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 = 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
##  30470.0  30629.6 -15207.0  30414.0     2180
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -18.7949  -0.9217  -0.1449   0.6262  27.1761
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 30.3676  5.5107
##  problem   (Intercept)                 10.0325  3.1674
```

```
##             interventioncharacter          1.3218  1.1497    0.12
##             interventionphrase             1.1350  1.0654   -0.03  0.37
##             interventionset_of_characters  1.7871  1.3368   -0.07  0.33  0.27
##             interventionstring             0.1794  0.4235    0.00  0.44  0.36
##             interventionword               0.5271  0.7260   -0.10  0.13  0.30
##
##
##
##
##
##
##   0.25
##   0.28  0.71
## Number of obs: 2208, groups:  prompt_id, 368; problem, 38
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -3.40577    0.61958  -5.497 3.87e-08 ***
## interventioncharacter         -0.58861    0.19355  -3.041  0.00236 **
## interventionphrase            -0.29135    0.17964  -1.622  0.10483
## interventionset_of_characters -1.07626    0.22457  -4.793 1.65e-06 ***
## interventionstring             0.11101    0.07528   1.475  0.14030
## interventionword              -0.09762    0.12419  -0.786  0.43185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnc intrvntnp intr__ intrvntns
## intrvntnchr  0.089
## intrvntnphr -0.034  0.372
## intrvntns__ -0.063  0.329     0.275
## intrvntnstr -0.009  0.436     0.362     0.258
## intrvntnwrd -0.090  0.142     0.310     0.280  0.686
```

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
##   3641.1   3683.2  -1810.6   3621.1      488
##
```

```
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.8781  -0.0357  -0.0201   0.0598   9.4022
##
## Random effects:
##  Groups    Name                    Variance Std.Dev. Corr
##  prompt_id (Intercept)             173.8146 13.1839
##  problem   (Intercept)               1.3266  1.1518
##            interventiondictionary    0.2223  0.4714  0.56
##            interventionmap           0.8628  0.9289  1.00 0.58
## Number of obs: 498, groups:  prompt_id, 166; problem, 15
##
## Fixed effects:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -12.4533     1.3874  -8.976   <2e-16 ***
## interventiondictionary  -0.1036     0.1415  -0.732    0.464
## interventionmap         -0.3289     0.2712  -1.213    0.225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd
## intrvntndct 0.124
## intervntnmp 0.233  0.574
```

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
##   5764.7   5833.5  -2867.3   5734.7      713
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -16.7869  -0.1885  -0.1118   0.2411  11.1255
##
## Random effects:
##  Groups    Name                     Variance Std.Dev. Corr
##  prompt_id (Intercept)              49.2171  7.0155
##  problem   (Intercept)              18.9032  4.3478
##            interventionint           0.3736  0.6112   -0.51
```

```
##             interventioninteger       0.6514  0.8071   -0.35  0.72
##             interventionwhole_number  0.9258  0.9622   -0.58  0.66  0.41
## Number of obs: 728, groups:  prompt_id, 182; problem, 27
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -5.60445    1.05511  -5.312 1.09e-07 ***
## interventionint            -0.09633    0.13625  -0.707    0.480
## interventioninteger         0.13735    0.17732   0.775    0.439
## interventionwhole_number -0.10337    0.20613  -0.501    0.616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intervntnnt intrvntnntg
## intervntnnt -0.397
## intrvntnntg -0.280  0.705
## intrvntnwh_ -0.462  0.646       0.416
```

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
##  62676.1  62903.2 -31302.1  62604.1     4017
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -30.484  -0.568  -0.116   0.241  42.779
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 27.9067  5.2827
##  problem   (Intercept)                 18.8623  4.3431
##            interventionarray            0.2364  0.4862    0.01
##            interventionarray_list       0.3312  0.5755   -0.24  0.66
##            interventionbrackets         0.9882  0.9941   -0.54  0.16  0.43
##            interventionlist             0.1032  0.3212    0.11  0.21  0.11
```

```
##             interventionset             11.2317  3.3514   -0.22  0.11  0.00
##             interventionset_of_brackets  6.1256  2.4750   -0.39 -0.09  0.07
##
##
##
##
##
##
##  -0.20
##   0.25  0.16
##   0.53  0.01  0.83
## Number of obs: 4053, groups:  prompt_id, 579; problem, 41
##
## Fixed effects:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -5.09169    0.74106  -6.871 6.38e-12 ***
## interventionarray           -0.05712    0.08302  -0.688 0.491406
## interventionarray_list       0.10636    0.09754   1.090 0.275531
## interventionbrackets        -0.54571    0.16482  -3.311 0.000930 ***
## interventionlist            -0.07447    0.05663  -1.315 0.188476
## interventionset             -2.27671    0.55368  -4.112 3.92e-05 ***
## interventionset_of_brackets -1.56617    0.40600  -3.858 0.000115 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnr intrv_ intrvntnb intrvntnl intrvntns
## intrvntnrry  0.001
## intrvntnrr_ -0.217  0.656
## intrvntnbrc -0.494  0.177     0.434
## intrvntnlst  0.084  0.227     0.138 -0.169
## intervntnst -0.202  0.108     0.004  0.240     0.157
## intrvntns__ -0.363 -0.077     0.069  0.521     0.022     0.819
## 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
##   8100.0   8322.6  -4005.0   8010.0      995
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -24.9995  -0.2914  -0.0730   0.1751  21.6062
##
## Random effects:
##  Groups    Name                  Variance Std.Dev. Corr
##  prompt_id (Intercept)           57.1197  7.5578
##  problem   (Intercept)           14.5209  3.8106
##            interventionattribute  1.8568  1.3626   -0.91
##            interventionelement    0.2459  0.4958   -0.79  0.92
##            interventionentry      0.3540  0.5949   -0.91  0.91  0.95
##            interventionitem       0.4848  0.6963   -0.61  0.84  0.93  0.83
##            interventionkey        0.1086  0.3296   -0.88  0.79  0.73  0.86
##            interventionpart       1.3186  1.1483   -0.68  0.91  0.96  0.84
##            interventionvariable   2.5351  1.5922   -0.55  0.79  0.84  0.74
##
##
##
##
##
##
##
##   0.67
##   0.97  0.62
##   0.96  0.68  0.92
## Number of obs: 1040, groups:  prompt_id, 130; problem, 12
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -6.4516     1.3677  -4.717 2.39e-06 ***
## interventionattribute  -0.8732     0.4261  -2.049   0.0404 *
## interventionelement    -0.2938     0.1594  -1.843   0.0653 .
## interventionentry      -0.2811     0.1891  -1.487   0.1370
## interventionitem       -0.2801     0.2220  -1.262   0.2070
## interventionkey        -0.1585     0.1090  -1.454   0.1460
## interventionpart       -0.4848     0.3622  -1.338   0.1808
## interventionvariable   -0.8480     0.5030  -1.686   0.0918 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##             (Intr) intrvntntt intrvntnl intrvntnn intrvntntm intrvntnk
## intrvntnttr -0.768
## intrvntnlmn -0.657  0.904
## intrvntnntr -0.759  0.898      0.933
## intervntntm -0.509  0.826      0.916     0.818
## intervntnky -0.705  0.750      0.697     0.825     0.646
## intrvntnprt -0.570  0.905      0.936     0.826     0.955      0.589
## intrvntnvrb -0.459  0.782      0.823     0.727     0.953      0.648
##             intrvntnp
## intrvntnttr
## intrvntnlmn
## intrvntnntr
## intervntntm
## intervntnky
## intrvntnprt
## intrvntnvrb  0.910
## 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$lower, control = control, :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
## Warning in optwrap(optimizer, devfun, start, rho$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
##   2170.6   2266.6  -1057.3   2114.6      200
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.6356 -0.1895 -0.0228  0.1803  9.7323
##
## Random effects:
##  Groups    Name                 Variance Std.Dev. Corr
##  prompt_id (Intercept)          39.35450 6.2733
```

```
##  problem   (Intercept)           85.81731 9.2638
##            interventioncast       2.36877 1.5391    0.92
##            interventionchange     0.04286 0.2070    0.52  0.82
##            interventionconvert    0.24348 0.4934   -1.00 -0.89 -0.46
##            interventiontype_cast  2.41372 1.5536    0.98  0.98  0.69 -0.96
##            interventiontypecast   4.40855 2.0997    0.98  0.97  0.67 -0.97
##
##
##
##
##
##
##
##   1.00
## Number of obs: 228, groups:  prompt_id, 38; problem, 6
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -7.6228     4.2634  -1.788   0.0738 .
## interventioncast       -0.8918     0.7360  -1.212   0.2257
## interventionchange      0.2663     0.1328   2.006   0.0448 *
## interventionconvert     0.3050     0.2347   1.300   0.1937
## interventiontype_cast  -0.6081     0.7188  -0.846   0.3975
## interventiontypecast   -1.0465     0.9651  -1.084   0.2782
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntncs intrvntnch intrvntncn intrv_
## intrvntncst  0.827
## intrvntnchn  0.336  0.722
## intrvntncnv -0.916 -0.772     -0.200
## intrvntnty_  0.912  0.970      0.586     -0.876
## intrvntntyp  0.926  0.958      0.549     -0.897      0.996
## 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
```

```
##   4050.1   4130.2  -2004.1   4008.1      314
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.0906  -0.8632  -0.1692   0.6438  10.6116
##
## Random effects:
##  Groups    Name                    Variance Std.Dev. Corr
##  prompt_id (Intercept)             17.3721  4.1680
##  problem   (Intercept)             24.5140  4.9512
##            interventionadd          0.4641  0.6813    0.18
##            interventioncombine      2.0774  1.4413   -0.55  0.59
##            interventionconcatenate  1.3898  1.1789   -0.78  0.12  0.71
##            interventionsplice       3.1394  1.7718   -0.95 -0.13  0.69  0.79
## Number of obs: 335, groups:  prompt_id, 67; problem, 13
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -1.47301    1.51610  -0.972    0.331
## interventionadd          -0.02441    0.20868  -0.117    0.907
## interventioncombine       0.37120    0.42660   0.870    0.384
## interventionconcatenate   0.28476    0.34795   0.818    0.413
## interventionsplice       -0.30888    0.51046  -0.605    0.545
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntncm intrvntncn
## intervntndd  0.126
## intrvntncmb -0.503  0.596
## intrvntncnc -0.712  0.149     0.701
## intrvntnspl -0.879 -0.089     0.690      0.783
```

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
##   3510.3   3598.0  -1734.2   3468.3      459
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.6882  -0.0733  -0.0227   0.1034  13.1130
```

```
##
## Random effects:
##  Groups    Name               Variance Std.Dev. Corr
##  prompt_id (Intercept)        121.915  11.0415
##  problem   (Intercept)         36.772   6.0640
##            interventionadd      0.875   0.9354   0.22
##            interventionappend   2.053   1.4327   0.49  0.73
##            interventionattach   2.193   1.4809   0.12  0.22 -0.17
##            interventioninsert   3.807   1.9513  -0.31  0.58  0.10  0.74
## Number of obs: 480, groups:  prompt_id, 96; problem, 26
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -9.9347     1.8651  -5.327    1e-07 ***
## interventionadd     -0.4049     0.2208  -1.833   0.0667 .
## interventionappend  -0.5091     0.3306  -1.540   0.1236
## interventionattach  -0.1099     0.3405  -0.323   0.7470
## interventioninsert  -0.7593     0.4460  -1.702   0.0887 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntnp intrvntnt
## intervntndd  0.140
## intrvntnppn  0.310  0.707
## intrvntnttc  0.082  0.231    -0.137
## intrvntnnsr -0.207  0.577     0.121     0.736
```

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
```

```
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
##    622.8    700.9   -283.4    566.8       92
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.6403 -0.0133 -0.0076 -0.0005  7.9380
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         168.7332 12.9897
##  problem   (Intercept)          96.8126  9.8393
##            interventionavoid     2.6676  1.6333  -0.83
##            interventionignore    0.1438  0.3792   0.93 -0.98
##            interventionneglect   0.6807  0.8250   0.07  0.49 -0.31
##            interventionremove   50.1648  7.0827  -0.41  0.85 -0.72  0.88
##            interventionskip      0.8164  0.9036   0.78 -0.31  0.49  0.67  0.25
## Number of obs: 120, groups:  prompt_id, 20; problem, 8
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -13.7938     6.3672  -2.166  0.03028 *
## interventionavoid     0.0415     0.8140   0.051  0.95934
## interventionignore   -0.9847     0.2214  -4.447 8.69e-06 ***
## interventionneglect  -1.1743     0.4342  -2.705  0.00684 **
## interventionremove   -6.6072     3.6266  -1.822  0.06848 .
## interventionskip     -0.6616     0.4678  -1.414  0.15727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnv intrvntng intrvntnn intrvntnr
## intervntnvd -0.590
## intrvntngnr  0.556 -0.776
## intrvntnngl  0.040  0.505    -0.183
## intrvntnrmv -0.285  0.832    -0.587     0.860
## intrvntnskp  0.528 -0.245     0.449     0.676     0.281
## 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
##  53332.4  53504.5 -26638.2  53276.4     3416
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -42.196  -0.420  -0.122   0.429 101.772
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         34.391   5.8644
##  problem   (Intercept)         11.010   3.3181
##            interventiondisplay  3.717   1.9279   -0.24
##            interventionoutput   1.288   1.1350   -0.16  0.83
##            interventionprint    7.798   2.7925   -0.20  0.89  0.63
##            interventionproduce  1.119   1.0579   -0.19  0.53  0.81  0.30
##            interventionreturn   0.407   0.6379    0.27  0.13  0.30  0.06  0.31
## Number of obs: 3444, groups:  prompt_id, 574; problem, 48
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -4.867606   0.556998  -8.739  < 2e-16 ***
## interventiondisplay -0.868442   0.284759  -3.050  0.00229 **
## interventionoutput  -0.070902   0.168125  -0.422  0.67323
## interventionprint   -2.815966   0.413367  -6.812 9.61e-12 ***
## interventionproduce  0.209757   0.156942   1.337  0.18138
## interventionreturn   0.006053   0.095679   0.063  0.94956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnd intrvntnt intrvntnprn intrvntnprd
## intrvntndsp -0.211
## intrvntntpt -0.139  0.821
## intrvntnprn -0.172  0.891     0.620
## intrvntnprd -0.168  0.529     0.806     0.296
## intrvntnrtr  0.222  0.136     0.308     0.068       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
```

```
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
##   2946.2   3136.2  -1428.1   2856.2      459
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.5859 -0.0948 -0.0282  0.0511  7.4759
##
## Random effects:
##  Groups    Name                                Variance Std.Dev. Corr
##  prompt_id (Intercept)                         46.078   6.788
##  problem   (Intercept)                         46.165   6.794
##            interventionexecute_a_for_loop_with 27.512   5.245    0.59
##            interventiongo_through               5.190   2.278    0.56 0.37
##            interventioniterate_through          1.721   1.312    0.70 0.14 0.79
##            interventionlook_through             3.021   1.738    0.36 0.57 0.78
##            interventionloop_through             1.803   1.343    0.63 0.88 0.58
##            interventionrun_a_for_loop_through   3.977   1.994    0.36 0.87 0.24
##            interventionrun_through              3.636   1.907    0.40 0.23 0.93
##
##
##
##
##
##
##
##  0.53
##  0.46 0.80
##  0.08 0.64 0.91
##  0.77 0.64 0.48 0.14
## Number of obs: 504, groups:  prompt_id, 63; problem, 27
##
## Fixed effects:
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -10.2966     1.7636  -5.839 5.27e-09 ***
## interventionexecute_a_for_loop_with  -4.6946     1.3388  -3.507 0.000454 ***
## interventiongo_through               -0.9165     0.5577  -1.643 0.100305
## interventioniterate_through          -1.3370     0.3276  -4.082 4.47e-05 ***
## interventionlook_through             -0.6214     0.4470  -1.390 0.164495
## interventionloop_through             -1.4104     0.3383  -4.169 3.05e-05 ***
## interventionrun_a_for_loop_through   -1.6921     0.5179  -3.267 0.001086 **
## interventionrun_through              -0.3393     0.4783  -0.709 0.478118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##               (Intr) intrvntnx____ intrvntng_ intrvntnt_ intrvntnlk_
## intrvntnx____ 0.418
## intrvntng_t   0.375  0.303
## intrvntntr_   0.484  0.044         0.734
## intrvntnlk_   0.208  0.497         0.752      0.490
## intrvntnlp_   0.426  0.828         0.530      0.404      0.764
## intrvntnr____ 0.233  0.825         0.199      0.044      0.613
## intrvntnrn_   0.249  0.182         0.906      0.731      0.609
##               intrvntnlp_ intrvntnr____
## intrvntnx____
## intrvntng_t
## intrvntntr_
## intrvntnlk_
## intrvntnlp_
## intrvntnr____ 0.902
## intrvntnrn_   0.444       0.119
## 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
##  10094.2  10241.6  -5019.1  10038.2     1400
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -9.6762 -0.1564 -0.0839  0.1782 16.5332
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         64.7902  8.0492
##  problem   (Intercept)         34.3224  5.8585
##            interventionaccept   0.6965  0.8345   -0.34
##            interventionbring_in 1.5618  1.2497    0.30  0.44
##            interventionget      0.6507  0.8067   -0.27  0.72  0.41
##            interventioninput    0.6317  0.7948    0.25  0.66  0.75  0.57
##            interventiontake     0.4971  0.7050   -0.29  0.75  0.43  0.65  0.57
```

```
## Number of obs: 1428, groups:  prompt_id, 238; problem, 41
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -5.23466    1.12958  -4.634 3.58e-06 ***
## interventionaccept   -0.02397    0.14499  -0.165    0.869
## interventionbring_in -0.21849    0.21538  -1.014    0.310
## interventionget       0.13540    0.14066   0.963    0.336
## interventioninput     0.05400    0.13870   0.389    0.697
## interventiontake     -0.06599    0.12414  -0.532    0.595
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnc intrv_ intrvntng intrvntnn
## intrvntnccp -0.274
## intrvntnbr_  0.228  0.440
## intervntngt -0.218  0.709     0.412
## intrvntnnpt  0.192  0.661     0.723  0.570
## intervntntk -0.235  0.733     0.435  0.634     0.572
```

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
##   7198.8   7302.7  -3578.4   7156.8     1024
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.6959 -0.1758 -0.0903  0.1878 11.1723
##
## Random effects:
##  Groups    Name                        Variance Std.Dev. Corr
##  prompt_id (Intercept)                 59.7041  7.7268
##  problem   (Intercept)                 23.9804  4.8970
##            interventionargument         0.5910  0.7687   -0.40
##            interventioninput            0.2839  0.5328   -0.62  0.62
##            interventionparameter        0.7025  0.8382   -0.72  0.56  0.60
##            interventionvalue_provided   0.8961  0.9466   -0.67  0.60  0.56  0.49
## Number of obs: 1045, groups:  prompt_id, 209; problem, 42
```

```
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -5.40993    0.98420  -5.497 3.87e-08 ***
## interventionargument        0.28413    0.13151   2.161   0.0307 *
## interventioninput          -0.03122    0.09427  -0.331   0.7405
## interventionparameter       0.22899    0.14124   1.621   0.1049
## interventionvalue_provided  0.27245    0.15886   1.715   0.0863 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnr intrvntnn intrvntnp
## intrvntnrgm -0.301
## intrvntnnpt -0.460  0.605
## intrvntnprm -0.551  0.560     0.580
## intrvntnvl_ -0.508  0.590     0.539     0.485
```

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
##    364.4    403.5   -167.2    334.4       85
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.15652 -0.01083 -0.00758  0.03034  1.42685
##
## Random effects:
##  Groups    Name                Variance Std.Dev. Corr
##  prompt_id (Intercept)         150.9413 12.2858
##  problem   (Intercept)         260.7392 16.1474
##            interventionenter     1.0291  1.0144   0.52
##            interventioninput     1.4466  1.2027  -0.76 -0.06
##            interventionprovide   0.9226  0.9605   0.13  0.72  0.53
## Number of obs: 100, groups:  prompt_id, 25; problem, 13
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -14.6827     6.5317  -2.248  0.02458 *
```

```
## interventionenter    -0.2712     0.4333  -0.626  0.53135
## interventioninput     0.2929     0.4861   0.602  0.54689
## interventionprovide  -1.1133     0.4226  -2.634  0.00843 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) intrvntnnt intrvntnnp
## intrvntnntr  0.339
## intrvntnnpt -0.549  0.078
## intrvntnprv  0.073  0.713      0.606
```