prompt analysis

2024-10-09

Data reformatting

```
data.raw = read.csv('llama3p1_70b_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 = 50)
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
##  11097.5  11257.1  -5520.8  11041.5     2180
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.1568  -0.4342  -0.1410   0.3311  13.3466
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 24.83432 4.9834
##  problem   (Intercept)                  9.06952 3.0116
```

```
##             interventioncharacter          1.03408 1.0169    0.03
##             interventionphrase             0.93198 0.9654   -0.22  0.13
##             interventionset_of_characters  2.12132 1.4565   -0.21  0.14  0.30
##             interventionstring             0.03042 0.1744    0.07  0.44  0.16
##             interventionword               0.56340 0.7506   -0.21  0.07  0.32
##
##
##
##
##
##
##
##   0.36
##   0.53  0.41
## Number of obs: 2208, groups:  prompt_id, 368; problem, 38
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -2.97446    0.58319  -5.100 3.39e-07 ***
## interventioncharacter         -0.62563    0.17970  -3.481 0.000499 ***
## interventionphrase            -0.39091    0.17121  -2.283 0.022416 *
## interventionset_of_characters -1.16534    0.25124  -4.638 3.51e-06 ***
## interventionstring             0.02787    0.05228   0.533 0.594015
## interventionword              -0.27028    0.13645  -1.981 0.047611 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnc intrvntnp intr__ intrvntns
## intrvntnchr  0.006
## intrvntnphr -0.192  0.163
## intrvntns__ -0.183  0.163     0.303
## intrvntnstr  0.002  0.343     0.193     0.271
## intrvntnwrd -0.184  0.118     0.333     0.510  0.355
```

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
##   1838.1   1880.2   -909.1   1818.1      488
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.7635 -0.0526 -0.0309  0.1299  3.8802
##
## Random effects:
##  Groups    Name                    Variance  Std.Dev. Corr
##  prompt_id (Intercept)             107.73971 10.3798
##  problem   (Intercept)               2.90528  1.7045
##            interventiondictionary    0.05561  0.2358  0.88
##            interventionmap           0.92478  0.9617  0.97 0.73
## Number of obs: 498, groups:  prompt_id, 166; problem, 15
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -10.14669    1.14154  -8.889   <2e-16 ***
## interventiondictionary   0.15709    0.09799   1.603    0.109
## interventionmap         -0.23915    0.28920  -0.827    0.408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd
## intrvntndct 0.251
## intervntnmp 0.395  0.584
```

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
##   2687.9   2756.8  -1329.0   2657.9      713
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -9.0783 -0.1850 -0.1256  0.2241  5.2865
##
## Random effects:
##  Groups    Name                    Variance Std.Dev. Corr
##  prompt_id (Intercept)             33.79220 5.8131
##  problem   (Intercept)             11.56616 3.4009
##            interventionint          0.15267 0.3907   -0.25
```

```
##            interventioninteger       0.02564 0.1601   -0.08 -0.02
##            interventionwhole_number  0.40671 0.6377   -0.20  0.17 -0.96
## Number of obs: 728, groups:  prompt_id, 182; problem, 27
##
## Fixed effects:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -4.49495    0.84768  -5.303 1.14e-07 ***
## interventionint           -0.01038    0.11129  -0.093    0.926
## interventioninteger        0.08438    0.07209   1.170    0.242
## interventionwhole_number -0.12415    0.15568  -0.797    0.425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intervntnnt intrvntnntg
## intervntnnt -0.179
## intrvntnntg -0.067  0.230
## intrvntnwh_ -0.155  0.246      -0.264
```

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
##  21300.1  21527.2 -10614.1  21228.1     4017
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -15.7088  -0.2900  -0.1158   0.1709  26.5921
##
## Random effects:
##  Groups    Name                        Variance  Std.Dev. Corr
##  prompt_id (Intercept)                 21.413412 4.62746
##  problem   (Intercept)                 14.335184 3.78618
##            interventionarray            0.155494 0.39433  -0.13
##            interventionarray_list       0.337290 0.58077  -0.15  0.72
##            interventionbrackets         1.257373 1.12133  -0.55  0.28  0.46
##            interventionlist             0.004573 0.06762  -0.84  0.34  0.47
```

```
##            interventionset              7.823522 2.79706  -0.22 -0.09 -0.03
##            interventionset_of_brackets  4.717545 2.17199  -0.40 -0.11  0.10
##
##
##
##
##
##
##   0.90
##   0.28  0.27
##   0.58  0.54  0.82
## Number of obs: 4053, groups:  prompt_id, 579; problem, 41
##
## Fixed effects:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -4.417401   0.648356  -6.813 9.54e-12 ***
## interventionarray            0.104692   0.074874   1.398   0.1620
## interventionarray_list       0.241663   0.103651   2.332   0.0197 *
## interventionbrackets        -0.472268   0.188751  -2.502   0.0123 *
## interventionlist             0.006358   0.033553   0.189   0.8497
## interventionset             -2.121760   0.465954  -4.554 5.27e-06 ***
## interventionset_of_brackets -1.627853   0.359610  -4.527 5.99e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnr intrv_ intrvntnb intrvntnl intrvntns
## intrvntnrry -0.118
## intrvntnrr_ -0.139  0.674
## intrvntnbrc -0.500  0.290     0.461
## intrvntnlst -0.283  0.304     0.294  0.373
## intervntnst -0.204 -0.054    -0.016  0.277     0.118
## intrvntns__ -0.370 -0.070     0.116  0.572     0.217     0.808
## 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 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
##   3197.7   3420.3  -1553.9   3107.7      995
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -9.4131 -0.0737 -0.0263  0.1135  4.7639
##
## Random effects:
##  Groups    Name                  Variance  Std.Dev. Corr
##  prompt_id (Intercept)           84.400255 9.18696
##  problem   (Intercept)           11.458482 3.38504
##            interventionattribute  1.991678 1.41127  -1.00
##            interventionelement    0.188359 0.43400  -0.97  0.98
##            interventionentry      0.229633 0.47920  -0.95  0.97  0.98
##            interventionitem       0.174488 0.41772  -0.81  0.85  0.89  0.95
##            interventionkey        0.004006 0.06329   0.57 -0.56 -0.53 -0.51
##            interventionpart       0.862614 0.92877  -0.90  0.94  0.97  0.97
##            interventionvariable   1.356182 1.16455  -0.60  0.65  0.67  0.78
##
##
##
##
##
##
##
##  -0.35
##   0.91 -0.61
##   0.89 -0.48  0.76
## Number of obs: 1040, groups:  prompt_id, 130; problem, 12
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -9.30816    1.42299  -6.541  6.1e-11 ***
## interventionattribute -0.52369    0.45118  -1.161    0.246
## interventionelement   -0.01410    0.15610  -0.090    0.928
## interventionentry      0.02501    0.16870   0.148    0.882
## interventionitem      -0.02901    0.15247  -0.190    0.849
## interventionkey       -0.01700    0.07784  -0.218    0.827
## interventionpart      -0.19543    0.30301  -0.645    0.519
## interventionvariable  -0.34230    0.37824  -0.905    0.365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntntt intrvntnl intrvntnn intrvntntm intrvntnk
## intrvntnttr -0.725
## intrvntnlmn -0.634  0.887
```

```
## intrvntnntr -0.635  0.892      0.876
## intervntntm -0.524  0.765      0.793     0.847
## intervntnky  0.081 -0.058      0.117     0.101     0.168
## intrvntnprt -0.646  0.907      0.883     0.891     0.819     -0.031
## intrvntnvrb -0.431  0.632      0.613     0.726     0.798     -0.024
##             intrvntnp
## intrvntnttr
## intrvntnlmn
## intrvntnntr
## intervntntm
## intervntnky
## intrvntnprt
## intrvntnvrb  0.737
## 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
##    886.8    982.8   -415.4    830.8      200
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6225 -0.1779 -0.0256  0.1662  5.2616
##
## Random effects:
##  Groups    Name                 Variance Std.Dev. Corr
##  prompt_id (Intercept)          40.8700  6.3930
##  problem   (Intercept)          79.0736  8.8923
##            interventioncast      0.8182  0.9045   0.69
##            interventionchange    0.1674  0.4092   0.95 0.88
##            interventionconvert   0.1121  0.3349   0.79 0.99 0.94
##            interventiontype_cast 1.0177  1.0088   0.96 0.87 1.00 0.93
##            interventiontypecast  1.7955  1.3400   0.92 0.92 1.00 0.97 0.99
## Number of obs: 228, groups:  prompt_id, 38; problem, 6
```

```
##
## Fixed effects:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -10.60862    4.40322  -2.409    0.016 *
## interventioncast       -0.57746    0.51057  -1.131    0.258
## interventionchange      0.15706    0.24089   0.652    0.514
## interventionconvert     0.02295    0.21947   0.105    0.917
## interventiontype_cast  -0.36878    0.51507  -0.716    0.474
## interventiontypecast   -0.43758    0.68529  -0.639    0.523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntncs intrvntnch intrvntncn intrv_
## intrvntncst 0.549
## intrvntnchn 0.736  0.768
## intrvntncnv 0.540  0.845      0.788
## intrvntnty_ 0.865  0.815      0.882      0.785
## intrvntntyp 0.825  0.881      0.874      0.814      0.964
## 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
##   1645.7   1725.8   -801.9   1603.7      314
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.2538 -0.5201 -0.1820  0.3585  4.9686
##
## Random effects:
##  Groups    Name                    Variance Std.Dev. Corr
##  prompt_id (Intercept)             15.1829  3.8965
##  problem   (Intercept)             11.2049  3.3474
##            interventionadd          0.2204  0.4694    0.16
##            interventioncombine      0.6129  0.7829    0.05  0.99
##            interventionconcatenate  1.1208  1.0587   -0.50  0.32  0.45
##            interventionsplice       0.7408  0.8607   -0.91  0.26  0.35  0.54
## Number of obs: 335, groups:  prompt_id, 67; problem, 13
```

```
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -1.82764    1.09760  -1.665   0.0959 .
## interventionadd          -0.05134    0.17174  -0.299   0.7650
## interventioncombine       0.19472    0.26088   0.746   0.4554
## interventionconcatenate   0.40230    0.32962   1.221   0.2223
## interventionsplice       -0.16168    0.27354  -0.591   0.5545
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntncm intrvntncn
## intervntndd  0.080
## intrvntncmb  0.010  0.887
## intrvntncnc -0.430  0.355     0.472
## intrvntnspl -0.765  0.332     0.406      0.543
```

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
##   1584.4   1672.1   -771.2   1542.4      459
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.2008 -0.1242 -0.0912  0.1432  4.8199
##
## Random effects:
##  Groups    Name               Variance Std.Dev. Corr
##  prompt_id (Intercept)        43.4993  6.5954
##  problem   (Intercept)        33.6405  5.8000
##            interventionadd     0.9876  0.9938   -0.71
##            interventionappend  0.8664  0.9308    0.10  0.62
##            interventionattach  2.3318  1.5270   -0.78  0.53 -0.09
##            interventioninsert  2.8963  1.7019   -0.70  0.91  0.51  0.77
## Number of obs: 480, groups:  prompt_id, 96; problem, 26
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)         -5.9502     1.4471  -4.112 3.92e-05 ***
## interventionadd     -0.1604     0.2455  -0.653    0.514
## interventionappend  -0.2440     0.2391  -1.021    0.307
## interventionattach   0.1159     0.3555   0.326    0.744
## interventioninsert  -0.2679     0.4001  -0.670    0.503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntnp intrvntnt
## intervntndd -0.565
## intrvntnppn  0.053  0.619
## intrvntnttc -0.610  0.488     0.006
## intrvntnnsr -0.549  0.849     0.529     0.737
```

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
```

```
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
##    383.4    461.4   -163.7    327.4       92
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2965 -0.0947 -0.0674 -0.0091  4.8776
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         58.7537  7.6651
##  problem   (Intercept)         32.4211  5.6940
##            interventionavoid    0.2929  0.5412   -1.00
##            interventionignore   0.1536  0.3919    0.29 -0.29
##            interventionneglect  1.3436  1.1591    0.44 -0.44  0.99
##            interventionremove  38.4374  6.1998   -0.66  0.66  0.52  0.38
##            interventionskip     0.7287  0.8536    0.18 -0.18  0.99  0.96  0.62
## Number of obs: 120, groups:  prompt_id, 20; problem, 8
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -4.87904    3.00493  -1.624    0.104
## interventionavoid    0.18259    0.33485   0.545    0.586
```

```
## interventionignore  -0.44063    0.29751  -1.481    0.139
## interventionneglect -0.25142    0.63060  -0.399    0.690
## interventionremove  -5.14219    3.47959  -1.478    0.139
## interventionskip     0.06715    0.49755   0.135    0.893
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnv intrvntng intrvntnn intrvntnr
## intervntnvd -0.550
## intrvntngnr  0.066  0.118
## intrvntnngl  0.206 -0.152     0.759
## intrvntnrmv -0.448  0.493     0.414     0.428
## intrvntnskp  0.041  0.060     0.767     0.883     0.594
```

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
```

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

```
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
##  18722.7  18894.8  -9333.4  18666.7     3416
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -16.597  -0.285  -0.126   0.227  43.417
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         26.1158  5.1104
##  problem   (Intercept)          8.3821  2.8952
##            interventiondisplay  3.8605  1.9648   -0.27
##            interventionoutput   1.1838  1.0880   -0.15  0.82
##            interventionprint    6.8435  2.6160   -0.19  0.88  0.59
##            interventionproduce  0.9708  0.9853   -0.12  0.51  0.82  0.22
##            interventionreturn   0.2697  0.5193    0.26  0.05  0.22 -0.03  0.37
## Number of obs: 3444, groups:  prompt_id, 574; problem, 48
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)         -4.146959   0.487527  -8.506  < 2e-16 ***
## interventiondisplay -0.792778   0.292232  -2.713  0.00667 **
## interventionoutput  -0.002916   0.164306  -0.018  0.98584
## interventionprint   -2.783894   0.390079  -7.137 9.56e-13 ***
## interventionproduce  0.301218   0.149865   2.010  0.04444 *
## interventionreturn   0.102713   0.084576   1.214  0.22457
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntnt intrvntnprn intrvntnprd
## intrvntndsp -0.236
## intrvntntpt -0.140  0.808
## intrvntnprn -0.168  0.870     0.576
## intrvntnprd -0.109  0.499     0.805     0.216
## intrvntnrtr  0.186  0.069     0.239    -0.005       0.373
```

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
```

```
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
##   1328.5   1518.5   -619.2   1238.5      459
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9470 -0.1146 -0.0147 -0.0021  3.1198
##
## Random effects:
```

```
##  Groups    Name                                Variance Std.Dev. Corr
##  prompt_id (Intercept)                         39.5403  6.2881
##  problem   (Intercept)                         75.9306  8.7138
##            interventionexecute_a_for_loop_with  6.9956  2.6449    0.88
##            interventiongo_through               2.1823  1.4773   -0.27 -0.11
##            interventioniterate_through          0.3227  0.5680    0.65  0.45
##            interventionlook_through             0.5157  0.7181    0.05  0.12
##            interventionloop_through             0.4438  0.6662    0.58  0.40
##            interventionrun_a_for_loop_through   1.4447  1.2019    0.91  0.88
##            interventionrun_through              1.5530  1.2462   -0.42 -0.18
##
##
##
##
##
##  -0.24
##   0.64  0.23
##  -0.13  0.94  0.08
##  -0.21  0.82  0.21  0.75
##   0.97 -0.41  0.52 -0.29 -0.35
## Number of obs: 504, groups:  prompt_id, 63; problem, 27
##
## Fixed effects:
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -11.8703     2.3194  -5.118 3.09e-07 ***
## interventionexecute_a_for_loop_with  -3.7673     0.7161  -5.260 1.44e-07 ***
## interventiongo_through                0.6814     0.4318   1.578   0.1145
## interventioniterate_through          -1.2159     0.2010  -6.049 1.46e-09 ***
## interventionlook_through             -0.2124     0.2454  -0.865   0.3868
## interventionloop_through             -1.1103     0.2290  -4.848 1.25e-06 ***
## interventionrun_a_for_loop_through   -2.1470     0.3386  -6.340 2.29e-10 ***
## interventionrun_through               0.6513     0.3667   1.776   0.0757 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) intrvntnx____ intrvntng_ intrvntnt_ intrvntnlk_
## intrvntnx____  0.707
## intrvntng_t   -0.230 -0.038
## intrvntntr_    0.382  0.320       -0.053
## intrvntnlk_    0.001  0.135        0.601      0.338
## intrvntnlp_    0.351  0.287        0.015      0.808      0.192
## intrvntnr____  0.694  0.795       -0.094      0.716      0.288
## intrvntnrn_   -0.343 -0.066        0.933     -0.163      0.514
##              intrvntnlp_ intrvntnr____
## intrvntnx____
## intrvntng_t
## intrvntntr_
## intrvntnlk_
## intrvntnlp_
## intrvntnr____  0.668
## intrvntnrn_   -0.085      -0.191
## 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
##   4608.1   4755.5  -2276.1   4552.1     1400
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.1976 -0.1533 -0.0944  0.1728  7.1692
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         44.84491 6.6966
##  problem   (Intercept)         20.23009 4.4978
##            interventionaccept   0.15002 0.3873   -0.45
##            interventionbring_in 0.37848 0.6152    0.04  0.22
##            interventionget      0.23702 0.4868   -0.04  0.54  0.62
##            interventioninput    0.24318 0.4931    0.30  0.21  0.74  0.45
##            interventiontake     0.04655 0.2157   -0.64  0.65  0.41  0.73  0.30
## Number of obs: 1428, groups:  prompt_id, 238; problem, 41
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -4.28835    0.89302  -4.802 1.57e-06 ***
## interventionaccept    0.01563    0.09048   0.173    0.863
## interventionbring_in -0.04799    0.12528  -0.383    0.702
## interventionget       0.07946    0.10607   0.749    0.454
## interventioninput     0.06935    0.10593   0.655    0.513
## interventiontake     -0.03495    0.06755  -0.517    0.605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnc intrv_ intrvntng intrvntnn
## intrvntnccp -0.299
## intrvntnbr_ -0.005  0.302
## intervntngt -0.050  0.527     0.579
## intrvntnnpt  0.168  0.331     0.667  0.462
## intervntntk -0.327  0.540     0.398  0.586     0.385
```

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
##   3426.4   3530.4  -1692.2   3384.4     1024
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4248 -0.1881 -0.1201  0.2113  4.8233
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 36.49260 6.0409
##  problem   (Intercept)                 10.92648 3.3055
##            interventionargument         0.24372 0.4937   -0.28
##            interventioninput            0.04926 0.2219   -0.25  0.58
##            interventionparameter        0.23113 0.4808   -0.28  0.73  0.76
##            interventionvalue_provided   0.37397 0.6115   -0.47  0.30 -0.43  0.10
## Number of obs: 1045, groups:  prompt_id, 209; problem, 42
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -3.74260    0.71874  -5.207 1.92e-07 ***
## interventionargument        0.15302    0.10578   1.447   0.1480
## interventioninput          -0.01698    0.07046  -0.241   0.8096
## interventionparameter       0.08853    0.10408   0.851   0.3950
## interventionvalue_provided  0.23316    0.12121   1.924   0.0544 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnr intrvntnn intrvntnp
## intrvntnrgm -0.192
## intrvntnnpt -0.146  0.498
## intrvntnprm -0.192  0.649     0.580
## intrvntnvl_ -0.315  0.358     0.002     0.228
```

Provide interventions

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

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

```r
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
##    254.6    293.7   -112.3    224.6       85
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.05263 -0.01145 -0.01007  0.09197  1.07298
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         216.1859 14.7033
##  problem   (Intercept)          63.8971  7.9936
##            interventionenter     0.6122  0.7824  -0.75
##            interventioninput     1.0334  1.0166  -0.89  0.89
##            interventionprovide   0.9749  0.9874  -0.79  0.94  0.98
## Number of obs: 100, groups:  prompt_id, 25; problem, 13
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -12.9873     4.9672  -2.615  0.00893 **
## interventionenter     0.2432     0.4020   0.605  0.54512
## interventioninput     0.1420     0.4803   0.296  0.76756
## interventionprovide  -0.6858     0.4745  -1.445  0.14840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnnt intrvntnnp
## intrvntnntr -0.300
## intrvntnnpt -0.377  0.805
## intrvntnprv -0.335  0.842      0.897
```