

Colour experiment

Contents

Introduction	2
Data	2
Poisson regression	3
Load libraries	4
Load data	4
ANOVA	4
LMER models	5
Results	7
Random slopes	10
Final model	16
Summary	18
Graphs	19
Teaching and Try marking	19
Teaching and checking	20

Introduction

This analysis looks at the sign variants used in a colour naming game between signers of different sign languages meeting after 1 week of interaction and after 3 weeks of interaction. The data was collected by Kang Suk Byun (Kang-Suk.Byun@mpi.nl).

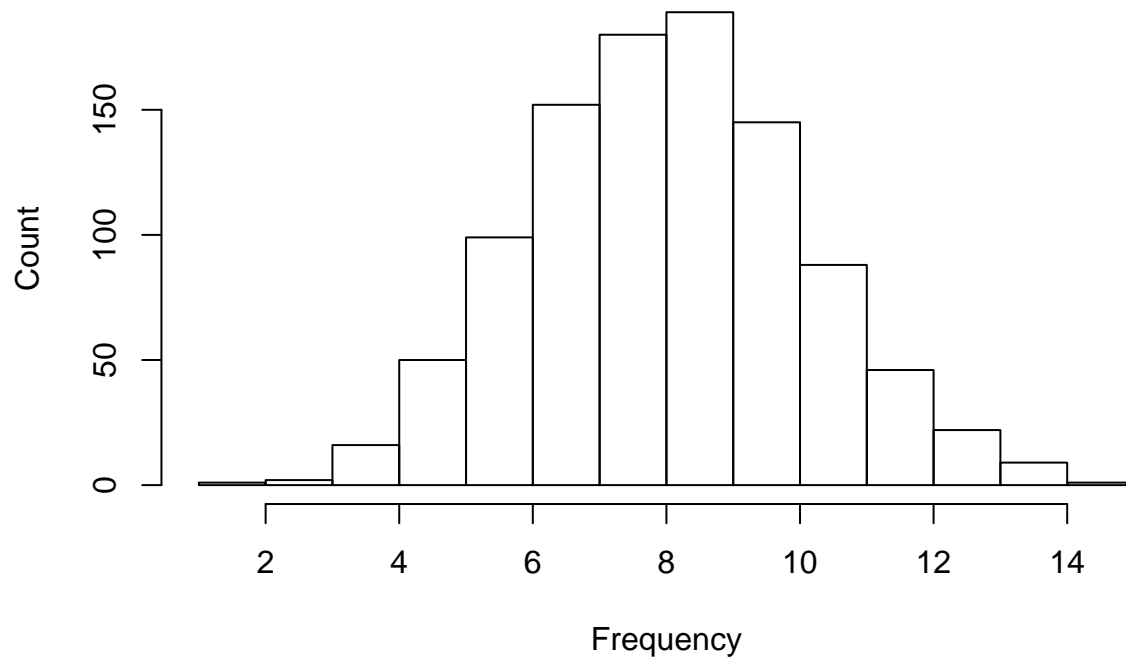
The analysis tries to predict the relative frequency of each variant within a colour category in week 3, based on measures from week 1.

Data

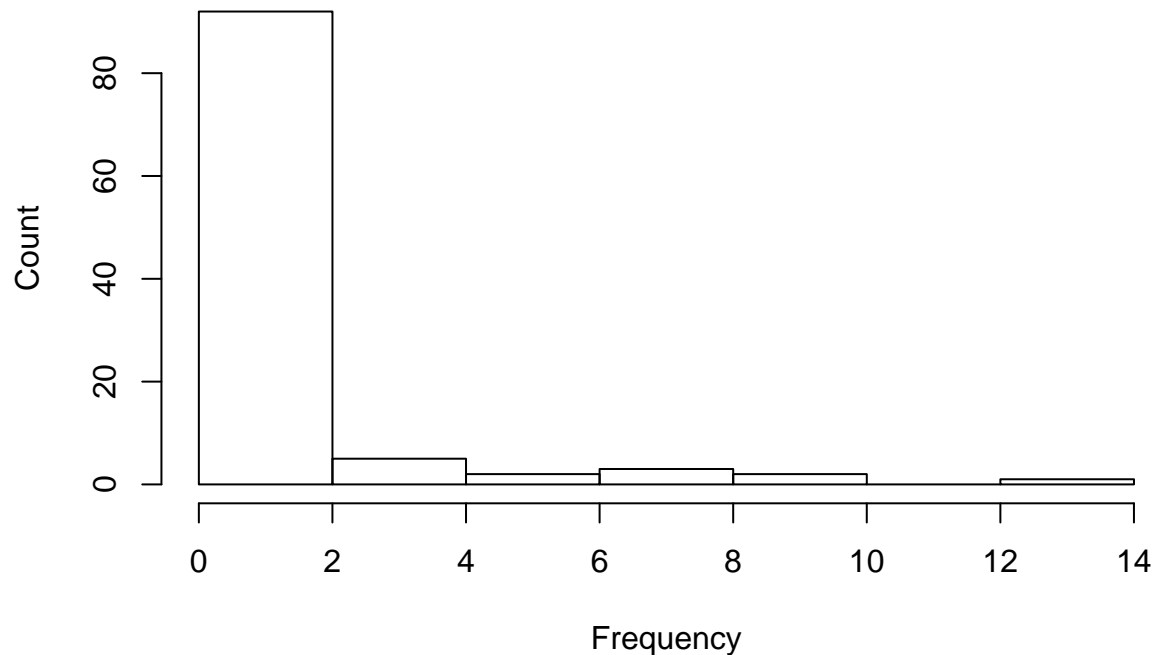
- colour: Code of the target colour
- colourName: English name of the target colour
- sign: label for the variant produced
- freq_week_1_total: Total number of occurrences of the variant in the first week, across all colour contexts.
- freq_week_4_total: Total number of occurrences of the variant in the final week, across all colour contexts.
- freq_week_1: Number of occurrences of the variant used during the given target colour context in week 1.
- freq_week_4: Number of occurrences of the variant used during the given target colour context in the final week.
- prop_week_1: Same as freq_week_1, but as a proportion of all variants used in the given colour context.
- prop_week_4: Same as freq_week_4, but as a proportion of all variants used in the given colour context.
- origin: The origin language of the sign. For many, identifying an origin is not possible, so is labelled “None”
- iconic: Old variable
- check: The number of times this variable was used in a checking turn.
- indexical: Is the variant non-indexical, indexical or indexical of the body?
- inventedBy: The name of the first signer to use this variant in the experiment.
- TryMarked: The number of times this sign was used in try-marking.
- Teach: The number of times this sign was explicitly taught.
- averageLength_week_1: Average time to produce the variant in milliseconds
- averageTrialLength_week_1: Average time for completing the trial for the given target colour.
- BodyAnchor: Is the variant body-anchored (redundant with ‘indexical’)

Poisson regression

This study uses a mixed effects regression model with poisson distributions. Most standard regression analyses assume that the values they are trying to model come from a normal distribution, like this:



However, the main variable for this study is the frequency of sign variants, with a strong skew and many zero values:



Instead of using a normal distribution as the basis for the statistical model or transforming the data (which is difficult anyway because of the large number of zero counts), we can use a poisson distribution. This also has the advantage of only predicting whole, non-negative numbers, which makes sense for this data because a variant can't be used half a time or a negative number of times.

Load libraries

```
library(ggplot2)
library(lme4)
library(party)
library(Rmisc)
library(dplyr)
```

Load data

```
variants = read.csv('../data/processedData/variants_summary.csv', stringsAsFactors = F)
```

There is only 1 variant for 'white'. Therefore, we remove it from this statistical analysis.

```
variants = variants[variants$colourName!='white',]
```

Transform some variables.

```
# The range of values for 'Teach' is very small:
table(variants$Teach)
```

```
##
##  0  1  2  3
## 86 12  3  3
```

```
# So we'll turn it into a binary category:
# variants that were never taught and variants that were
variants$Teach = variants$Teach > 0
```

```
# Similar for checking
variants$check.any = variants$check > 0
```

```
# Transform total frequency
variants$freq_week_1_total.logcenter =
  log(variants$freq_week_1_total + 1)
variants$freq_week_1_total.logcenter =
  variants$freq_week_1_total.logcenter - mean(variants$freq_week_1_total.logcenter)
```

```
# cut TryMarking into two categories
variants$TryMarked.cat = cut(variants$TryMarked,
                             c(-Inf, 3, Inf),
                             labels = c("Low", "High"))
```

```
# transform length
variants$averageLength_week_1.logcenter = log(variants$averageLength_week_1)
variants$averageLength_week_1.logcenter =
  variants$averageLength_week_1.logcenter -
  mean(variants$averageLength_week_1.logcenter)
```

ANOVA

Perform a straightforward ANOVA analysis for comparison:

```

m1 = aov(prop_week_4+1 ~
  indexical +
  Teach * TryMarked*check.any+
  freq_week_1_total.logcenter +
  averageLength_week_1.logcenter+
  inventedBy + colourName,
  data=variants)
summary(m1)

```

```

##              Df Sum Sq Mean Sq F value    Pr(>F)
## indexical      2  0.2569   0.1284    6.575 0.00222 **
## Teach          1  0.0046   0.0046    0.233 0.63059
## TryMarked      1  0.4342   0.4342   22.228 9.52e-06 ***
## check.any      1  0.0529   0.0529    2.710 0.10347
## freq_week_1_total.logcenter 1  0.0307   0.0307    1.570 0.21366
## averageLength_week_1.logcenter 1  0.0046   0.0046    0.234 0.62994
## inventedBy     3  0.0402   0.0134    0.686 0.56321
## colourName     5  0.2064   0.0413    2.114 0.07165 .
## Teach:TryMarked 1  0.0623   0.0623    3.189 0.07773 .
## Teach:check.any 1  0.0640   0.0640    3.278 0.07378 .
## TryMarked:check.any 1  0.0161   0.0161    0.826 0.36613
## Teach:TryMarked:check.any 1  0.0465   0.0465    2.381 0.12660
## Residuals     84  1.6408   0.0195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

LMER models

Each model predicts the frequency of a variant in week 4, with a random intercept by colourName. The random intercept allows some colours to have higher variant frequencies than others. This is useful because we know that signs for some colours are converged on quickly, making their frequencies within those colours potentially higher. In other words, the use of a particular variant to refer to a given colour is not entirely independent of the use of another variant to refer to the same colour.

We begin with a null model and gradually add predictor variables, using model comparison to judge the significance of each variable.

```

# Null model
m0 = glmer(freq_week_4 ~
  1
  + (1 | colourName) ,
  data=variants, family=poisson)

# add frequency in week 1
m1 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  + (1 | colourName) ,
  data=variants, family=poisson)

# add indexicality
m2 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +

```

```

      + (1 | colourName) ,
      data=variants, family=poisson)

# add whether the variant is explicitly taught
m3 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach)
  + (1 | colourName) ,
  data=variants, family=poisson)

# Add try marking
m4 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach) + (TryMarked)
  + (1 | colourName) ,
  data=variants, family=poisson)

# Add the interaction between teaching and try marking
m5 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  Teach * TryMarked
  + (1 | colourName) ,
  data=variants, family=poisson)

# Add the average length of the sign
m6 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter +
  (1 | colourName) ,
  data=variants, family=poisson)

# Add checks
m7 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any
  + (1 | colourName) ,
  data=variants, family=poisson)

# Add the identity of the first signer to use the variant
m8 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy

```

```

+ (1 | colourName) ,
data=variants, family=poisson)

# Add interaction between teaching and checking
m9 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any
+ (1 | colourName) ,
data=variants, family=poisson)

# Add interaction between try marking and checking
m10 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any +
  check.any : TryMarked
+ (1 | colourName) ,
data=variants, family=poisson)

# Add 3 way interaction for teaching, checking and try marking
m11 = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach * check.any * TryMarked
+ (1 | colourName) ,
data=variants, family=poisson)

```

Results

Model comparison test:

```
anova(m0,m1,m2,m3,m4,m5, m6,m7,m8, m9, m10, m11)
```

```

## Data: variants
## Models:
## m0: freq_week_4 ~ 1 + (1 | colourName)
## m1: freq_week_4 ~ 1 + freq_week_1_total + +(1 | colourName)
## m2: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + +(1 | colourName)
## m3: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach) +
## m3:      (1 | colourName)
## m4: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach) +

```

```

## m4:      (TryMarked) + (1 | colourName)
## m5: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + Teach * TryMarked +
## m5:      (1 | colourName)
## m6: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m6:      TryMarked) + averageLength_week_1.logcenter + (1 | colourName)
## m7: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m7:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m7:      (1 | colourName)
## m8: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8:      inventedBy + (1 | colourName)
## m9: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m9:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m9:      inventedBy + Teach:check.any + (1 | colourName)
## m10: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m10:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m10:      inventedBy + Teach:check.any + check.any:TryMarked + (1 |
## m10:      colourName)
## m11: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m11:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m11:      inventedBy + Teach * check.any * TryMarked + (1 | colourName)
##      Df      AIC      BIC    logLik deviance   Chisq Chi Df Pr(>Chisq)
## m0    2 374.62 379.91 -185.309   370.62
## m1    3 344.77 352.70 -169.383   338.77 31.8534      1 1.663e-08 ***
## m2    5 305.74 318.96 -147.869   295.74 43.0270      2 4.537e-10 ***
## m3    6 292.79 308.66 -140.397   280.79 14.9431      1 0.0001108 ***
## m4    7 245.78 264.29 -115.889   231.78 49.0159      1 2.539e-12 ***
## m5    8 227.55 248.70 -105.774   211.55 20.2316      1 6.861e-06 ***
## m6    9 226.82 250.62 -104.411   208.82  2.7256      1 0.0987531 .
## m7   10 224.48 250.93 -102.241   204.48  4.3404      1 0.0372172 *
## m8   13 219.75 254.13  -96.877   193.75 10.7274      3 0.0132950 *
## m9   14 216.91 253.93  -94.457   188.91  4.8405      1 0.0277992 *
## m10  15 217.82 257.49  -93.912   187.82  1.0897      1 0.2965421
## m11  16 216.62 258.93  -92.310   184.62  3.2032      1 0.0734930 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Inside the chosen final model (m10 and 11 add many parameters and do not improve the model, so we choose m9):

```
summary(m9)
```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
##      TryMarked) + averageLength_week_1.logcenter + check.any +
##      inventedBy + Teach:check.any + (1 | colourName)
## Data: variants
##
##      AIC      BIC    logLik deviance df.resid
##    216.9    253.9    -94.5    188.9      90
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max

```

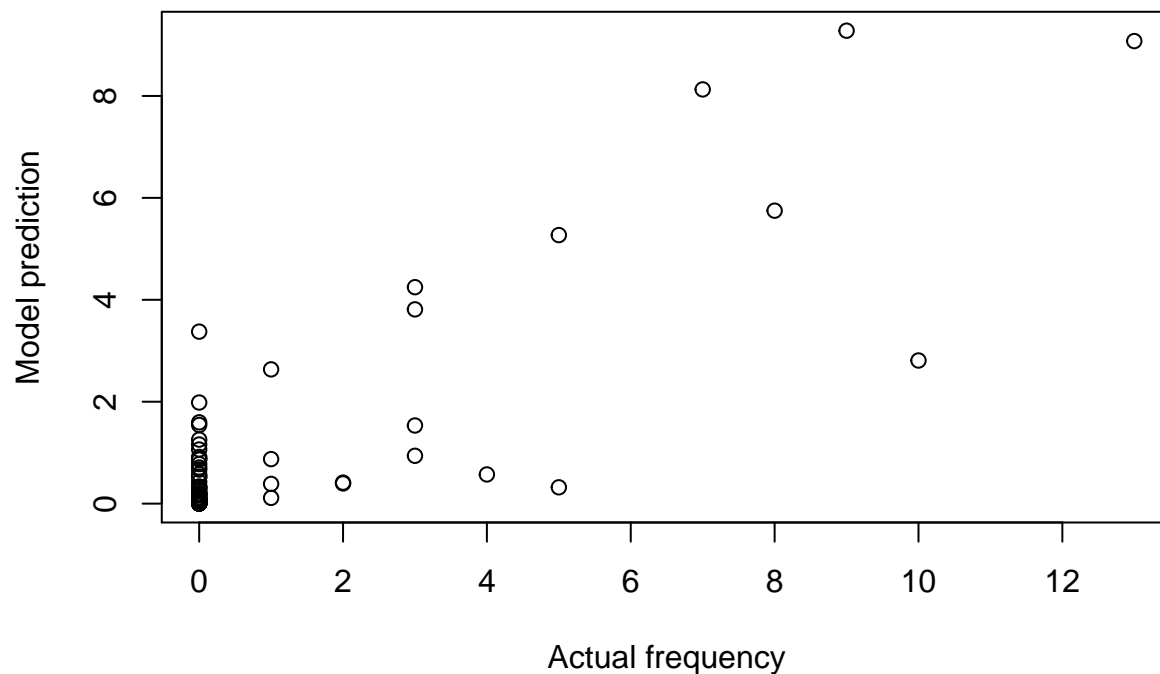


```
## -1.8372 -0.4471 -0.2450 -0.1015 8.2693
##
## Random effects:
## Groups Name Variance Std.Dev.
## colourName (Intercept) 3.204 1.79
## Number of obs: 104, groups: colourName, 6
##
## Fixed effects:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.54877 0.90421 -3.925 8.68e-05 ***
## freq_week_1_total -0.01242 0.02246 -0.553 0.58033
## indexicalYes 1.34306 0.68878 1.950 0.05119 .
## indexicalYes-body 1.04189 0.32792 3.177 0.00149 **
## TeachTRUE 2.52779 0.92965 2.719 0.00655 **
## TryMarked 0.96935 0.13895 6.976 3.04e-12 ***
## averageLength_week_1.logcenter -0.01335 0.19040 -0.070 0.94410
## check.anyTRUE 1.11633 0.38408 2.906 0.00365 **
## inventedByIndonesia 0.25399 0.48605 0.523 0.60128
## inventedByJordan 1.45726 0.55910 2.606 0.00915 **
## inventedByNepal -0.32466 0.47785 -0.679 0.49687
## TeachTRUE:TryMarked -0.58263 0.14613 -3.987 6.69e-05 ***
## TeachTRUE:check.anyTRUE -1.75427 0.81901 -2.142 0.03220 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it
```

See how well the model predictions match the real data:

```
plot(variants$freq_week_4, exp(predict(m9)),
     xlab="Actual frequency",
     ylab="Model prediction")
```



Random slopes

For each of the predictors, we see if random slopes help improve the model. Random slopes allow the strenght of the effect of a factor to be different for each colour concept.

```
m8R = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy+
  Teach : check.any
  + (1 | colourName) ,
  data=variants, family=poisson)
```

*# Some convergence issues with this model, so use
bobyqa algorithm for both steps.*

```
m8R.indexical = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any +
  (1 | colourName) +
  (0 + indexical | colourName),
  data=variants, family=poisson,
  control=glmerControl(
    optimizer = "bobyqa",
    optCtrl = list(maxfun=500000)))
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
```

```

## $checkConv, : unable to evaluate scaled gradient
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 2 negative
## eigenvalues
anova(m8R,m8R.indexical)

## Data: variants
## Models:
## m8R: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R:      inventedBy + Teach:check.any + (1 | colourName)
## m8R.indexical: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R.indexical:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R.indexical:      inventedBy + Teach:check.any + (1 | colourName) + (0 + indexical |
## m8R.indexical:      colourName)
##           Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m8R           14 216.91 253.94 -94.457   188.91
## m8R.indexical 20 210.38 263.27 -85.189   170.38 18.535      6 0.005026
##
## m8R
## m8R.indexical **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m8R.Teach = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any +
  (1 + Teach| colourName),
  data=variants, family=poisson,
  control=glmerControl(
    optimizer = "bobyqa",
    optCtrl = list(maxfun=500000)))
anova(m8R,m8R.Teach)

## Data: variants
## Models:
## m8R: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R:      inventedBy + Teach:check.any + (1 | colourName)
## m8R.Teach: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R.Teach:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R.Teach:      inventedBy + Teach:check.any + (1 + Teach | colourName)
##           Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m8R           14 216.91 253.94 -94.457   188.91
## m8R.Teach 16 209.06 251.37 -88.532   177.06 11.85      2 0.002671 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m8R.TryMark = glmer(freq_week_4 ~
  1 + freq_week_1_total +

```

```

      (indexical) +
      (Teach * TryMarked) +
      averageLength_week_1.logcenter+
      check.any + inventedBy +
      Teach : check.any +
      (1 | colourName) +
      (0 + TryMarked| colourName),
      data=variants, family=poisson,
      control=glmerControl(
        optimizer = "bobyqa",
        optCtrl = list(maxfun=500000)))
anova(m8R,m8R.TryMark)

```

```

## Data: variants
## Models:
## m8R: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R:      inventedBy + Teach:check.any + (1 | colourName)
## m8R.TryMark: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R.TryMark:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R.TryMark:      inventedBy + Teach:check.any + (1 | colourName) + (0 + TryMarked |
## m8R.TryMark:      colourName)
##           Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m8R           14 216.91 253.94 -94.457   188.91
## m8R.TryMark 15 214.54 254.20 -92.268   184.54 4.378      1   0.03641 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

m8R.Freq = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any +
  (1 | colourName) +
  (0 + freq_week_1_total| colourName),
  data=variants, family=poisson,
  control=glmerControl(
    optimizer = "bobyqa",
    optCtrl = list(maxfun=500000)))
anova(m8R,m8R.Freq)

```

```

## Data: variants
## Models:
## m8R: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R:      inventedBy + Teach:check.any + (1 | colourName)
## m8R.Freq: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R.Freq:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R.Freq:      inventedBy + Teach:check.any + (1 | colourName) + (0 + freq_week_1_total |
## m8R.Freq:      colourName)
##           Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m8R           14 216.91 253.94 -94.457   188.91

```

```
## m8R.Freq 15 217.09 256.76 -93.545 187.09 1.8226 1 0.177
```

```
m8R.Length = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any +
  (1 | colourName) +
  (0 + averageLength_week_1.logcenter| colourName),
data=variants, family=poisson,
control=glmerControl(
  optimizer = "bobyqa",
  optCtrl = list(maxfun=500000)))
anova(m8R,m8R.Length)
```

```
## Data: variants
```

```
## Models:
```

```
## m8R: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
```

```
## m8R: TryMarked) + averageLength_week_1.logcenter + check.any +
```

```
## m8R: inventedBy + Teach:check.any + (1 | colourName)
```

```
## m8R.Length: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
```

```
## m8R.Length: TryMarked) + averageLength_week_1.logcenter + check.any +
```

```
## m8R.Length: inventedBy + Teach:check.any + (1 | colourName) + (0 + averageLength_week_1.logcenter
```

```
## m8R.Length: colourName)
```

```
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
## m8R 14 216.91 253.94 -94.457 188.91
```

```
## m8R.Length 15 218.91 258.58 -94.457 188.91 0 1 0.9998
```

```
m8R.Check = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any +
  (1 | colourName) +
  (0 + check.any| colourName),
data=variants, family=poisson,
control=glmerControl(
  optimizer = "Nelder_Mead",
  optCtrl = list(maxfun=500000)))
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
```

```
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
```

```
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
```

```
## eigenvalues
```

```
anova(m8R,m8R.Check)
```

```
## Data: variants
```

```
## Models:
```

```
## m8R: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
```

```
## m8R: TryMarked) + averageLength_week_1.logcenter + check.any +
```

```
## m8R:      inventedBy + Teach:check.any + (1 | colourName)
## m8R.Check: freq_week_4 ~ 1 + freq_week_1_total + (indexical) + (Teach *
## m8R.Check:      TryMarked) + averageLength_week_1.logcenter + check.any +
## m8R.Check:      inventedBy + Teach:check.any + (1 | colourName) + (0 + check.any |
## m8R.Check:      colourName)
##          Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m8R          14 216.91 253.94 -94.457   188.91
## m8R.Check    17 215.15 260.10 -90.574   181.15 7.7646     3 0.05113 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation of random slopes

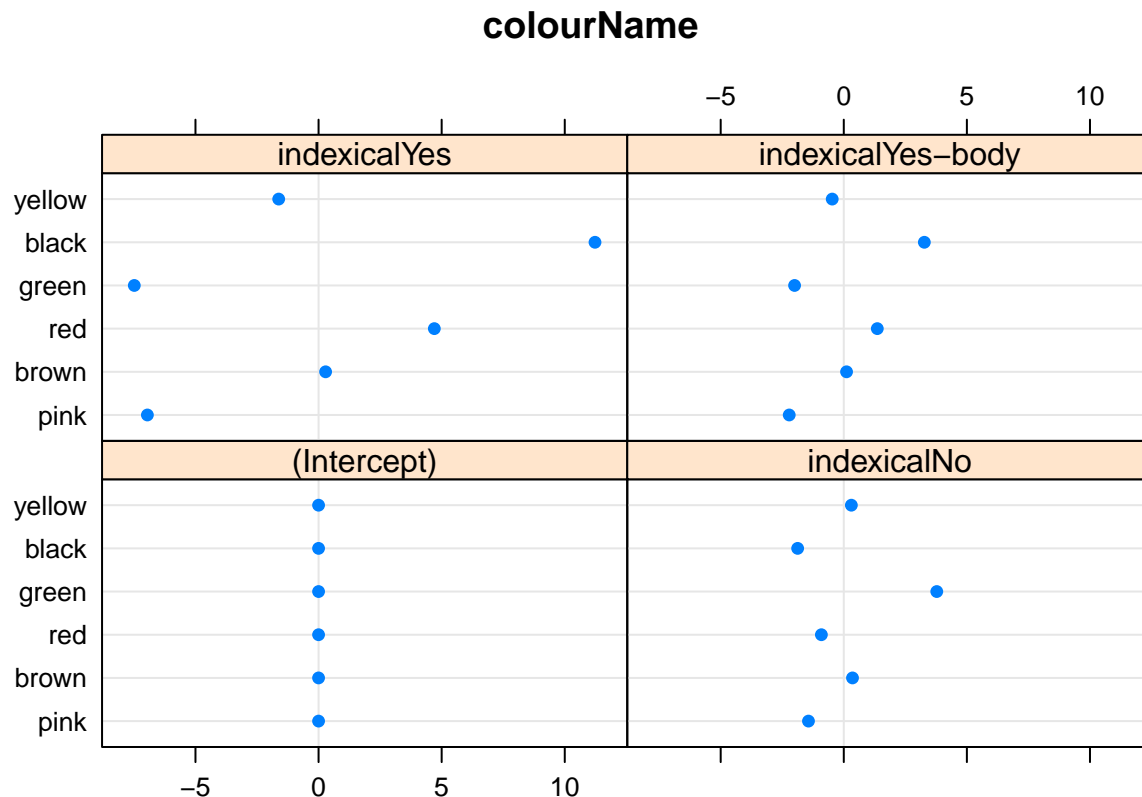
We see that indexicality and checking improves the model as a random slope.

- Indexicality

For some colours, indexicality matters more than for others. We can plot the random slopes below:

```
dotplot(ranef(m8R.indexical))
```

```
## $colourName
```



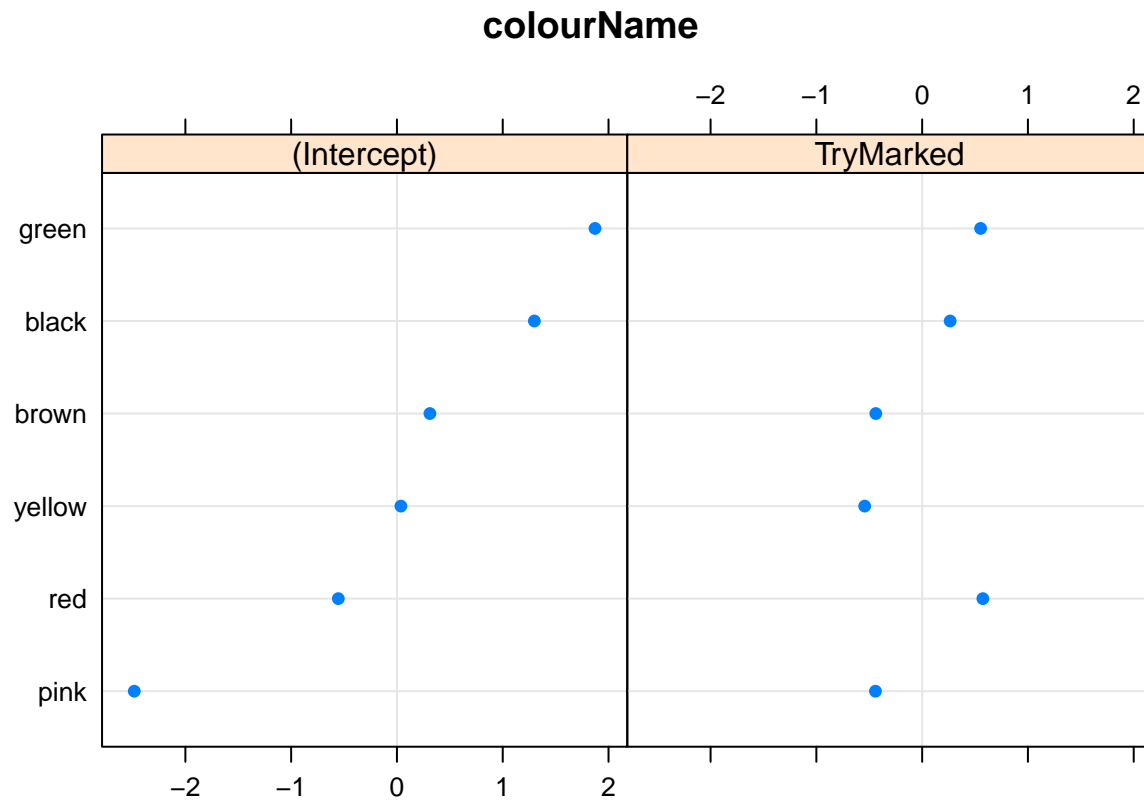
We see that indexicality improves the chance of selection of variants for black and red, but reduces the chance of selection for green and pink.

- Try Marking

For some colours, try marking has a bigger effect than others. Here are the random slopes.

```
dotplot(ranef(m8R.TryMark))
```

```
## $colourName
```

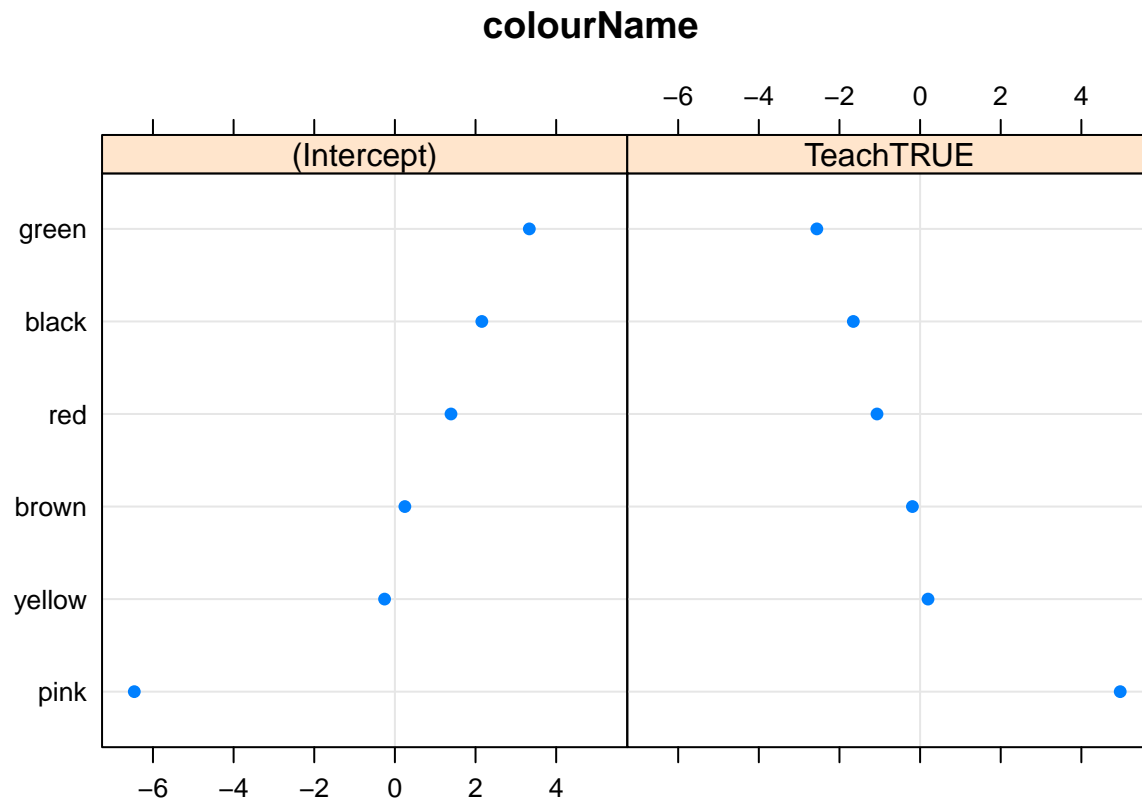


Try marking improves the probability of selection for red, green and black, but decreases the probability of selection for brown, yellow and pink.

- Teaching

```
dotplot(ranef(m8R.Teach))
```

```
## $colourName
```



Explicit teaching improves the likelihood of teaching for pink, but not for other colours.

Final model

Choose a final model for the beta values. Both *indexical* and *Teach* are included as random slopes.

```
finalModel = glmer(freq_week_4 ~
  1 + freq_week_1_total +
  (indexical) +
  (Teach * TryMarked) +
  averageLength_week_1.logcenter+
  check.any + inventedBy +
  Teach : check.any +
  (1 + indexical + TryMarked + Teach | colourName),
  data=variants, family=poisson,
  control=glmerControl(
    optimizer = "bobyqa",
    optCtrl = list(maxfun=500000)))
```

Fixed effects:

```
coef(summary(finalModel))[1:3]
```

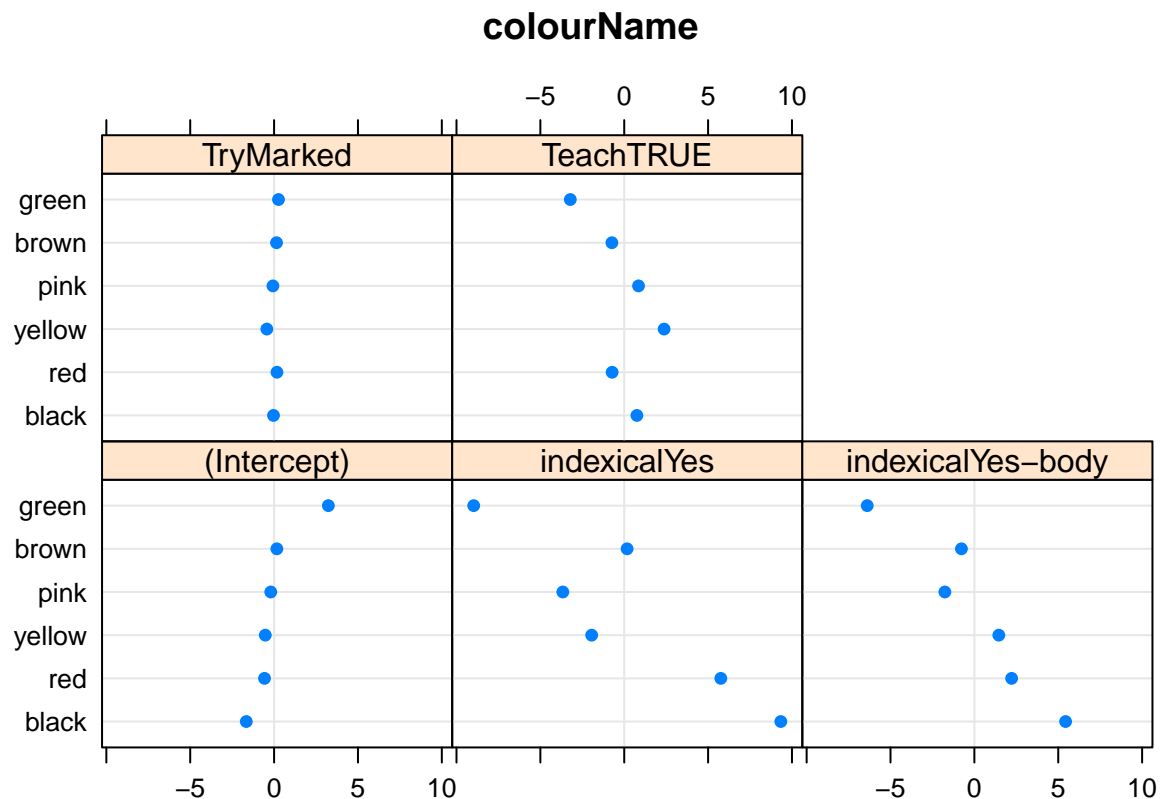
	Estimate	Std. Error	z value
## (Intercept)	-4.62845415	1.3277929	-3.48582522
## freq_week_1_total	-0.03654130	0.0405721	-0.90065094
## indexicalYes	3.07039606	3.5890358	0.85549330
## indexicalYes-body	0.85307908	1.9291485	0.44220498
## TeachTRUE	1.31300570	2.4908961	0.52712183


```
## TryMarked 0.81665673 0.3447713 2.36869141
## averageLength_week_1.logcenter 0.21784996 0.2696942 0.80776646
## check.anyTRUE 0.05057627 0.7065868 0.07157828
## inventedByIndonesia 1.29500734 0.8892337 1.45631833
## inventedByJordan 2.08451592 1.0210934 2.04145477
## inventedByNepal 0.93877558 0.7874702 1.19214102
## TeachTRUE:TryMarked -0.22562534 0.3470725 -0.65008135
## TeachTRUE:check.anyTRUE -0.35330907 1.1057324 -0.31952494
```

Random effects:

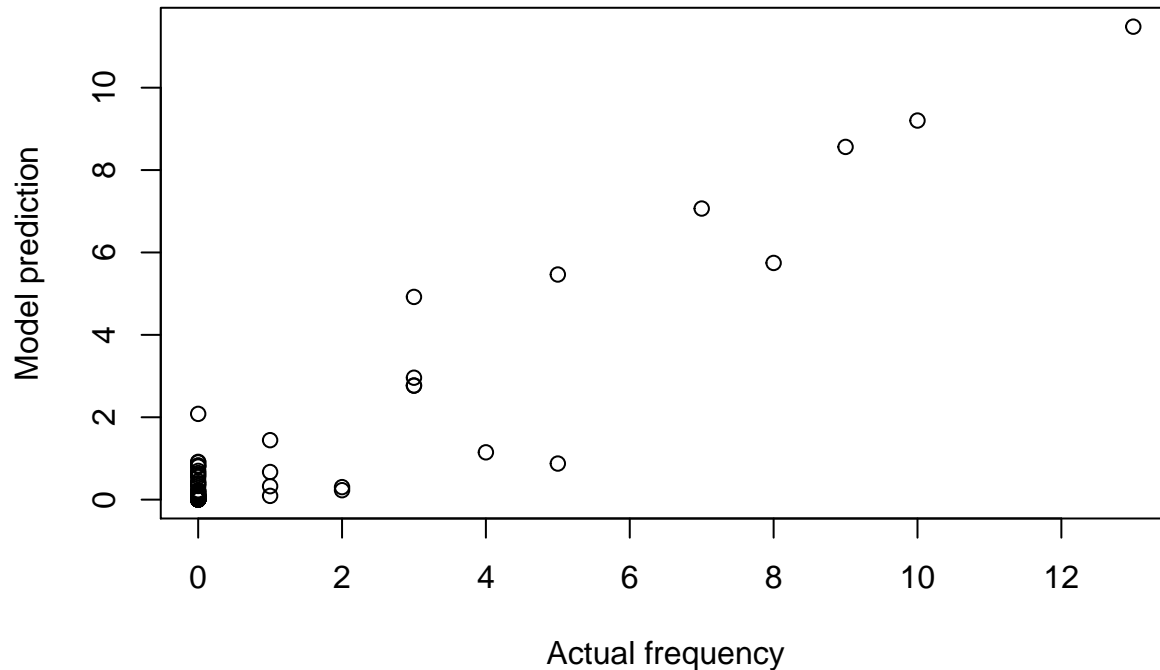
```
dotplot(ranef(finalModel))
```

```
## $colourName
```



Check the predictions:

```
plot(variants$freq_week_4, exp(predict(finalModel)),
     xlab="Actual frequency",
     ylab="Model prediction")
```



Summary

Here is a summary of the main results:

There was a significant main effect of frequency in week 1 ($\beta = -0.037$, $\text{std.err} = 0.041$, $\text{Wald } t = -0.9$; $\log \text{ likelihood difference} = 16$, $df = 1$, $\text{Chi Squared} = 31.85$, $p = 1.7e-08$). On its own, more frequent variants in week 1 are also more frequent in the final week. However, when considering this variable with the other variables the relationship is reversed: variants used a lot in week 1 are slightly less likely to be used in the final week. An explanation may be the following: A poor variant may be repeated many times before it is understood, while a good variant only needs to be used once. That is, frequent use in the first week may be an indication of communication problems.

There was a significant main effect of indexicality ($\beta = 3.1$, $\text{std.err} = 3.6$, $\text{Wald } t = 0.86$; $\log \text{ likelihood difference} = 22$, $df = 2$, $\text{Chi Squared} = 43.03$, $p = 4.5e-10$). (β for body-indexical = 0.853). Indexical variants were more likely to be selected.

There was a significant main effect of teaching ($\beta = 1.3$, $\text{std.err} = 2.5$, $\text{Wald } t = 0.53$; $\log \text{ likelihood difference} = 7.5$, $df = 1$, $\text{Chi Squared} = 14.94$, $p = 0.00011$). Teaching increased the likelihood of selection.

There was a significant main effect of try marking ($\beta = 0.82$, $\text{std.err} = 0.34$, $\text{Wald } t = 2.4$; $\log \text{ likelihood difference} = 25$, $df = 1$, $\text{Chi Squared} = 49.02$, $p = 2.5e-12$). Variants that were try marked more often were more likely to be selected.

There was no significant main effect of sign length ($\beta = 0.22$, $\text{std.err} = 0.27$, $\text{Wald } t = 0.81$; $\log \text{ likelihood difference} = 1.4$, $df = 1$, $\text{Chi Squared} = 2.73$, $p = 0.099$).

There was a significant main effect of checking ($\beta = 0.051$, $\text{std.err} = 0.71$, $\text{Wald } t = 0.072$; $\log \text{ likelihood difference} = 2.2$, $df = 1$, $\text{Chi Squared} = 4.34$, $p = 0.037$). A variant was more likely to be selected if it had been used in a checking context.

There was a significant main effect of first user (inventedBy) ($\log \text{ likelihood difference} = 5.4$, $df = 3$, $\text{Chi Squared} = 10.73$, $p = 0.013$). Compared to the Indian signer, signs first used by the Jordanian signer were slightly more likely to be selected and signs first used by the Indonesian and Nepalese signers were slightly less likely to be selected.

There was a significant interaction between try marking and teaching ($\beta = -0.23$, $\text{std.err} = 0.35$, Wald $t = -0.65$; log likelihood difference = 10 , $\text{df} = 1$, Chi Squared = 20.23 , $p = 6.9\text{e-}06$). The effect of teaching was bigger when the variant was also often try marked (see graphs below).

There was a significant interaction between teaching and checking ($\beta = -0.35$, $\text{std.err} = 1.1$, Wald $t = -0.32$; log likelihood difference = 2.4 , $\text{df} = 1$, Chi Squared = 4.84 , $p = 0.028$). (see graphs below).

We also found that certain factors are more important for particular colours (see above).

Graphs

Teaching and Try marking

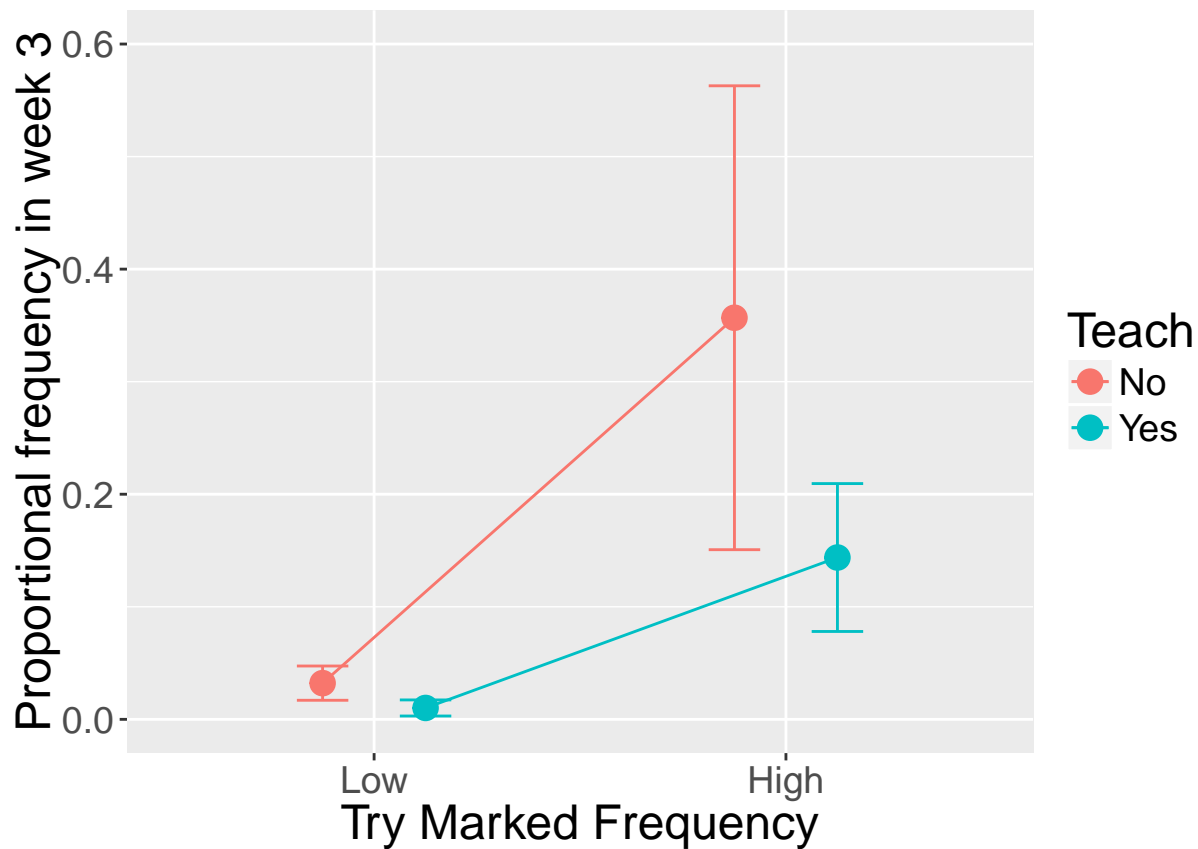
Plot the interaction between teaching and try marking.

Overall, the model suggests that teaching a variant improves its chances of being selected. However, this effect is mainly due to the interaction between teaching and try marking.

```
sumStats2 = summarySE(variants, measurevar="prop_week_4",
                      groupvars=c("TryMarked.cat", "Teach"))

dodge <- position_dodge(width=0.5)

main.plot <- ggplot(sumStats2,
  aes(x = TryMarked.cat, y = prop_week_4, colour=Teach)) +
  geom_point(position=dodge, size=4) + geom_line(aes(group=Teach), position=dodge) +
  geom_errorbar(aes(ymax=prop_week_4+se, ymin=prop_week_4-se), width=0.25, position=dodge) +
  xlab("Try Marked Frequency") +
  ylab("Proportional frequency in week 3") +
  coord_cartesian(ylim=c(0,0.6)) +
  scale_color_discrete(breaks=c(FALSE, TRUE),
                      labels=c("No", "Yes"),
                      name="Teach") +
  theme(text=element_text(size=18))
main.plot
```



```
## pdf
## 2
```

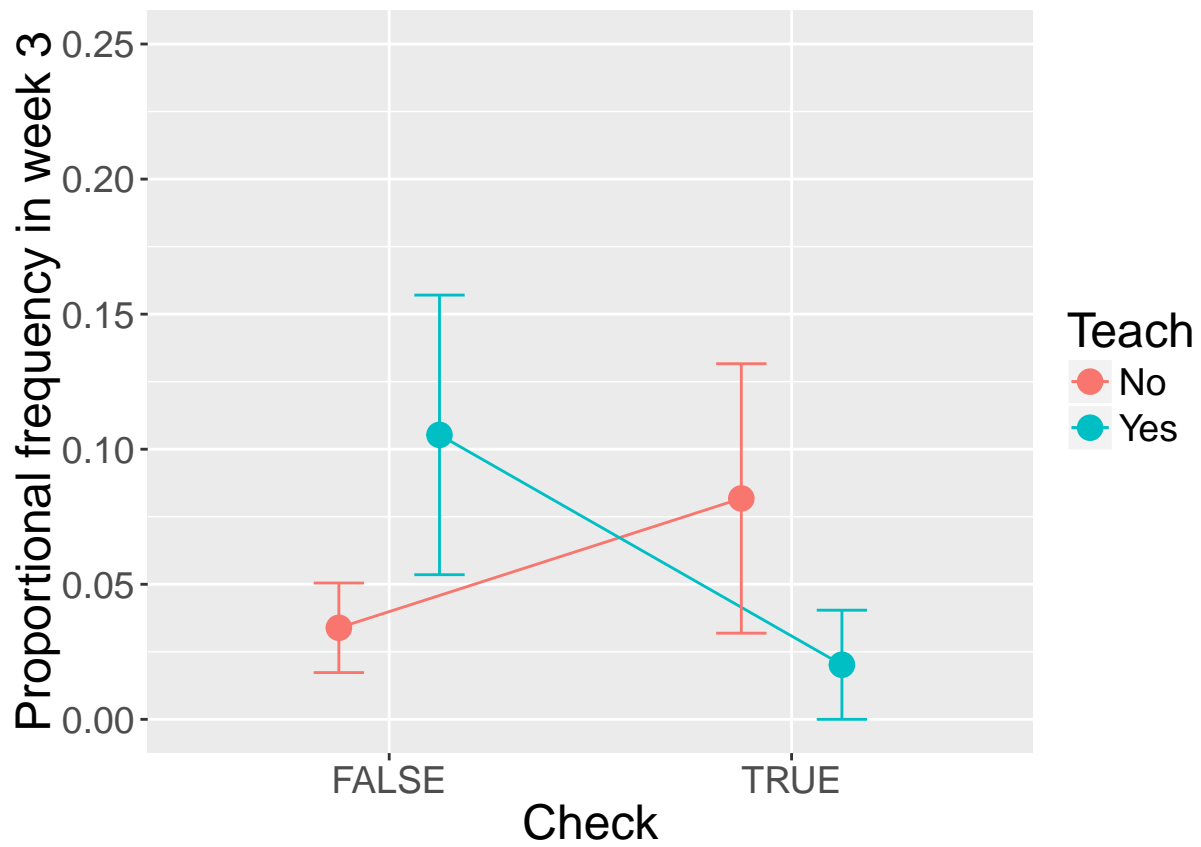
Teaching and checking

Plot the interaction between teaching and checking.

```
sumStats2 = summarySE(variants, measurevar="prop_week_4",
                      groupvars=c("Teach", "check.any"))

dodge <- position_dodge(width=0.5)

main.plot2 <- ggplot(sumStats2,
  aes(x = check.any, y = prop_week_4, colour=Teach)) +
  geom_point(position=dodge, size=4) + geom_line(aes(group=Teach), position=dodge) +
  geom_errorbar(aes(ymax=prop_week_4+se, ymin=prop_week_4-se), width=0.25, position=dodge) +
  xlab("Check") +
  ylab("Proportional frequency in week 3") +
  coord_cartesian(ylim=c(0,0.25)) +
  scale_color_discrete(breaks=c(FALSE,TRUE),
    labels=c("No", "Yes"),
    name="Teach") +
  theme(text=element_text(size=18))
main.plot2
```



```
## pdf
## 2
## pdf
## 2
```