# prompt analysis

2024-10-09

## Data reformatting

```
data.raw = read.csv('llama3p1_8b_base.csv',header=TRUE)
orig <- unique(subset(data.raw,select=c(1,2,8:10))) %>% rename(success = n_successes_original,
                                                               pass = pass1_original) %>% mutate(category = "Ba
                                                                                                 intervention =
                                                                                                 original_word =

subs <- subset(data.raw,select=-c(8:9)) %>% rename(success = n_successes_intervened,
                                                   pass = pass1_intervened,
                                                 category = intervention_category)

data <- rbind(orig,subs) %>% mutate(n = 200)
data$category <- fct_relevel(data$category,"Baseline")
data$intervention <- fct_relevel(data$intervention,"None")
```

## String interventions

```
string.ps <- subset(data,category=="string")$prompt_id
strings <- subset(data,(category=="string"|category=="Baseline")&prompt_id %in% string.ps)
```

```
string.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=s
```

```
summary(string.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: strings
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##  29685.5  29845.1 -14814.7  29629.5     2180
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -47.929  -0.738  -0.128   0.388  56.464
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 29.42812 5.4248
##  problem   (Intercept)                 13.47838 3.6713
```

```
##             interventioncharacter          3.59203 1.8953   -0.04
##             interventionphrase             1.41755 1.1906   -0.59  0.06
##             interventionset_of_characters  4.47041 2.1143   -0.45  0.28  0.30
##             interventionstring             0.07603 0.2757    0.22 -0.14  0.26
##             interventionword               1.21081 1.1004   -0.30  0.49  0.46
##
##
##
##
##
##
##   0.02
##   0.20  0.28
## Number of obs: 2208, groups:  prompt_id, 368; problem, 38
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -4.35762    0.68688  -6.344 2.24e-10 ***
## interventioncharacter         -1.13982    0.32053  -3.556 0.000376 ***
## interventionphrase            -0.68844    0.20130  -3.420 0.000626 ***
## interventionset_of_characters -1.44950    0.35532  -4.079 4.52e-05 ***
## interventionstring             0.04751    0.05431   0.875 0.381684
## interventionword              -0.62409    0.18767  -3.325 0.000883 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnc intrvntnp intr__ intrvntns
## intrvntnchr -0.036
## intrvntnphr -0.511  0.066
## intrvntns__ -0.390  0.277     0.296
## intrvntnstr  0.156 -0.094     0.264     0.045
## intrvntnwrd -0.258  0.488     0.455     0.203  0.277
```

## Dict interventions

```
dict.ps <- subset(data,category=="dictionary")$prompt_id
dicts <- subset(data,(category=="dictionary"|category=="Baseline")&prompt_id %in% dict.ps)
```

```
dict.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=dic
```

```
summary(dict.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: dicts
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   3526.0   3568.1  -1753.0   3506.0      488
##
```

```
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -12.1924  -0.0326  -0.0213  -0.0112   9.0706
##
## Random effects:
##  Groups    Name                    Variance  Std.Dev. Corr
##  prompt_id (Intercept)             139.57329 11.8141
##  problem   (Intercept)               1.72609  1.3138
##            interventiondictionary    0.01858  0.1363   0.77
##            interventionmap           1.88807  1.3741  -0.86 -0.33
## Number of obs: 498, groups:  prompt_id, 166; problem, 15
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -12.78117    1.29146  -9.897   <2e-16 ***
## interventiondictionary  -0.09900    0.05556  -1.782   0.0748 .
## interventionmap         -0.06637    0.40710  -0.163   0.8705
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd
## intrvntndct  0.166
## intervntnmp -0.253 -0.202
```

## Integer interventions

```
int.ps <- subset(data,category=="integer")$prompt_id
ints <- subset(data,(category=="integer"|category=="Baseline")&prompt_id %in% int.ps)
```

```
int.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=ints
```

```
summary(int.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: ints
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   4451.1   4519.9  -2210.5   4421.1      713
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -12.1437  -0.2193  -0.1257   0.2480   8.0524
##
## Random effects:
##  Groups    Name                     Variance  Std.Dev. Corr
##  prompt_id (Intercept)              39.426073 6.27902
##  problem   (Intercept)              13.016898 3.60789
##            interventionint           0.160966 0.40121   0.34
```

```
##             interventioninteger       0.009252 0.09619  -0.04 -0.43
##             interventionwhole_number  1.199782 1.09535  -0.01  0.76 -0.61
## Number of obs: 728, groups:  prompt_id, 182; problem, 27
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -5.93868    0.90125  -6.589 4.42e-11 ***
## interventionint            -0.10463    0.09067  -1.154   0.2485
## interventioninteger         0.05727    0.03769   1.519   0.1286
## interventionwhole_number -0.44797    0.23035  -1.945   0.0518 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intervntnnt intrvntnntg
## intervntnnt  0.232
## intrvntnntg -0.029 -0.063
## intrvntnwh_ -0.015  0.713      -0.259
```

## List interventions

```
list.ps <- subset(data,category=="list")$prompt_id
lists <- subset(data,(category=="list"|category=="Baseline")&prompt_id %in% list.ps)
```

```
list.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=lis
```

```
## Warning in commonArgs(par, fn, control, environment()): maxfun < 10 *
## length(par)^2 is not recommended.
```

```
summary(list.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: lists
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##  59532.8  59759.8 -29730.4  59460.8     4017
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -21.947  -0.410  -0.118   0.245 102.961
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 27.42467 5.2369
##  problem   (Intercept)                 19.05812 4.3656
##            interventionarray            0.24883 0.4988    0.21
##            interventionarray_list       0.38913 0.6238    0.13  0.72
##            interventionbrackets         1.87457 1.3691   -0.17  0.27  0.19
##            interventionlist             0.04482 0.2117    0.07  0.30  0.01
```

```
##            interventionset              9.12200 3.0203   -0.18  0.09 -0.19
##            interventionset_of_brackets  4.72779 2.1743   -0.16  0.09  0.05
##
##
##
##
##
##
##   0.06
##   0.61  0.23
##   0.70  0.07  0.83
## Number of obs: 4053, groups:  prompt_id, 579; problem, 41
##
## Fixed effects:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -5.538747   0.746860  -7.416 1.21e-13 ***
## interventionarray            0.102479   0.086892   1.179    0.238
## interventionarray_list       0.110117   0.107576   1.024    0.306
## interventionbrackets        -1.010364   0.231968  -4.356 1.33e-05 ***
## interventionlist            -0.003201   0.041047  -0.078    0.938
## interventionset             -2.390728   0.510457  -4.684 2.82e-06 ***
## interventionset_of_brackets -1.915900   0.367154  -5.218 1.81e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnr intrv_ intrvntnb intrvntnl intrvntns
## intrvntnrry  0.175
## intrvntnrr_  0.109  0.702
## intrvntnbrc -0.162  0.272     0.198
## intrvntnlst  0.044  0.310     0.052  0.070
## intervntnst -0.162  0.099    -0.176  0.601     0.212
## intrvntns__ -0.146  0.093     0.055  0.693     0.072     0.828
## optimizer (bobyqa) convergence code: 0 (OK)
## maxfun < 10 * length(par)^2 is not recommended.
```

# Key interventions

```
key.ps <- subset(data,category=="key")$prompt_id
keys <- subset(data,(category=="key"|category=="Baseline")&prompt_id %in% key.ps)
```

```
key.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=keys
```

```
## Warning in commonArgs(par, fn, control, environment()): maxfun < 10 *
## length(par)^2 is not recommended.
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
```

```
## Warning in commonArgs(par, fn, control, environment()): maxfun < 10 *
## length(par)^2 is not recommended.
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
```

```
## evaluations exceeded
```

```
summary(key.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: keys
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   7387.2   7609.9  -3648.6   7297.2      995
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -12.4988  -0.1045  -0.0359  -0.0006  22.3280
##
## Random effects:
##  Groups    Name                  Variance Std.Dev. Corr
##  prompt_id (Intercept)           78.5681  8.8639
##  problem   (Intercept)            7.9541  2.8203
##            interventionattribute  0.9820  0.9910   -0.48
##            interventionelement    0.2008  0.4481   -0.62  0.23
##            interventionentry      0.1752  0.4186   -0.55  0.35  0.82
##            interventionitem       0.1873  0.4327   -0.53  0.39  0.84  0.89
##            interventionkey        0.2196  0.4686   -0.03  0.30  0.26  0.66
##            interventionpart       0.4216  0.6493   -0.09  0.55  0.62  0.63
##            interventionvariable   0.1887  0.4344   -0.62  0.68  0.61  0.64
##
##
##
##
##
##
##
##   0.60
##   0.67  0.61
##   0.74  0.62  0.69
## Number of obs: 1040, groups:  prompt_id, 130; problem, 12
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -9.82477    1.25695  -7.816 5.44e-15 ***
## interventionattribute -0.69695    0.31737  -2.196  0.02809 *
## interventionelement   -0.25206    0.14749  -1.709  0.08746 .
## interventionentry     -0.39181    0.13909  -2.817  0.00485 **
## interventionitem      -0.28474    0.14330  -1.987  0.04692 *
## interventionkey        0.04586    0.15426   0.297  0.76623
## interventionpart      -0.59187    0.21068  -2.809  0.00496 **
## interventionvariable  -0.56200    0.14437  -3.893 9.90e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##             (Intr) intrvntntt intrvntnl intrvntnn intrvntntm intrvntnk
## intrvntnttr -0.327
## intrvntnlmn -0.417  0.235
## intrvntnntr -0.367  0.347      0.798
## intervntntm -0.358  0.382      0.813     0.852
## intervntnky -0.030  0.308      0.276     0.648     0.599
## intrvntnprt -0.064  0.544      0.610     0.618     0.660      0.597
## intrvntnvrb -0.415  0.656      0.601     0.625     0.711      0.610
##             intrvntnp
## intrvntnttr
## intrvntnlmn
## intrvntnntr
## intervntntm
## intervntnky
## intrvntnprt
## intrvntnvrb  0.670
## optimizer (bobyqa) convergence code: 1 (bobyqa -- maximum number of function evaluations exceeded)
## maxfun < 10 * length(par)^2 is not recommended.
```

## Typecast interventions

```
cast.ps <- subset(data,category=="typecast")$prompt_id
casts <- subset(data,(category=="typecast"|category=="Baseline")&prompt_id %in% cast.ps)
```

```
cast.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=cast
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
```

```
summary(cast.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: casts
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   2341.8   2437.8  -1142.9   2285.8      200
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -9.0539 -0.5304 -0.1427  0.2090 14.0520
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         19.8342  4.4536
##  problem   (Intercept)          8.8208  2.9700
##            interventioncast     0.7296  0.8542    0.39
##            interventionchange   3.5410  1.8818    0.33  0.46
```

```
##             interventionconvert    0.3355  0.5792   -0.62 -0.40  0.08
##             interventiontype_cast  1.4078  1.1865    0.42  0.56  0.99 -0.06
##             interventiontypecast   3.7162  1.9277    0.48  0.59  0.98 -0.05
##
##
##
##
##
##
##
##   0.99
## Number of obs: 228, groups:  prompt_id, 38; problem, 6
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -4.9993     1.5212  -3.287  0.00101 **
## interventioncast       -0.8626     0.3927  -2.196  0.02807 *
## interventionchange     -1.4541     0.8490  -1.713  0.08677 .
## interventionconvert    -0.0475     0.2707  -0.176  0.86068
## interventiontype_cast  -0.7289     0.5353  -1.362  0.17335
## interventiontypecast   -1.4163     0.8649  -1.637  0.10153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntncs intrvntnch intrvntncn intrv_
## intrvntncst  0.297
## intrvntnchn  0.252  0.443
## intrvntncnv -0.503 -0.324      0.109
## intrvntnty_  0.324  0.538      0.983     -0.020
## intrvntntyp  0.390  0.571      0.974     -0.016      0.985
## optimizer (bobyqa) convergence code: 1 (bobyqa -- maximum number of function evaluations exceeded)
```

## Concatenation interventions

```
concat.ps <- subset(data,category=="concatenate")$prompt_id
concats <- subset(data,(category=="concatenate"|category=="Baseline")&prompt_id %in% concat.ps)
```

```
concat.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=c
```

```
summary(concat.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: concats
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   2639.9   2720.0  -1298.9   2597.9      314
##
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -11.1303  -0.2916  -0.1244   0.3146   8.0186
##
## Random effects:
##  Groups    Name                    Variance Std.Dev. Corr
##  prompt_id (Intercept)             35.5048  5.9586
##  problem   (Intercept)             20.7866  4.5592
##            interventionadd          0.7849  0.8859   -0.83
##            interventioncombine      0.1296  0.3600   -0.04  0.15
##            interventionconcatenate  0.1757  0.4192   -0.91  0.54 -0.12
##            interventionsplice       0.3127  0.5592    0.33 -0.33 -0.01 -0.27
## Number of obs: 335, groups:  prompt_id, 67; problem, 13
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -4.3624     1.5456  -2.823  0.00477 **
## interventionadd            0.1068     0.2739   0.390  0.69670
## interventioncombine       -0.1406     0.1324  -1.062  0.28823
## interventionconcatenate    0.2441     0.1399   1.745  0.08101 .
## interventionsplice        -0.5639     0.1901  -2.966  0.00301 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntncm intrvntncn
## intervntndd -0.687
## intrvntncmb -0.062  0.184
## intrvntncnc -0.723  0.479    -0.010
## intrvntnspl  0.219 -0.232     0.073     -0.138
```

## Insert/Append interventions

```
append.ps <- subset(data,category=="insert")$prompt_id
appends <- subset(data,(category=="insert"|category=="Baseline")&prompt_id %in% append.ps)
```

```
append.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=ap
```

```
summary(append.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: appends
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   2389.6   2477.3  -1173.8   2347.6      459
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -9.853 -0.105 -0.064  0.076 41.583
##
```

```
## Random effects:
##  Groups    Name               Variance Std.Dev. Corr
##  prompt_id (Intercept)        96.2070  9.8085
##  problem   (Intercept)        44.3087  6.6565
##            interventionadd     0.1737  0.4168   -0.30
##            interventionappend  3.9368  1.9841   -0.81  0.58
##            interventionattach  3.6896  1.9208   -0.68  0.60  0.97
##            interventioninsert 21.7105  4.6595   -0.89  0.41  0.97  0.93
## Number of obs: 480, groups:  prompt_id, 96; problem, 26
##
## Fixed effects:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -5.45781    1.80917  -3.017  0.00256 **
## interventionadd    -0.09062    0.12412  -0.730  0.46533
## interventionappend -0.37637    0.45955  -0.819  0.41279
## interventionattach -0.52940    0.45227  -1.171  0.24179
## interventioninsert -1.39991    1.05080  -1.332  0.18278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntnp intrvntnt
## intervntndd -0.207
## intrvntnppn -0.605  0.528
## intrvntnttc -0.500  0.536     0.955
## intrvntnnsr -0.683  0.345     0.957     0.907
```

## Skip interventions

```
skip.ps <- subset(data,category=="skip")$prompt_id
skips <- subset(data,(category=="skip"|category=="Baseline")&prompt_id %in% skip.ps)
```

```
skip.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=skip
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
## Warning in optwrap(optimizer, devfun, start, rho$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
```

```
summary(skip.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: skips
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##    447.8    525.8   -195.9    391.8       92
##
```

```
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.90129 -0.02365 -0.01427 -0.00756  3.00566
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         95.19749 9.7569
##  problem   (Intercept)         24.59051 4.9589
##            interventionavoid    0.23755 0.4874    0.51
##            interventionignore   0.04565 0.2137    0.33  0.79
##            interventionneglect  0.26352 0.5133    0.97  0.67  0.54
##            interventionremove  15.12744 3.8894   -0.17  0.45 -0.06 -0.14
##            interventionskip     0.78943 0.8885    0.61  0.34 -0.26  0.50  0.51
## Number of obs: 120, groups:  prompt_id, 20; problem, 8
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -12.52369    3.81199  -3.285  0.00102 **
## interventionavoid    -0.76067    0.27697  -2.746  0.00603 **
## interventionignore    0.01387    0.16806   0.083  0.93425
## interventionneglect  -0.17591    0.28400  -0.619  0.53565
## interventionremove   -4.17516    2.09257  -1.995  0.04602 *
## interventionskip     -0.21022    0.45845  -0.459  0.64657
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnv intrvntng intrvntnn intrvntnr
## intervntnvd  0.244
## intrvntngnr  0.100  0.613
## intrvntnngl  0.484  0.616     0.470
## intrvntnrmv -0.116  0.400    -0.009    -0.100
## intrvntnskp  0.325  0.341    -0.062     0.474     0.470
## optimizer (bobyqa) convergence code: 1 (bobyqa -- maximum number of function evaluations exceeded)
```

## Return interventions

```
return.ps <- subset(data,category=="return")$prompt_id
returns <- subset(data,(category=="return"|category=="Baseline")&prompt_id %in% return.ps)
```

```
return.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=r
```

```
summary(return.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: returns
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##  40728.4  40900.5 -20336.2  40672.4     3416
```

```
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -17.116  -0.431  -0.117   0.323  85.154
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         31.2243  5.5879
##  problem   (Intercept)         14.6522  3.8278
##            interventiondisplay  2.8905  1.7002   -0.44
##            interventionoutput   0.7218  0.8496   -0.20  0.67
##            interventionprint    6.1269  2.4753   -0.17  0.88  0.64
##            interventionproduce  0.6598  0.8123    0.01  0.49  0.79  0.36
##            interventionreturn   0.1478  0.3844    0.25 -0.09  0.27 -0.02  0.30
## Number of obs: 3444, groups:  prompt_id, 574; problem, 48
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -4.97240    0.61672  -8.063 7.47e-16 ***
## interventiondisplay -1.28157    0.25303  -5.065 4.09e-07 ***
## interventionoutput  -0.14127    0.12808  -1.103   0.2700
## interventionprint   -2.87002    0.36904  -7.777 7.43e-15 ***
## interventionproduce -0.15787    0.12274  -1.286   0.1983
## interventionreturn   0.11118    0.06101   1.822   0.0684 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntnt intrvntnprn intrvntnprd
## intrvntndsp -0.391
## intrvntntpt -0.180  0.661
## intrvntnprn -0.153  0.875     0.625
## intrvntnprd  0.003  0.491     0.777     0.353
## intrvntnrtr  0.204 -0.064     0.284    -0.007       0.309
```

## Loop through interventions

```
loop.ps <- subset(data,category=="loop_through")$prompt_id
loops <- subset(data,(category=="loop_through"|category=="Baseline")&prompt_id %in% loop.ps)
```

```
loop.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=loop
```

```
## Warning in commonArgs(par, fn, control, environment()): maxfun < 10 *
## length(par)^2 is not recommended.
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
```

```
## Warning in commonArgs(par, fn, control, environment()): maxfun < 10 *
## length(par)^2 is not recommended.
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
```

```r
summary(loop.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: loops
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   2105.1   2295.2  -1007.6   2015.1      459
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.1405 -0.1344 -0.0290  0.0295  7.1066
##
## Random effects:
##  Groups    Name                                Variance Std.Dev. Corr
##  prompt_id (Intercept)                         32.8346  5.7302
##  problem   (Intercept)                         81.6284  9.0348
##            interventionexecute_a_for_loop_with  5.9259  2.4343    0.98
##            interventiongo_through               0.1843  0.4293    0.42  0.39
##            interventioniterate_through          0.1532  0.3914   -0.31 -0.27
##            interventionlook_through             0.8727  0.9342    0.60  0.55
##            interventionloop_through             0.7999  0.8943   -0.08 -0.01
##            interventionrun_a_for_loop_through   3.6901  1.9210    0.94  0.97
##            interventionrun_through              0.3632  0.6027    0.20  0.19
##
##
##
##
##
##  -0.56
##   0.38 -0.67
##  -0.50  0.70 -0.70
##   0.29 -0.07  0.36  0.20
##   0.51  0.16  0.10  0.12  0.23
## Number of obs: 504, groups:  prompt_id, 63; problem, 27
##
## Fixed effects:
##                                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -10.665842   2.149100  -4.963 6.94e-07 ***
## interventionexecute_a_for_loop_with  -3.526788   0.585623  -6.022 1.72e-09 ***
## interventiongo_through               -0.563307   0.147398  -3.822 0.000133 ***
## interventioniterate_through           0.004471   0.138621   0.032 0.974270
## interventionlook_through             -0.411786   0.265542  -1.551 0.120964
## interventionloop_through             -0.383890   0.272791  -1.407 0.159348
## interventionrun_a_for_loop_through   -2.855362   0.476153  -5.997 2.01e-09 ***
## interventionrun_through              -0.373536   0.203594  -1.835 0.066548 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##               (Intr) intrvntnx____ intrvntng_ intrvntnt_ intrvntnlk_
## intrvntnx____  0.838
## intrvntng_t    0.242  0.279
## intrvntntr_   -0.192 -0.124        -0.270
## intrvntnlk_    0.433  0.436         0.282     -0.428
## intrvntnlp_   -0.060  0.036        -0.350      0.596     -0.607
## intrvntnr____  0.775  0.953         0.189      0.080      0.213
## intrvntnrn_    0.109  0.150         0.482      0.255      0.110
##               intrvntnlp_ intrvntnr____
## intrvntnx____
## intrvntng_t
## intrvntntr_
## intrvntnlk_
## intrvntnlp_
## intrvntnr____  0.275
## intrvntnrn_    0.167       0.198
## optimizer (bobyqa) convergence code: 1 (bobyqa -- maximum number of function evaluations exceeded)
## maxfun < 10 * length(par)^2 is not recommended.
```

# Input interventions

## Take interventions

```
take.ps <- subset(data,category=="take")$prompt_id
takes <- subset(data,(category=="take"|category=="Baseline")&prompt_id %in% take.ps)
```

```
take.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=take
```

```
summary(take.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: takes
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   7070.7   7218.1  -3507.3   7014.7     1400
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.9187  -0.1623  -0.0868   0.1790  10.1529
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         51.91159 7.2050
##  problem   (Intercept)         24.44728 4.9444
##            interventionaccept   0.16360 0.4045    0.57
##            interventionbring_in 0.58238 0.7631   -0.04  0.03
##            interventionget      0.34461 0.5870    0.37  0.13  0.60
##            interventioninput    0.31551 0.5617    0.02  0.03  0.55  0.16
##            interventiontake     0.07434 0.2727    0.63  0.58  0.02  0.55 -0.25
```

```
## Number of obs: 1428, groups:  prompt_id, 238; problem, 41
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -5.715687   0.970661  -5.888  3.9e-09 ***
## interventionaccept   -0.061032   0.077673  -0.786    0.432
## interventionbring_in -0.051220   0.135654  -0.378    0.706
## interventionget       0.008606   0.106653   0.081    0.936
## interventioninput    -0.147679   0.102633  -1.439    0.150
## interventiontake      0.028506   0.056376   0.506    0.613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnc intrv_ intrvntng intrvntnn
## intrvntnccp  0.402
## intrvntnbr_ -0.032  0.090
## intervntngt  0.270  0.161     0.591
## intrvntnnpt  0.013  0.102     0.542  0.196
## intervntntk  0.409  0.543     0.082  0.519    -0.107
```

### Parameter interventions

```
param.ps <- subset(data,category=="parameter")$prompt_id
params <- subset(data,(category=="parameter"|category=="Baseline")&prompt_id %in% param.ps)
```

```
param.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=pa
```

```
summary(param.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: params
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##   5476.7   5580.7  -2717.3   5434.7     1024
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.3896 -0.1927 -0.1059  0.2122  8.0852
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 44.10896 6.6415
##  problem   (Intercept)                 15.33421 3.9159
##            interventionargument         0.30904 0.5559   -0.12
##            interventioninput            0.07847 0.2801   -0.16  0.19
##            interventionparameter        0.35621 0.5968   -0.05  0.71 -0.14
##            interventionvalue_provided   0.51665 0.7188   -0.36  0.61 -0.06  0.41
## Number of obs: 1045, groups:  prompt_id, 209; problem, 42
```

```
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -5.56213    0.82559  -6.737 1.62e-11 ***
## interventionargument       -0.04480    0.10413  -0.430    0.667
## interventioninput           0.08819    0.06037   1.461    0.144
## interventionparameter      -0.09542    0.11053  -0.863    0.388
## interventionvalue_provided -0.17126    0.13006  -1.317    0.188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnr intrvntnn intrvntnp
## intrvntnrgm -0.094
## intrvntnnpt -0.112  0.252
## intrvntnprm -0.051  0.677    -0.021
## intrvntnvl_ -0.253  0.598     0.030     0.414
```

### Provide interventions

```
provide.ps <- subset(data,category=="provide")$prompt_id
provides <- subset(data,(category=="provide"|category=="Baseline")&prompt_id %in% provide.ps)
```

```
provide.model <- glmer(pass ~ intervention + (1+intervention|problem) + (1|prompt_id), weights=n, data=p
```

```
summary(provide.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pass ~ intervention + (1 + intervention | problem) + (1 | prompt_id)
##    Data: provides
## Weights: n
## Control: glmerControl(optimizer = "bobyqa", calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##    365.9    405.0   -168.0    335.9       85
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.43252 -0.03487 -0.00953  0.05077  1.55524
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         208.4504 14.4378
##  problem   (Intercept)          22.3676  4.7294
##            interventionenter     0.6467  0.8042  -0.87
##            interventioninput     0.2082  0.4563  -0.06  0.55
##            interventionprovide   9.1976  3.0328  -0.55  0.84  0.76
## Number of obs: 100, groups:  prompt_id, 25; problem, 13
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -14.50003    4.05645  -3.575 0.000351 ***
```

```
## interventionenter     0.31400    0.36114   0.869 0.384585
## interventioninput     0.07856    0.23938   0.328 0.742778
## interventionprovide   1.77527    1.17748   1.508 0.131634
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnnt intrvntnnp
## intrvntnntr -0.289
## intrvntnnpt -0.043  0.475
## intrvntnprv -0.220  0.751      0.602
```