

Samantekt

Ingunn Lilja Bergsdóttir

7/27/2020

Samantekt úr cuenocue.

Lýsandi tölfræði

Stöplarit og tíðnitafla sem sýnir fjölda réttra og rangra svara eftir cue eða ekki cue spurningu.

```
ggplot(data = cue_nocue,
  aes(x=cue,
    fill=correct=="1")) +
geom_bar() +
labs(y="Fjöldi",
  x="cue",
  fill="Rétt/Rangt",
  title = "Fjöldi svara fyrir cue eða ekki cue")
```

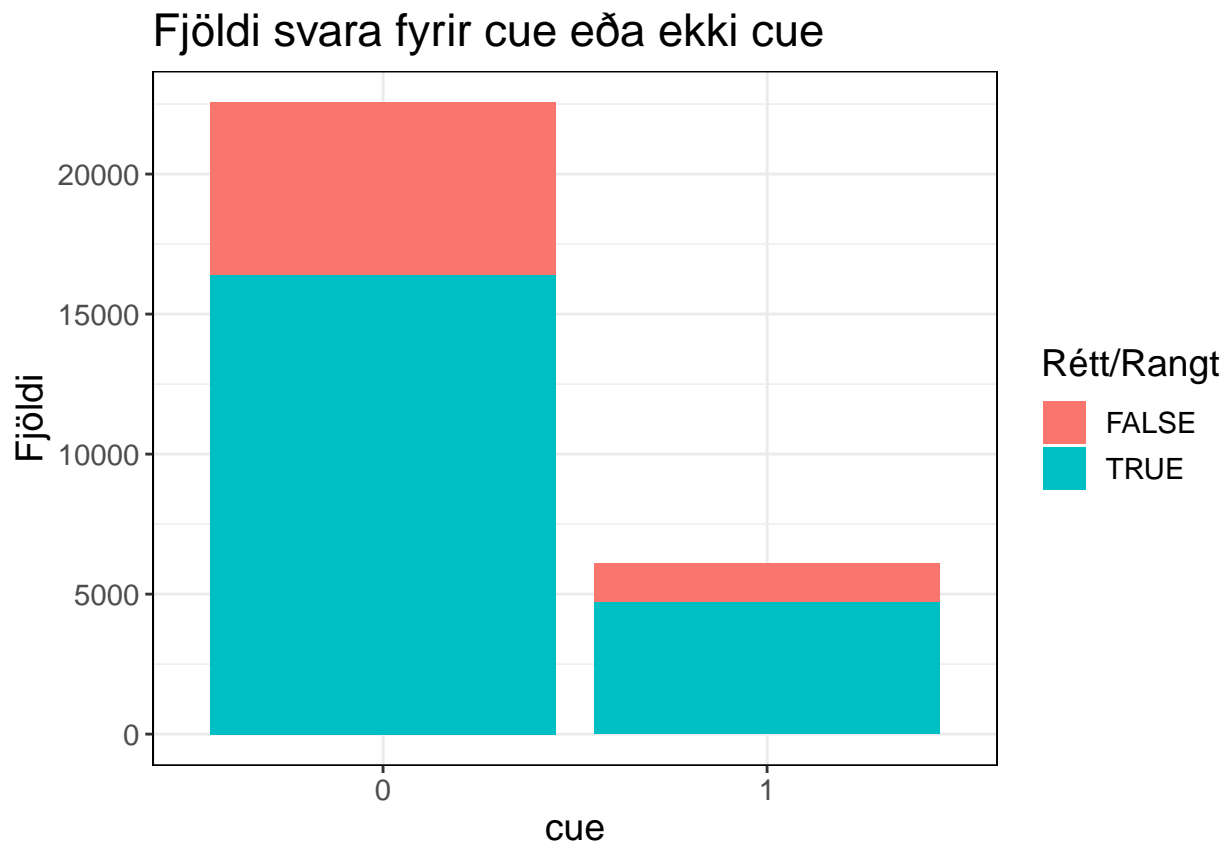


Table 2: Meðaleinkunn

Cue	Meðaleinkunn
0	7.271317
1	7.744534

Tíðnitafla

```
cue_nocue$correct <- with(cue_nocue, ifelse(correct==1, "Rétt", "Rangt"))
Heild <- sum
addmargins(table(correct=cue_nocue$correct, cue=cue_nocue$cue), FUN = Heild) %>%
  kable(col.names = c("Án cue", "Með cue", "Heild"),
        align = c('cccc'),
        caption = "Tíðnitafla") %>%
  kable_styling(bootstrap_options = c("striped", "hover"), "H")

## Margins computed over dimensions
## in the following order:
## 1: correct
## 2: cue
```

Table 1: Tíðnitafla

	Án cue	Með cue	Heild
Rangt	6154	1372	7526
Rétt	16399	4711	21110
Heild	22553	6083	28636

```
cue_nocue$correct <- with(cue_nocue, ifelse(correct=="Rétt", 1, 0))
cue_nocue$correct <- as.integer(cue_nocue$correct)
```

Tafla sem sýnir meðaleinkunn fyrir hvorn flokk þar sem einkunnin byggir á fjölda réttra svara deilt með fjölda svara.

```
cue_nocue %>%
  group_by(cue) %>%
  summarise(Medal_einkunn=mean(correct, na.rm = T)*10) %>%
  kable(col.names = c("Cue",
                    "Meðaleinkunn"),
        align = c('cc'),
        caption = "Meðaleinkunn") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```

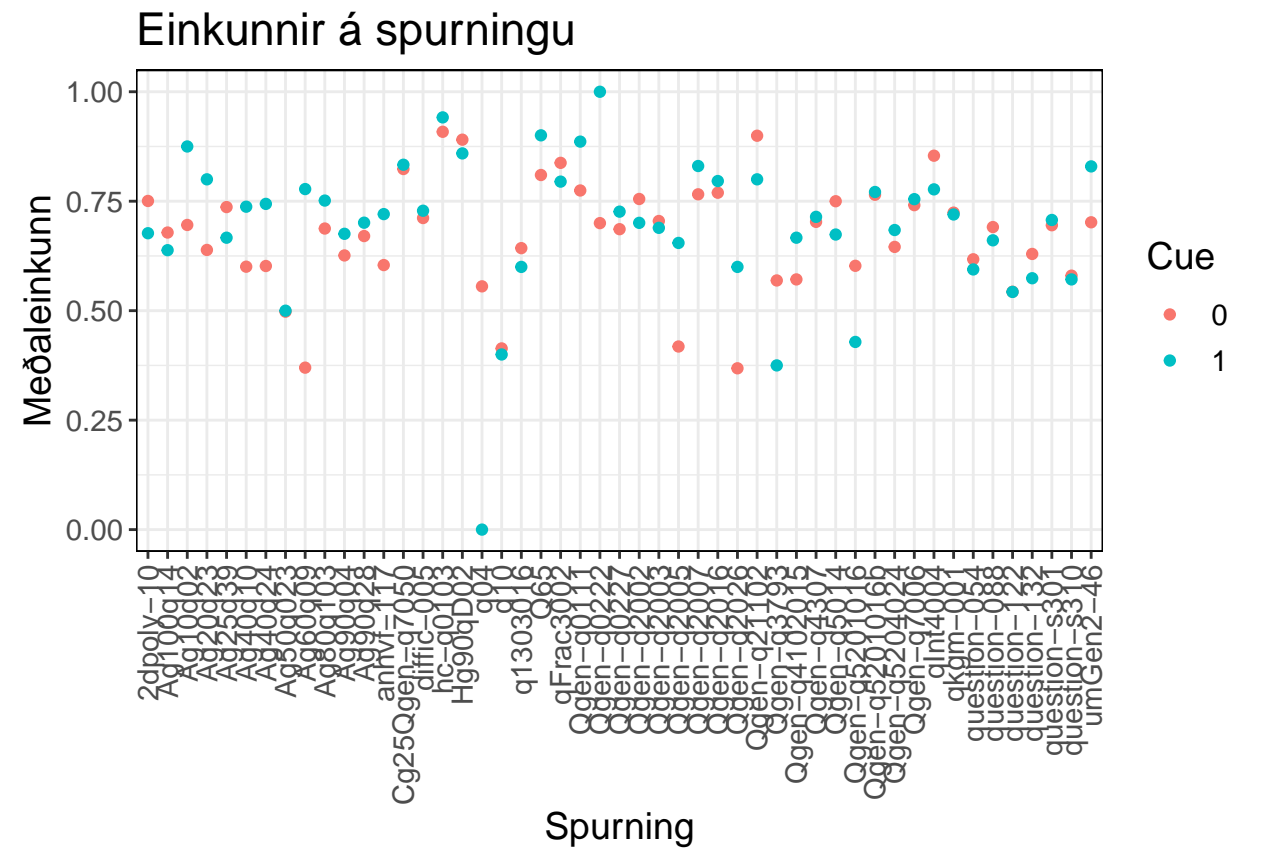
```
## `summarise()` ungrouping output (override with `.groups` argument)
#cue = 0 er spurning án vísbendingar og cue = 1 er vísbendingaspurning.
```

Mynd sem sýnir meðaltöl fyrir cue og ekki cue fyrir hverja spurningu.

```
cue_nocue %>%
  group_by(qName, cue) %>%
  summarize(mean = mean(correct)) %>%
  ggplot(aes(x= qName, y=mean, color=cue)) +
  geom_point() +
```

```
theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+
labs(x="Spurning", y="Meðaleinkunn", color="Cue", title = "Einkunnir á spurningu" )
```

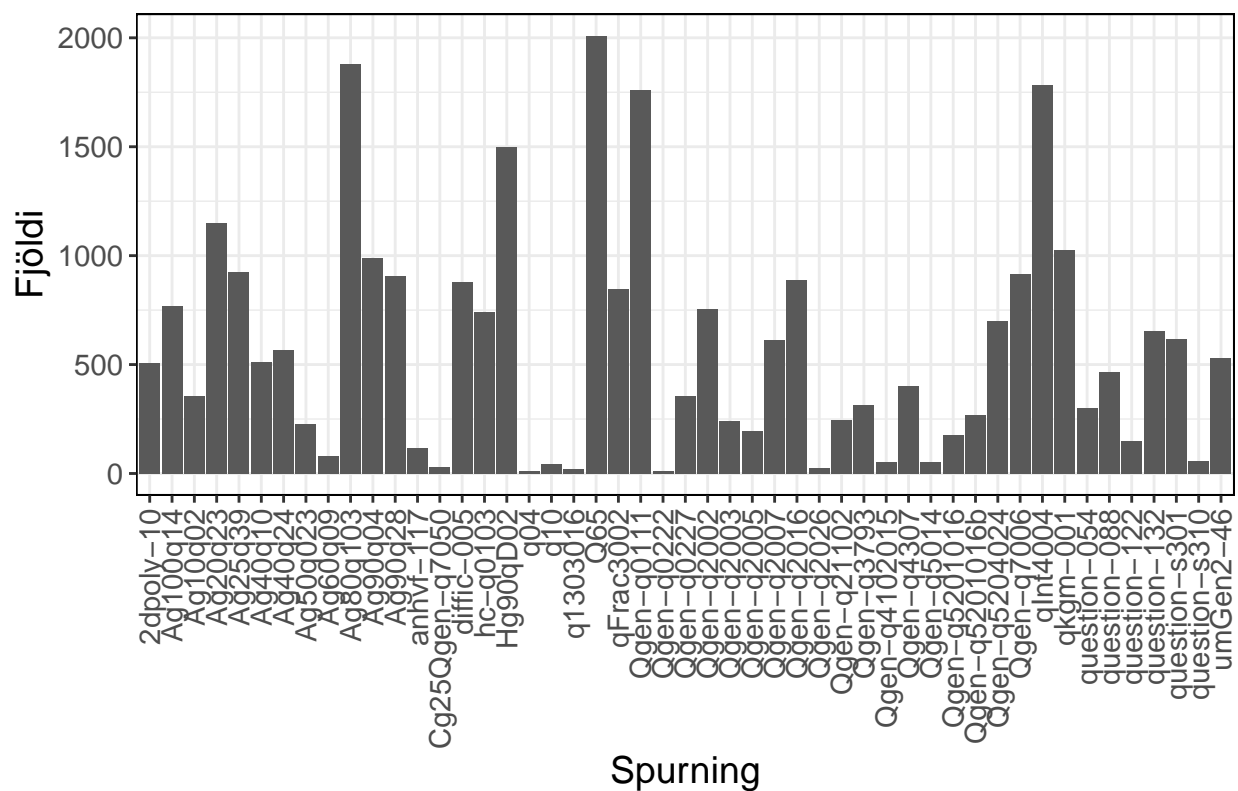
```
## `summarise()` regrouping output by 'qName' (override with `.groups` argument)
```



Fjöldi svara fyrir hverja spurningu.

```
cue_nocue %>%
  ggplot(aes(x= qName)) +
  geom_bar() +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) +
  labs(x="Spurning", y="Fjöldi", title = "Fjöldi svara á hverja spurningu")
```

Fjöldi svara á hverja spurningu



Kassarit til að sýna meðaltölin og miðgildi á milli hópa.

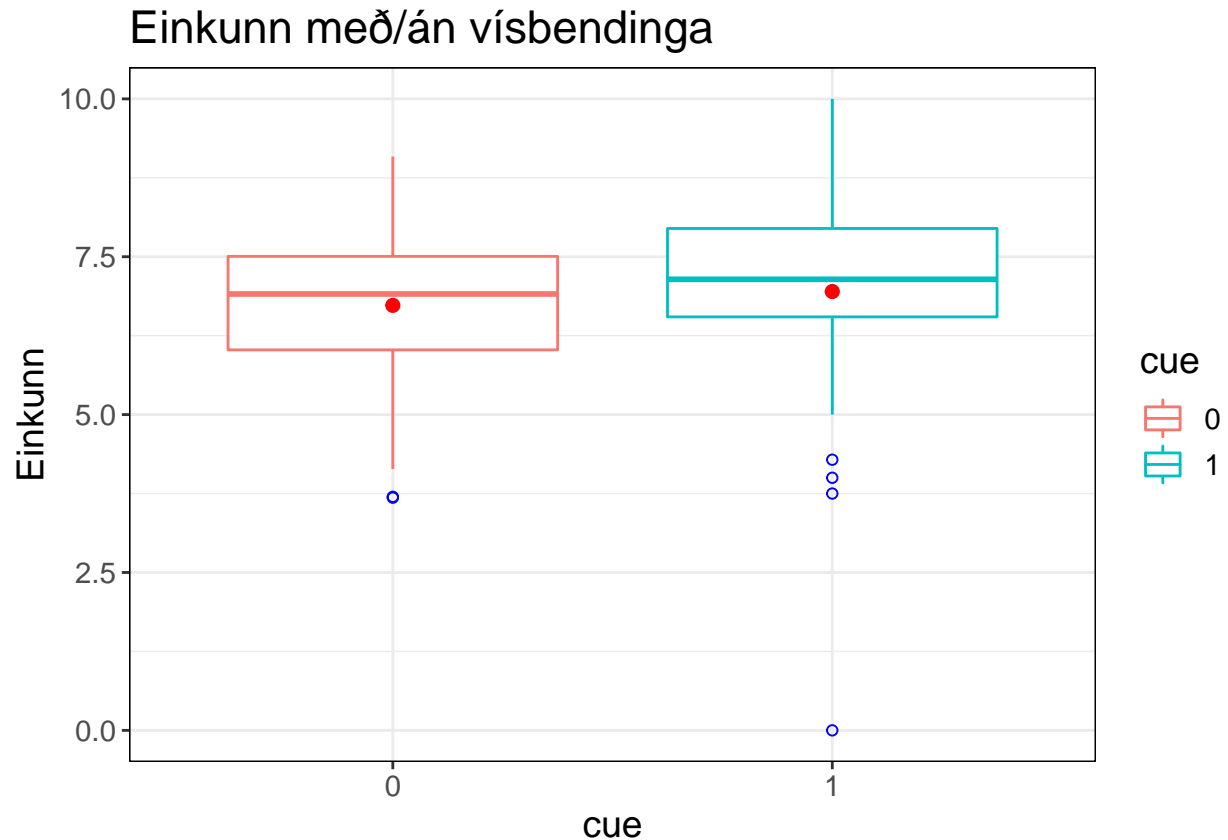
```
f <- medal %>%
  group_by(cue) %>%
  summarise(
    medal = mean(medal)
  )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
medal$cue <- as.factor(medal$cue)
ggplot(data=medal,
  aes(x=cue,
    y=medal,
    color=cue)) +
  geom_boxplot(outlier.colour = "blue",
    outlier.shape = 1) +
  xlab("cue") +
  ylab("Einkunn") +
  ggtitle("Einkunn með/án vísbendinga") +
  geom_point(data=f,
    aes(x=cue,
      y=medal),
    col="red",
    size=2)
```

Table 3: Fjórðungamörk fyrir cue spurningar

	Meðaleinkunn
0%	0.000000
25%	6.547619
50%	7.142857
75%	7.947368
100%	10.000000



Tölur sem sýna fjórðungamörk einkunna fyrir svör með og án cue.

Cue svör:

```
med_cue <-
  filter(medal, cue==1)

med_cue$medal %>%
  quantile() %>%
  kable(col.names = c( "Meðaleinkunn"),
        align = c('cc'),
        caption = "Fjórðungamörk fyrir cue spurningar") %>%
  kable_styling(bootstrap_options = c("striped",
                                       "hover"))
```

Svör án cue:

Table 4: Fjórðungamörk fyrir án cue spurningar

	Meðaleinkunn
0%	3.684210
25%	6.024845
50%	6.909976
75%	7.505669
100%	9.085714

Table 5: Hlutföll

	Án cue	Með cue
Rangt	0.273	0.226
Rétt	0.727	0.774

```
an_cue <-
  filter(medal,
    cue==0)

an_cue$medal %>%
  quantile() %>%
  kable(col.names = c( "Meðaleinkunn"),
    align = c('cc'),
    caption = "Fjórðungamörk fyrir án cue spurningar") %>%
  kable_styling(bootstrap_options = c("striped",
    "hover"))
```

Hutfalla tafla.

```
cue_nocue$correct <-
  with(cue_nocue,
    ifelse(correct==1,
      "Rétt",
      "Rangt"))

kable(prop.table(table(cue_nocue$correct,
  cue_nocue$cue),
  2),
  col.names = c( "Án cue",
    "Með cue"),
  align = c('rr'),
  caption = "Hlutföll",
  digits = 3)
```

```
cue_nocue$correct <-
  with(cue_nocue,
    ifelse(correct=="Rétt",
      1,
      0))
cue_nocue$correct <-
  as.integer(cue_nocue$correct)
```

Kannað hvort það sé rétt að það séu hlutfallslega fleiri rétt svör í cue hópnum með prop.test.

```
# Hér gæti farið inn scientific gildi
prop.test(table(cue_nocue$correct,
               cue_nocue$cue)) %>%
  broom::tidy() %>%
  kable(digits = 4) %>%
  kable_styling(bootstrap_options = c("striped",
                                       "hover"))
```

estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method
0.8177	0.7768	55.1297	0	1	0.0304	0.0513	2-sample test for equality of proportions

Líkanagerð

Einfalt línulegt aðhvarfsgreiningar líkan fyrir medal töfluna. Breytur eru meðaleinkunn á spurningu og cue.

```
fit.lm <-
  lm(medal ~ cue,
     data = medal)

Anova(fit.lm,
      type = "III") %>%
  broom::tidy() %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped",
                                       "hover"))
```

term	sumsq	df	statistic	p.value
(Intercept)	2219.4895	1	1033.2083689	0.0000000
cue	1.1740	1	0.5465159	0.4615486
Residuals	206.2227	96	NA	NA

Mixed effect líkan

Breytur eru cue, qName og víxlhrif þar á milli

```
glmer3 <- glmer(correct ~ cue + qName + cue:qName + (1|studentId),
               data = cue_nocue,
               nAGQ = 0,
               control = glmerControl(optimizer = 'nloptwrap'),
               family = 'binomial')
```

Anova tafla.

```
# setja inn scientific
Anova(glmer3,
      type = "III") %>%
  broom::tidy() %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped",
                                       "hover"))
```

term	statistic	df	p.value
(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

Marktækt fyrir qName og interaction á milli cue og qName.

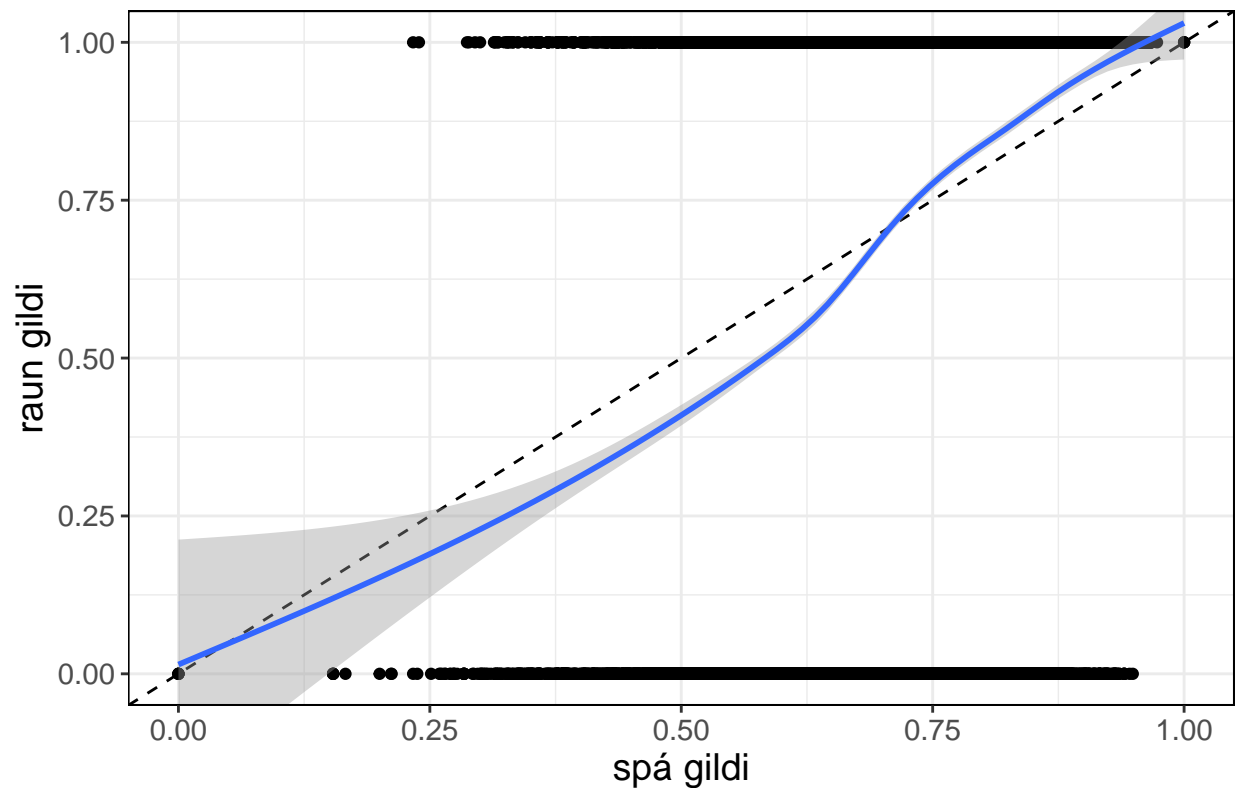
Gæði mixed effect líkans 3

Kvörðun

```
tibble(y = cue_nocue$correct,
        pred = predict(glmer3,
                        type = 'response')) %>%
  ggplot(aes(x = pred,
             y = y)) +
  geom_point() +
  geom_abline(intercept = 0,
              slope = 1,
              lty = 2) +
  geom_smooth() +
  xlab("spá gildi") +
  ylab("raun gildi") +
  ggtitle("Kvörðun mixed effect líkan 3") +
  coord_cartesian(xlim = c(0, 1),
                  ylim = c(0, 1))
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```


Kvörðun mixed effect líkan 3



Spáir of lágt fyrir mest allt en engu að síður nokkuð gott og best hingað til.

ROC kúrf, AUC og brier gildi.

```
phats <- fitted(glmer3)
```

```
auc(cue_nocue$correct,
    phats) -> auc
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
b <- mean((cue_nocue$correct - predict(glmer3,
                                         type = 'response'))^2)
bm <- mean(predict(glmer3,
                   type = 'response'))*(1-mean(predict(glmer3,
                                                         type = 'response')))
```

```
bs <- 1 - b/bm
```

```
pred <- prediction(phats,
                   cue_nocue$correct)
```

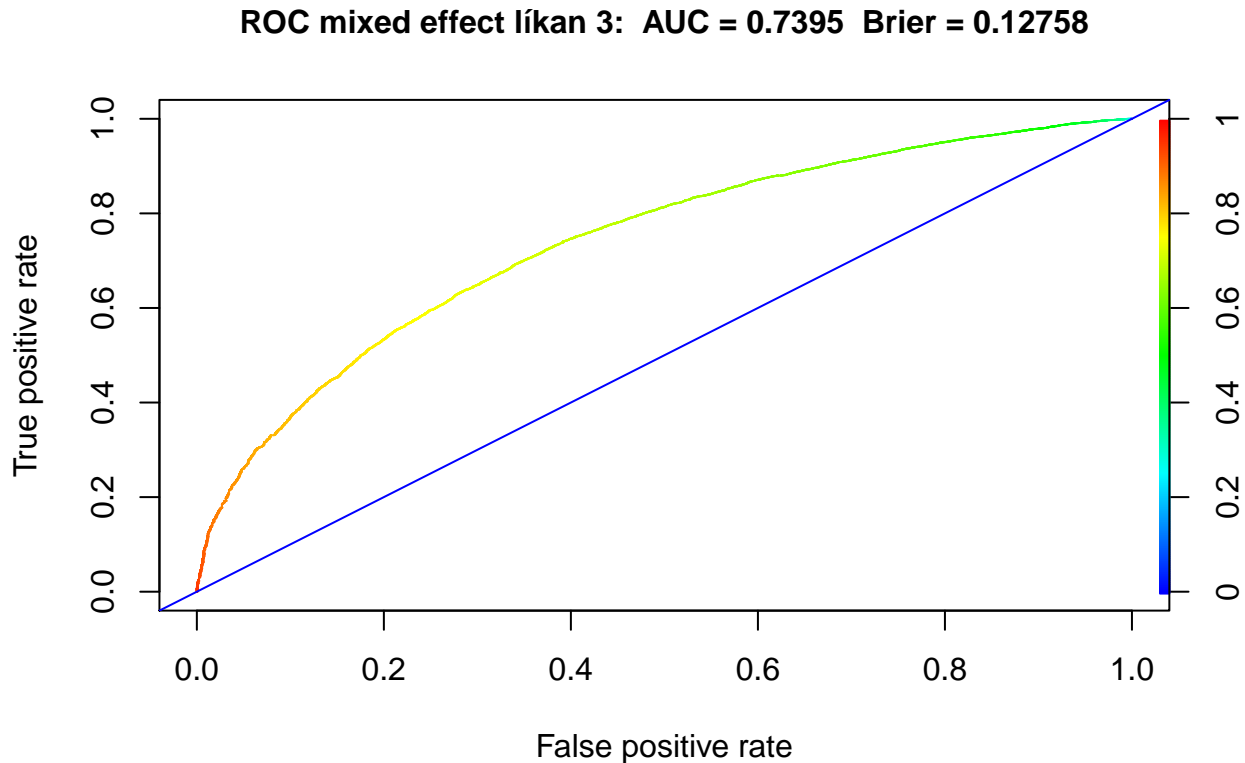
```
perf <- performance(pred,
                    "tpr",
                    "fpr")
```

```
plot(perf,
     colorize=T,
```

```

cex.main=1,
main= paste("ROC mixed effect líkan 3:  AUC =",
            round(auc,4),
            " Brier =",
            round(bs,5))
abline(a=0,
       b = 1,
       col='blue')

```



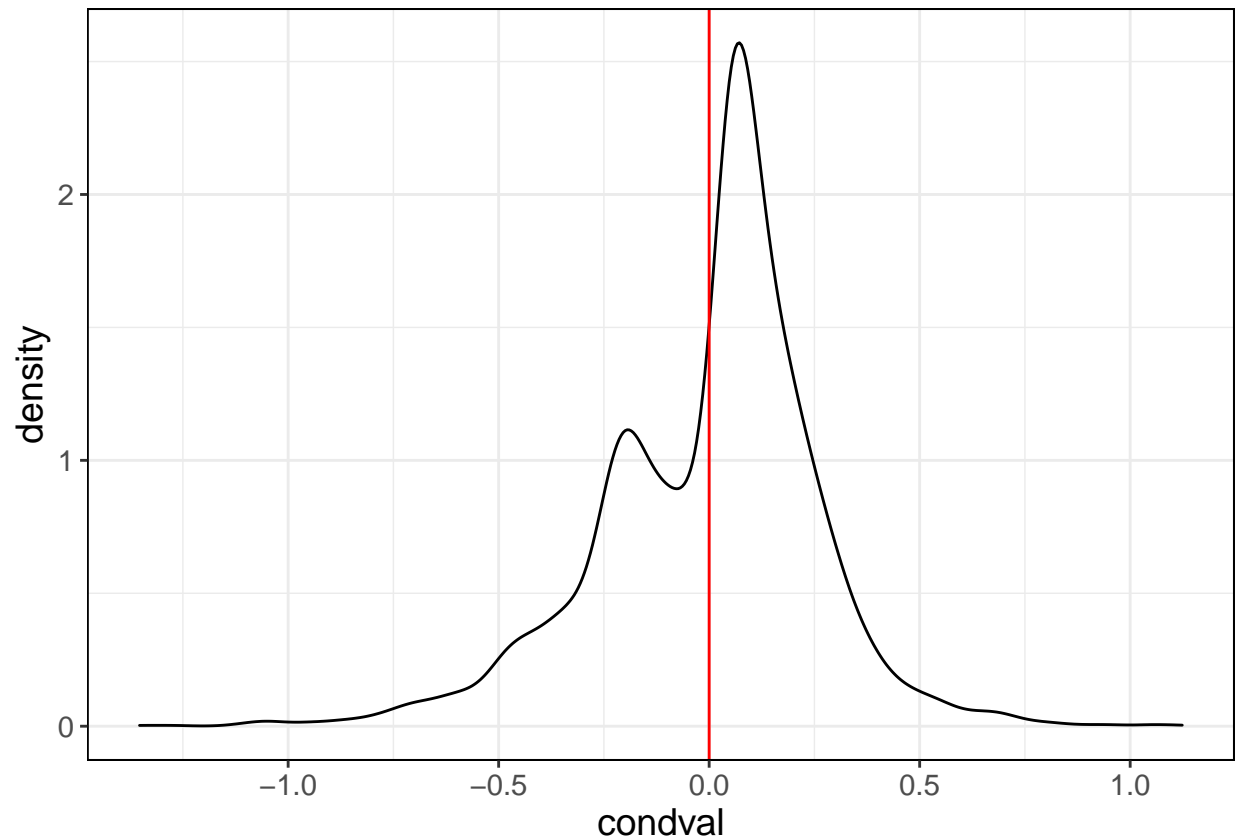
Mjög gott AUC gildi og langbesta brier gildi sem komið hefur upp hingað til.

Dreifni skurðpunkta við y-ás fyrir hvern nemanda í mixed effect líkaninu. Sýnir random effect.

```

ranef(glmer3) %>%
  as_tibble() -> ranef_fitful
ranef_fitful %>%
  ggplot(aes(x = condval)) +
  geom_density() +
  geom_vline(xintercept = mean(ranef_fitful$condval),
            col = 'red')

```



Bootstrap

Bootstrap með cue, qName og interaction. 100 ítranir.

```
# Taflan gefur okkur auc bjartsýni fyrir hverja ítrun þar sem bjartsýnin er mismunurinn á bootstrap gagni
boot <- read.csv("drasl_cue_allt.csv")
```

Myndræn framsetning bootstrap töflunnar.

```
# density = þéttleiki ?

boot %>%
  dplyr::select(1,2,3) %>%
  gather(type,
    Score) %>%
  mutate(type = factor(type,
    levels = c('auc_b',
               'auc_afgangs',
               'auc_opt'),
    labels = c('Bootstrap',
               'Afgangs',
               'Bjartsýni')))) %>%

  ggplot(aes(x = Score)) +
  geom_density() +
  facet_wrap(~type,
    scales = 'free') +
```

```

labs(x="Gildi", title = "AUC")-> p1

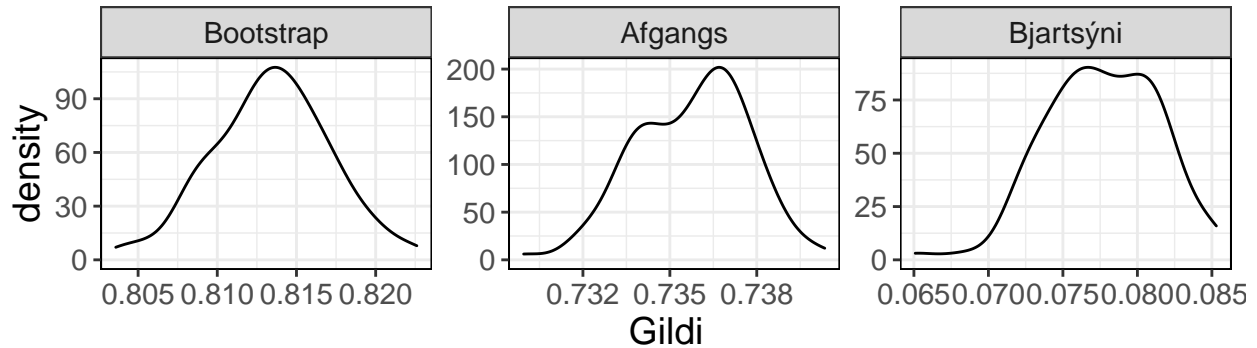
boot %>%
  dplyr::select(4,5,6) %>%
  gather(type,
          Score) %>%
  mutate(type = factor(type,
                        levels = c('brier_afgangs',
                                   'brier_b',
                                   'brier_opt'),
                        labels = c('Bootstrap',
                                   'Afgangs',
                                   'Bjartsýni')) %>%

ggplot(aes(x = Score)) +
  geom_density() +
  facet_wrap(~type,
             scales = 'free') +
  labs(x="Gildi", title = "Brier") -> p2

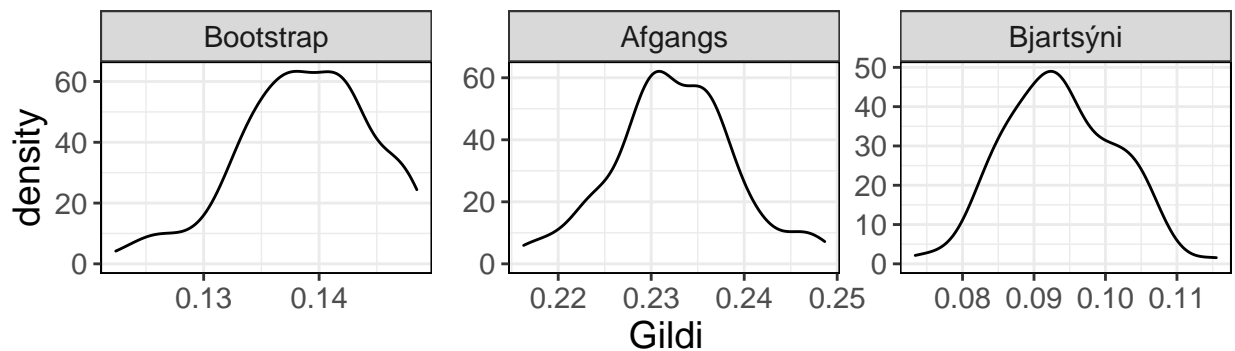
plot_grid(p1, p2,
          align = 'v',
          ncol = 1)

```

AUC



Brier



ROC kúrfra með leiðréttum gildum af AUC og brier eftir bootstrap.

Brier gildi án bootstrap.

```
b <- mean((cue_nocue$correct - predict(glmer3,
                                     type = 'response'))^2)
bm <- mean(predict(glmer3,
                  type = 'response')) * (1 - mean(predict(glmer3,
                                                          type = 'response'))))
bs <- 1 - b/bm

brier_medal_opt <- mean(boot$brier_opt)

brier_leidrett <- bs - brier_medal_opt

auc_likan <- auc(roc(cue_nocue$correct,
                   predict(glmer3,
                           type = 'response'),
                   quiet = T))[[1]]

auc_medal_opt <- mean(boot$auc_opt)

auc_leidrett <- auc_likan - auc_medal_opt

phats <- fitted(glmer3)

pred <- prediction(phats,
                  cue_nocue$correct)

perf <- performance(pred,
                    "tpr",
                    "fpr")

plot(perf,
     colorize=T,
     cex.main=1,
     main= paste("ROC kúrfra - leiðrétt AUC og brier
                  AUC =",
                  round(auc_leidrett,4),
                  " Brier =",
                  round(brier_leidrett,5)),
     cex.sub=1,)
abline(a=0, b = 1, col='blue')
```

ROC kúrfra – leiðrétt AUC og brier
AUC = 0.6619 Brier = 0.0339

