

Sumarverkefni

Skýrsla um vísbendingaspurningar í Tutor-web

1 Inngangur

Markmið verkefnisins var að greina hvort og hversu mikil áhrif það hefði fyrir nemendur að fá cue spurningar frekar en venjulegar spurningar. Einnig var ákveðið að greina með power útreikningum hversu stórt hlutfall af spurningum og hversu mörg svör þyrfti til að setja upp rannsókn þar sem gögnin væri enn sambærilegri og hægt væri að tryggja marktækni á niðurstöðum.

1.1 Rannsóknarspurning

Hafa vísbendingaspurningar marktæk áhrif á frammistöðu nemenda í tutor-web kennslukerfinu.

1.2 Kóði

Allur kóði sem var smíðaður til að skapa myndir, töflur og líkön má finna á [Github](#)

2 Lýsandi tölfræði

2.1 Gögn og breytur

Gagnarammi af svörum nemenda var smíðaður út frá tutor-web gagnasafninu sem innihélt breytur cue, qName, studentId, grade og correct.

1. Cue

- Tvíkostabreyta sem segir til hvort spurning hafi verið vísbendingaspurning eða ekki.
- Allar spurningar í gagnarammanum voru að minnsta kosti með eitt svar af hvorri gerð.
- Ef breytan var jafngild 0 var spurningin ekki með vísbendingu en ef breytan var jafngild 1 var spurningin með vísbendingu.

2. qName

- Heitið á spurningu. Farið var í gegnum tutor-web gagnagrunninn og fundið allar spurningar sem bornar höfðu verið fram bæði með og án vísbendingu. 49 slíkar spurnningar fundust.

3. StudentId

- Tilgreinir nemanda sem svaraði fyrir hvert svar.

4. Grade

- Áunninn einkunn nemenda fyrir öll svör til og með svarinu sem grade er skráð við.

5. Correct.

- Tvíkösta breyta sem segir til um hvort svarið hafi verið rétt eða rangt.

2.2 Fjöldi svara

Til þess að sjá hversu mörg svör eru til staðar við spurningunum 98, þar af helmingur cue, er sett upp stuðlarit sem sýnir fjölda svara við bæði cue og ekki cue spurningunum. Inná myndina er svo merkt hversu mörg svör þar af voru rétt eða röng.



Mynd 1: Fjöldi svara með/án cue

Sjáum að umtalsvert meira af án cue svörum er til staðar í gagnasettinu.

Tölulegar upplýsingar myndar 1 voru settar fram í töflu.

Tafla 1: Hér má sjá fjölda svara eftir cue og correct

	Nocue	Cue	Heild
Rangt	6154	1372	7526
Rétt	16399	4711	21110
Heild	22553	6083	28636

Sjá má að langstærsti flokkurinn er rétt svar án cue með 16399 svör, á móti eru aðeins 1372 röng svör án cue. Einnig er ljóst að cue svör eru umtalsvert færri en án cue svörin. Mögulegt er að ekki sé nægilega mikið af cue svörum til staðar og því var farið í power útreikninga í lok skýrslunnar til þess að kanna hvort og hversu mikið vantar uppá.

2.2.1 Hlutföll

Miðað við mynd 1 og töflu 1 er mjög mikill munur á fjölda svara af hvorri gerð og til að sjá betur hversu mikill munurinn er í gagnasettninu var sett upp tafla sem sýnir hlutfall cue og ekki cue svara.

Tafla 2: Hlutföll á milli svara

Cue	Hlutfall
0	0.7875751
1	0.2124249

Sjáum að um 80% gagnasettsins eru án cue svör.

Til þess að sjá betur hver hlutfallslegi munurinn á milli réttra og rangra svara eftir cue var sett upp sambærileg tafla og ?? sem sýnir hlutföll af heild eftir cue og án cue svara miðað við rétt eða rangt.

Tafla 3: Hlutföll á milli svara

	NoCue	Cue
Rangt	0.2728684	0.2255466
Rétt	0.7271316	0.7744534

Samkvæmt töflunni eru hlutfallslega fleiri cue spurningar réttar eða 77% á móti 73%. Munurinn er þó ekki mikill og var sett upp hlutfallapróf til að kanna hvort munurinn sé marktækur.

Tafla 4: Hlutfallapróf með prop.test

Mat1	Mat2	Prófstærð	P-gildi	Frígráða	Neðra mark öryggisb.	Efra mark öryggisb.
0.8176986	0.7768356	55.12973	0	1	0.0303982	0.0513279

Tilgátuprófið var „2-sample test for equality of proportions with continuity correction“ með gagntilgátuna „two-sided“. Munurinn er marktækur með p-gildi u.þ.b. 0 og 95% öryggisbil fyrir mismun hlutfallanna frá þremur prósentum uppí fimm prósent. Niðurstaðan er vísbending um að mögulega standist það að nemendur sem fá cue spurningu séu líklegri til að svara spurningu rétt.

2.3 Meðaleinkunn

Til þess að sjá hver munurinn er á meðaleinkunn á milli hópanna tveggja var fjölda réttra svara deilt með fjölda svara innan hvors hóps fyrir sig. Niðurstöðurnar voru svo settar fram í töflu.

Tafla 5: Meðaleinkunn eftir cue eða ekki cue

Cue	Meðaleinkunn
0	7.271317
1	7.744534

Meðaleinkunnin er u.þ.b. 0.5 hærri fyrir svör við vísbendinga spurningum en spurningum án þeirra.

Tafla 5 miðar við fjölda réttra svara deilt með fjölda svara en til þess að geta sett upp kassarit sem sýnir fjórðungamörk fyrir hvorn flokk fyrir sig var ákveðið að flokka niður svörin á hverja spurningu. Nýr gagnarammi var búin til sem innihélt eftirfarandi breytur.

1. QName

- Tilgreindi spurninguna. Hvert qName kemur tvisvar sinnum. 49 ólík gildi.

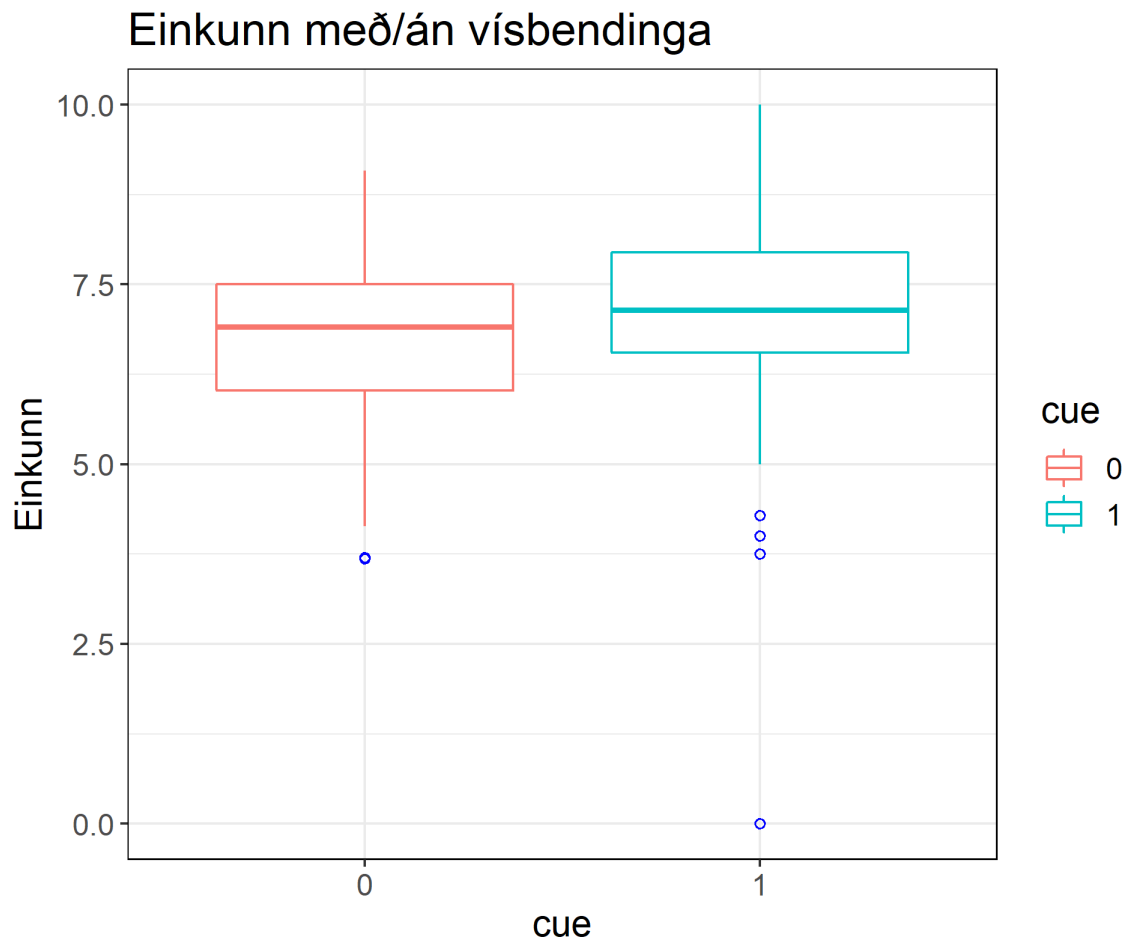
2. Cue

- Merkti hvort um var að ræða cue eða ekki cue. Hvert qName kemur fyrir einu sinni með cue=0 og einu sinni með cue=1

3. Meðaleinkun

- Nýja breytan sem er fjölda réttra svara á spurningu deilt með fjölda svara á spurningu.

Út frá þessum upplýsingum var hægt að setja upp kassarit fyrir meðaleinkunn á spurningu eftir cue eða ekki.



Mynd 2: Meðaleinkunn á spurningu eftir cue eða ekki

Það er ekki mikill munur á milli hópanna þó cue sé með aðeins hærri fjórðungamörk. Eftirtektarvert að það eru mun fleiri útlagar hjá cue en ekki cue.

Tölulegar upplýsingar myndar 2 voru settar fram í töflum.

Tafla 6: Fjórðungamörk meðaleinkunnar fyrir ekki cue

Fjórðungamork	Medaleinkunn
0	3.684210
25	6.024845
50	6.909976
75	7.505669
100	9.085714

Tafla 7: Fjórðungamörk meðaleinkunnar fyrir ekki cue

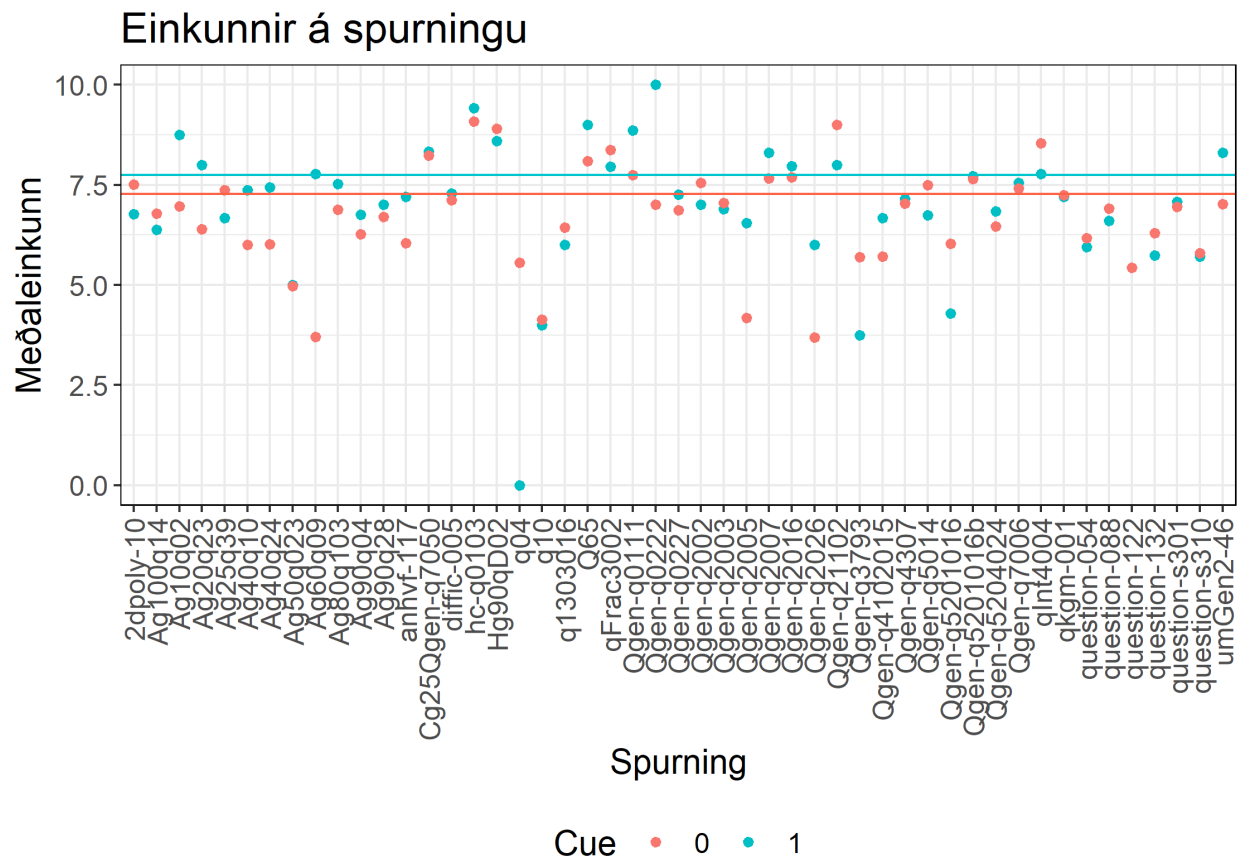
Fjórðungamork	Medaleinkunn
0	0.000000
25	6.547619
50	7.142857
75	7.947368
100	10.000000

Eins og á kassaritinu miðast töflurnar við meðaleinkunn á spurningar. Það munar ekki miklu á milli miðgilda hópanna en efri og neðri fjórðungamörk fyrir cue eru þó nokkuð hærri.

2.4 Spurningar

Til þess að sjá hvort og hversu mikill munur var á milli spurninganna 49 sem unnið var með var sett upp punktarit sem sýnir meðaleinkunn á spurningu fyrir cue og ekki cue. Settar voru inn línur fyrir meðaleinkunn með og án cue í sama lit og samsvarandi punktar.

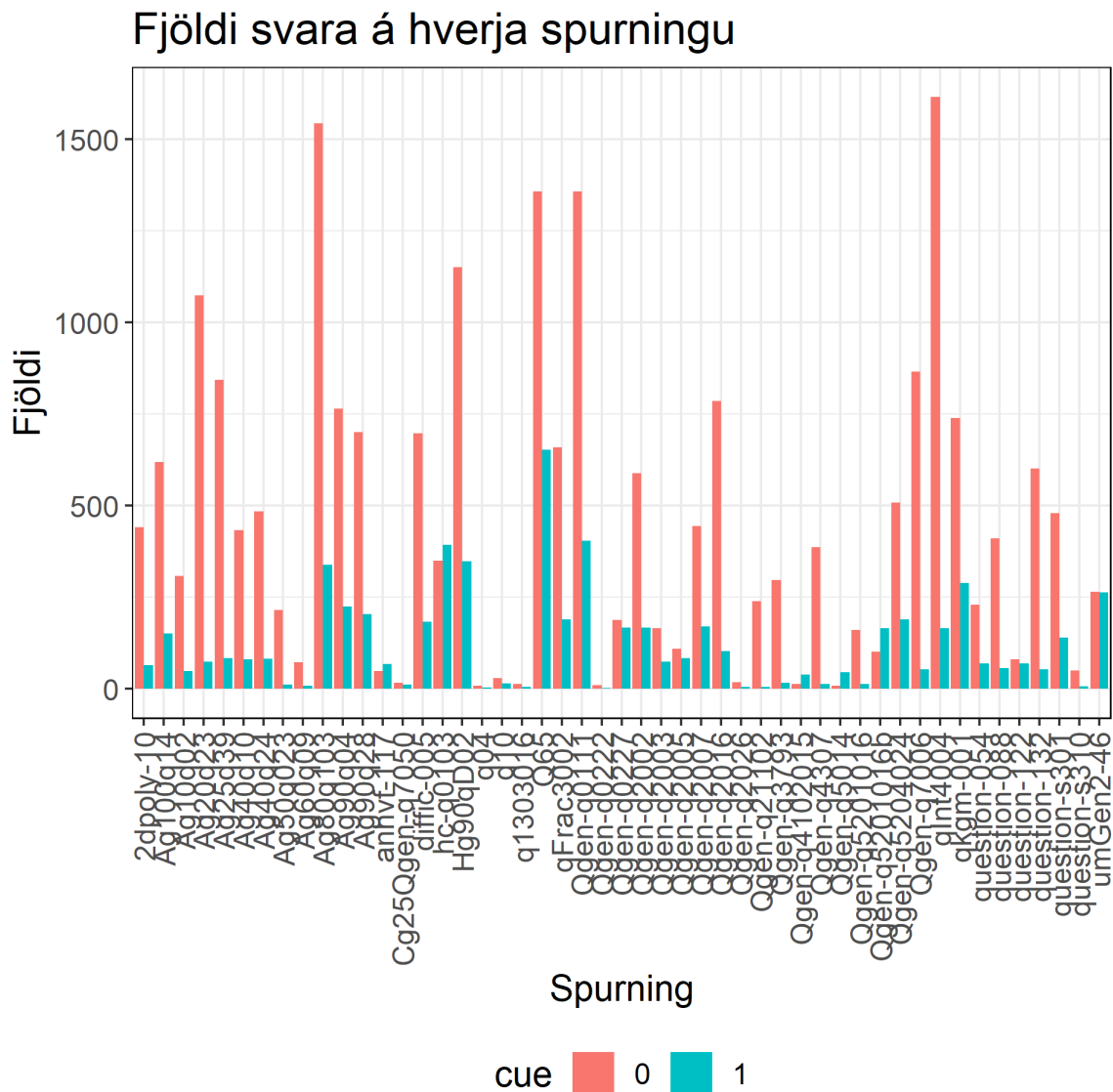
Myndin var búin til útfrá sama gagnasetti og lýst er að ofan þar sem einungis er unnið með spurningarnar ekki svör.



Mynd 3: Meðaleinkunn á spurningu eftir cue eða ekki

Fyrir utan spurningu q04 þar sem meðaleinkunn cue svara er 0 eru spurningarnar allar með nokkuð sambærilegar meðaleinkunnir þar sem bláu punktarnir fyrir cue virðast að staðaldri vera staðsettur skör hærra. Bláa línan er svo sýnilega hærra staðsett en appelsínugula sem staðfestir að bláu punkatanir eru hærri.

Til þess að sjá betur hvernig munurinn á milli hvernar spurningar er er sett upp stöplarit fyrir hverja spurningu eins og áður en núna settur fram fjöldi svara á hverja spurningu eftir cue og ekki cue.

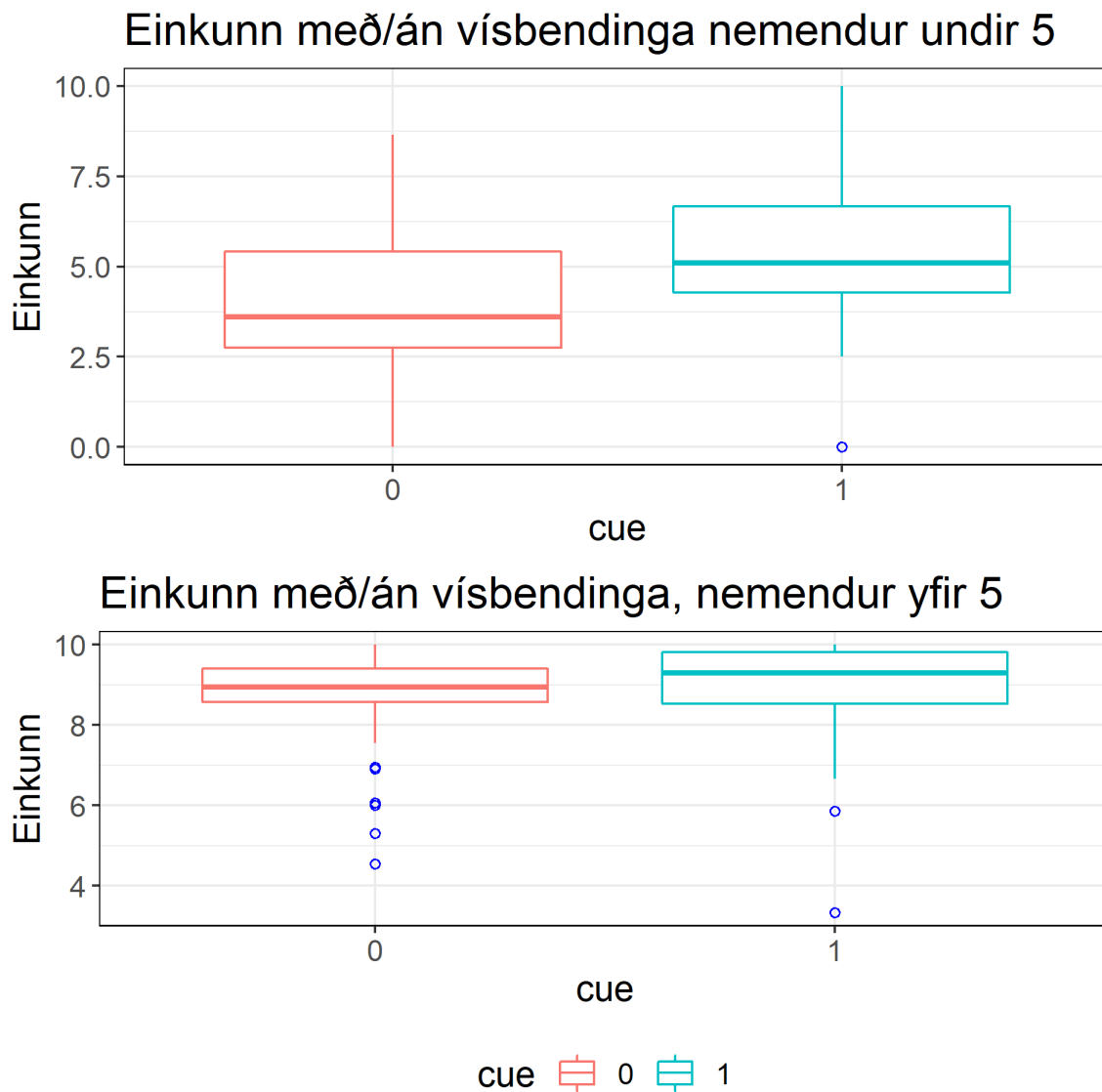


Mynd 4: Meðaleinkunn á spurningu eftir cue eða ekki

Verulega mikill munur á fjölda svara á milli spurninga. Eftirtektavert er að það eru greinilega mjög fá svör við spurningu q04 sem hafði meðaleinkunnina 0 fyrir cue spurningar.

2.5 Nemendur flokkaðir niður eftir grade yfir og undir fimm.

Við hvert svar er skráð grade breyta sem tilgreinir einkunn nemenda þegar hann fær spurninguna. Einkunnin byggist því á fyrri svörum og frammistöðu nemenda. Þótti áhugavert að skoða hvort cue hefði meiri eða minni áhrif fyrir nemendur yfir eða undir fimm.



Mynd 5: Meðaleinkunn eftir cue eða ekki fyrir nemendur yfir og undir fimm

Áberandi mikill munur á miðgildum hópanna tveggja cue og ekki cue fyrir nemendur undir fimm. Miðgildi meðaleinkunnar cue svara er miklu hærra en ekki cue svara.

Til þess að skilja betur hvorn flokk fyrir sig, nemendur yfir og undir fimm voru settar upp tvær töflur sem sýna fjölda og meðaleinkunn innan hvors hóps.

Tafla 8: Meðaltal nemenda undir grade fimm eftir cue

Cue	Fjoldi	Medaleinkunn
0	10809	5.309464
1	2797	6.049339

Sjáum að það er svipað mikið af svörum í sitthvorum hópnum. Munurinn á milli hópanna er 0.7 fyrir undir fimm nemendur sem gæti bent til þess að vísbendingasurningar séu frekar stuðningur við nemendur með lægri einkunn frekar en háa.

Tafla 9: Meðaltal nemenda yfir grade fimm eftir cue

Cue	Fjoldi	Medaleinkunn
0	12341	8.992788
1	3406	9.133881

Medaleinkunn nemenda yfir fimm er mjög há fyrir báða hópa og munar litlu sem engu á milli hópanna.

3 Líkan

Útfrá upplýsingunum í lýsandi tölfræði kaflanum var sett upp slembipátta tvíkosta aðhvarfsgreiningarlíkan til að greina hvort cue hefði marktæk áhrif eða ekki. Breyturnar sem voru settar inn eru cue, qName og studentId sem slembipáttur. Auka breyta var sett inn sem víxlhrif á milli cue og qName. Formúla líkansins var eftirfarandi:

$$\text{correct} = \text{cue} + \text{qName} + \text{cue} \times \text{qName} + (1 \mid \text{studentId})$$

Niðurstöður líkansins voru settar fram í Anova töflu.

Tafla 10: Anova tafla

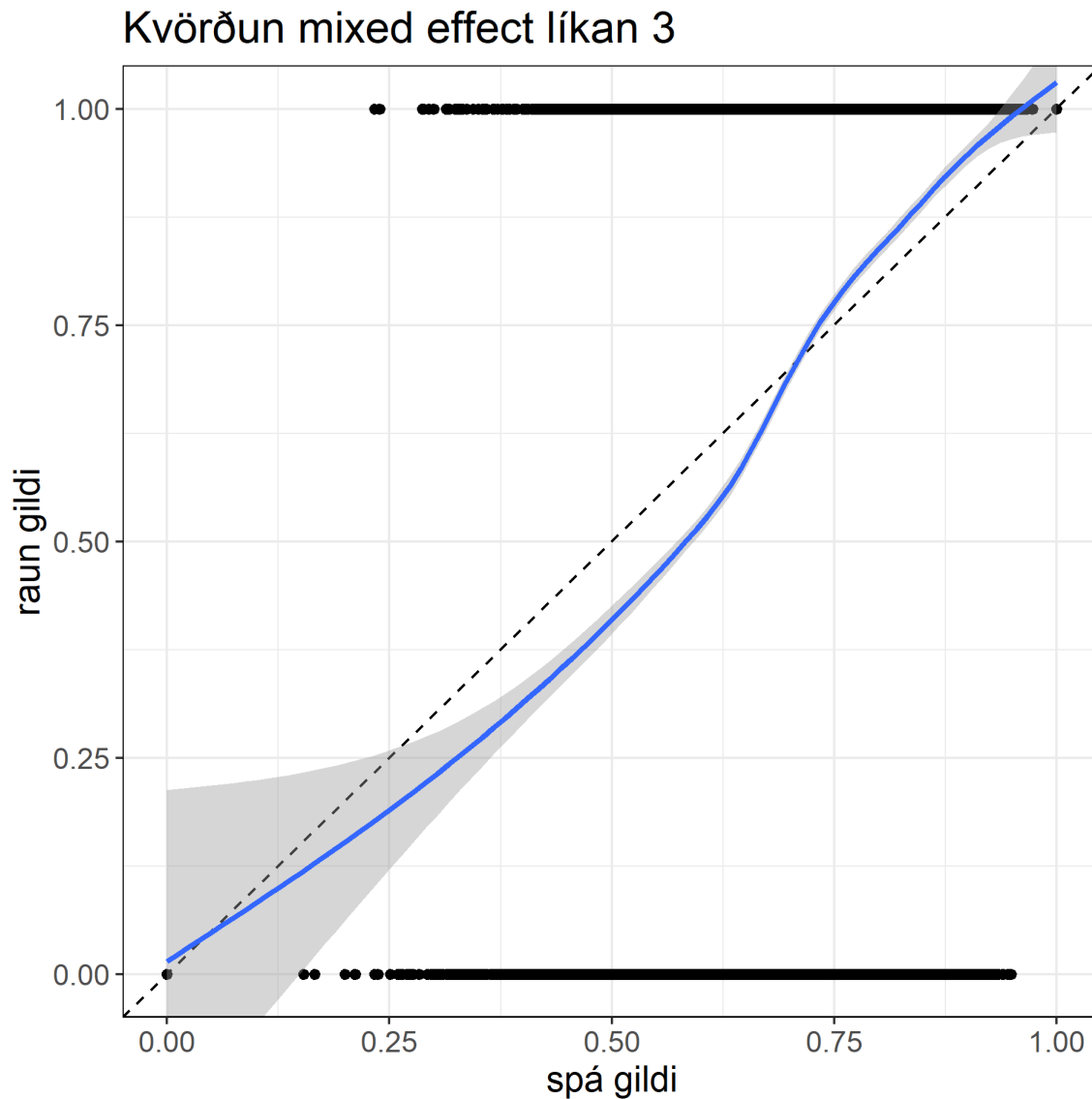
Breytur	Prófstærð	Frígráður	P-gildi
(Intercept)	102.53477	1	0.0000000
cue	3.04862	1	0.0808058
qName	947.73285	48	0.0000000
cue:qName	119.74546	48	0.0000000

cue er ekki marktækt með p-gildið 0.08 en víxlhrif á milli cue og qName eru marktæk sem og qname.

3.1 Mátgæði líkans

3.1.1 Kvörðun

Til þess að kanna hversu vel líkaninu gengur að spá fyrir um raungildi í gögnunum var sett upp kvörðunarlíkan þar sem x-ásinn er spá gildi líkansins og y-ásinn raun gildi gagnanna.

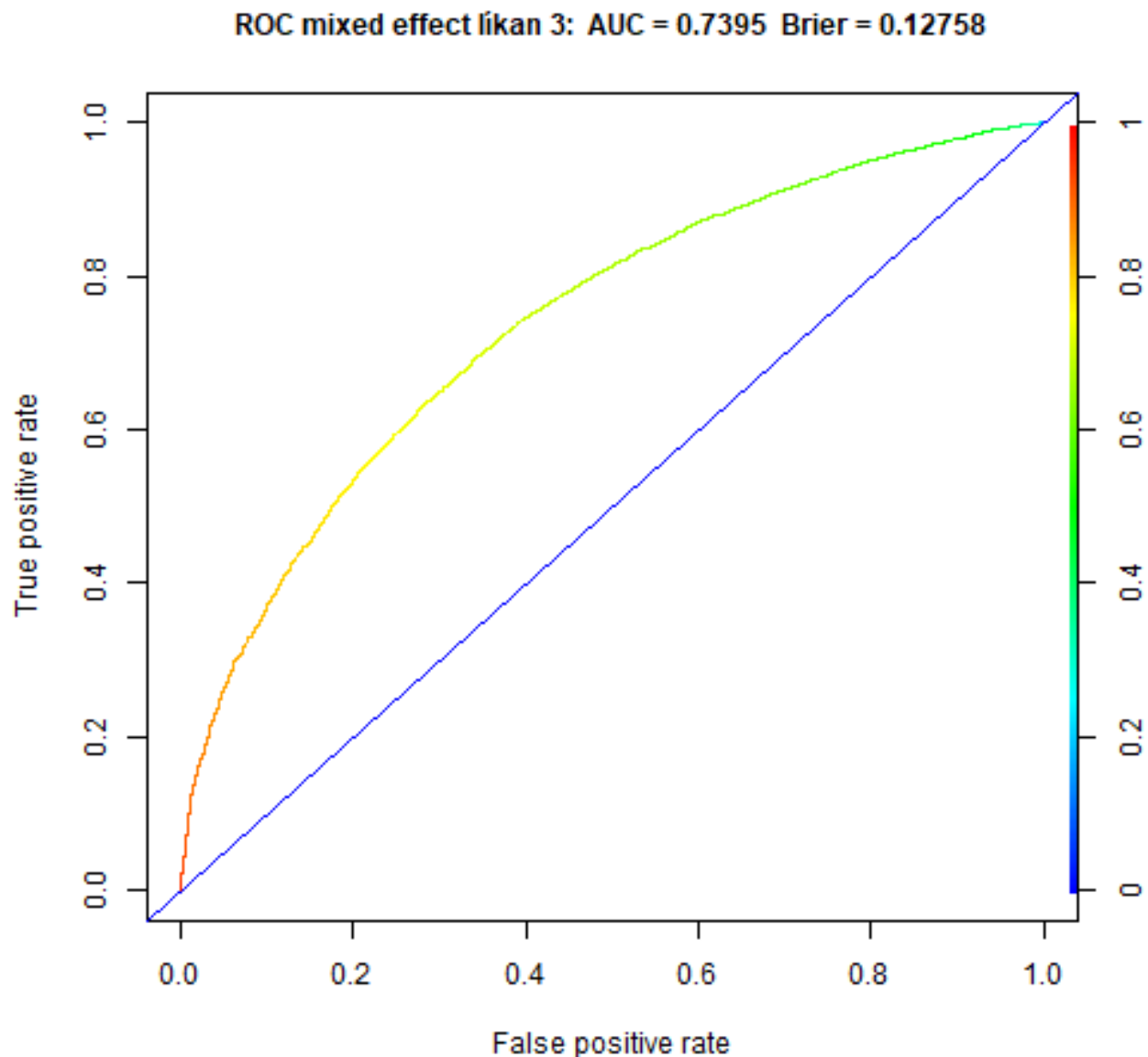


Mynd 6: Kvörðun

Sjáum að líkaninu gengur að staðaldri vel að spá fyrir um gildi líkansins. Upp að 0.7 eru spá gildin að staðaldri of lág og fyrir ofan 0.7 er spáin að staðaldri of há en skekkjan er ekki mikil.

3.1.2 ROC kúrfra, AUC og brier gildi.

Sett var upp ROC kúrfra sem sýnir false positive rate á móti true positive rate líkansins. False positive ásin merkir þegar líkanið segir að svar sé rétt en svarið var rangt og true positive ásin stendur fyrir þegar líkanið segir að svar sé rétt og það er rétt. Markmiðið er þar með að ROC kúrfan verði sem stærst þannig svæðið undir henni AUC gildið, aðgreiningargeta líkansins, sé sem næst 1. Fullkomin spá væri með AUC gildið 1 þar sem ROC kúrfan skildi ekkert svæði eftir.



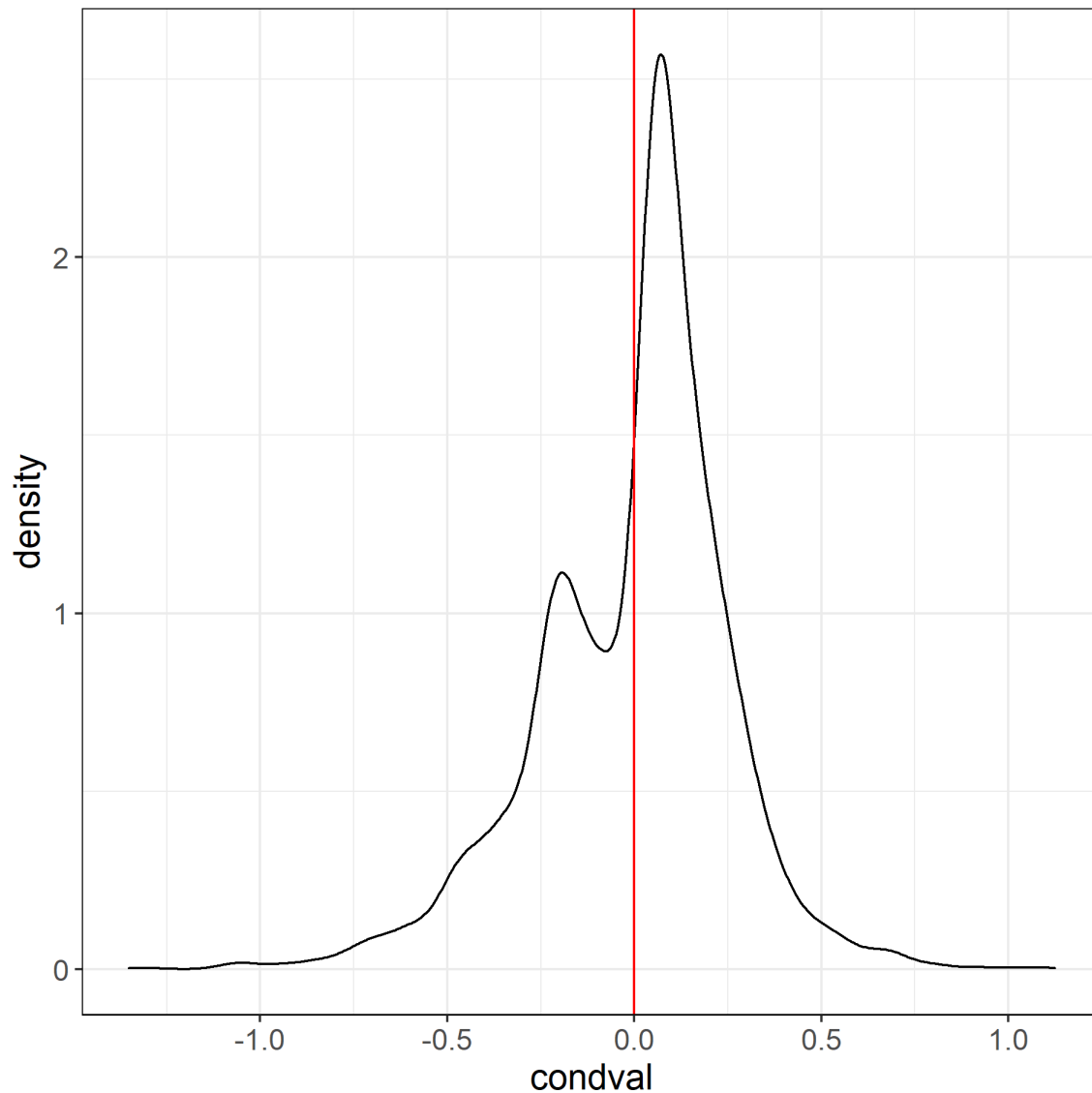
Mynd 7: ROC

AUC gildið er 0.7395 sem er allt í lagi. Líkanið hefur þá rétt fyrir sér í 74% tilvika sem er hátt.

Inná mynd 7 var einnig sett Brier gildi líkansins sem segir til um hversu vel líkaninu tókst til að spá um raungildi gagnanna. Brier gildi er fullomið ef það er einn og bendir svo til gagnslauss líkans ef það er 0. Brier gildið 0.13 er mjög slappt miðað við þetta.

3.2 Ranef

Til þess að sýna hversu mikil áhrif slemipátturinn nemandi hafði var sett upp ranef mynd sem sýnir ólíka skurðpunkta nemenda við y ásinn í líkaninu. Mælingar fyrir línulegt líkan eiga að vera áháðar sem svörin í gagnasettinu væru ef hver nemandi svaraði aðeins einu sinni en nemendur hafa ólíkan fjölda svara og eru svör hvers nemenda innbyrðis háð því þarf að setja nemenda upp sem slemipátt og teikna upp mynd sem skýrir hver dreifni skurðpunkta við y-ás er.



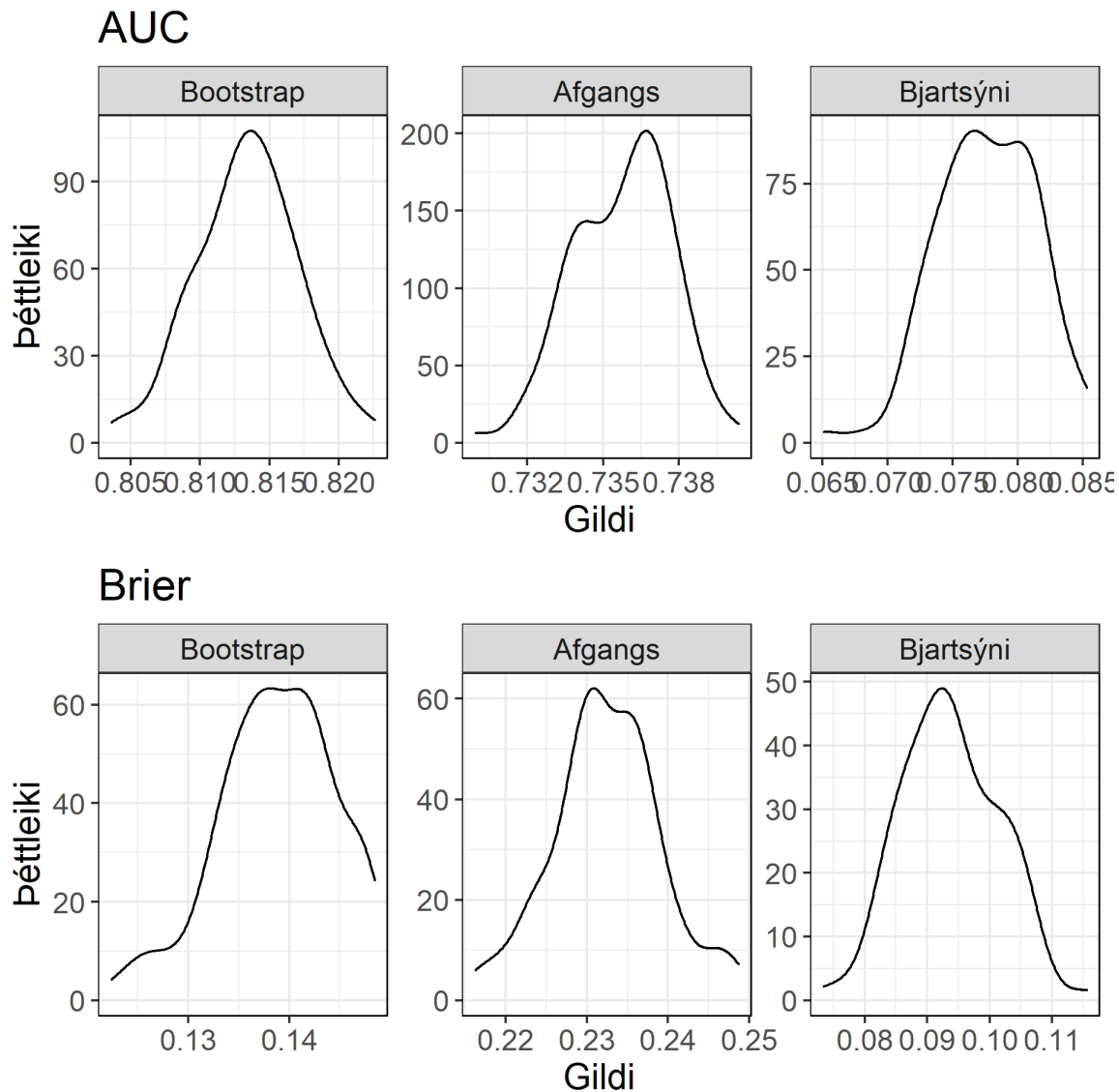
Mynd 8: Ranef

Sjáum að langflestir nemendur eru nálægt skurðpunktinum 0. Miðað er við að nemendur séu normaldreifðir og því koma þessar niðurstöður ekki á óvart.

4 Bootstrap

Til þess að kanna gæði líkans betur er búið til bootstrap þar sem gögnin voru hermd til þess að finna bjartsýni fyrir brier gildið og aðgreiningargetu líkansins og svo leiðrétt gildi fyrir bæði fundin með því að draga bjartsýnina frá gildunum sem voru sett fram í 7. Í greiningunni fólst að bootstrap gildi voru fundin fyrir auc og brier fyrir 80% gagnanna, valin af handahófi, og svo voru gildin einnig fundin fyrir uppsetningu líkansins á 20% gagnanna sem eftir voru. Bjartsýnin var þá mismunur þessara tveggja ólíku gilda.

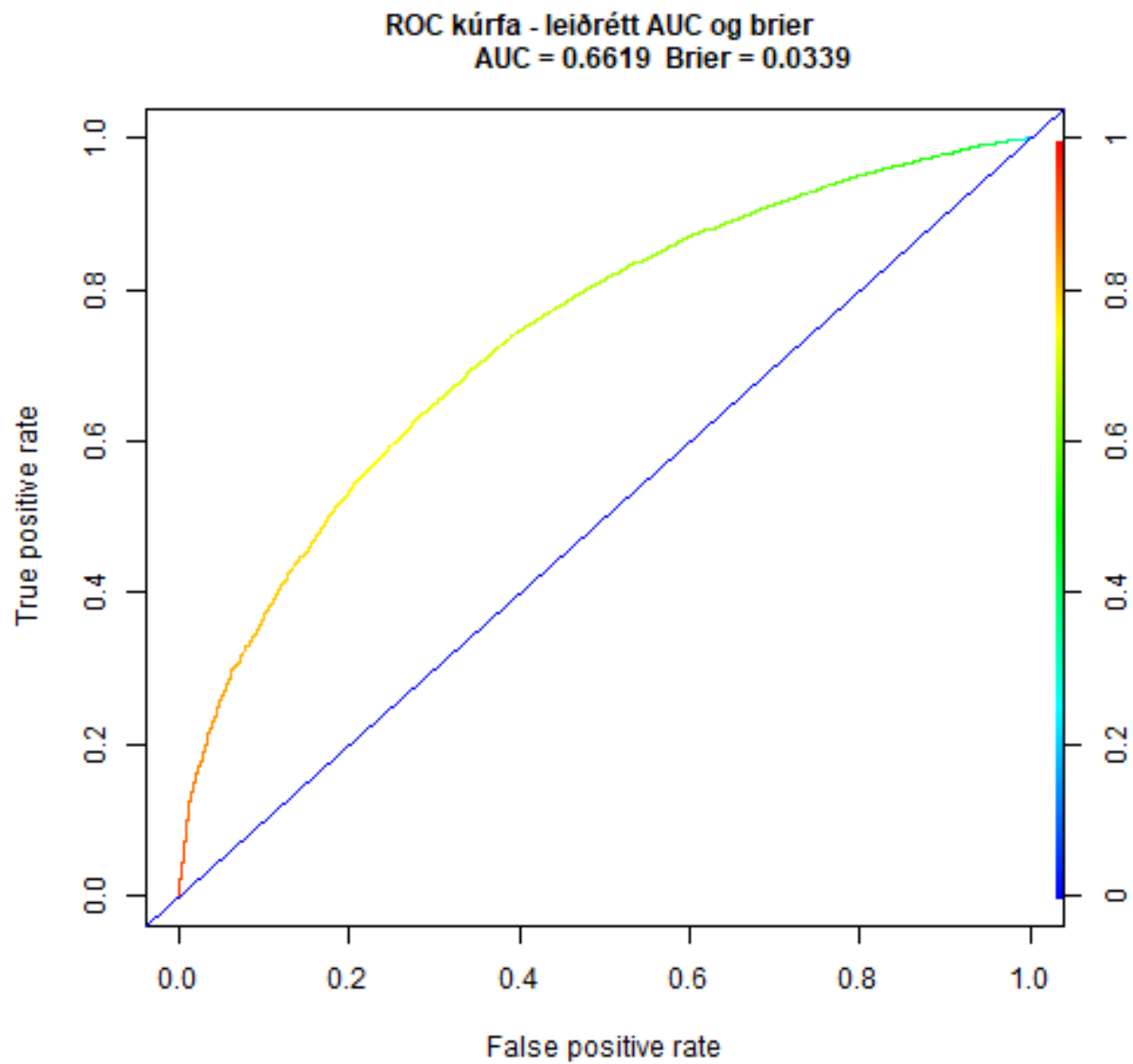
Niðurstöðru 100 bootstrap ítranan voru settar fram myndrænt.



Mynd 9: ROC

Bjartsýni beggja gilda er mjög svipuð en þar sem brier gildið var lágt til að byrja með kemur niðurstaðan mjög illa út fyrir brier. Leiðrétt aðgreiningargetu og brier gildi voru svo fundin út frá upprunalegum gildum að frádreginni bjartsýninni.

Niðurstaða bootstrap greiningarinnar var sett fram á sama hátt og mynd 7



Mynd 10: ROC

AUC gildið er ennþá ásættanlega hátt en brier gildið er mjög lágt. Bendir til þess að líkanið sé allt of viðkvæmt fyrir gögnunum sem unnið var með.

5 Power útreikningar

Fyrir power útreikninga þurfti að nota grade breytuna frekar en correct til þess að hafa samfellda breytu. Sett var upp slembipáttalíkan með cue og studentID sem slembipætti:

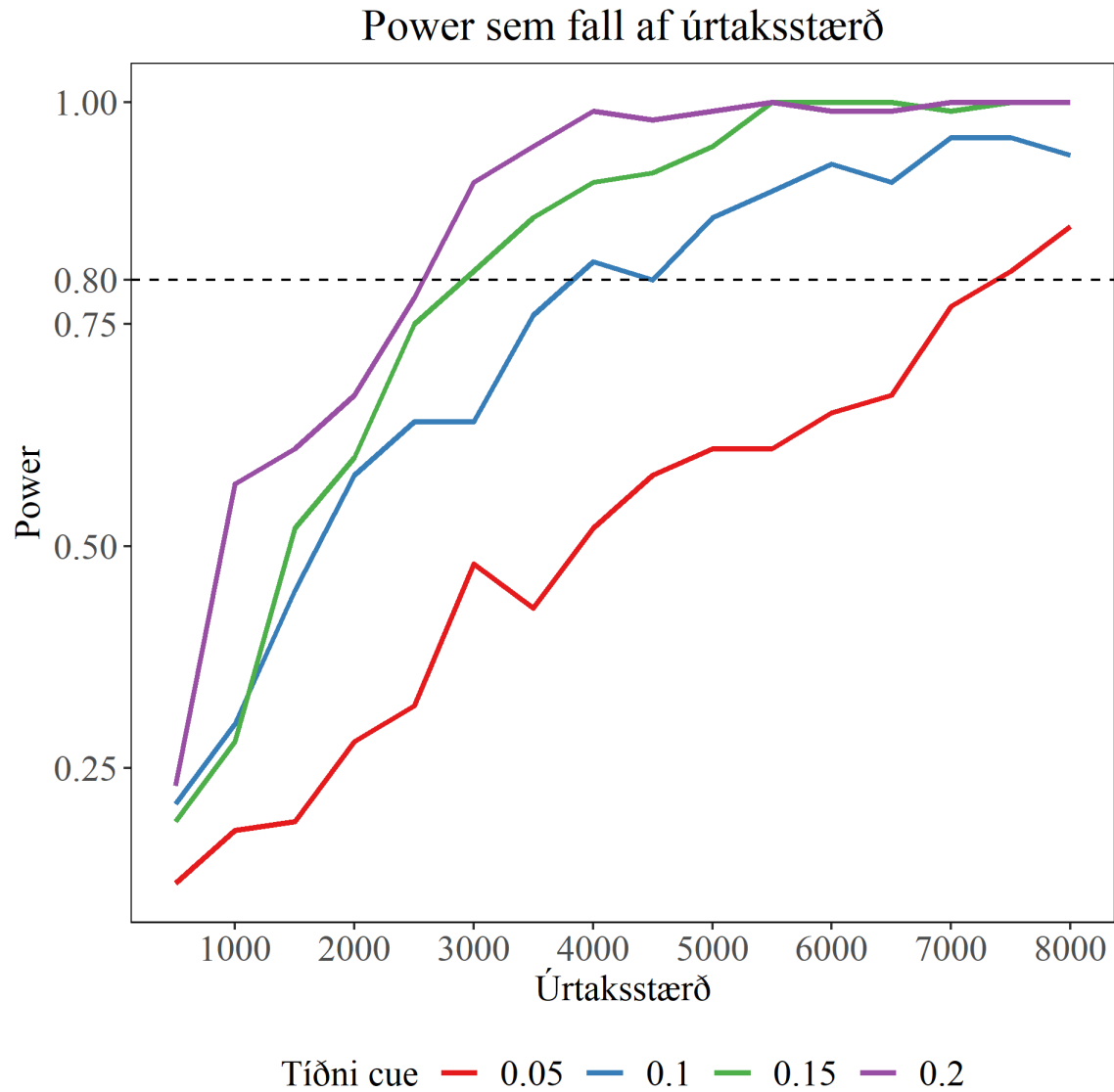
$$\text{grade} = \text{cue} + (1 \mid \text{studentId})$$

Markmiðið var að finna úrtaksstærð svara sem gefði power uppá 0.8 sem merkir að við getum hafnað núlltilgátunni að cue hafi ekki áhrif í 80% tilvika. Notuð var hermun til þess að síendurtekið skapa handahófskennd gögn byggð á líkaninu. Greina svo hvert gagnasett og taka hlutfallið á milli niðurstaðanna sem eru marktækar og ekki. Hlutfallið er þá powerið fyrir líkanið.

Aðferðin var eftirfarandi.

1. Stikamöt líkansins geymd.
2. Vigrar $n = (500, 100, 1500, \dots, 8000)$ og $p = (0.05, 0.1, 0.15, 0.2)$, fyrir úrtaksstærð og tíðni cue skilgreindir.
3. Fyrir hvert par $n \times p$ voru $B = 100$ ítranir framkvæmdar þar sem:
 - student vigur af stærð n_i var hermdur.
 - cue vigur af lengd n_i með stika p_k var hermdur.
 - leifa vigur af lengd n_i fyrir skekkjuna sem fæst af slembipættinum student var hermdur.
 - leifa vigur af lengd n_i fyrir leifar líkansins var hermdur.
 - grade vigur af lengd n_i var hermdur útfrá stikamötunum þar sem hallatalan var margfölduð við cue vigurinn að viðbættum báðum leifa vigrum.
 - Vigrarnir cue, student og grade voru sameinaðir í ramma og tvö slembipátta líkön smíðuð. Bæði með nemenda sem slembipátt en annað með cue sem breytu og hitt með fastan 1 í stað cue.
 - Líkönin þá borin saman með anova og ef cue var marktæk breyta þá skilaði ítrunin gildinu 1, annars 0.
 - Fjöldi marktækra líkana var þá deilt með $B = 100$ ítrunum og skráð við hvert par af $n \times p$.

Greint er frá niðurstöðunum með línuriti þar sem sýnt er power á móti úrtaksstærð eftir tíðni cue og töflu fyrir þörin sem voru næst 0.8



Mynd 11: Power

Tafla 11: Pör af úrtaksstærð n og tíðni cue p sem gefa power um 0.8

Úrtaksstaerd	Hlutfall cue	Power
2500	0.20	0.78
3000	0.15	0.81
4500	0.10	0.80
7500	0.05	0.81

Því hærri tíðni cue sem spurningarnar eru látnar hafa því færri svör þarf til að hafa power um 0.8 Miðað við 20% svara sem cue dugar að hafa 2500 svör.

6 Umræða

Brier gildi nálægt núlli fyrir annars gott líkan benti til þess að gögnin sem væru til staðar væru ekki nógu góð. Það væri mjög áhugavert að setja upp tilraun útfrá power útreikningunum og endurtaka rannsóknina.