga svarmöguleikana. Eftir sameiningu, þá var haldið eftirfarandi breytum:

Tafla 1: Breytur sem voru eftir í gögnunum

Breyta	Tegund	Skýring
lectureId	Merkibreyta	Númer fyrirlesturs
studentId	Merkibreyta	Númer nemenda
questionId	Merkibreyta	Númer spurningar
correct	Flokkabreyta	Hvort svarað var rétt eða rangt
hash	Flokkabreyta	Hver svarmöguleikinn er
fsfat	Samfelld breyta	Fjöldi spurninga svarað fram að þessari spurningu
hsta	Flokkabreyta	Hef séð þetta rétta svar áður
hluta	Samfelld breyta	hlutfall rangra svarmöguleika sem hafa sést áður
timeDif	Samfelld breyta	Tímamunur séðan rétta svarið sást seinast
nicc	Flokkabreyta	Fjöldi vitlausa svarmöguleika
gpow	Samfelld breyta	Erfileika hraði fyrir uppkomandi spurningar
hluta2	Flokkabreyta	Discretized hlutfall rangra svarmöguleika sem hafa sést áður

2.2.3 Gagnavinnsla

Til að fá gagnasettið yfir í loka formið, þá var framkvæmt eftirfarandi aðgerðir.

Fyrsta skrefið var að tengja saman öll gagnasettin. Byrjandi með því að tengja svörinn og spurningarnar, svo hægt væri að tengja við rétту og röngu svörinn seinna. Eftir það var tengt við fyrirlestra stillingar nemenda til að finna *gpow* þeirra. Svo var tengt við röngu svörinn og reiknað var *hluta*, en þar var tapað u.þ.b. 6023 línur. Svo í lokinn var tengt rétту svörinn og reiknað hvort nemandinn hafi séð svarið áður eða ekki fyrir hverja spurningu.

Ákveðið var að taka minna en allt gagnasafnið, með því að skoða ekki spurningar sem koma eftir að nemandinn hefur svarað 100 sinnum. Þetta kemur frá þeirri hugsun að eftir að hafa svarað svona margar spurningar, þá er nemandinn ekki að sjá mikið nýtt, svo ekki er hægt að sjá framfarir við sambærilegar spurningar þar. (DJÚPARI RÖK FYRIR ÞETTA KEMUR SEINNA, SPURNING HVORT ÞAÐ ER EINHVER ÞÖRF FYRIR ÞETTA ÞÁ?)

2.2.4 Breytur

Það voru nokkrar breytur sem var þurft að búa til, þær voru *fsfat*, *hsta*, *hluta* og *hluta2*.

Fyrir *fsfat*, sem sagt “Fjöldi spurninga fram að þessu”, sem telur spurningarnar sem hafa verið svarað hingað til. Aðferðin til að búa til *fsfat* fór svona:

1. Raða safninu eftir tíma sem spurninginn byrjaði
2. fyrir hvern nemenda í hverjum fyrirlestri, telja upp frá 0 eftir röðinni.

Þar sem einhverjar detta út þegar bætt er við *hluta* og það eru 3 línur sem koma þegar AOTA+ spurningu er að ræða, þá er búið til *fsfat* áður sameinað við .txt skránnar með rétту og röngu svörinn.

Að næstu fyrir *hsta*, semsagt “hef séð þetta rétta svar áður”. Fyrir hvern nemenda, þá var fundið fyrsta skiptið sem nemandinn sá rétta svarmöguleikann, svo var sett að ef nemandinn sá það í fyrsta skiptið, þá hafði hann ekki séð svarmöguleikann áður, annars hefur nemandinn séð svarmöguleikann áður. Nema í tilvikum þar sem NOTA+ spurning er að ræða, semsagt “None of the above” er rétta svarið. Þá var sett að ef hlutfall rangra svarmöguleika sem nemandinn hafi séð áður væri 100%, þá er talið að nemandinn hafi séð rétta svarið áður.

Næst var sett upp hluta, semsagt “hlutfall rangra svarmöguleika sem hafa sést áður”. Þar var fyrst fundið fyrir hvern nemenda, hvort ranga svarið hafi sést áður og sett upp eins og *hsta*. Eftir það var tekið meðaltal rangra svara sem hafa sést áður til að fá hlutfallið. Sér hugsun þurfti að koma tengt spurningu með “all of the above” sem rangur möguleiki og “None of the above” sem rangur möguleiki. Fyrir fyrra tilvikið var skoðað hvort eitthvað af hinum röngu valmöguleikunum hafa sést áður, ef svo þá var hugsað eins og “all of the above” ranga svarið hafi sést áður. Fyrir “NOTA-” þá var skoðað hvort rétta svarið hafi sést áður, ef svo þá var hugsað eins og “NOTA-” svarmöguleikinn hafi sést áður.

Að lokum fyrir *hluta2* þá var skipt *hluta* upp í 5 jafn langa parta, semsagt tilvikin þegar 0%-20% af röngu svarmöguleikunum hafa sést áður, tilvikin fyrir 20%-40% og svo framvegis upp að 80%-100%.

2.2.5 Aðferðarfræði við líkanasmið

Það voru gerð þrjú “mixed effect logistic regression” líkön, þar sem aðalmunur þeirra er að:

- Fyrsta líkanið inniheldur víxláhrif milli *fsfat* og *hsta*, en inniheldur ekki *hluta2*.
- Annað líkanið inniheldur *hluta2*, en inniheldur ekki *fsfat*.
- Þriðja líkanið inniheldur *fsfat* og *hluta2*, en ekki víxláhrif milli *fsfat* og *hsta*.

Með þessum líkönum, væri hægt að skoða hvort það koma framfarir. Þar sem hugsunin byggist á því að ef *fsfat* er ennþá marktækt, þá eru framfarir að sjást. Því sem fleiri spurningar eru svarað, þá eru líkurnar á að næstu spurningu er svarað rétt að hækka. Hægt er að sjá það sem framfarir. Á móti kemur áhrif utanbókarlærdóms, sem kemur frá réttu og röngu svarmöguleikunum sem hafa komið áður, semsagt *hsta* og *hluta*. Svo hægt er að sjá hvort *fsfat* er ennþá marktækt ef tekið er með áhrif utanbókarlærdóms.

Líkonin þrjú eru hér að neðan, þar sem y stendur fyrir líkurnar á því að fá rétt svar og studentId_n er “mixed logistic effect” í þessu tilviki.

- Fyrsta líkanið: $\text{logit}(y) = \beta_0 + \text{fsfat} + \text{hsta}_k + \text{fsfat:hsta} + \text{nicc}_l + \text{gpow} + \text{lectureId}_m + \text{studentId}_n$
- Annað líkanið: $\text{logit}(y) = \beta_0 + \text{hluta2}_s + \text{hsta}_k + \text{nicc}_l + \text{gpow} + \text{lectureId}_m + \text{studentId}_n$
- Þriðja líkanið: $\text{logit}(y) = \beta_0 + \text{fsfat} + \text{hsta}_k + \text{hluta2}_s + \text{nicc}_l + \text{gpow} + \text{lectureId}_m + \text{studentId}_n$



3 Niðurstöður

3.1 lýsandi tölfræði



3.1.1 öll gögn

Tafla 2 inniheldur fjölda og hlutfall hvernar flokkabreytu fyrir öll gögninn.

Tafla 2: Hlutföll flokkabreytanna

	fjöldi	hlutfall
correct		
Rangt	16811	16.5%
Rétt	85183	83.5%
hsta		
Sést í fyrsta skipti	36915	36.2%
Hef séð svarið áður	65079	63.8%
hluta2		
0% - 20%	9698	9.5%
20% - 40%	6404	6.3%
40% - 60%	5362	5.3%
60% - 80%	16347	16%
80% - 100%	64183	62.9%
lectureId		
3082	23651	23.2%
3201	7109	7%
3202	4825	4.7%
3203	10982	10.8%
3204	9117	8.9%
3208	7287	7.1%
3209	5904	5.8%
3210	4712	4.6%
3211	4438	4.4%
3212	6553	6.4%
3213	5704	5.6%
3214	5480	5.4%
3215	6232	6.1%

Stórt hlutfall gagnanna sýnist vera spurningar sem voru svaraðar rétt, spurningar með rétta svarmöguleika sem nemendur hafa séð áður og spurningar þar sem nemandurnir hafa séð 80%-100% af röngu svarmöguleikunum áður. Samsagt ef skoðað eru öll gögninn, þá sést að nemendur eru ekki að sjá mikið af nýju efni, heldur mikið af endurtekningum.

Þar sem markmiðið er að skoða framfarir, þá er erfitt að sjá beinar framfarir yfir utan-

Tafla 3: hlutfall gagnanna eftir x spurningar svaraðar

limit	FY	HY
50	29990	29.4%
100	9155	9%
150	3072	3%
200	1127	1.1%
250	331	0.3%
300	86	0.1%

bókarlærdóm þegar nemendur hafa svarað mörgum spurningum, þar sem flestir svarmögu-
leikar hafa sést áður. Tafla 3 inniheldur hlutfall gagnanna sem eru eftir að hafa svarað
ákveðið mikið af spurningum, til að fá innsæi hve mikið af gögnunum er eftir x spurningar
svaraðar.

Það sést að 29.4% af gögnunum eru svör eftir 50 spurningar svarað, en bara 9% gagnanna
eftir 100 spurningum. Svo til að geta fengið betri skoðun á gögnunum, þá var ákveðið að
skipta gögnunum upp í tvö gagnasöfn, í fyrri er ekki leift meiri 100 fsfat og í seinni er ekki
leift meira en 50 fsfat. Þau gagnasöfn eru notuð héðan í frá.

3.1.2 Stytt gögninn

Tafla 4 og 5 eru alveg eins hlutfallstöflur og 2, nema með styttri gögn.

Tafla 4: Hlutfall flokkabreytana fyrir fyrstu 100 svör

	fjoldi	hlutfall
correct		
Rangt	15954	17.2%
Rétt	76856	82.8%
hsta		
Sést í fyrsta skipti	36290	39.1%
Hef séð svarið áður	56520	60.9%
hluta2		
0% - 20%	9652	10.4%
20% - 40%	6392	6.9%
40% - 60%	5331	5.7%
60% - 80%	15956	17.2%
80% - 100%	55479	59.8%
lectureId		
3082	17648	19%
3201	7033	7.6%
3202	4825	5.2%

3203	10266	11.1%
3204	8429	9.1%
3208	7023	7.6%
3209	5751	6.2%
3210	4709	5.1%
3211	4432	4.8%
3212	5892	6.3%
3213	5575	6%
3214	5288	5.7%
3215	5939	6.4%

Tafla 5: Hlutfall flokkabreytana fyrir fyrstu 50 svör

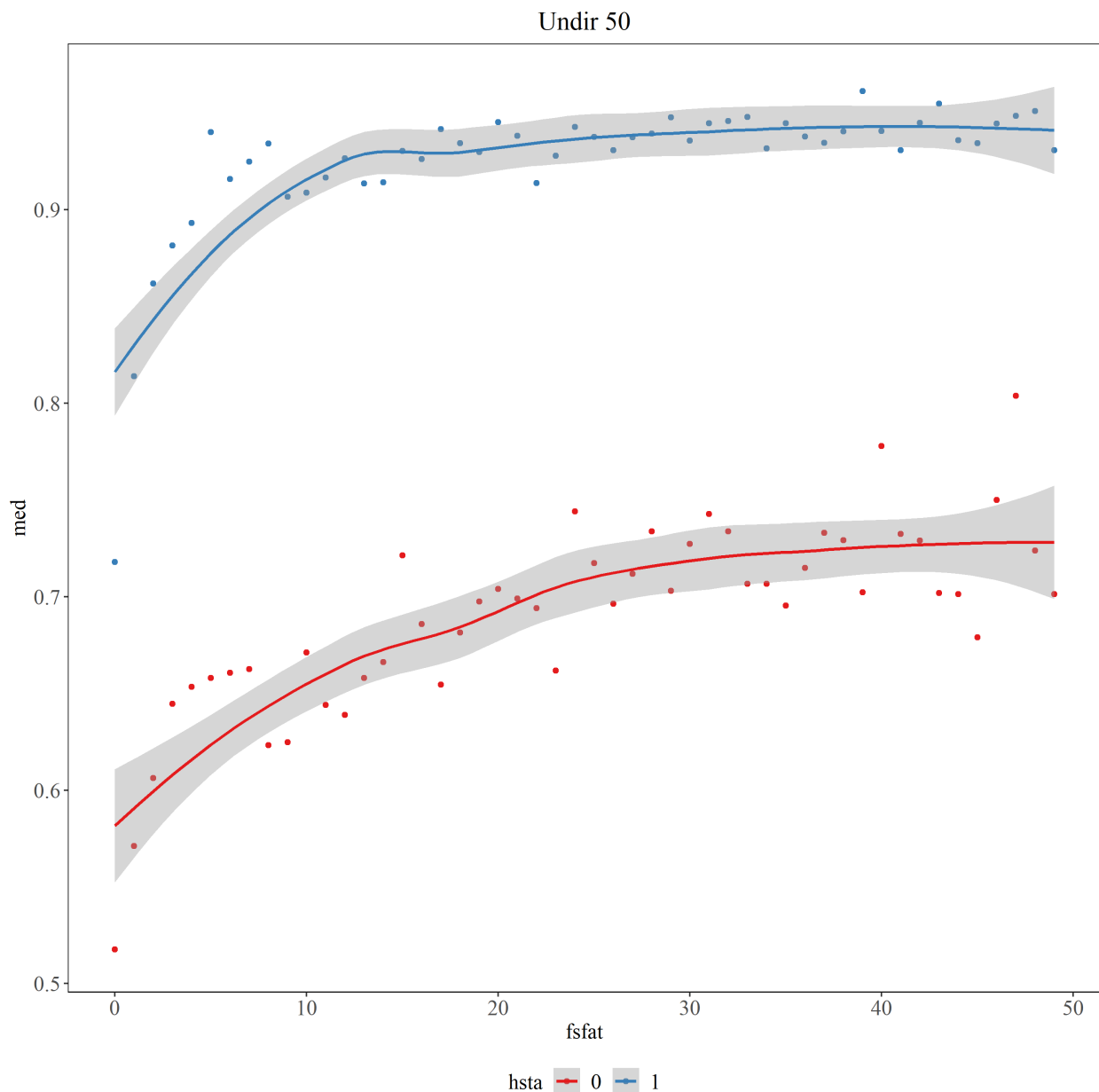
	fjoldi	hlutfall
correct		
Rangt	13979	19.4%
Rétt	57996	80.6%
hsta		
Sést í fyrsta skipti	32933	45.8%
Hef séð svarið áður	39042	54.2%
hluta2		
0% - 20%	9570	13.3%
20% - 40%	6266	8.7%
40% - 60%	5163	7.2%
60% - 80%	13751	19.1%
80% - 100%	37225	51.7%
lectureId		
3082	10549	14.7%
3201	5991	8.3%
3202	4494	6.2%
3203	7555	10.5%
3204	6498	9%
3208	5521	7.7%
3209	4543	6.3%
3210	4249	5.9%
3211	3959	5.5%
3212	4628	6.4%
3213	4747	6.6%
3214	4407	6.1%
3215	4834	6.7%

Frá töflu 4 þá sést að hlutfall rangra svarmöguleika hækkar aðeins, sömuleiðis með rétt svör

sem sést í fyrsta skiptið. En ekki endilega það mikið.

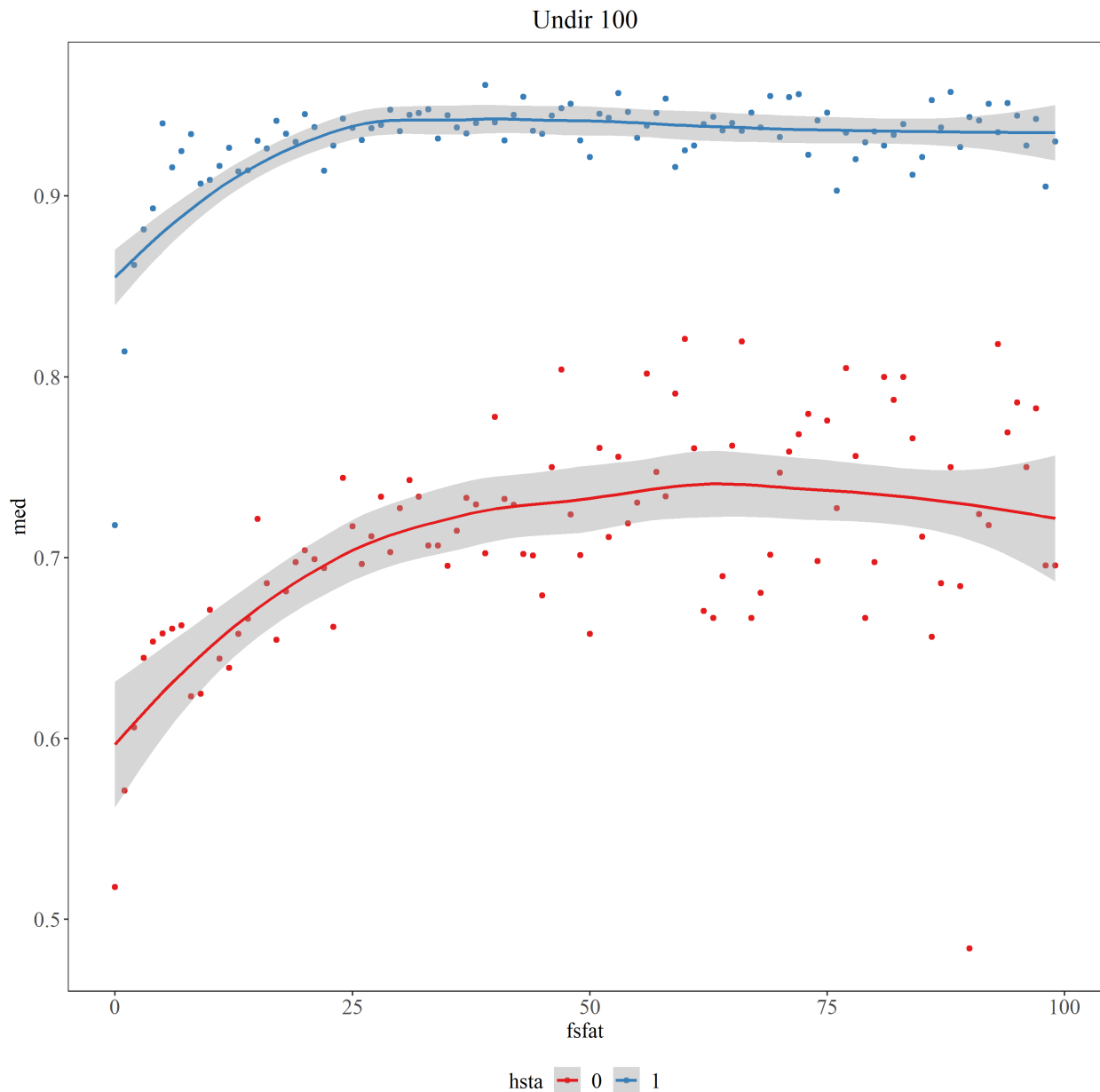
Fyrir töflu 5 þá er hlutfall rangra svarmöguleika aftur aðeins stærra, svo er hvort rétta svarið hafi sést áður eða ekki er miklu nær helmingi en áður. Nemendur eru enn að sjá mikið af eindurtekningum, en það er orðið jafnara.

Til að geta séð hvernig árangur kemur með tíma, þá var skoðað meðaltal réttra svara fyrir hvert *fsfat* upp að 50 og svo upp að 100. Semsagt hvað er prósentu spurninga sem var svarað rétt hjá spurningu nr. *x*, þetta er svo skipt fyrir bæði réttu svör sem eru að sjást í fyrsta sinn og þau sem hafa sést áður. Mynd 1 sýnir tilvikið þegar skoðað er bara upp að 50 spurningum svarað, á meðan mynd 2 skoðar tilvikið þegar upp að 100 spurningum er svarað.



Mynd 1: Upp að 50 meðaltal





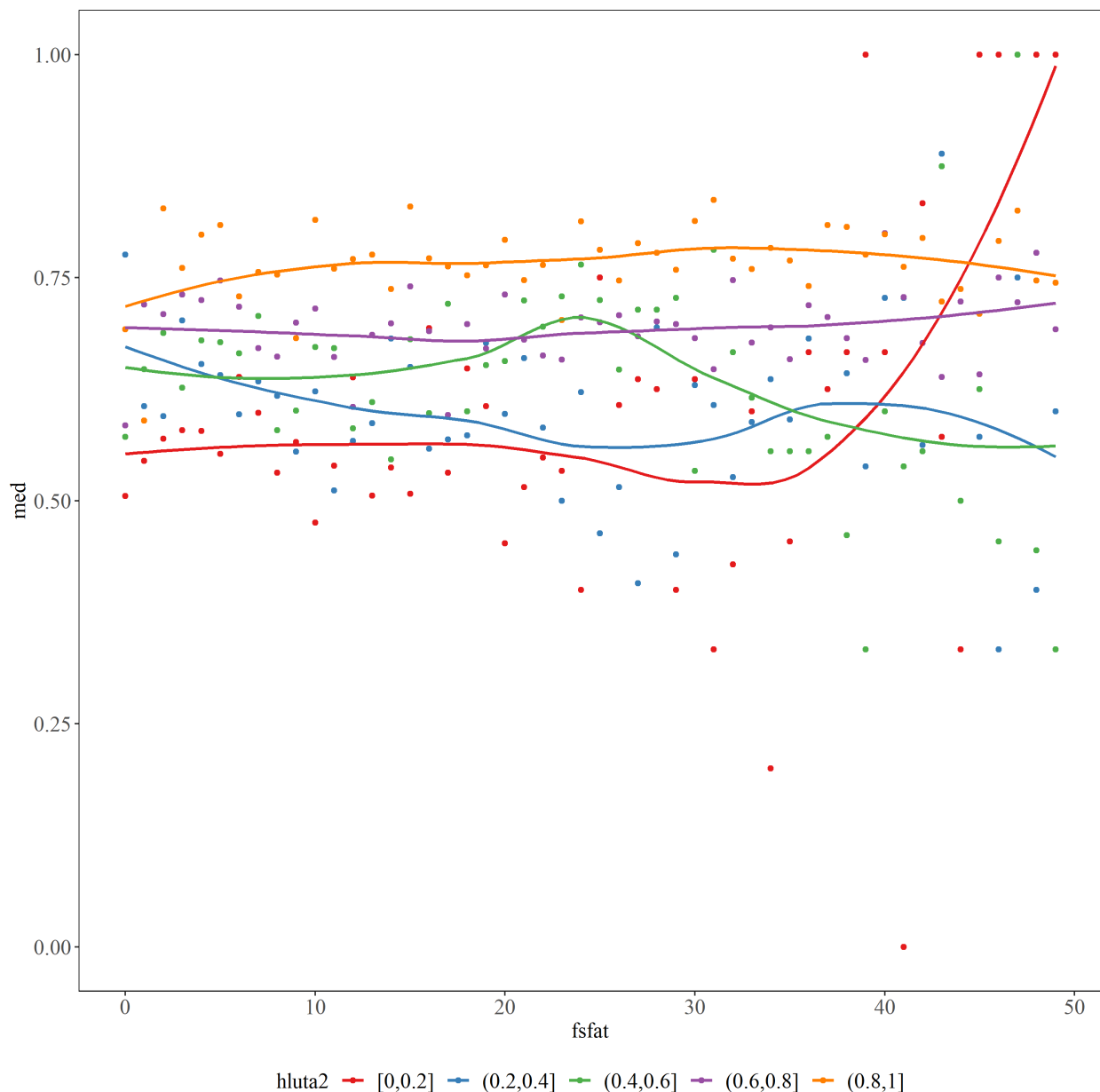
Mynd 2: Upp að 100 meðaltal



Frá mynd 1 þá sést greinilega að einhverjar framfarir eru að koma fram þegar rétti svarmöguleikinn er að koma í fyrsta skiptið, þar sem meðaltölinn eru að hækka því lengra sem er farið. Fyrir mynd 2 þá sést að þegar það er farið lengra, þá fara tilvikin þar sem rétta svarið sést í fyrsta skiptið að jafnast út og byrjar að falla aðeins niður. Afhverju þetta jafnast svona út þegar lengra er farið gæti verið útaf því að dreifingin er óstöðugri þar sem lengra er farið, þar sem færri spurningar með réttu svarmöguleikana sem ekki hafa sést áður koma upp. Annað sem gæti verið hér að baki er að það sem sást á mynd 1 var ekki endilega beinar framfarir, en frekar það að nemandurnir eru að setja röngu svarmöguleikana á minnið. Þá er línan í myndinni að jafnast því að nemandirnir eru ekki að sjá mikið af nýjum möguleikum

röngum svarmöguleikum lengur, svo línan jafnast út.

Hugmynd er að skoða líka sambærilega mynd, nema bara fyrir tilvikið þegar þetta eru ný svör og skipt upp eftir hlutfall rangra svarmöguleika sem hafa sést áður, þetta er gert til að sjá hvort innan við hvers hlutfallshóp það eru framfarir eða eru aðall framfarirnar stökkin frá einum hlutfallshópi yfir á næsta. Mynd 3 sýnir tilvikið þegar skoðað er upp að 50 spurningum svarað.

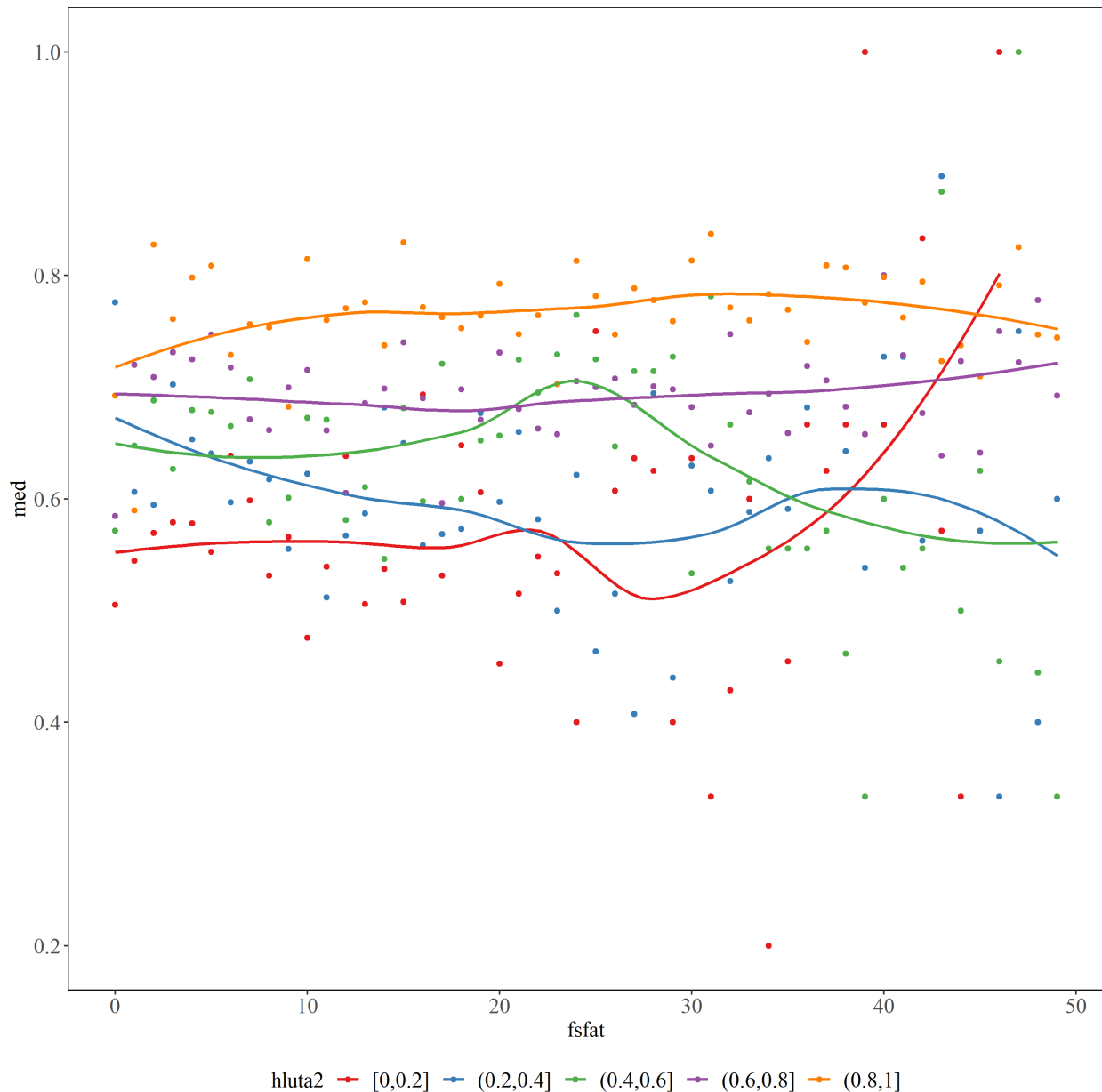


Mynd 3: meðaltal fyrir hlutfall, upp að 50



Hér sést að þegar við skoðum eftir hlutföllum rangra svara, þá eru línunar orðnar jafnari en vöxturinn sem sást fyrst, það er smá vöxtur hjá 60%-80% og 80%-100%, eitthvað fall hjá

20%-40% og 40%-60% og svo rosalegt stökk hjá 0%-20%. Óstöðugleikinn gæti verið að koma frá því að það eru mjög fáar breytur fyrir aftan sumt af punktunum, til að jafna það út þá endurgerði ég myndina með þeim mun að fjarlægja punkta sem eru bara með 1 “observation” fyrir aftan þá. Mynd 4 er því sama og mynd 3 nema með fjarlægða punkta með ekki meira en 1 “observation” fyrir aftan það.

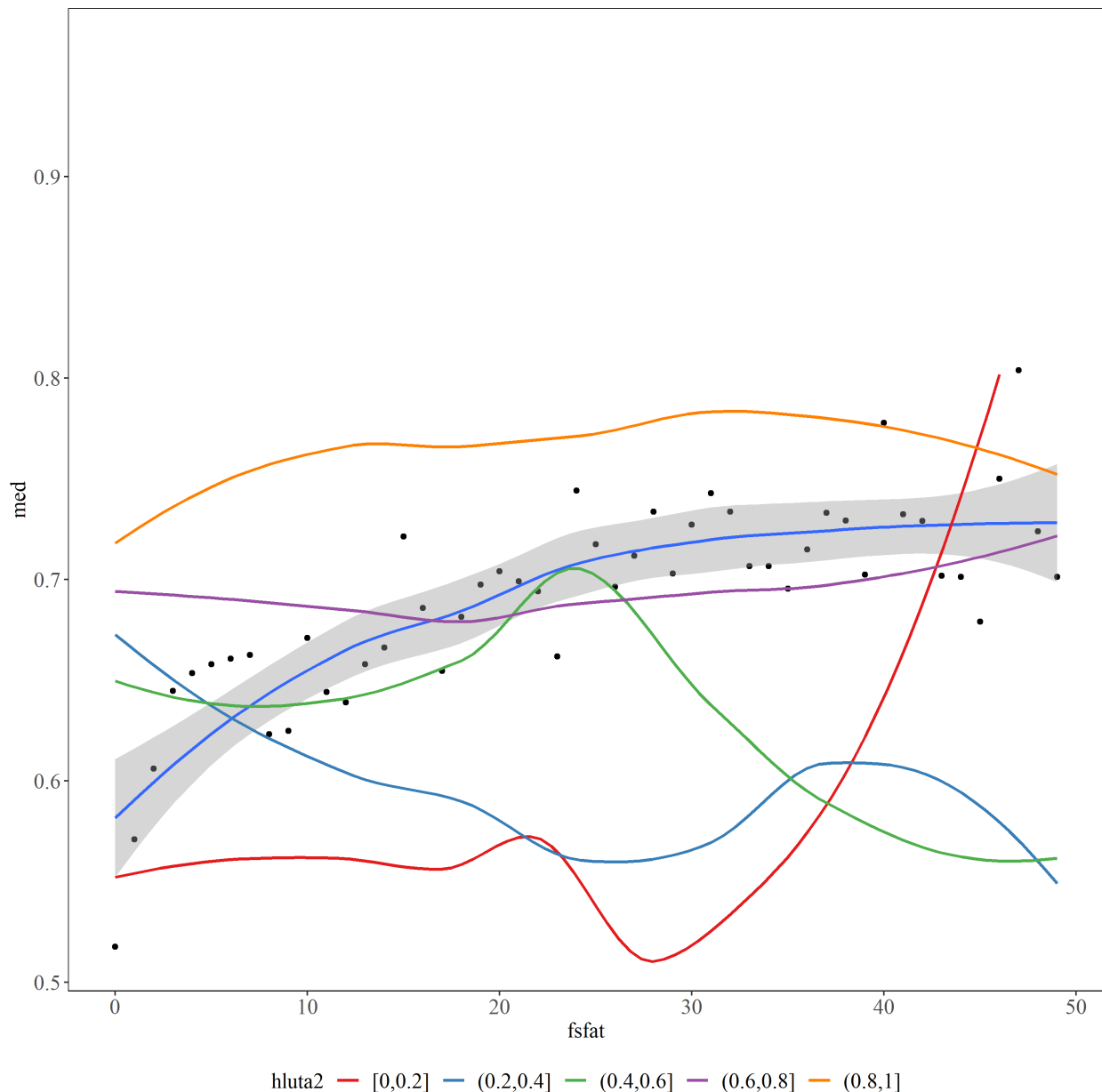


Mynd 4: meðaltal fyrir hlutfall, upp að 50. Með minnsta kosti 2 svör á hverjum punkti

Stóra stökkið hjá 0%-20% tilvikinu er aðeins stöðugra, en fyrir utan það, þá er ekki nein breyting hér.

Til að fá betri sjónarhorn á hlutföllin, þá gæti verið gott að bæta við línurnar hjá hlutföllunum

við mynd 1, sem sagt hafa línuna fyrir tilvikið þegar nemandinn hefur ekki séð rétta svarið áður úr mynd 1 með línurnar hjá hlutfall röngu svarmöguleikana fram að þessu úr mynd 4 hliðiná. Dökkbláa línun er upprunalega línun frá mynd 1, á meðan hinar línurnar eru fyrir sitthvora hlutfalls hópana.

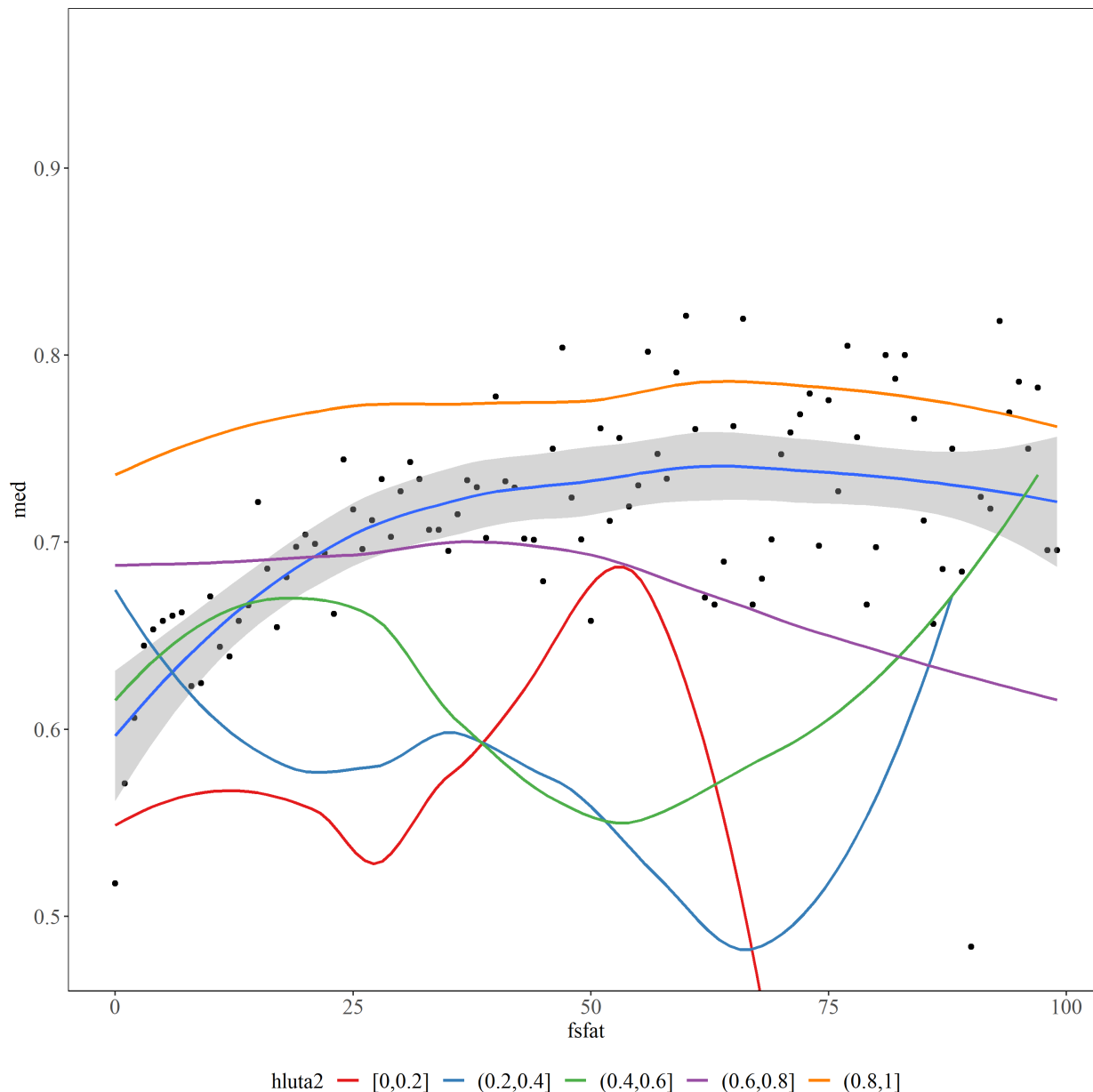


Mynd 5: meðaltal fyrir hlutfall, upp að 50. Með meðaltal hjá fyrsta skiptið séð rétta svar

Hér sést að á meðan upprunalega línun úr mynd 1 er vaxandi, þá eru hinar línurnar ekki að vaxa jafn mikið. Miklar sveiflur sjást hjá minni hlutföllunum, sem eru líklegast að koma frá því að ekki eru mikið af spurningum sem eru að koma upp með svona fáa ranga svarmöguleika. Það sést samt hjá 60%-80% að það er smá fall en svo byrja að koma framfarir í gang og hjá

80%-100% þá eru einhverjar framfarir í gangi en það jafnar sig svo aðeins út í lokinn.

Að sömu leiti getum við skoðað sömu mynd fyrir tilvikið þegar farið er upp að 100 spurningum svarað



Mynd 6: meðaltal fyrir hlutfall, upp að 100. Með meðaltal hjá fyrsta skiptið séð rétta svarið

Það lítur út fyrir að vera einhverjar framfarir í gangi þar, hjá tilvikunum með 20%-60% þá er eitthvað að vaxa, spurning hvort þetta er að vaxa almennilega eða hvort þetta er að koma frá því tilviki að mjög fáar breytur eru að toga þeim upp. Fyrir tilvikin þegar 60%-80% af röngu svarmöguleikunum hafi sést áður þá er það að mesta lagi stöðugt með hægt fall niður og 80%-100% er mjög jafnt yfir allar línurnar.

3.1.3 Smá samantekt fyrir myndirnar

Þegar við skoðuðum fyrstu myndina, þá leit út fyrir að einhverjir jákvæðir vextir voru að koma með fjölda spurninga svarað, þá kom spurningin hvort þessi vöxtur eru framfarir eða utanbókarlærdómur röngu svarana, þar sem fyrir því fleiri spurningum sem nemendarnir hafa svarað, því fleiri rangra svarmöguleika þeir hafa séð áður, það gæti bent til þess að vöxturinn er aðallega tengur utanbókdarlærdómi. Farið var því að skoða sambærilegt fyrir hlutföllin og þar sást að línurnar voru jafnari en upprunalega línán, en með einhverjum litlum framförum þar ennþá til staðar.

Til að sjá hvort framfarirnar eru ennþá marktækar þá eru skoðuð þrjú “logistic mixed effect” líkönn, þar sem nokkrar breytur eru alltaf til staðar hjá þeim öllum, en það er skipt upp þannig að fyrsta líkanið inniheldur fjölda spurninga svarað fram að þessu, næsta líkanið inniheldur hlutfall rangra svara fram að þessu flokkana og í lokinn er þriðja líkanið sem skoðar bæði fjölda spurninga fram að þessu og hlutfall rangra svara fram að þessu. Hugmyndin er að ef fjöldi spurninga fram að þessu er enn marktæk eftir að bæta við hlutfall rangra svarmöguleika, þá gæti það verið að benda til þess að hægt sé að sjá framfarir nemenda.

3.2 líkanasmið

3.2.1 breytur

Það voru gerð þrjú líkönn til að skoða, þar sem aðal munurinn er hvort notað er *fsfat*, *hluta2* eða bæði. Hér sýnir tafla 6 allar breytturnar sem eru í líkönunum, frá þeim sem eru í öllum líkönunum og þeim sem er skipt út á milli þeirra.

Tafla 6: Breytur notað í líkonunum

Breyta	Tegund	Flokkar	
Fixed effect fyrir öll líkön			
lectureId	Flokka	3082, 3201 - 3204, 3208 - 3215	
hsta	Flokka	Sést í fyrsta skipti, Hef séð áður	
nicc	Flokka	1, 2, 3, 4, 5, 6, 7	
gpow	Samfelld		
Mixed effect fyrir öll líkön			
studentId	Flokka	VEIT EKKI HVAÐ ÉG SET HÉR	
Fixed effect breytist eftir líkani			
fsfat	Samfelld		
hluta2	Flokka	[0-0.2], (0.2-0.4], (0.4-0.6], (0.6-0.8], (0.8-1]	

Þar sem aðal munur líkananna þriggja er:

- Fyrsta: $fsfat + fsfat:hsta + \dots$

Tafla 7: anova milli fyrsta og annað líkanið

term	npars	AIC	BIC	logLik	deviance	statistic	df	p.value
Fyrsta	24	66158.84	66385.36	-33055.42	66110.84			
Annað	26	65783.86	66029.25	-32865.93	65731.86	378.9829	2	0



Tafla 8: anova milli annað og þriðja líkanið

term	npars	AIC	BIC	logLik	deviance	statistic	df	p.value
Annað	26	65783.86	66029.25	-32865.93	65731.86			
Þriðja	27	65338.01	65592.85	-32642.01	65284.01	447.8451	1	0

- Annað: hluta2 + ...
- Þriðja: fsfat + hluta2 + ...

3.2.2 val á líkani

Líkanið var byggt ofaná líkani hans Þórarinnss Jónmundsson (SPURNING HVERNIG HÆGT SÉ AÐ BENDA TIL ÞESS), með auka breytturnar *hluta2* og tímaröðina *fsfat*



Fyrsta líkanið: Til að byrja með var skoðað hvernig væri að bæta við *fsfat* og víxláhrif á milli *fsfat* og *hsta*. Til að sjá hvernig framfarir eru í gangi. Þar sem hugsað er að jákvæð áhrif hjá *fsfat* gæti verið séð sem framfarir nemenda.

Vandi kom upp frá hugsuninni hvort þetta voru framfarir, eða bara utanbókarlærdómur sem var fyrir aftan *fsfat* svo því var prufað að skipta út *hluta2* og *fsfat* til að sjá hvor hefur meiri áhrif.

Annað líkanið: Hér var notað *hluta2* í stað *fsfat* og víxláhrif *hsta* og *fsfat*. Það var svo skoðað muninn á milli þeirra í gegnum anova til að sjá hvort væri sterkari, þar sést í töflu 7 að samkvæmt anova þá er annað líkanið sterkari en fyrri líkanið.

Þriðja líkanið: Hér var notað *hluta2* og *fsfat* saman, til að sjá hvort enn er hægt að sjá framfarir með *fsfat* þegar *hluta2* er innifallið enn í líkaninu. Að sömu leiti var gert anova milli þriðja og annars líkansins. Þá sést skv. töflu 8 að þriðja líkanið er sterkari en annað líkanið.

3.2.3 Matsgæði líkana



Tafla 9 sýnir hér frammistöðu líkannana þriggja eftir AUC og standardized brier skorinu, eftir að hafa gert “optimism correction” með bootstrappi um 2500 bootstrap ítranir. Það sést hér að munurinn á milli líkannana er ekki það mikill, þar sem stærsti munurinn er hjá fyrsta og þriðja líkaninu og það er ekki meira en 0.0056 hjá AUC og 0.01 hjá standardized brier skorinu. Þá lítur út fyrir að það er ekki mikill munur þar með tilliti til AUC og standardized

Tafla 9: Optimism corrected AUC og Standardized brier skorið hjá líkonunum þremur

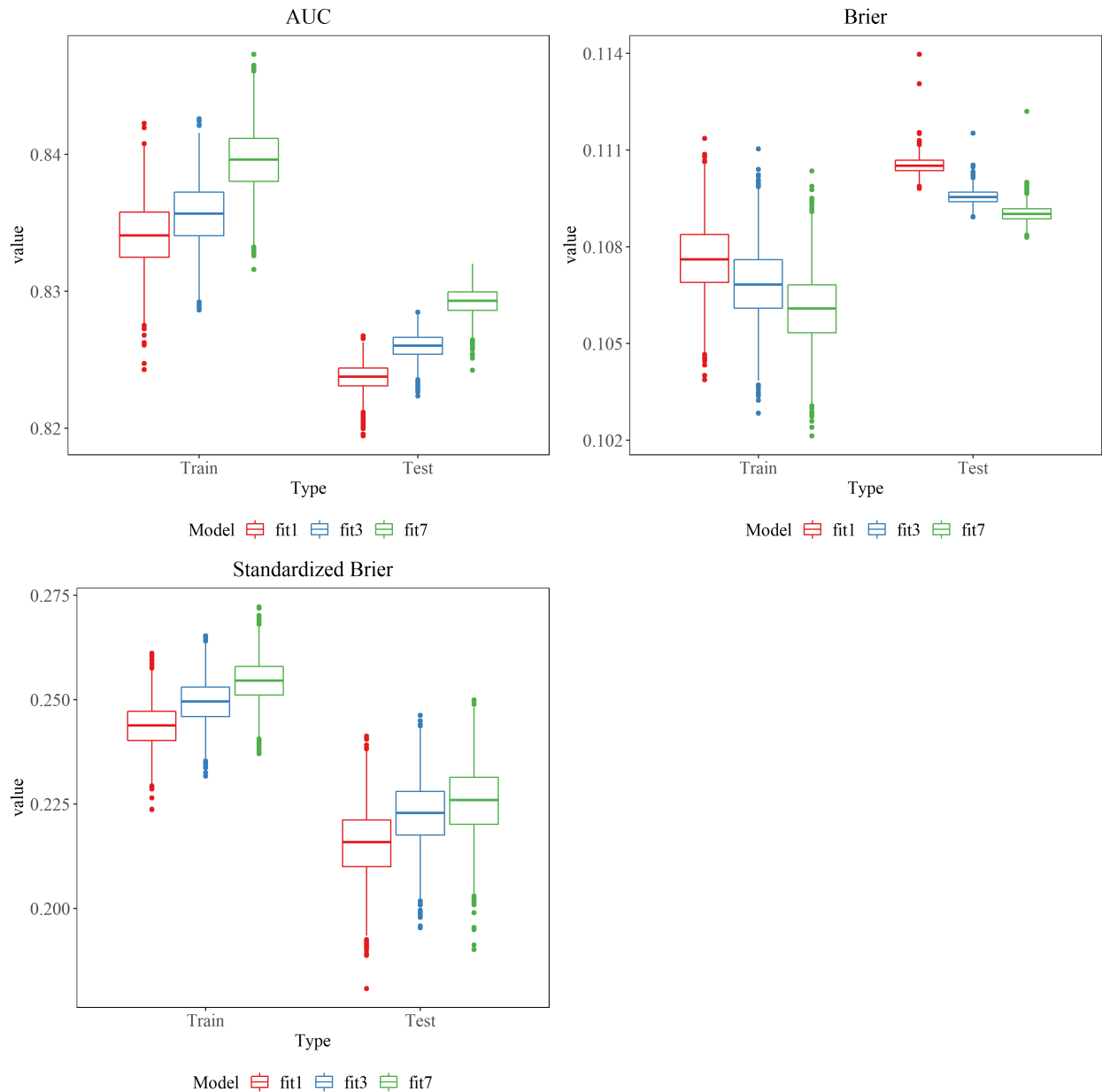
	AUC	stBrier
fit1	0.8197	0.2073
fit3	0.8223	0.2147
fit7	0.8254	0.2174

Tafla 10: optimism-ið fyrir AUC og standardized brier skorið hjá líkönunum þremur

	AUC	stBrier
fit1	0.0104	0.0283
fit3	0.0096	0.0268
fit7	0.0104	0.0289

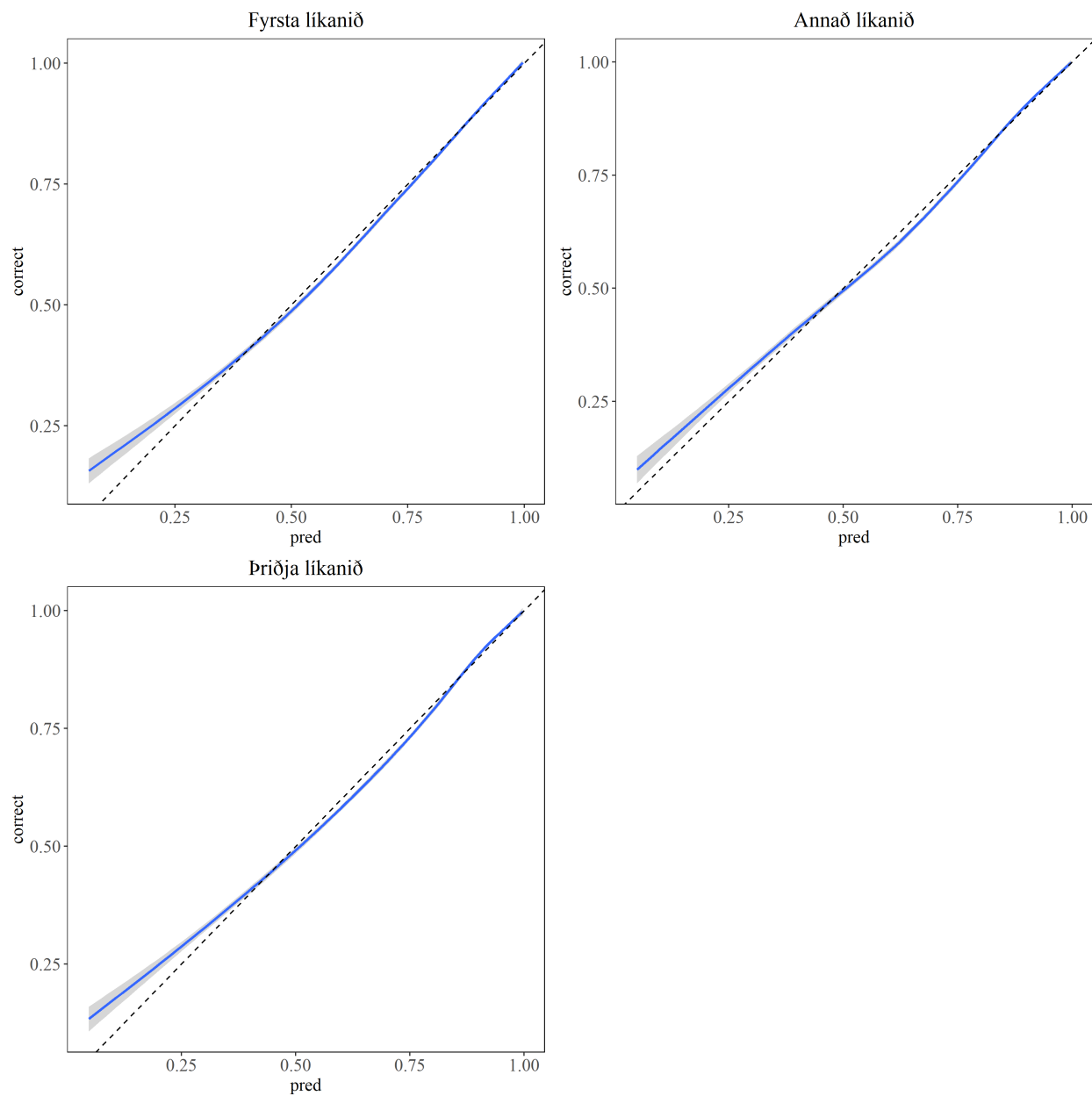
Brier. Til að bæta þá inniheldur tafla 10 “optimism” fyrir líkönin 3, þar sést að “optimism” er rosalega líkt hjá þeim öllum þremur.

Á mynd 7 sést svo hvernig AUC og standardized brier skorið dreifðist fyrir hverja ítrun. Þar sést að hvert líkann er aðeins hærra en það fyrra og að fyrir AUC þá er dreifingin á “Train” gildunum miklu meiri en hjá “Test” gildunum, en það er svo öfugt fyrir standardized brier, því þar er dreifingin hjá “Test” meiri en hjá “Train” gildunum.



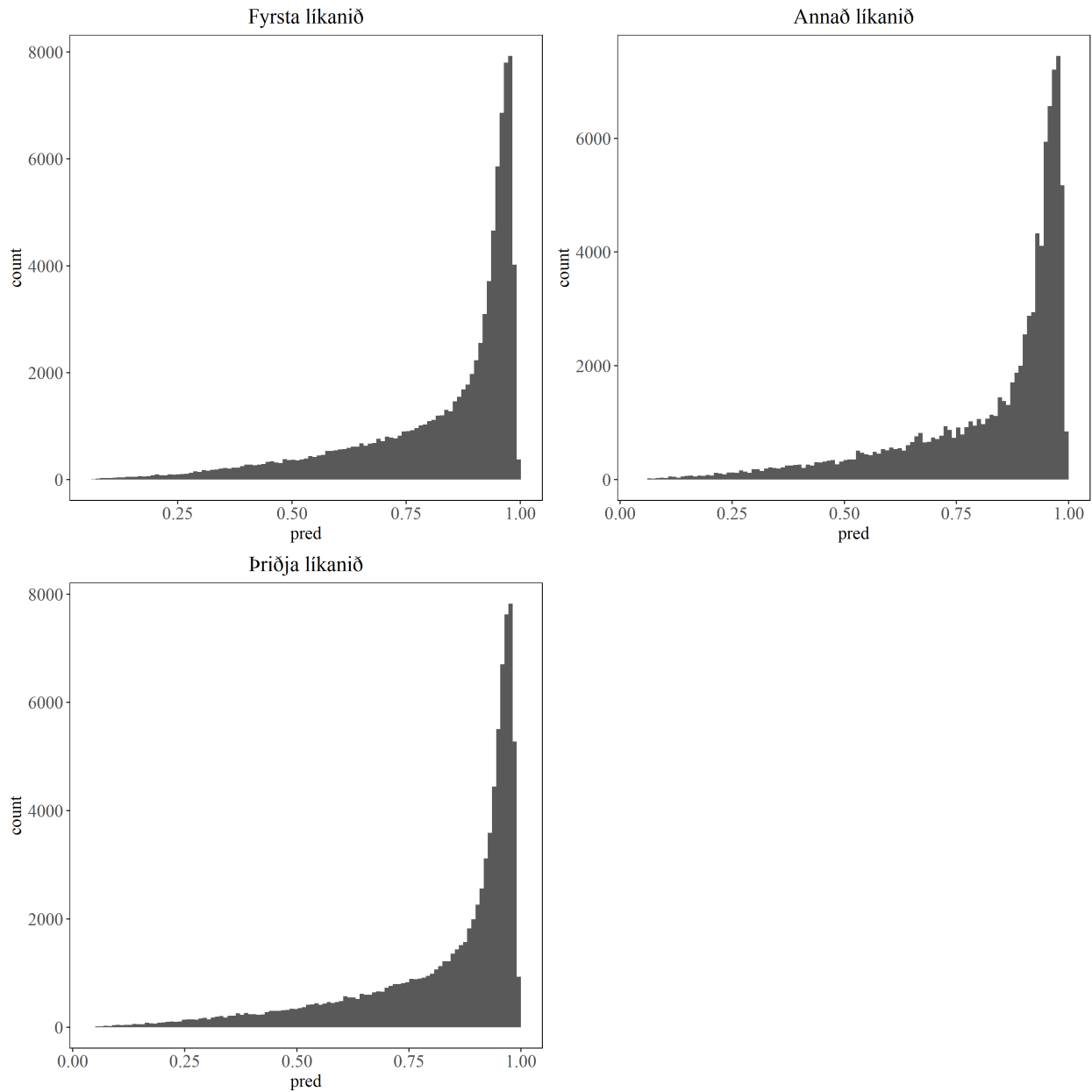
Mynd 7: Dreifing AUC og Standardized brier fyrir hvert líkann, hjá “Train“ og “Test“ bootstrap gildunum

Mynd 8 er “calibration” mynd fyrir líköninn þrjú, þar sem sést að líkönin þrjú eru öll ágætlega bein þar og mjög lík.



Mynd 8: Calibration plot fyrir líköninn þrjú

Mynd 9 sýnir svo hvernig “prediction” lítur út fyrir líkönin þrjú.



Mynd 9: Histogram fyrir predictionin fyrir líköninn þrjú

Að bættu þá eru allir stikarnir marktækir miðað við Anova III, semsagt frá:

3.2.4 Mat stika lokalíkana

4 Umræða

Markmið verkefnisins er að sjá hvort hægt sé að mæla framfarir, en ekki bara utanbókarlærdóm. Til að gera það var byrjað að sjá hvort hægt væri að finna jákvæðar breytingar þegar

Tafla 11: Anova fyrir fyrsta líkanið

term	statistic	df	p.value
(Intercept)	0.5933	1	0.4411
fsfat	1270.7712	1	0.0000
hsta	3108.9751	1	0.0000
nicc	144.7159	6	0.0000
gpow	12.9246	1	0.0003
lectureId	1984.8236	12	0.0000
fsfat:hsta	81.2268	1	0.0000

Tafla 12: Anova fyrir annað líkanið

term	statistic	df	p.value
(Intercept)	0.0971	1	0.7553
hluta2	2057.7959	4	0.0000
hsta	4158.8876	1	0.0000
nicc	248.7301	6	0.0000
gpow	11.1665	1	0.0008
lectureId	1343.5644	12	0.0000

Tafla 13: Anova fyrir þriðja líkanið

term	statistic	df	p.value
(Intercept)	0.4518	1	0.5015
hluta2	889.1891	4	0.0000
fsfat	430.7083	1	0.0000
hsta	3603.8987	1	0.0000
nicc	200.9407	6	0.0000
gpow	10.9312	1	0.0009
lectureId	1681.5027	12	0.0000

Tafla 14: Stuðlatafla lokalíkans fyrir fyrsta líkanið

	Gagnalíkindahlutfall	Hallastuðull	SE	z-próf	p
Skurðpunktur	1.1393	0.1304	0.1693	0.7703	0.4411
fjöldi spurninga fram að þessu	1.0252	0.0249	0.0007	35.6479	0.0000
Hef séð rétta svarmöguleikann áður					
hef ekki séð rétta svarmöguleikann áður (Viðmið)					
Hef séð rétta svarimöguleikann áður	8.1243	2.0949	0.0376	55.7582	0.0000
Fjöldi rangra svarmöguleika					
1 rangur svarmöguleiki (Viðmið)					
2 rangir svarmöguleikar	0.7428	-0.2973	0.1682	-1.7672	0.0772
3 rangir svarmöguleikar	0.4767	-0.7408	0.1591	-4.6570	0.0000
4 rangir svarmöguleikar	0.5294	-0.6361	0.1615	-3.9387	0.0001
5 rangir svarmöguleikar	0.4437	-0.8125	0.1611	-5.0447	0.0000
6 rangir svarmöguleikar	0.4165	-0.8758	0.1610	-5.4398	0.0000
7 rangir svarmöguleikar	0.4217	-0.8634	0.1601	-5.3940	0.0000
gpow	1.1339	0.1256	0.0349	3.5951	0.0003
hvaða fyrirlestur					
Fyrirlestur a (Viðmið)					
Fyrirlestur b	3.2156	1.1680	0.0454	25.7062	0.0000
Fyrirlestur c	5.2541	1.6590	0.0564	29.3990	0.0000
Fyrirlestur d	1.6582	0.5058	0.0378	13.3756	0.0000
Fyrirlestur e	2.3876	0.8703	0.0408	21.3331	0.0000
Fyrirlestur f	3.2017	1.1637	0.0471	24.6997	0.0000
Fyrirlestur g	3.8987	1.3607	0.0513	26.5127	0.0000
Fyrirlestur h	4.5382	1.5125	0.0590	25.6203	0.0000
Fyrirlestur i	3.9299	1.3686	0.0591	23.1653	0.0000
Fyrirlestur j	2.6342	0.9686	0.0499	19.4148	0.0000
Fyrirlestur k	3.1906	1.1602	0.0527	22.0129	0.0000
Fyrirlestur l	3.3745	1.2162	0.0524	23.2282	0.0000
Fyrirlestur m	2.6292	0.9667	0.0491	19.6750	0.0000
víxláhrif fjöldi spurninga og hafa séð rétta svarmöguleikann	0.9910	-0.0090	0.0010	-9.0126	0.0000

Tafla 15: Stuðlatafla loka líkans fyrir annað líkanið

	Gagnalíkindahlutfall	Hallastuðull	SE	z-próf	p
Skurðpunktur	1.0549	0.0535	0.1716	0.3116	0.7553
Hlutfall rangra svarmöguleika sem hafa sést áður					
0%-20% (viðmið)					
20%-40%	1.3998	0.3364	0.0372	9.0303	0.0000
40%-60%	1.7783	0.5756	0.0423	13.6224	0.0000
60%-80%	2.2378	0.8055	0.0326	24.7372	0.0000
80%-100%	4.1602	1.4256	0.0328	43.4913	0.0000
Hef séð rétta svarmöguleikann áður					
hef ekki séð rétta svarmöguleikann áður (Viðmið)					
Hef séð rétta svarmöguleikann áður	5.2177	1.6521	0.0256	64.4894	0.0000
Fjöldi rangra svarmöguleika					
1 rangur svarmöguleikar (Viðmið)					
2 rangir	0.8126	-0.2075	0.1711	-1.2128	0.2252
svarmöguleikar					
3 rangir	0.5294	-0.6359	0.1620	-3.9258	0.0001
svarmöguleikar					
4 rangir	0.5412	-0.6140	0.1644	-3.7353	0.0002
svarmöguleikar					
5 rangir	0.4794	-0.7352	0.1639	-4.4856	0.0000
svarmöguleikar					
6 rangir	0.4055	-0.9025	0.1637	-5.5123	0.0000
svarmöguleikar					
7 rangur	0.4000	-0.9164	0.1629	-5.6253	0.0000
svarmöguleikar					
gpow	1.1253	0.1180	0.0353	3.3416	0.0008
hvaða fyrirlestur					
Fyrirlestur a (Viðmið)					
Fyrirlestur b	2.7013	0.9938	0.0452	21.9735	0.0000
Fyrirlestur c	3.0992	1.1312	0.0555	20.3642	0.0000
Fyrirlestur d	1.5008	0.4060	0.0378	10.7339	0.0000
Fyrirlestur e	2.0314	0.7087	0.0406	17.4496	0.0000
Fyrirlestur f	2.5027	0.9174	0.0468	19.6123	0.0000
Fyrirlestur g	3.4851	1.2485	0.0513	24.3572	0.0000
Fyrirlestur h	3.1516	1.1479	0.0584	19.6488	0.0000
Fyrirlestur i	2.7760	1.0210	0.0585	17.4657	0.0000
Fyrirlestur j	1.9545	0.6702	0.0494	13.5788	0.0000
Fyrirlestur k	2.6130	0.9605	0.0523	18.3552	0.0000
Fyrirlestur l	2.6495	0.9744	0.0519	18.7746	0.0000
Fyrirlestur m	2.0978	0.7409	0.0487	15.2210	0.0000

Tafla 16: Stuðlatafla lokalíkans fyrir þriðja líkanið

	Gagnalíkindahlutfall	Hallastuðull	SE	z-próf	p
Skurðpunktur	0.8907	-0.1157	0.1722	-0.6722	0.5015
Hlutfall rangra svarmöguleika sem hafa sést áður					
0%-20% (viðmið)					
20%-40%	1.3027	0.2645	0.0376	7.0381	0.0000
40%-60%	1.5874	0.4621	0.0428	10.8062	0.0000
60%-80%	1.8009	0.5883	0.0343	17.1750	0.0000
80%-100%	2.9125	1.0690	0.0368	29.0737	0.0000
fjöldi spurninga fram að þessu	1.0126	0.0125	0.0006	20.7535	0.0000
Hef séð rétta svarmöguleikann áður					
hef ekki séð rétt svarmöguleikann áður (Viðmið)					
Hef séð rétta svarimöguleikann áður	4.7510	1.5583	0.0260	60.0325	0.0000
Fjöldi rangra svarmöguleika					
1 rangur svarmöguleikar (Viðmið)					
2 rangir	0.7732	-0.2573	0.1711	-1.5038	0.1326
svarmöguleikar					
3 rangir	0.5169	-0.6598	0.1619	-4.0746	0.0000
svarmöguleikar					
4 rangir	0.5375	-0.6208	0.1643	-3.7777	0.0002
svarmöguleikar					
5 rangir	0.4736	-0.7475	0.1639	-4.5615	0.0000
svarmöguleikar					
6 rangir	0.4093	-0.8933	0.1637	-5.4570	0.0000
svarmöguleikar					
7 rangur	0.4073	-0.8983	0.1629	-5.5156	0.0000
svarmöguleikar					
gpow	1.1252	0.1180	0.0357	3.3062	0.0009
hvaða fyrirlestur					
Fyrirlestur a (Viðmið)					
Fyrirlestur b	3.1947	1.1615	0.0460	25.2607	0.0000
Fyrirlestur c	4.1588	1.4252	0.0572	24.9200	0.0000
Fyrirlestur d	1.6452	0.4978	0.0383	13.0053	0.0000
Fyrirlestur e	2.2939	0.8302	0.0412	20.1404	0.0000
Fyrirlestur f	2.8899	1.0612	0.0475	22.3524	0.0000
Fyrirlestur g	3.9402	1.3712	0.0518	26.4807	0.0000
Fyrirlestur h	3.9953	1.3851	0.0595	23.2734	0.0000
Fyrirlestur i	3.5074	1.2549	0.0595	21.0775	0.0000
Fyrirlestur j	2.3166	0.8401	0.0501	16.7596	0.0000
Fyrirlestur k	3.1239	1.1391	0.0531	21.4355	0.0000
Fyrirlestur l	3.1574	1.1497	0.0527	21.8017	0.0000
Fyrirlestur m	2.4809	0.9086	0.0495	18.3400	0.0000

svarað eru fleiri og fleiri spurningar með því að skoða myndir með meðaltali spurninga sem var svarað rétt eftir því hve margar spurningar nemandarnir hafði svarað hingað til og skipt upp eftir því hvort nemandarnir höfðu séð rétta svarmöguleikann áður eða ekki. Þar sást einhverjar framfarir í gangi, en óvissa var hvort þetta voru beinar framfarir eða utanbókarlærdómur á röngu svarmöguleikunum. Þá var skoðað líka mynd með tilliti til hlutfall rangra svarmöguleika sem hafa sést áður. Þar leit út fyrir að vera eitthverjar framfarir, sumar jákvæðar og sumar neikvæðar. En annars miklu beinari línur en fyrri. Til að gá hvort þetta voru framfarir í gangi þar, þá voru skoðuð nokkur líkunn, með þá hugmynd að baki að ef fjöldi spurninga fram að þessu er enn marktæk með öllum utanbókarlærdómi með í líkaninu, þá gæti það verið að benda til þess að hægt er að mæla framfarir nemendans. Þessi þrjú líkunn voru lík að mörgu leiti nema það var smá breyting milli tveggja breyta. Fyrsta líkanið innihélt fjöldi spurninga fram að þessu, til að sjá framfarir nemenda, annað líkanið innihélt hlutfall rangra svarmöguleika sem hafa sést áður og að lokum þriðja líkanið sem innihélt bæði fjölda spurninga fram að þessu og hlutfall rangra svarmöguleika sem hafa sést áður.

4.1 fyrsta líkanið

Til að byrja með, þá var gert fyrsta líkanið til að sjá hvort það væri hægt að sjá framfarir eftir því hve margar spurningar nemandinn hafði svarað hingað til innan í fyrirlestrinum. Þar var breyta yfir því hvort rétta svarið hafi sést áður og víxláhrif á milli þess og fjölda spurninga svarað. Þar sást að fjöldi spurninga svarað var marktæk og framfarir sást þar.

Vandi kemur upp að þegar svarað hefur mikið, þá er stærri og stærri hlutfall af röngu svarmöguleikunum sem hafa sést áður, því getur vel verið að framfarirnar sem eru að sjást hér, eru í raun bara utanbókarlærdómur rangra svara sem er að toga upp líkurnar á því að svara rétt. Því var skoðað hlutfall rangra svarmöguleika án fjölda spurninga fram að þessu.

4.2 Annað líkanið

Þegar skoðað var annað líkanið, þá var séð að hluta2 er marktæk breyta miðað við Anova III. Til að sjá muninn milli fyrra líkansins og þessara þá var gert anova á milli þeim til að sjá hvort af líkönunum voru marktækari, skv. anova þá er annað líkanið sterkara en fyrra líkanið. Svo var skoðað aðeins standardized brier score og AUC þeirra tveggja, þar leit út fyrir að annað líkanið er aðeins sterkara en það fyrra, með muninn milli þeirra sem 0.0026 hjá AUC og 0.0074 hjá standardized brier skorinu. Svo á meðan annað líkara er sterkara miðað við anova, þá er munurinn ekki endilega það mikill. (ÓVISS UM EFTIRFARANDI:) Þetta bendir til þess að fjöldi spurninga svarað fram að þessu er jafn sterkt og að skoða hlutföllin þeirra, eru þá framfarirnar sem sáust áður bara útaf hlutfall röngu svara, eða voru framfarir þar?

Til að skoða það, þá er skoðað þriðja líkanið, þá er innihaldið bæði fjöldi spurninga fram að þessu og hlutfall rangra svarmöguleika fram að þessu, en víxláhrifin á milli fjölda spurninga fram að þessu og hvort rétta svarið hafi sést áður er horfið.

4.3 Þriðja líkanið

Að lokum var sett saman þriðja líkanið, í líkaninu er bæði fjöldi spurninga fram að þessu og hlutfall rangra svarmöguleika sem hafa sést áður. Þar var bæði fjöldi spurninga og hlutfall rangra svarmöguleika marktækt, svo var þriðja líkanið sterkara en annað og fyrsta líkönin, ef skoðað er til “optimism correct” standardized brier og AUC þá lítur út fyrir að munurinn milli þriðja og annars er líka mjög lítill, þar sem munurinn milli þriðja og Annars líkansins er 0.003 hjá AUC og 0.0026 hjá standardized brier skorinu. Þetta bendir til að öll þrjú líkönin eru lík í styrkleika miðað við standardized brier og AUC á meðan þriðja líkanið er sterkast af þeim miðað við anova.

Hvað þýðir þetta? Frá þessu sjáum við að fjöldi spurninga fram að þessu er ennþá marktækt, þótt það sé bætt við utanbókaráhrifum hjá því hvort rétti svarmöguleikinn hafi sést áður og hlutfall rangra svarmöguleika sem hafa sést áður, þar bendir til þess að hægt sé að mæla áhrif nemenda. Þar sem stikinn fyrir fsfat er séð sem framfarir.

4.4 Smá loka niðurstöður

Svo frá líkönunum er hægt að finna rök fyrir bæði að nógu sterkt er að skoða bara hlutfall rangra svarmöguleika sem hafa sést áður og að skoða fjölda spurninga svarað miðað við að AUC og standardized brier skorið er mjög líkt milli líkannana þriggja, þetta gæti verið að benda til rosalegu sterku áhrif utanbókarlærdóms til að sjá hvernig svarað er næst.

Á sömu mótum þá er fyrsta og þriðja líkönin líka í sama styrkleika miðað við AUC og standardized brier skorið, semsagt hægt er að fá stika fyrir fjölda spurninga fram að þessu. Sem leyfir því að mæla framfarir nemenda frá spurningu til spurninga, að sjá hverjar framfarirnar eru þegar nemandinn byrjaði til þann punkts þar sem nemandinn kláraði.

4.5 Smá upphát hugsanir um hugmyndir til að skoða í framtíðinni

Spurning kemur samt að því hvernig hægt væri að skoða framfarir hvers nemenda fyrir sig, hugmynd að því væri að bæta við fjöldi spurninga fram að þessu í mixed effect í líkaninu, semsagt hafa $(1 + \text{fsfat} \mid \text{studentId})$ í líkaninu í staðinn fyrir $(1 \mid \text{studentId})$. Þá myndi fást framfarir hvers nemanda, ef þetta er rétt hugsun. Vandinn við svona líkann er að það tekur miklu lengur að keyra miðað við upprunalega líkanið.

Annað væri að skoða meiri sambærilegar spurningar, með því að hafa fjölda spurninga fram að þessu vera telja eftir spurningahaus, frekar enn að skoða innan við fyrirlesturs.