

phishing_links_analysis

Yixin Zou

September 13, 2018

```
#Recode and format variables for the model
#Start with categorical variables with regards to participants
#For gender, we code 0 (N/A) and non-binary as other, and exclude other in the final analysis because t
data_cleaned$DQ_Gender[data_cleaned$DQ_Gender=="0" | data_cleaned$DQ_Gender=="Non binary/third gender"]
data_cleaned$gender.f <- factor(data_cleaned$DQ_Gender,levels=c("Female","Male","Other"), exclude = "Other")

#For education, we divide it into no bachelor's degree, bachelor's degree, and graduate degree
data_cleaned$DQ_Educ[data_cleaned$DQ_Educ=="0"] <- "Other"
data_cleaned$DQ_Educ[data_cleaned$DQ_Educ=="High school or equivalent (e.g., GED)" |
  data_cleaned$DQ_Educ=="Some college but no degree"] <- "No Bachelor's degree"
data_cleaned$DQ_Educ[data_cleaned$DQ_Educ=="Associate's degree" |
  data_cleaned$DQ_Educ=="Trade, technical, or vocational training" |
  data_cleaned$DQ_Educ=="Bachelor's degree"] <- "Bachelor's degree"
data_cleaned$DQ_Educ[data_cleaned$DQ_Educ=="Professional degree (JD, MD etc.)" |
  data_cleaned$DQ_Educ=="Doctoral's degree" |
  data_cleaned$DQ_Educ=="Master's degree"] <- "Graduate degree"
data_cleaned$education.f <- factor(data_cleaned$DQ_Educ,
  levels=c("No Bachelor's degree","Bachelor's degree","Graduate degree"), exclude = "Other")

#For occupation, we divide it into technical vs. non-technical
data_cleaned$DQ_Occ[data_cleaned$DQ_Occ == "Computers (Hardware, Desktop Software)" |
  data_cleaned$DQ_Occ == "Engineering / Architecture" |
  data_cleaned$DQ_Occ == "Internet"] <- "Technical occupations"
data_cleaned$DQ_Occ[data_cleaned$DQ_Occ != "Technical occupations"] <- "Non-technical occupations"
data_cleaned$occupation.f <- factor(data_cleaned$DQ_Occ,
  levels=c("Technical occupations","Non-technical occupations"))

#For each email and link, we categorize if it's a phish or not, and factor the related variables
data_cleaned$p_email.f <- factor(data_cleaned$p_email)
data_cleaned$p_link.f <- factor(data_cleaned$p_link)

#Factor condition groups and placement, give them labels
data_cleaned$placement.f <- factor(data_cleaned$placement,
  level=c("inmail","banner","browser","no warning"))
data_cleaned$condition_group.f <- factor(data_cleaned$condition_group,
  levels=c(1,2,3,4,5,6,0),
  labels=c("on-load, no forced","on-load, forced",
    "on-click, no forced","on-click, forced",
    "banner","browser","no warning"))

#Factor dependent variables: click and hover actions
data_cleaned$click_action.f <- factor(data_cleaned$click_action, labels=c("no","yes"))
data_cleaned$hover_action.f <- factor(data_cleaned$hover_action, labels=c("no","yes"))
data_cleaned$click_email_action.f <- factor(data_cleaned$email_p_link_click,levels=c("False","True",""))
data_cleaned$hover_email_action.f <- factor(data_cleaned$email_p_link_hover,levels=c("False","True",""))
data_cleaned$click_warning_action.f <- factor(data_cleaned$warn_p_link_click,levels=c("False","True",""))
data_cleaned$hover_warning_action.f <-
```

```
factor(data_cleaned$warn_p_link_hover, levels=c("False", "True", ""), labels=c("no", "yes", "N/A"))
```

```
#Standardize continuous variables
```

```
library(standardize)
```

```
data_cleaned$age_scaled <- scale(as.numeric(data_cleaned$DQ_Age))[, 1]
```

```
data_cleaned$cyber_quiz_score_scaled <- scale(data_cleaned$cyber_quiz_score)[, 1]
```

```
data_cleaned$PE_score_scaled <- scale(data_cleaned$PE_score)[, 1]
```

```
data_cleaned$brand_usage_scaled <- scale(data_cleaned$brand_usage)[, 1]
```

```
data_cleaned$hover_time_scaled <- scale(data_cleaned$hover_time)[, 1]
```

```
#rename time series variable
```

```
colnames(data_cleaned)[which(names(data_cleaned) == "time_series")] <- "warnings_seen"
```

```
data_cleaned$warnings_seen_scaled <- scale(data_cleaned$warnings_seen)[, 1]
```

```
#Factor random effect variables
```

```
data_cleaned$ref_id.f <- factor(data_cleaned$ref_id)
```

```
data_cleaned$username.f <- factor(data_cleaned$username)
```

```
data_cleaned$adj_link_id.f <- factor(data_cleaned$adj_link_id)
```

```
#Filter the dataset to include phishing link entries only
```

```
library(tidyverse)
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.1.0      v purrr  0.2.5
```

```
## v tibble  1.4.2      v dplyr  0.7.8
```

```
## v tidyr   0.8.2      v stringr 1.3.1
```

```
## v readr   1.3.1      v forcats 0.3.0
```

```
## -- Conflicts -----
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
data_phish_only <- data_cleaned %>%
```

```
  filter(data_cleaned$p_link == "True")
```

```
#We conduct a one-way ANOVA on phishing links CTR between groups
```

```
# library(dplyr)
```

```
data_phish_ctr <- data_phish_only %>%
```

```
  select(username, condition_group.f, true_phish_ctr) %>%
```

```
  distinct(username, condition_group.f, true_phish_ctr)
```

```
#descriptive stats for each group, look into normality
```

```
library(pastecs)
```

```
##
```

```
## Attaching package: 'pastecs'
```

```
## The following objects are masked from 'package:dplyr':
```

```
##
```

```
##     first, last
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##     extract
```

```
stat.desc(data_phish_ctr$true_phish_ctr, norm = TRUE)
```

```
##                                nbr.val
```

```
## data_phish_ctr$condition_group.f: on-load, no forced
##          nbr.val          nbr.null          nbr.na
## 103.0000000000000000  55.0000000000000000  0.0000000000000000
##          min          max          range
##  0.0000000000000000  1.0000000000000000  1.0000000000000000
##          sum          median          mean
## 34.6666666666666643  0.0000000000000000  0.3365695792880259
##          SE.mean          CI.mean.0.95          var
##  0.0412629135596473  0.0818447934000504  0.1753706876493855
##          std.dev          coef.var          skewness
##  0.4187728353766342  1.2442385204940380  0.7163744548479584
##          skew.2SE          kurtosis          kurt.2SE
##  1.5053797199909145 -1.2258442841269990 -1.2995901259609870
```

```

##          normtest.W          normtest.p
## 0.7165088296742375 0.0000000000008282
## -----
## data_phish_ctr$condition_group.f: on-load, forced
##          nbr.val          nbr.null          nbr.na
## 98.000000000000000000 72.000000000000000000 0.000000000000000000
##          min          max          range
## 0.000000000000000000 1.000000000000000000 1.000000000000000000
##          sum          median          mean
## 15.33333333333332149 0.000000000000000000 0.156462585034013613
##          SE.mean          CI.mean.0.95          var
## 0.030273349902971345 0.060084219470741433 0.089814620006078025
##          std.dev          coef.var          skewness
## 0.299690874078738068 1.915415586503238954 1.817622150041585227
##          skew.2SE          kurtosis          kurt.2SE
## 3.728318289836493715 2.068238606214739761 2.141198696291315429
##          normtest.W          normtest.p
## 0.575722762467548721 0.00000000000002201
## -----
## data_phish_ctr$condition_group.f: on-click, no forced
##          nbr.val          nbr.null          nbr.na
## 101.0000000000000000 58.000000000000000000 0.000000000000000000
##          min          max          range
## 0.000000000000000000 1.000000000000000000 1.000000000000000000
##          sum          median          mean
## 30.6666666666666643 0.000000000000000000 0.3036303630363036
##          SE.mean          CI.mean.0.95          var
## 0.0398288444661270 0.0790192930365003 0.1602200220022002
##          std.dev          coef.var          skewness
## 0.4002749330175451 1.3182967902643061 0.8212075583362094
##          skew.2SE          kurtosis          kurt.2SE
## 1.709310588823166 -1.0128813267067129 -1.0638061547663276
##          normtest.W          normtest.p
## 0.7111717937706055 0.0000000000008435
## -----
## data_phish_ctr$condition_group.f: on-click, forced
##          nbr.val          nbr.null          nbr.na
## 100.0000000000000000 86.000000000000000000 0.000000000000000000
##          min          max          range
## 0.000000000000000000 1.000000000000000000 1.000000000000000000
##          sum          median          mean
## 8.6666666666666607455 0.00000000000000000000 0.0866666666666666963
##          SE.mean          CI.mean.0.95          var
## 0.0244444444444444225 0.04850308103877908061 0.05975308641975308338
##          std.dev          coef.var          skewness
## 0.2444444444444443643 2.82051282051282026231 2.90881893313298256487
##          skew.2SE          kurtosis          kurt.2SE
## 6.02539894745985993296 7.46472697903148763032 7.80288639409179918260
##          normtest.W          normtest.p
## 0.39807414473956137524 0.000000000000000226
## -----
## data_phish_ctr$condition_group.f: banner
##          nbr.val          nbr.null          nbr.na
## 100.0000000000000000 43.0000000000000000 0.0000000000000000

```

```

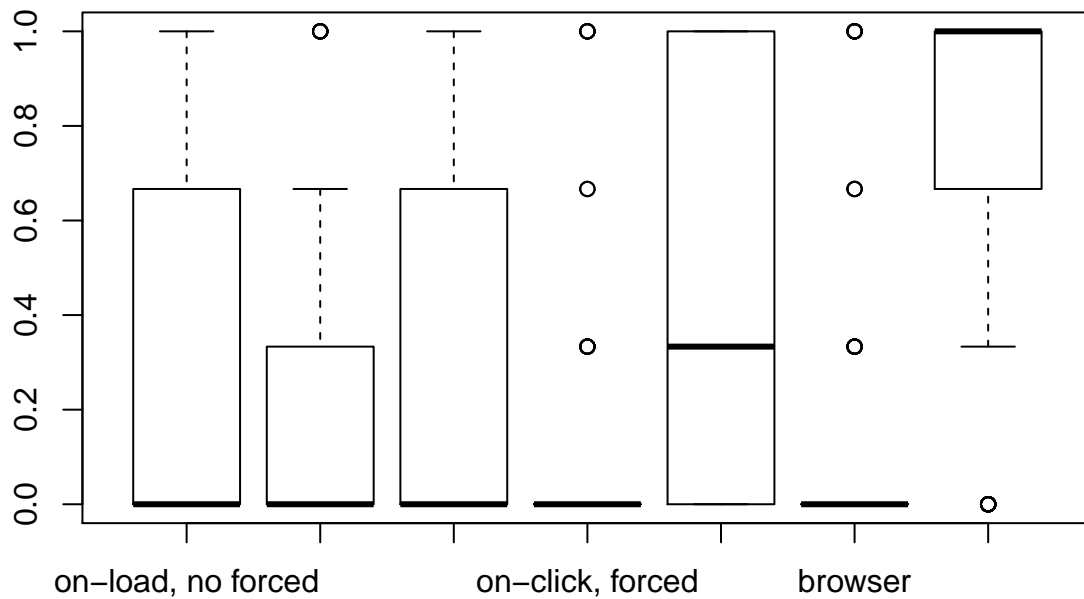
##           min           max           range
## 0.0000000000000000 1.0000000000000000 1.0000000000000000
##           sum           median           mean
## 44.000000000000000 0.3333333333333333 0.4400000000000000
##           SE.mean      CI.mean.0.95      var
## 0.044414131076513 0.088127271771999 0.197261503928171
##           std.dev      coef.var          skewness
## 0.444141310765133 1.009412069920757 0.270526160279621
##           skew.2SE      kurtosis          kurt.2SE
## 0.560374529621733 -1.721230981006292 -1.799204423753195
##           normtest.W      normtest.p
## 0.749477033467484 0.0000000000009069
## -----
## data_phish_ctr$condition_group.f: browser
##           nbr.val           nbr.null           nbr.na
## 99.000000000000000000 81.000000000000000000 0.000000000000000000
##           min           max           range
## 0.000000000000000000 1.000000000000000000 1.000000000000000000
##           sum           median           mean
## 11.666666666666667455 0.00000000000000000000 0.11784511784511783605
##           SE.mean      CI.mean.0.95      var
## 0.02844010678036456435 0.05643846630837947437 0.08007512769417531040
##           std.dev      coef.var          skewness
## 0.28297548956433543621 2.40124915430307517283 2.33981790810587897411
##           skew.2SE      kurtosis          kurt.2SE
## 4.82316098245310254100 4.15855249990630237988 4.32614566391064681028
##           normtest.W      normtest.p
## 0.46556028981711150561 0.00000000000000002695
## -----
## data_phish_ctr$condition_group.f: no warning
##           nbr.val           nbr.null           nbr.na
## 100.0000000000000000 15.000000000000000000 0.000000000000000000
##           min           max           range
## 0.000000000000000000 1.000000000000000000 1.000000000000000000
##           sum           median           mean
## 76.66666666666667140 1.00000000000000000000 0.766666666666666661
##           SE.mean      CI.mean.0.95      var
## 0.03715469526818085 0.07372297618215209 0.13804713804713806
##           std.dev      coef.var          skewness
## 0.37154695268180854 0.48462646001975029 -1.25211499875638599
##           skew.2SE      kurtosis          kurt.2SE
## -2.59366174692748386 -0.06630785246876991 -0.06931166287841116
##           normtest.W      normtest.p
## 0.64052835783764395 0.000000000000002712

```

```

boxplot(data_phish_ctr$true_phish_ctr~data_phish_ctr$condition_group.f)

```



```
#since the data is non-normal, we do Kruskal-Wallis Test
kruskal.test(true_phish_ctr ~ condition_group.f, data = data_phish_ctr)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  true_phish_ctr by condition_group.f
## Kruskal-Wallis chi-squared = 180, df = 6, p-value
## <0.00000000000000002
```

```
#Pairwise Mann-Whitney U-tests, use bonferroni as the p-value adjust method
```

```
PT_click_phish = pairwise.wilcox.test(data_phish_ctr$true_phish_ctr,
                                     data_phish_ctr$condition_group.f,
                                     p.adjust.method="bonferroni")
```

```
PT_click_phish = PT_click_phish$p.value
```

```
# Make it into a nicely formatted table
```

```
library(rcompanion)
```

```
PT_click_phish1 <- fullPTable(PT_click_phish)
```

```
round(PT_click_phish1, digits=4)
```

```
##           on-load, no forced on-load, forced on-click, no forced
## on-load, no forced          1.0000          0.0231          1.0000
## on-load, forced             0.0231          1.0000          0.1528
## on-click, no forced         1.0000          0.1528          1.0000
## on-click, forced            0.0000          0.6719          0.0001
## banner                     1.0000          0.0000          0.4501
## browser                     0.0003          1.0000          0.0030
```

```
## no warning                0.0000      0.0000      0.0000
##               on-click, forced banner browser no warning
## on-load, no forced        0.0000 1.0000  0.0003      0
## on-load, forced          0.6719 0.0000  1.0000      0
## on-click, no forced       0.0001 0.4501  0.0030      0
## on-click, forced         1.0000 0.0000  1.0000      0
## banner                   0.0000 1.0000  0.0000      0
## browser                  1.0000 0.0000  1.0000      0
## no warning               0.0000 0.0000  0.0000      1
```

```
#Mixed-effect logistic regression model on phishing links click action
#For all group comparison we include placement as the only warning-related predictor
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##      expand
```

```
glm_click_phish <- glmer(click_action.f ~
  placement.f #warning-related factors
  + warnings_seen_scaled #number of warnings seen before and during clicking on the link
  + benign_ctr #tendency to click on benign links
  + cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled # phishing-related individual characteristics
  + gender.f + education.f + age_scaled + occupation.f #participants demographics
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analysis
  data=data_phish_only,
  family=binomial(link=logit),
  control=glmerControl(optCtrl=list(maxfun=2e4)))
```

```
## singular fit
```

```
summary(glm_click_phish)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
```

```
## Approximation) [glmerMod]
```

```
## Family: binomial ( logit )
```

```
## Formula:
```

```
## click_action.f ~ placement.f + warnings_seen_scaled + benign_ctr +
##   cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled +
##   gender.f + education.f + age_scaled + occupation.f + (1 |
##   ref_id.f) + (1 | adj_link_id.f)
```

```
## Data: data_phish_only
```

```
## Control: glmerControl(optCtrl = list(maxfun = 20000))
```

```
##
```

```
##      AIC      BIC    logLik deviance df.resid
##    2111     2202     -1040     2079     2063
```

```
##
```

```
## Scaled residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -3.302 -0.590 -0.380  0.494  4.390
```

```
##
```

```
## Random effects:
```

```

## Groups      Name      Variance Std.Dev.
## ref_id.f    (Intercept) 0      0
## adj_link_id.f (Intercept) 0      0
## Number of obs: 2079, groups:  ref_id.f, 3; adj_link_id.f, 3
##
## Fixed effects:
##
##              Estimate Std. Error z value
## (Intercept)    -1.84635    0.21137   -8.74
## placement.fbanner      1.18194    0.14415    8.20
## placement.fbrowser    -0.78842    0.20159   -3.91
## placement.fno warning   2.35181    0.20954   11.22
## warnings_seen_scaled  -0.14253    0.07982   -1.79
## benign_ctr           1.46869    0.15800    9.30
## cyber_quiz_score_scaled -0.11313    0.05746   -1.97
## PE_score_scaled      -0.23719    0.05883   -4.03
## brand_usage_scaled     0.04766    0.02354    2.02
## gender.fMale         -0.04730    0.11368   -0.42
## education.fBachelor's degree -0.20574    0.11956   -1.72
## education.fGraduate degree -0.17203    0.19123   -0.90
## age_scaled          -0.04851    0.05586   -0.87
## occupation.fNon-technical occupations 0.00354    0.15624    0.02
##
##              Pr(>|z|)
## (Intercept)    < 0.0000000000000002 ***
## placement.fbanner      0.0000000000000024 ***
## placement.fbrowser    0.00009190704600325 ***
## placement.fno warning   < 0.0000000000000002 ***
## warnings_seen_scaled    0.074 .
## benign_ctr           < 0.0000000000000002 ***
## cyber_quiz_score_scaled    0.049 *
## PE_score_scaled      0.00005533075049282 ***
## brand_usage_scaled     0.043 *
## gender.fMale          0.677
## education.fBachelor's degree 0.085 .
## education.fGraduate degree 0.368
## age_scaled            0.385
## occupation.fNon-technical occupations 0.982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE) or
##      vcov(x)      if you need it

## convergence code: 0
## singular fit

#Print correlation tables
print(glm_click_phish, correlation=TRUE)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## click_action.f ~ placement.f + warnings_seen_scaled + benign_ctr +
##      cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled +

```



```
##      gender.f + education.f + age_scaled + occupation.f + (1 |
##      ref_id.f) + (1 | adj_link_id.f)
##      Data: data_phish_only
##      AIC      BIC    logLik deviance df.resid
##      2111     2202     -1040     2079     2063
## Random effects:
##      Groups      Name      Std.Dev.
##      ref_id.f      (Intercept) 0
##      adj_link_id.f (Intercept) 0
## Number of obs: 2079, groups:  ref_id.f, 3; adj_link_id.f, 3
## Fixed Effects:
##
##      (Intercept)
##      -1.84635
##      placement.fbanner
##      1.18194
##      placement.fbrowser
##      -0.78842
##      placement.fno warning
##      2.35181
##      warnings_seen_scaled
##      -0.14253
##      benign_ctr
##      1.46869
##      cyber_quiz_score_scaled
##      -0.11313
##      PE_score_scaled
##      -0.23719
##      brand_usage_scaled
##      0.04766
##      gender.fMale
##      -0.04730
##      education.fBachelor's degree
##      -0.20574
##      education.fGraduate degree
##      -0.17203
##      age_scaled
##      -0.04851
##      occupation.fNon-technical occupations
##      0.00354
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
```

```
#Get confidence intervals
```

```
se_click_phish <- sqrt(diag(vcov(glm_click_phish)))
```

```
# table of estimates with 95% CI
```

```
(tab_click_phish <- cbind(Est = fixef(glm_click_phish),
  LL = fixef(glm_click_phish) - 1.96 * se_click_phish,
  UL = fixef(glm_click_phish) + 1.96 * se_click_phish))
```

	Est	LL	UL
## (Intercept)	-1.84635	-2.260630	-1.4320673
## placement.fbanner	1.18194	0.899405	1.4644820
## placement.fbrowser	-0.78842	-1.183540	-0.3933065
## placement.fno warning	2.35181	1.941103	2.7625121
## warnings_seen_scaled	-0.14253	-0.298974	0.0139053
## benign_ctr	1.46869	1.159022	1.7783655

```
## cyber_quiz_score_scaled -0.11313 -0.225758 -0.0005008
## PE_score_scaled -0.23719 -0.352490 -0.1218847
## brand_usage_scaled 0.04766 0.001526 0.0937949
## gender.fMale -0.04730 -0.270120 0.1755158
## education.fBachelor's degree -0.20574 -0.440087 0.0285972
## education.fGraduate degree -0.17203 -0.546841 0.2027846
## age_scaled -0.04851 -0.157985 0.0609711
## occupation.fNon-technical occupations 0.00354 -0.302681 0.3097613
```

#Odds ratio

```
exp(tab_click_phish)
```

##	Est	LL	UL
## (Intercept)	0.1578	0.1043	0.2388
## placement.fbanner	3.2607	2.4581	4.3253
## placement.fbrowser	0.4546	0.3062	0.6748
## placement.fno warning	10.5045	6.9664	15.8396
## warnings_seen_scaled	0.8672	0.7416	1.0140
## benign_ctr	4.3436	3.1868	5.9202
## cyber_quiz_score_scaled	0.8930	0.7979	0.9995
## PE_score_scaled	0.7888	0.7029	0.8853
## brand_usage_scaled	1.0488	1.0015	1.0983
## gender.fMale	0.9538	0.7633	1.1919
## education.fBachelor's degree	0.8140	0.6440	1.0290
## education.fGraduate degree	0.8420	0.5788	1.2248
## age_scaled	0.9527	0.8539	1.0629
## occupation.fNon-technical occupations	1.0035	0.7388	1.3631

#Can do the same analysis with email-placed phishing links separated from warning-placed phishing links

#Logistic regression on click actions regarding email-placed phishing links

```
glm_click_phish_email <- glmer(click_email_action.f ~
  placement.f #warning-related factors
  + warnings_seen_scaled #number of warnings seen before and during clicking on the link
  + benign_ctr #tendency to click on benign links
  + cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled # phishing-related individual characteristics
  + gender.f + education.f + age_scaled + occupation.f #participants demographics
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analysis
  data=data_phish_only,
  family=binomial(link=logit),
  control=glmerControl(optCtrl=list(maxfun=2e4)))
```

singular fit

```
summary(glm_click_phish_email)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## click_email_action.f ~ placement.f + warnings_seen_scaled + benign_ctr +
## cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled +
## gender.f + education.f + age_scaled + occupation.f + (1 |
## ref_id.f) + (1 | adj_link_id.f)
## Data: data_phish_only
## Control: glmerControl(optCtrl = list(maxfun = 20000))
##
```

```

##      AIC      BIC   logLik deviance df.resid
##    1780.7   1871.0   -874.4   1748.7     2063
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.157 -0.545 -0.186  0.431 10.831
##
## Random effects:
##   Groups             Name             Variance Std.Dev.
##  ref_id.f             (Intercept)      0         0
##  adj_link_id.f        (Intercept)      0         0
## Number of obs: 2079, groups:  ref_id.f, 3; adj_link_id.f, 3
##
## Fixed effects:
##                                     Estimate Std. Error z value
## (Intercept)                       -3.45718    0.25797  -13.40
## placement.fbanner                   0.60364    0.17093   3.53
## placement.fbrowser                  3.98569    0.22431  17.77
## placement.fno warning                1.83676    0.22770   8.07
## warnings_seen_scaled               -0.11112    0.08769  -1.27
## benign_ctr                         3.95779    0.22229  17.80
## cyber_quiz_score_scaled             -0.28379    0.06530  -4.35
## PE_score_scaled                    -0.18607    0.06523  -2.85
## brand_usage_scaled                  0.02548    0.02641   0.96
## gender.fMale                       0.03071    0.12258   0.25
## education.fBachelor's degree       -0.06065    0.13042  -0.46
## education.fGraduate degree         -0.00887    0.20918  -0.04
## age_scaled                         -0.04945    0.05846  -0.85
## occupation.fNon-technical occupations -0.36217    0.17701  -2.05
##                                     Pr(>|z|)
## (Intercept)                       < 0.0000000000000002 ***
## placement.fbanner                   0.00041 ***
## placement.fbrowser                  < 0.0000000000000002 ***
## placement.fno warning                0.00000000000000072 ***
## warnings_seen_scaled                0.20507
## benign_ctr                         < 0.0000000000000002 ***
## cyber_quiz_score_scaled             0.00001387738418016 ***
## PE_score_scaled                    0.00434 **
## brand_usage_scaled                  0.33464
## gender.fMale                       0.80218
## education.fBachelor's degree        0.64193
## education.fGraduate degree          0.96618
## age_scaled                         0.39761
## occupation.fNon-technical occupations 0.04075 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE) or
##      vcov(x)          if you need it
##
## convergence code: 0
## singular fit

```

```

#Print correlation tables
print(glm_click_phish_email, correlation=TRUE)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## click_email_action.f ~ placement.f + warnings_seen_scaled + benign_ctr +
##   cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled +
##   gender.f + education.f + age_scaled + occupation.f + (1 |
##   ref_id.f) + (1 | adj_link_id.f)
## Data: data_phish_only
##      AIC      BIC    logLik deviance df.resid
## 1780.7   1871.0   -874.4   1748.7     2063
## Random effects:
## Groups          Name          Std.Dev.
## ref_id.f         (Intercept)  0
## adj_link_id.f    (Intercept)  0
## Number of obs: 2079, groups:  ref_id.f, 3; adj_link_id.f, 3
## Fixed Effects:
##
##              (Intercept)
##                -3.45718
##           placement.fbanner
##                0.60364
##           placement.fbrowser
##                3.98569
##           placement.fno warning
##                1.83676
##           warnings_seen_scaled
##               -0.11112
##                benign_ctr
##                3.95779
##           cyber_quiz_score_scaled
##               -0.28379
##                PE_score_scaled
##               -0.18607
##           brand_usage_scaled
##                0.02548
##                gender.fMale
##                0.03071
##           education.fBachelor's degree
##               -0.06065
##           education.fGraduate degree
##               -0.00887
##                age_scaled
##               -0.04945
##           occupation.fNon-technical occupations
##               -0.36217
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
#Get confidence intervals
se_click_phish_email <- sqrt(diag(vcov(glm_click_phish_email)))
# table of estimates with 95% CI
(tab_click_phish_email <- cbind(Est = fixef(glm_click_phish_email),

```

```
LL = fixef(glm_click_phish_email) - 1.96 * se_click_phish_email,
UL = fixef(glm_click_phish_email) + 1.96 * se_click_phish_email))
```

```
##               Est      LL      UL
## (Intercept)   -3.457183 -3.96280 -2.95157
## placement.fbanner 0.603639 0.26862 0.93866
## placement.fbrowser 3.985686 3.54604 4.42533
## placement.fno warning 1.836762 1.39047 2.28305
## warnings_seen_scaled -0.111120 -0.28299 0.06075
## benign_ctr      3.957791 3.52210 4.39349
## cyber_quiz_score_scaled -0.283794 -0.41179 -0.15580
## PE_score_scaled -0.186067 -0.31392 -0.05821
## brand_usage_scaled 0.025482 -0.02628 0.07725
## gender.fMale      0.030710 -0.20954 0.27096
## education.fBachelor's degree -0.060646 -0.31627 0.19498
## education.fGraduate degree -0.008869 -0.41886 0.40113
## age_scaled       -0.049452 -0.16404 0.06513
## occupation.fNon-technical occupations -0.362167 -0.70910 -0.01523
```

#Odds ratio

```
exp(tab_click_phish_email)
```

```
##               Est      LL      UL
## (Intercept)    0.03152 0.01901 0.05226
## placement.fbanner 1.82876 1.30816 2.55654
## placement.fbrowser 53.82217 34.67577 83.54037
## placement.fno warning 6.27618 4.01674 9.80657
## warnings_seen_scaled 0.89483 0.75353 1.06263
## benign_ctr      52.34160 33.85538 80.92195
## cyber_quiz_score_scaled 0.75292 0.66246 0.85573
## PE_score_scaled 0.83022 0.73057 0.94345
## brand_usage_scaled 1.02581 0.97406 1.08031
## gender.fMale     1.03119 0.81095 1.31123
## education.fBachelor's degree 0.94116 0.72886 1.21529
## education.fGraduate degree 0.99117 0.65779 1.49351
## age_scaled       0.95175 0.84871 1.06730
## occupation.fNon-technical occupations 0.69617 0.49209 0.98489
```

#Logistic regression on click actions regarding warning-placed phishing links

#Exclude participants in the no warning group, since they did not see any warning

```
data_warning_only <- data_phish_only %>%
```

```
  filter(data_phish_only$condition_group != 0)
```

```
glm_click_phish_warning <- glmer(click_warning_action.f ~
```

```
  placement.f #warning-related factors
```

```
  + warnings_seen_scaled #number of warnings seen before and during clicking on the link
```

```
  + benign_ctr #tendency to click on benign links
```

```
  + cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled # phishing-related individual characteristics
```

```
  + gender.f + education.f + age_scaled + occupation.f #participants demographics
```

```
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analysis
```

```
  data=data_warning_only,
```

```
  family=binomial(link=logit),
```

```
  control=glmerControl(optCtrl=list(maxfun=2e4)))
```

```
## singular fit
```

```
summary(glm_click_phish_warning)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: click_warning_action.f ~ placement.f + warnings_seen_scaled +
##         benign_ctr + cyber_quiz_score_scaled + PE_score_scaled +
##         brand_usage_scaled + gender.f + education.f + age_scaled +
##         occupation.f + (1 | ref_id.f) + (1 | adj_link_id.f)
## Data: data_warning_only
## Control: glmerControl(optCtrl = list(maxfun = 20000))
##
##      AIC      BIC    logLik deviance df.resid
##    840.3    922.6   -405.2    810.3     1767
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -1.013 -0.312 -0.204 -0.121  11.666
##
## Random effects:
##  Groups           Name      Variance Std.Dev.
##  ref_id.f      (Intercept) 0          0
##  adj_link_id.f (Intercept) 0          0
## Number of obs: 1782, groups:  ref_id.f, 3; adj_link_id.f, 3
##
## Fixed effects:
##                                     Estimate Std. Error z value
## (Intercept)                       -4.0860    0.3892  -10.50
## placement.fbanner                   0.4663    0.2511   1.86
## placement.fbrowser                 0.7324    0.2427   3.02
## warnings_seen_scaled               -0.0221    0.1284  -0.17
## benign_ctr                        2.8029    0.3426   8.18
## cyber_quiz_score_scaled            -0.2562    0.1061  -2.41
## PE_score_scaled                   -0.1013    0.1107  -0.92
## brand_usage_scaled                 0.1012    0.0368   2.75
## gender.fMale                      0.1506    0.2005   0.75
## education.fBachelor's degree       0.0971    0.2172   0.45
## education.fGraduate degree         0.2349    0.3380   0.70
## age_scaled                       -0.2019    0.1056  -1.91
## occupation.fNon-technical occupations -0.7587    0.2412  -3.15
##                                     Pr(>|z|)
## (Intercept)                       < 0.0000000000000002 ***
## placement.fbanner                  0.0633 .
## placement.fbrowser                 0.0025 **
## warnings_seen_scaled              0.8632
## benign_ctr                        0.00000000000000028 ***
## cyber_quiz_score_scaled            0.0158 *
## PE_score_scaled                   0.3600
## brand_usage_scaled                 0.0059 **
## gender.fMale                      0.4524
## education.fBachelor's degree       0.6549
## education.fGraduate degree         0.4870
## age_scaled                       0.0559 .
## occupation.fNon-technical occupations 0.0017 **
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

## convergence code: 0
## singular fit

#Print correlation tables
print(glm_click_phish_warning, correlation=TRUE)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: click_warning_action.f ~ placement.f + warnings_seen_scaled +
##   benign_ctr + cyber_quiz_score_scaled + PE_score_scaled +
##   brand_usage_scaled + gender.f + education.f + age_scaled +
##   occupation.f + (1 | ref_id.f) + (1 | adj_link_id.f)
## Data: data_warning_only
##      AIC      BIC   logLik deviance df.resid
##  840.3   922.6  -405.2   810.3     1767
## Random effects:
## Groups          Name          Std.Dev.
## ref_id.f         (Intercept)  0
## adj_link_id.f    (Intercept)  0
## Number of obs: 1782, groups:  ref_id.f, 3; adj_link_id.f, 3
## Fixed Effects:
##
##              (Intercept)
##                -4.0860
##           placement.fbanner
##                0.4663
##           placement.fbrowser
##                0.7324
##           warnings_seen_scaled
##               -0.0221
##                benign_ctr
##                2.8029
##           cyber_quiz_score_scaled
##               -0.2562
##                PE_score_scaled
##               -0.1013
##           brand_usage_scaled
##                0.1012
##                gender.fMale
##                0.1506
##           education.fBachelor's degree
##                0.0971
##           education.fGraduate degree
##                0.2349
##                age_scaled
##               -0.2019
##           occupation.fNon-technical occupations
##               -0.7587

```

```
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
```

```
#Get confidence intervals
```

```
se_click_phish_warning <- sqrt(diag(vcov(glm_click_phish_warning)))
```

```
# table of estimates with 95% CI
```

```
(tab_click_phish_warning <- cbind(Est = fixef(glm_click_phish_warning),
  LL = fixef(glm_click_phish_warning) - 1.96 * se_click_phish_warning,
  UL = fixef(glm_click_phish_warning) + 1.96 * se_click_phish_warning))
```

##	Est	LL	UL
## (Intercept)	-4.08597	-4.84888	-3.323056
## placement.fbanner	0.46633	-0.02578	0.958434
## placement.fbrowser	0.73245	0.25671	1.208183
## warnings_seen_scaled	-0.02211	-0.27369	0.229474
## benign_ctr	2.80289	2.13137	3.474407
## cyber_quiz_score_scaled	-0.25625	-0.46429	-0.048205
## PE_score_scaled	-0.10133	-0.31830	0.115647
## brand_usage_scaled	0.10121	0.02911	0.173307
## gender.fMale	0.15062	-0.24227	0.543511
## education.fBachelor's degree	0.09706	-0.32856	0.522686
## education.fGraduate degree	0.23493	-0.42755	0.897404
## age_scaled	-0.20187	-0.40883	0.005085
## occupation.fNon-technical occupations	-0.75873	-1.23156	-0.285904

```
#Odds ratio
```

```
exp(tab_click_phish_warning)
```

##	Est	LL	UL
## (Intercept)	0.01681	0.007837	0.03604
## placement.fbanner	1.59413	0.974546	2.60761
## placement.fbrowser	2.08016	1.292670	3.34740
## warnings_seen_scaled	0.97813	0.760565	1.25794
## benign_ctr	16.49224	8.426436	32.27867
## cyber_quiz_score_scaled	0.77395	0.628584	0.95294
## PE_score_scaled	0.90364	0.727385	1.12260
## brand_usage_scaled	1.10651	1.029538	1.18923
## gender.fMale	1.16256	0.784848	1.72204
## education.fBachelor's degree	1.10193	0.719961	1.68655
## education.fGraduate degree	1.26482	0.652105	2.45323
## age_scaled	0.81720	0.664429	1.00510
## occupation.fNon-technical occupations	0.46826	0.291837	0.75133

```
#Logistic regression on hover actions regarding phishing links
```

```
glm_hover_phish <- glmer(hover_action.f ~
```

```
  placement.f #warning-related factors
```

```
  + warnings_seen_scaled #number of warnings seen before and during hovering on the link
```

```
  + benign_hover_rate #tendency to hover over benign links
```

```
  + cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled # phishing-related individual characteristics
```

```
  + gender.f + education.f + age_scaled + occupation.f #participants demographics
```

```
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analysis
```

```
  data=data_phish_only,
```

```
  family=binomial(link=logit),
```

```
  control=glmerControl(optCtrl=list(maxfun=2e4)))
```

```
## singular fit
```



```
summary(glm_hover_phish)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## hover_action.f ~ placement.f + warnings_seen_scaled + benign_hover_rate +
##   cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled +
##   gender.f + education.f + age_scaled + occupation.f + (1 |
##   ref_id.f) + (1 | adj_link_id.f)
## Data: data_phish_only
## Control: glmerControl(optCtrl = list(maxfun = 20000))
##
##      AIC      BIC    logLik deviance df.resid
##    2052    2142    -1010    2020    2063
##
## Scaled residuals:
##   Min      1Q  Median      3Q      Max
## -5.024 -0.553  0.239  0.656  5.235
##
## Random effects:
##   Groups      Name      Variance      Std.Dev.
##   ref_id.f      (Intercept) 0.0000000000000104 0.0000000323
##   adj_link_id.f (Intercept) 0.00000000000003896 0.0000001974
## Number of obs: 2079, groups:  ref_id.f, 3; adj_link_id.f, 3
##
## Fixed effects:
##                                     Estimate Std. Error z value
## (Intercept)                       -2.0917    0.3157   -6.62
## placement.fbanner                    1.8939    0.2203    8.60
## placement.fbrowser                  -2.2393    0.1779  -12.59
## placement.fno warning                2.3913    0.2889    8.28
## warnings_seen_scaled                -0.0644    0.0762   -0.84
## benign_hover_rate                   3.4167    0.3253   10.50
## cyber_quiz_score_scaled             -0.1035    0.0631   -1.64
## PE_score_scaled                    -0.1762    0.0576   -3.06
## brand_usage_scaled                  0.0145    0.0242    0.60
## gender.fMale                       -0.1318    0.1133   -1.16
## education.fBachelor's degree       -0.1200    0.1214   -0.99
## education.fGraduate degree         -0.2501    0.1887   -1.33
## age_scaled                         0.0787    0.0572    1.38
## occupation.fNon-technical occupations -0.0277    0.1525   -0.18
##                                     Pr(>|z|)
## (Intercept)                       0.000000000000035 ***
## placement.fbanner                  < 0.0000000000000002 ***
## placement.fbrowser                  < 0.0000000000000002 ***
## placement.fno warning                < 0.0000000000000002 ***
## warnings_seen_scaled                0.3985
## benign_hover_rate                   < 0.0000000000000002 ***
## cyber_quiz_score_scaled              0.1006
## PE_score_scaled                     0.0022 **
## brand_usage_scaled                  0.5483
## gender.fMale                       0.2449
## education.fBachelor's degree        0.3230
```

```

## education.fGraduate degree          0.1850
## age_scaled                          0.1689
## occupation.fNon-technical occupations 0.8559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)           if you need it

## convergence code: 0
## singular fit

#Print correlation tables
print(glm_hover_phish, correlation=TRUE)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## hover_action.f ~ placement.f + warnings_seen_scaled + benign_hover_rate +
##   cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled +
##   gender.f + education.f + age_scaled + occupation.f + (1 |
##   ref_id.f) + (1 | adj_link_id.f)
## Data: data_phish_only
##      AIC      BIC   logLik deviance df.resid
##    2052    2142    -1010     2020     2063
## Random effects:
## Groups      Name              Std.Dev.
## ref_id.f    (Intercept) 0.0000000323
## adj_link_id.f (Intercept) 0.0000001974
## Number of obs: 2079, groups:  ref_id.f, 3; adj_link_id.f, 3
## Fixed Effects:
##
##              (Intercept)
##                -2.0917
##           placement.fbanner
##                1.8939
##           placement.fbrowser
##               -2.2393
##           placement.fno warning
##                2.3913
##           warnings_seen_scaled
##               -0.0644
##           benign_hover_rate
##                3.4167
##           cyber_quiz_score_scaled
##               -0.1035
##                PE_score_scaled
##               -0.1762
##           brand_usage_scaled
##                0.0145
##                gender.fMale
##               -0.1318
##           education.fBachelor's degree
##               -0.1200

```

```
##           education.fGraduate degree
##                               -0.2501
##           age_scaled
##                               0.0787
## occupation.fNon-technical occupations
##                               -0.0277
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings

#Get confidence intervals
se_hover_phish <- sqrt(diag(vcov(glm_hover_phish)))
# table of estimates with 95% CI
(tab_hover_phish <- cbind(Est = fixef(glm_hover_phish),
                          LL = fixef(glm_hover_phish) - 1.96 * se_hover_phish,
                          UL = fixef(glm_hover_phish) + 1.96 * se_hover_phish))

##                               Est      LL      UL
## (Intercept)                -2.09168 -2.71051 -1.47286
## placement.fbanner             1.89389  1.46214  2.32565
## placement.fbrowser           -2.23935 -2.58810 -1.89059
## placement.fno warning         2.39132  1.82497  2.95766
## warnings_seen_scaled         -0.06436 -0.21378  0.08506
## benign_hover_rate             3.41666  2.77911  4.05422
## cyber_quiz_score_scaled       -0.10354 -0.22712  0.02005
## PE_score_scaled              -0.17621 -0.28902 -0.06341
## brand_usage_scaled            0.01455 -0.03295  0.06204
## gender.fMale                 -0.13176 -0.35383  0.09031
## education.fBachelor's degree -0.12000 -0.35800  0.11799
## education.fGraduate degree   -0.25008 -0.61989  0.11973
## age_scaled                   0.07867 -0.03340  0.19075
## occupation.fNon-technical occupations -0.02769 -0.32662  0.27124

#Odds ratio
exp(tab_hover_phish)

##                               Est      LL      UL
## (Intercept)                  0.1235  0.06650  0.2293
## placement.fbanner             6.6452  4.31518 10.2333
## placement.fbrowser            0.1065  0.07516  0.1510
## placement.fno warning        10.9279  6.20264 19.2528
## warnings_seen_scaled          0.9377  0.80752  1.0888
## benign_hover_rate            30.4676 16.10461 57.6402
## cyber_quiz_score_scaled        0.9016  0.79682  1.0203
## PE_score_scaled               0.8384  0.74900  0.9386
## brand_usage_scaled            1.0147  0.96759  1.0640
## gender.fMale                  0.8766  0.70200  1.0945
## education.fBachelor's degree  0.8869  0.69907  1.1252
## education.fGraduate degree    0.7787  0.53800  1.1272
## age_scaled                    1.0819  0.96715  1.2102
## occupation.fNon-technical occupations 0.9727  0.72136  1.3116

#Linear regression on hover time regarding phishing links
library(lmerTest)

##
## Attaching package: 'lmerTest'
## The following object is masked from 'package:lme4':
```

```

##
## lmer
## The following object is masked from 'package:stats':
##
## step
glm_hovertime_phish <- lmerTest::lmer(hover_time_scaled ~
  placement.f #warning-related factors
  + warnings_seen_scaled #number of warnings seen before and during hovering on the link
  + benign_hover_rate #tendency to hover over benign links
  + cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled # phishing-related individual characteristics
  + gender.f + education.f + age_scaled + occupation.f #participants demographics
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analysis
  data=data_phish_only)

## singular fit
summary(glm_hovertime_phish)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## hover_time_scaled ~ placement.f + warnings_seen_scaled + benign_hover_rate +
##   cyber_quiz_score_scaled + PE_score_scaled + brand_usage_scaled +
##   gender.f + education.f + age_scaled + occupation.f + (1 |
##   ref_id.f) + (1 | adj_link_id.f)
## Data: data_phish_only
##
## REML criterion at convergence: 6186
##
## Scaled residuals:
##   Min      1Q  Median      3Q      Max
## -1.109 -0.433 -0.200  0.122  21.657
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##   ref_id.f      (Intercept) 0.00      0.00
##   adj_link_id.f (Intercept) 0.00      0.00
##   Residual              1.12      1.06
## Number of obs: 2079, groups:  ref_id.f, 3; adj_link_id.f, 3
##
## Fixed effects:
##
##              Estimate Std. Error      df
## (Intercept)    -0.48839    0.13068 2065.00000
## placement.fbanner      0.18246    0.06957 2065.00000
## placement.fbrowser   -0.64662    0.07179 2065.00000
## placement.fno warning -0.26324    0.08939 2065.00000
## warnings_seen_scaled -0.25060    0.03360 2065.00000
## benign_hover_rate      0.86359    0.13046 2065.00000
## cyber_quiz_score_scaled -0.00121    0.02682 2065.00000
## PE_score_scaled      -0.01890    0.02408 2065.00000
## brand_usage_scaled      0.01404    0.01018 2065.00000
## gender.fMale          0.07047    0.04805 2065.00000
## education.fBachelor's degree 0.03221    0.05186 2065.00000
## education.fGraduate degree -0.14715    0.08016 2065.00000

```

```
## age_scaled 0.11943 0.02374 2065.00000
## occupation.fNon-technical occupations -0.03937 0.06537 2065.00000
## t value Pr(>|t|)
## (Intercept) -3.74 0.00019 ***
## placement.fbanner 2.62 0.00879 **
## placement.fbrowser -9.01 < 0.0000000000000002 ***
## placement.fno warning -2.94 0.00327 **
## warnings_seen_scaled -7.46 0.000000000000013 ***
## benign_hover_rate 6.62 0.000000000004577 ***
## cyber_quiz_score_scaled -0.05 0.96393
## PE_score_scaled -0.79 0.43253
## brand_usage_scaled 1.38 0.16802
## gender.fMale 1.47 0.14266
## education.fBachelor's degree 0.62 0.53461
## education.fGraduate degree -1.84 0.06655 .
## age_scaled 5.03 0.00000053055603 ***
## occupation.fNon-technical occupations -0.60 0.54705
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it
## convergence code: 0
## singular fit
```

```
# anova(glm_hovertime_email_phish)
```

```
#Calculate effect size
```

```
r2.corr.mer <- function(m) {
  lmfit <- lm(model.response(model.frame(m)) ~ fitted(m))
  summary(lmfit)$r.squared
}
r2.corr.mer(glm_hovertime_phish)
```

```
## [1] 0.09548
```

```
#Now we further filter the dataset to only include entries in the link-focused groups, and examine the
# library(dplyr)
```

```
phish_inemail_only <- data_phish_only %>%
  filter(data_phish_only$condition_group == 1 | data_phish_only$condition_group == 2 | data_phish_only$
#Factor the activation and forced attention variables
phish_inemail_only$activation.f <- factor(phish_inemail_only$activation,
  level=c("on load", "on click"))
phish_inemail_only$forced_attention.f <- factor(phish_inemail_only$forced_attention,
  level=c("no", "yes"))
#Rescale continuous variables
library(standardize)
phish_inemail_only$age_rescaled <- scale(as.numeric(phish_inemail_only$DQ_Age))[, 1]
phish_inemail_only$cyber_quiz_score_rescaled <- scale(phish_inemail_only$cyber_quiz_score)[, 1]
phish_inemail_only$PE_score_rescaled <- scale(phish_inemail_only$PE_score)[, 1]
phish_inemail_only$brand_usage_rescaled <- scale(phish_inemail_only$brand_usage)[, 1]
phish_inemail_only$hover_time_rescaled <- scale(phish_inemail_only$hover_time)[, 1]
phish_inemail_only$warnings_seen_rescaled <- scale(phish_inemail_only$warnings_seen)[, 1]
```

#Now we include forced attention and activation as well as their interactions for in-email conditions i
#Logistics regression on phishing link click action for link-focused groups

```
library(lme4)
glm_click_phish_inemail <- glmer(click_action.f ~
  activation.f * forced_attention.f #warning-related factors
  + warnings_seen_rescaled #number of warnings seen before and during clicking on the link
  + benign_ctr #tendency to click on benign links
  + cyber_quiz_score_rescaled + PE_score_rescaled + brand_usage_rescaled # phishing-related individu
  + gender.f + education.f + age_rescaled + occupation.f #participants demographics
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analy
  data=phish_inemail_only,
  family=binomial(link=logit),
  control=glmerControl(optCtrl=list(maxfun=2e4)))
```

singular fit

```
summary(glm_click_phish_inemail)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## click_action.f ~ activation.f * forced_attention.f + warnings_seen_rescaled +
##   benign_ctr + cyber_quiz_score_rescaled + PE_score_rescaled +
##   brand_usage_rescaled + gender.f + education.f + age_rescaled +
##   occupation.f + (1 | ref_id.f) + (1 | adj_link_id.f)
## Data: phish_inemail_only
## Control: glmerControl(optCtrl = list(maxfun = 20000))
##
##      AIC      BIC    logLik deviance df.resid
##  1141.4   1222.7   -554.7   1109.4     1172
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.465 -0.567 -0.364 -0.181  5.585
##
## Random effects:
##   Groups             Name             Variance          Std.Dev.
##  ref_id.f      (Intercept)  0.000000000000001515  0.0000000389
##  adj_link_id.f (Intercept)  0.000000000000000787  0.0000000281
## Number of obs: 1188, groups:  ref_id.f, 3; adj_link_id.f, 3
##
## Fixed effects:
##                                     Estimate Std. Error z value
## (Intercept)                       -0.9694     0.2924   -3.32
## activation.fon click                -0.1730     0.1859   -0.93
## forced_attention.fyes               -1.1218     0.2101   -5.34
## warnings_seen_rescaled              -0.1582     0.0758   -2.09
## benign_ctr                          1.2723     0.2205    5.77
## cyber_quiz_score_rescaled           -0.1172     0.0783   -1.50
## PE_score_rescaled                  -0.1237     0.0809   -1.53
## brand_usage_rescaled                0.1995     0.0730    2.73
## gender.fMale                       -0.1367     0.1549   -0.88
## education.fBachelor's degree       -0.5253     0.1602   -3.28
```

```

## education.fGraduate degree          -0.2195      0.2640   -0.83
## age_rescaled                        -0.2051      0.0809  -2.54
## occupation.fNon-technical occupations  0.0224      0.2203   0.10
## activation.fon click:forced_attention.fyes -0.5451      0.3255  -1.67
##                                     Pr(>|z|)
## (Intercept)                        0.00091 ***
## activation.fon click                0.35192
## forced_attention.fyes              0.000000094 ***
## warnings_seen_rescaled             0.03690 *
## benign_ctr                        0.000000008 ***
## cyber_quiz_score_rescaled          0.13463
## PE_score_rescaled                 0.12644
## brand_usage_rescaled              0.00627 **
## gender.fMale                      0.37760
## education.fBachelor's degree       0.00104 **
## education.fGraduate degree         0.40576
## age_rescaled                      0.01122 *
## occupation.fNon-technical occupations  0.91886
## activation.fon click:forced_attention.fyes 0.09403 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE) or
##     vcov(x)         if you need it

## convergence code: 0
## singular fit

#Print correlation tables
print(glm_click_phish_inemail, correlation=TRUE)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## click_action.f ~ activation.f * forced_attention.f + warnings_seen_rescaled +
##     benign_ctr + cyber_quiz_score_rescaled + PE_score_rescaled +
##     brand_usage_rescaled + gender.f + education.f + age_rescaled +
##     occupation.f + (1 | ref_id.f) + (1 | adj_link_id.f)
## Data: phish_inemail_only
##      AIC      BIC    logLik deviance df.resid
## 1141.4  1222.7   -554.7   1109.4     1172
## Random effects:
## Groups      Name          Std.Dev.
## ref_id.f      (Intercept) 0.0000000389
## adj_link_id.f (Intercept) 0.0000000281
## Number of obs: 1188, groups:  ref_id.f, 3; adj_link_id.f, 3
## Fixed Effects:
##
##              (Intercept)
##              -0.9694
##              activation.fon click
##              -0.1730
##              forced_attention.fyes
##              -1.1218

```

```
## warnings_seen_rescaled
## -0.1582
## benign_ctr
## 1.2723
## cyber_quiz_score_rescaled
## -0.1172
## PE_score_rescaled
## -0.1237
## brand_usage_rescaled
## 0.1995
## gender.fMale
## -0.1367
## education.fBachelor's degree
## -0.5253
## education.fGraduate degree
## -0.2195
## age_rescaled
## -0.2051
## occupation.fNon-technical occupations
## 0.0224
## activation.fon click:forced_attention.fyes
## -0.5451
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
```

```
#Get confidence intervals
```

```
se_click_phish <- sqrt(diag(vcov(glm_click_phish)))
```

```
# table of estimates with 95% CI
```

```
(tab_click_phish <- cbind(Est = fixef(glm_click_phish),
  LL = fixef(glm_click_phish) - 1.96 * se_click_phish,
  UL = fixef(glm_click_phish) + 1.96 * se_click_phish))
```

	Est	LL	UL
## (Intercept)	-1.84635	-2.260630	-1.4320673
## placement.fbanner	1.18194	0.899405	1.4644820
## placement.fbrowser	-0.78842	-1.183540	-0.3933065
## placement.fno warning	2.35181	1.941103	2.7625121
## warnings_seen_scaled	-0.14253	-0.298974	0.0139053
## benign_ctr	1.46869	1.159022	1.7783655
## cyber_quiz_score_scaled	-0.11313	-0.225758	-0.0005008
## PE_score_scaled	-0.23719	-0.352490	-0.1218847
## brand_usage_scaled	0.04766	0.001526	0.0937949
## gender.fMale	-0.04730	-0.270120	0.1755158
## education.fBachelor's degree	-0.20574	-0.440087	0.0285972
## education.fGraduate degree	-0.17203	-0.546841	0.2027846
## age_scaled	-0.04851	-0.157985	0.0609711
## occupation.fNon-technical occupations	0.00354	-0.302681	0.3097613

```
#Odds ratio
```

```
exp(tab_click_phish)
```

	Est	LL	UL
## (Intercept)	0.1578	0.1043	0.2388
## placement.fbanner	3.2607	2.4581	4.3253
## placement.fbrowser	0.4546	0.3062	0.6748
## placement.fno warning	10.5045	6.9664	15.8396
## warnings_seen_scaled	0.8672	0.7416	1.0140


```
## benign_ctr                                4.3436 3.1868 5.9202
## cyber_quiz_score_scaled                   0.8930 0.7979 0.9995
## PE_score_scaled                           0.7888 0.7029 0.8853
## brand_usage_scaled                        1.0488 1.0015 1.0983
## gender.fMale                             0.9538 0.7633 1.1919
## education.fBachelor's degree              0.8140 0.6440 1.0290
## education.fGraduate degree                0.8420 0.5788 1.2248
## age_scaled                               0.9527 0.8539 1.0629
## occupation.fNon-technical occupations     1.0035 0.7388 1.3631
```

#Logistic regression on phishing link hover actions for link-focused groups

```
library(optimx)
```

```
glm_hover_phish_inemail <- glmer(hover_action.f ~
  activation.f * forced_attention.f #warning-related factors
  + warnings_seen_rescaled #number of warnings seen before and during hovering on the link
  + benign_hover_rate #tendency to hover over benign links
  + cyber_quiz_score_rescaled + PE_score_rescaled + brand_usage_rescaled # phishing-related individual
  + gender.f + education.f + age_rescaled + occupation.f #participants demographics
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analysis
  data=phish_inemail_only,
  family=binomial(link=logit),
  # REML = FALSE,
  control = glmerControl(optimizer = 'optimx', optCtrl=list(method='L-BFGS-B')))
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00116603
## (tol = 0.001, component 1)
```

```
summary(glm_hover_phish_inemail)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## hover_action.f ~ activation.f * forced_attention.f + warnings_seen_rescaled +
##   benign_hover_rate + cyber_quiz_score_rescaled + PE_score_rescaled +
##   brand_usage_rescaled + gender.f + education.f + age_rescaled +
##   occupation.f + (1 | ref_id.f) + (1 | adj_link_id.f)
## Data: phish_inemail_only
## Control:
## glmerControl(optimizer = "optimx", optCtrl = list(method = "L-BFGS-B"))
##
##      AIC      BIC   logLik deviance df.resid
##  955.7   1036.9   -461.8   923.7     1172
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -7.921 -0.433  0.164  0.362  4.592
##
## Random effects:
## Groups      Name                Variance Std.Dev.
## ref_id.f    (Intercept) 0.0126   0.112
## adj_link_id.f (Intercept) 0.0126   0.112
## Number of obs: 1188, groups:  ref_id.f, 3; adj_link_id.f, 3
##
```

```

## Fixed effects:
##
##               Estimate Std. Error z value
## (Intercept)      -1.9994    0.4810   -4.16
## activation.fon click      1.0059    0.3298    3.05
## forced_attention.fyes     -1.8717    0.2228   -8.40
## warnings_seen_rescaled   -0.2784    0.0849   -3.28
## benign_hover_rate        4.9137    0.4972    9.88
## cyber_quiz_score_rescaled -0.1126    0.0954   -1.18
## PE_score_rescaled        -0.1842    0.0853   -2.16
## brand_usage_rescaled      0.0775    0.0864    0.90
## gender.fMale             -0.0458    0.1700   -0.27
## education.fBachelor's degree -0.4411    0.1858   -2.37
## education.fGraduate degree -0.3041    0.2790   -1.09
## age_rescaled             0.1529    0.0920    1.66
## occupation.fNon-technical occupations 0.4308    0.2506    1.72
## activation.fon click:forced_attention.fyes -3.2438    0.3960   -8.19
##
##               Pr(>|z|)
## (Intercept)      0.00003224669594420 ***
## activation.fon click      0.0023 **
## forced_attention.fyes     < 0.00000000000000002 ***
## warnings_seen_rescaled   0.0010 **
## benign_hover_rate        < 0.00000000000000002 ***
## cyber_quiz_score_rescaled 0.2379
## PE_score_rescaled        0.0308 *
## brand_usage_rescaled      0.3696
## gender.fMale             0.7874
## education.fBachelor's degree 0.0176 *
## education.fGraduate degree 0.2757
## age_rescaled             0.0964 .
## occupation.fNon-technical occupations 0.0857 .
## activation.fon click:forced_attention.fyes 0.000000000000000026 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE) or
##     vcov(x)         if you need it

## convergence code: 0
## Model failed to converge with max|grad| = 0.00116603 (tol = 0.001, component 1)

#Print correlation tables
print(glm_hover_phish_inemail, correlation=TRUE)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
## Formula:
## hover_action.f ~ activation.f * forced_attention.f + warnings_seen_rescaled +
##   benign_hover_rate + cyber_quiz_score_rescaled + PE_score_rescaled +
##   brand_usage_rescaled + gender.f + education.f + age_rescaled +
##   occupation.f + (1 | ref_id.f) + (1 | adj_link_id.f)
## Data: phish_inemail_only
##      AIC      BIC    logLik deviance df.resid
##  955.7   1036.9   -461.8    923.7     1172

```

```
## Random effects:
## Groups      Name      Std.Dev.
## ref_id.f    (Intercept) 0.112
## adj_link_id.f (Intercept) 0.112
## Number of obs: 1188, groups: ref_id.f, 3; adj_link_id.f, 3
## Fixed Effects:
##              (Intercept)
##              -1.9994
##              activation.fon click
##              1.0059
##              forced_attention.fyes
##              -1.8717
##              warnings_seen_rescaled
##              -0.2784
##              benign_hover_rate
##              4.9137
##              cyber_quiz_score_rescaled
##              -0.1126
##              PE_score_rescaled
##              -0.1842
##              brand_usage_rescaled
##              0.0775
##              gender.fMale
##              -0.0458
##              education.fBachelor's degree
##              -0.4411
##              education.fGraduate degree
##              -0.3041
##              age_rescaled
##              0.1529
##              occupation.fNon-technical occupations
##              0.4308
## activation.fon click:forced_attention.fyes
##              -3.2438
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
```

```
#Get confidence intervals
se_hover_phish_inemail <- sqrt(diag(vcov(glm_hover_phish_inemail)))
# table of estimates with 95% CI
(tab_hover_phish_inemail <- cbind(Est = fixef(glm_hover_phish_inemail),
                                  LL = fixef(glm_hover_phish_inemail) - 1.96 * se_hover_phish_inemail,
                                  UL = fixef(glm_hover_phish_inemail) + 1.96 * se_hover_phish_inemail))
```

	Est	LL	UL
## (Intercept)	-1.99942	-2.94213	-1.05670
## activation.fon click	1.00590	0.35943	1.65237
## forced_attention.fyes	-1.87170	-2.30833	-1.43507
## warnings_seen_rescaled	-0.27844	-0.44479	-0.11208
## benign_hover_rate	4.91375	3.93914	5.88835
## cyber_quiz_score_rescaled	-0.11260	-0.29961	0.07441
## PE_score_rescaled	-0.18425	-0.35142	-0.01707
## brand_usage_rescaled	0.07751	-0.09180	0.24682
## gender.fMale	-0.04584	-0.37895	0.28727
## education.fBachelor's degree	-0.44112	-0.80533	-0.07691
## education.fGraduate degree	-0.30408	-0.85089	0.24273

```
## age_rescaled                0.15287 -0.02736  0.33310
## occupation.fNon-technical occupations  0.43075 -0.06046  0.92196
## activation.fon click:forced_attention.fyes -3.24377 -4.01991 -2.46762
```

#Odds ratio

```
exp(tab_hover_phish_inemail)
```

```
##                               Est      LL      UL
## (Intercept)                  0.13541  0.05275  0.34760
## activation.fon click          2.73437  1.43251  5.21936
## forced_attention.fyes         0.15386  0.09943  0.23810
## warnings_seen_rescaled       0.75697  0.64096  0.89397
## benign_hover_rate            136.14834 51.37449 360.80885
## cyber_quiz_score_rescaled     0.89350  0.74111  1.07724
## PE_score_rescaled             0.83173  0.70369  0.98308
## brand_usage_rescaled         1.08059  0.91229  1.27995
## gender.fMale                 0.95519  0.68458  1.33278
## education.fBachelor's degree  0.64332  0.44694  0.92598
## education.fGraduate degree    0.73780  0.42703  1.27472
## age_rescaled                 1.16517  0.97301  1.39529
## occupation.fNon-technical occupations 1.53842  0.94134  2.51422
## activation.fon click:forced_attention.fyes 0.03902  0.01795  0.08479
```

#Linear regression on phishing link hover time for link-focused groups

```
library(lmerTest)
```

```
glm_hovertime_phish_inemail <- lmerTest::lmer(hover_time_rescaled ~
  activation.f * forced_attention.f #warning-related factors
  + warnings_seen_rescaled #number of warnings seen before and during hovering on the link
  + benign_hover_rate #tendency to hover over benign links
  + cyber_quiz_score_rescaled + PE_score_rescaled + brand_usage_rescaled # phishing-related individual
  + gender.f + education.f + age_rescaled + occupation.f #participants demographics
  + (1 | ref_id.f) + (1 | adj_link_id.f), #random effect, dropped username.f for phishing link analysis
  data=phish_inemail_only)
```

```
summary(glm_hovertime_phish_inemail)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## hover_time_rescaled ~ activation.f * forced_attention.f + warnings_seen_rescaled +
##   benign_hover_rate + cyber_quiz_score_rescaled + PE_score_rescaled +
##   brand_usage_rescaled + gender.f + education.f + age_rescaled +
##   occupation.f + (1 | ref_id.f) + (1 | adj_link_id.f)
## Data: phish_inemail_only
##
## REML criterion at convergence: 3184
##
## Scaled residuals:
##   Min      1Q  Median      3Q      Max
## -1.517 -0.435 -0.142  0.180  22.024
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##   ref_id.f    (Intercept) 0.000680 0.0261
##   adj_link_id.f (Intercept) 0.000296 0.0172
##   Residual                0.821549 0.9064
## Number of obs: 1188, groups:  ref_id.f, 3; adj_link_id.f, 3
```

```

##
## Fixed effects:
##
##               Estimate Std. Error
## (Intercept)    -0.47642    0.15620
## activation.fon click      0.18821    0.07570
## forced_attention.fyes    -0.41813    0.07447
## warnings_seen_rescaled  -0.20016    0.02669
## benign_hover_rate        0.79731    0.14695
## cyber_quiz_score_rescaled -0.00627    0.03009
## PE_score_rescaled       -0.03055    0.02691
## brand_usage_rescaled     0.01759    0.02693
## gender.fMale            0.08883    0.05422
## education.fBachelor's degree -0.04879    0.05757
## education.fGraduate degree -0.12602    0.09169
## age_rescaled            0.11669    0.02731
## occupation.fNon-technical occupations 0.04035    0.07598
## activation.fon click:forced_attention.fyes -0.41990    0.10577
##
##               df t value
## (Intercept)    604.56671  -3.05
## activation.fon click    1172.01231    2.49
## forced_attention.fyes    1172.00123   -5.62
## warnings_seen_rescaled    1173.94300   -7.50
## benign_hover_rate        1172.05635    5.43
## cyber_quiz_score_rescaled  1172.00231   -0.21
## PE_score_rescaled        1172.01165   -1.14
## brand_usage_rescaled     1172.55388    0.65
## gender.fMale            1172.00360    1.64
## education.fBachelor's degree 1172.00090   -0.85
## education.fGraduate degree 1172.00325   -1.37
## age_rescaled            1172.00374    4.27
## occupation.fNon-technical occupations 1172.00041    0.53
## activation.fon click:forced_attention.fyes 1172.00030   -3.97
##
##               Pr(>|t|)
## (Intercept)          0.0024 **
## activation.fon click      0.0131 *
## forced_attention.fyes    0.00000002452119 ***
## warnings_seen_rescaled  0.000000000000013 ***
## benign_hover_rate        0.00000007010623 ***
## cyber_quiz_score_rescaled    0.8350
## PE_score_rescaled        0.2565
## brand_usage_rescaled     0.5138
## gender.fMale            0.1016
## education.fBachelor's degree 0.3969
## education.fGraduate degree 0.1696
## age_rescaled            0.00002088348176 ***
## occupation.fNon-technical occupations    0.5955
## activation.fon click:forced_attention.fyes 0.00007632015359 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE) or
##       vcov(x)           if you need it

```

```
# anova(glm_hovertime_email_phish)

#Calculate effect size
r2.corr.mer(glm_hovertime_phish_inemail)

## [1] 0.1912
```