All Interactions Are Wrong: Experimental Edition, Binomial Outcomes

toffa @ psicostat

2024-09-17

Table of contents

0.1 Introduction

In this document, we will fit a **Generalized Linear Mixed Model (GLMM)** with a binomial response. The model aims to predict a dichotomous outcome (binary response) as a function of two main effects, **X1** and **X2** (both are factors featuring 2 levels), while accounting for random intercepts associated with individual respondents.

0.2 Model Specification

The model is specified as follows:

$$\text{logit}(p_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i$$

Where:

- p_i is the probability of a positive response for individual i.
- β_0 is the fixed intercept.
- β_1 and β_2 are the fixed effect coefficients for predictors X_1 and X_2, respectively.
- $u_i \sim N(0, \sigma_u^2)$ represents the random intercept for individual i (respondent).

Thus, the model incorporates fixed effects of X_1 and X_2 and allows for random variability across respondents, improving model flexibility and accuracy.

0.3 Two Scenarios

Let's set a **Scenario A** in which:

- β_0 = -3.0
- $\beta_1 = 1.5$
- $\beta_2 = 1.5$
- $\sigma_u^2 = 1.0$

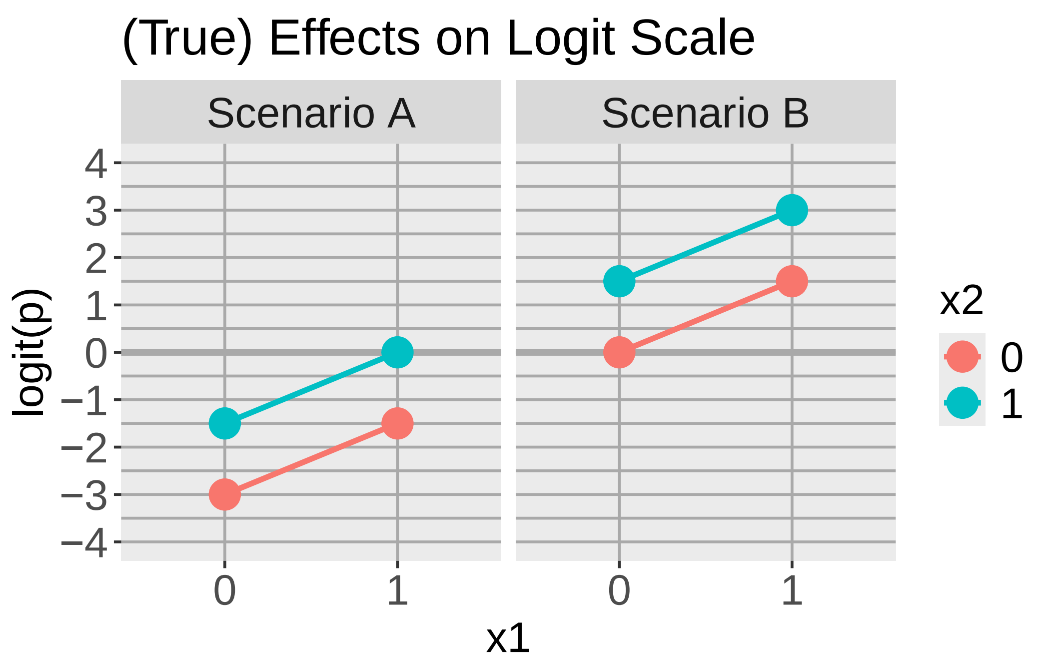
Scenario B is exactly identical to **scenario A** except:

- $\beta_0 = 1$

This is a visual depiction of the expected effects in the two scenarios:

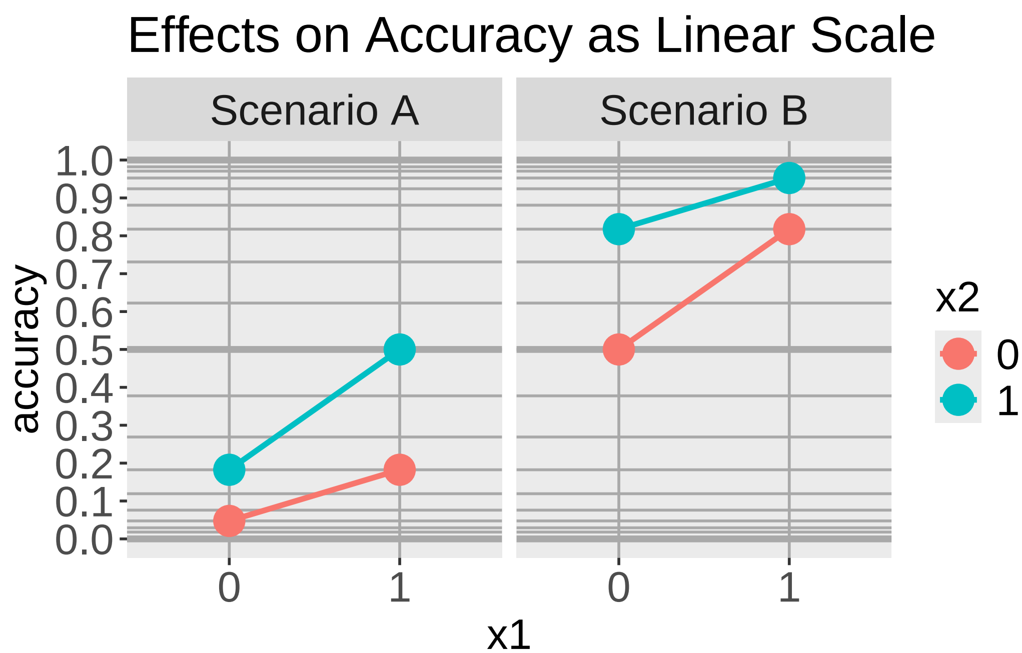
```
# Empty workspace and load needed libraries
rm(list=ls())
library(lme4)
library(lmerTest)
library(ggplot2)
options(round=3)

beta_0_A = -3
beta_0_B = 0
beta_1 = 1.5
beta_2 = 1.5
sigma_u = 1

dfA = data.frame(scenario="Scenario A",expand.grid(x1=c(0,1),
                                                   x2=c(0,1)),logit_p=NA)
dfA$logit_p = beta_0_A + beta_1*dfA$x1 + beta_2*dfA$x2
dfB = data.frame(scenario="Scenario B",expand.grid(x1=c(0,1),
                                                   x2=c(0,1)),logit_p=NA)
dfB$logit_p = beta_0_B + beta_1*dfB$x1 + beta_2*dfB$x2
df = rbind(dfA,dfB)
df$x1 = as.factor(df$x1)
df$x2 = as.factor(df$x2)

hlines = seq(-4,4,0.5)
vlines = c(0,1,2)
```

```r
ggplot(df,aes(y=logit_p,x=x1,group=x2,color=x2))+
  ggtitle("(True) Effects on Logit Scale")+
  geom_hline(yintercept=hlines,color="darkgray")+
  geom_hline(yintercept=0,color="darkgray",size=1.2)+
  geom_vline(xintercept=vlines,color="darkgray")+
  geom_point(size=5)+
  geom_line(size=1)+
  scale_y_continuous(breaks=seq(-10,10,1))+
  facet_wrap(.~scenario)+
  theme(text=element_text(size=20),title=element_text(size=16),
        panel.grid=element_blank())+
  ylab("logit(p)")
```




```r
ggplot(df,aes(y=plogis(logit_p),x=x1,group=x2,color=x2))+
  ggtitle("Effects on Accuracy as Linear Scale")+
  geom_hline(yintercept=plogis(hlines),color="darkgray")+
  geom_hline(yintercept=plogis(c(-Inf,0,Inf)),color="darkgray",size=1.2)+
  geom_vline(xintercept=vlines,color="darkgray")+
  geom_point(size=5)+
  geom_line(size=1)+
  scale_y_continuous(breaks=seq(0,1,.1),limits=c(0,1))+
  facet_wrap(.~scenario)+
```

```
  theme(text=element_text(size=20),title=element_text(size=16),
        panel.grid=element_blank())+
  ylab("accuracy")
```



0.4 Scenario A

```
set.seed(0)
n = 100 # individual respondents
k = 20 # trials
id = rep(1:n,each=k*2*2)
ui = rep(rnorm(n,0,1),each=k*2*2)
x1 = rep(0:1,each=k,times=n*2)
x2 = rep(0:1,each=k*2,times=n)

logit_p = beta_0_A + beta_1*x1 + beta_2*x2 + ui
y = rbinom(length(logit_p),1,plogis(logit_p))
dfA_binom = data.frame(id,x1,x2,y)
dfA_binom$x1 = as.factor(dfA_binom$x1); dfA_binom$x2 = as.factor(dfA_binom$x2)
fitA_logit = glmer(y~x1*x2+(1|id),data=dfA_binom,family="binomial")
summary(fitA_logit)$coefficients
```

```
              Estimate Std. Error      z value      Pr(>|z|)
(Intercept) -2.97585238  0.1271433 -23.4055070 3.755935e-121
x11          1.47878153  0.1102256  13.4159486  4.876530e-41
x21          1.52832188  0.1099172  13.9043000  5.964756e-44
x11:x21      0.03356495  0.1328391   0.2526737  8.005204e-01
```

```
dfA_averag = aggregate(y~id*x1*x2,data=dfA_binom,FUN=mean)
fitA_linear = lmer(y~x1*x2+(1|id),data=dfA_averag)
summary(fitA_linear)$coefficients
```

```
            Estimate Std. Error       df   t value     Pr(>|t|)
(Intercept)   0.0640 0.01558040 186.3997  4.107726 5.977034e-05
x11           0.1475 0.01372088 297.0000 10.750037 5.518934e-23
x21           0.1550 0.01372088 297.0000 11.296649 7.414149e-25
x11:x21       0.1470 0.01940426 297.0000  7.575657 4.565943e-13
```

0.5 Scenario B

```
logit_p = 0 + 1*x1 + 1*x2 + ui
y = rbinom(length(logit_p),1,plogis(logit_p))
dfB_binom = data.frame(id,x1,x2,y)
dfB_binom$x1 = as.factor(dfB_binom$x1); dfB_binom$x2 = as.factor(dfB_binom$x2)
fitB_logit = glmer(y~x1*x2+(1|id),data=dfB_binom,family="binomial")
summary(fitB_logit)$coefficients
```

```
               Estimate Std. Error     z value     Pr(>|z|)
(Intercept) -0.005579199 0.09730041 -0.05733993 9.542744e-01
x11          1.024261475 0.07158521 14.30828361 1.942485e-46
x21          1.107751930 0.07221779 15.33904472 4.193202e-53
x11:x21     -0.239348003 0.10919159 -2.19200032 2.837948e-02
```

```
dfB_averag = aggregate(y~id*x1*x2,data=dfB_binom,FUN=mean)
fitB_linear = lmer(y~x1*x2+(1|id),data=dfB_averag)
summary(fitB_linear)$coefficients
```

```
            Estimate Std. Error       df   t value     Pr(>|t|)
(Intercept)   0.4985 0.01819819 169.1209 27.392829 4.460052e-64
x11           0.2090 0.01481315 297.0000 14.109089 5.997864e-35
x21           0.2240 0.01481315 297.0000 15.121703 1.054408e-38
x11:x21      -0.0910 0.02094895 297.0000 -4.343893 1.923468e-05
```