Homework 11

Psych 5068

Emorie Beck

April 23, 2018

Contents

Workspace

Packages

Data

```
source("https://raw.githubusercontent.com/emoriebeck/homeworks/master/table_fun.R")
data_url <- "https://raw.githubusercontent.com/emoriebeck/homeworks/master/homework11/Safety_Binary(2).c
dat      <- data_url %>% read.csv %>% tbl_df
```

Level 1:
$\eta_{ij} = \beta_{0j} + \beta_{1j} * GMC\_Age + \beta_{2j} * sex$

Level 2:
$\beta_{0j} = \gamma_{00} + \gamma_{01} * crowded\_GMC_j + \gamma_{02} * economic_j + u_{0j}$
$\beta_{1j} = \gamma_{10} + \gamma_{11} * crowded\_GMC_j + \gamma_{12} * economic_j + u_{1j}$
$\beta_{2j} = \gamma_{20} + \gamma_{21} * crowded\_GMC_j + \gamma_{22} * economic_j$

For this assignment you will extend these analyses to test nonlinearity and interactions. The data for this assignment are contained in the file, Safety_Binary.csv.

If you have problems getting some of the models to converge, you can try alternative optimizers and increase the number of iterations that are used before the algorithm gives up. The easiest way to do this is to create several alternative control lists and then substitute them into the model statement as needed:

```
cl1 <- glmerControl(optimizer = "bobyqa", optCtrl=list(maxfun=10000))
cl2 <- glmerControl(optimizer = "Nelder_Mead", optCtrl=list(maxfun=10000))
cl3 <- glmerControl(optimizer = "optimx",optCtrl=list(method="nlminb",maxiter=10000))
```

You will need the `optimx` package for the last one.
Switch to a different optimizer like this:

`Safety\_Fit\_1 <- glmer(unsafe ~ 1 + Z\_Age + Z\_Age\_SQ + sex + (1 + Z\_Age + Z\_Age\_SQ + sex|street), data=Safety\_Data, binomial("logit"), control=cl1)`

Question 1

In the data file, two of the variables are named age and crowded. Standardize them and name them Z_Age and Z_Crowded. Create a new variable that is the square of Z_Age; name it Z_Age_SQ.

```r
dat <- dat %>%
  mutate_at(vars(age, crowded), funs(z = as.numeric(scale(.)))) %>%
  rename(Z_Age = age_z, Z_Crowded = crowded_z) %>%
  mutate(Z_Age_SQ = Z_Age^2)
```

Question 2

Test a Level 1 model that contains Z_Age, Z_Age_SQ, and sex. Level 2 should be unconditional (no predictors) with all residual variances estimated.

```r
fit1 <- glmer(unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ + sex | street),
              data = dat, family = binomial("logit"), control=cl1)
tab1 <- table_fun(fit1)

tab1 %>%
  select(-type) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("", "b", "OR", "CI")) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,4) %>%
  group_rows("Random", 5,8) %>%
  group_rows("Model", 9,10)
```

	b	OR	CI
Fixed			
Intercept	-0.64	0.53	[0.36, 0.73]
Z_Age	0.77	2.16	[1.62, 3.02]
Z_Age_SQ	0.11	1.11	[0.96, 1.46]
sex	1.07	2.92	[2.03, 4.17]
Random			
τ_{00}	0.85	2.35	[1.42, 8.32]
τ_{11}	0.50	1.66	[1.05, 2.66]
τ_{22}	0.13	1.14	[1.01, 1.70]
τ_{33}	0.56	1.75	[1.05, 2.87]
R^2_m	0.14	1.15	[,]
R^2_c	0.47	1.60	[,]

Part A

Is there a nonlinear relationship between age and feeling unsafe?
No, there is no nonlinear relationship between Age and feeling unsafe (b = 0.11, OR = 1.11, 95% CI = [0.96, 1.46]).

Part B

Is sex still a significant predictor?
Sex is still a significant preditor of feeling unsafe (b = 1.07, OR = 2.92, 95% CI = [2.03, 4.17]). The odds of a woman feeling unsafe is 2.92 times higher than a man, on average.

Part C

Can this model be simplified by eliminating any Level 2 variances? Test the following simplifications (the intercept is retained in all random effects specifications):

Part i

Eliminate Z_Age_SQ from the random effects.

```
fit1i <- glmer(unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + sex | street),
               data = dat, family = binomial("logit"), control=cl1)
tab1i <- table_fun(fit1i)
```

Part ii

Eliminate Z_Age and Z_Age_SQ from the random effects.

```
fit1ii <- glmer(unsafe ~ Z_Age + Z_Age_SQ + sex + (sex | street),
                data = dat, family = binomial("logit"))
tab1ii <- table_fun(fit1ii)
```

Part iii

Eliminate sex from the random effects.

```
fit1iii <- glmer(unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ | street),
                 data = dat, family = binomial("logit"))
tab1iii <- table_fun(fit1iii)
```

Part iv

Eliminate sex and Z_Age_SQ from the random effects.

```
fit1iv <- glmer(unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age | street),
                data = dat, family = binomial("logit"))
tab1iv <- table_fun(fit1iv)
```

Part v

Eliminate sex, Z_Age, and Z_Age_SQ from the random effects.

```
fit1v <- glmer(unsafe ~ Z_Age + Z_Age_SQ + sex + (1 | street),
               data = dat, family = binomial("logit"))
tab1v <- table_fun(fit1v)
```

```
tab1i %>% mutate(Model = "1i") %>%
  full_join(tab1ii %>% mutate(Model = "1ii")) %>%
  full_join(tab1ii %>% mutate(Model = "1iii")) %>%
  full_join(tab1ii %>% mutate(Model = "1iv")) %>%
  full_join(tab1ii %>% mutate(Model = "1v")) %>%
  mutate(b = sprintf("\\makecell{%s \\\\ {%s}}", OR, CI)) %>%
  select(type:b, Model) %>%
```

```r
  spread(key = Model, value = b) %>%
  select(-type) %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c("", rep("OR [CI]", 5))) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1, 4) %>%
  group_rows("Model", 5, 6) %>%
  group_rows("Random", 7,9)
```

	OR [CI]	OR [CI]	OR [CI]	OR [CI]	OR [CI]
Fixed					
Intercept	0.54 [0.35, 0.79]	0.57 [0.41, 0.80]	0.57 [0.41, 0.80]	0.57 [0.41, 0.80]	0.57 [0.41, 0.80]
sex	2.87 [1.98, 4.28]	2.61 [1.91, 3.56]	2.61 [1.91, 3.56]	2.61 [1.91, 3.56]	2.61 [1.91, 3.56]
Z_Age	2.03 [1.64, 2.52]	1.89 [1.62, 2.35]	1.89 [1.62, 2.35]	1.89 [1.62, 2.35]	1.89 [1.62, 2.35]
Z_Age_SQ	1.08 [0.85, 1.39]	1.05 [0.90, 1.25]	1.05 [0.90, 1.25]	1.05 [0.90, 1.25]	1.05 [0.90, 1.25]
R^2_c	1.54 [,]	1.43 [,]	1.43 [,]	1.43 [,]	1.43 [,]
R^2_m	1.14 [,]	1.13 [,]	1.13 [,]	1.13 [,]	1.13 [,]
Random					
τ_{00}	2.95 [1.52, 8.37]	2.79 [1.37, 6.70]	2.79 [1.37, 6.70]	2.79 [1.37, 6.70]	2.79 [1.37, 6.70]
τ_{11}	1.48 [1.07, 2.01]	1.42 [1.00, 3.52]	1.42 [1.00, 3.52]	1.42 [1.00, 3.52]	1.42 [1.00, 3.52]
τ_{22}	1.51 [1.00, 3.08]				

Compared to the original model, identify what you consider to be the simplest model using the likelihood ratio test as well as AIC.

```r
anova(fit1, fit1i)
```

```
## Data: dat
## Models:
## fit1i: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + sex | street)
## fit1: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ + sex | street)
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit1i 10 1229.2 1278.2 -604.58   1209.2
## fit1  14 1231.9 1300.6 -601.93   1203.9 5.3084      4     0.2571
```

```r
anova(fit1, fit1ii)
```

```
## Data: dat
## Models:
## fit1ii: unsafe ~ Z_Age + Z_Age_SQ + sex + (sex | street)
## fit1: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ + sex | street)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit1ii  7 1234.4 1268.8 -610.21   1220.4
## fit1   14 1231.9 1300.6 -601.93   1203.9 16.571      7    0.02038 *
```

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

```
anova(fit1, fit1iii)
```

```
## Data: dat
## Models:
## fit1iii: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ | street)
## fit1: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ + sex | street)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit1iii 10 1227.8 1276.9 -603.92   1207.8                         
## fit1    14 1231.9 1300.6 -601.93   1203.9 3.9874      4     0.4077
```

```
anova(fit1, fit1iv)
```

```
## Data: dat
## Models:
## fit1iv: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age | street)
## fit1: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ + sex | street)
##        Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## fit1iv  7 1224.8 1259.1 -605.38   1210.8                        
## fit1   14 1231.9 1300.6 -601.93   1203.9 6.912      7     0.4381
```

```
anova(fit1, fit1v)
```

```
## Data: dat
## Models:
## fit1v: unsafe ~ Z_Age + Z_Age_SQ + sex + (1 | street)
## fit1: unsafe ~ Z_Age + Z_Age_SQ + sex + (Z_Age + Z_Age_SQ + sex | street)
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)  
## fit1v  5 1231.8 1256.3 -610.90   1221.8                           
## fit1  14 1231.9 1300.6 -601.93   1203.9 17.934      9    0.03595 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The best model is the model that only includes a random slope for sex.

Question 3

Carry out a similar series of analyses, but substitute the sex:Z_age interaction for the nonlinear age effect.

```
fit3 <- glmer(unsafe ~ Z_Age*sex + (Z_Age * sex | street),
              data = dat, family = binomial("logit"), control=cl1)
tab3 <- table_fun(fit3)

tab3 %>%
  select(-type) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("", "b", "OR", "CI")) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,4) %>%
  group_rows("Random", 5,8) %>%
  group_rows("Model", 9,10)
```

	b	OR	CI
Fixed			
Intercept	-0.58	0.56	[0.38, 0.79]
Z_Age	0.83	2.29	[1.65, 3.08]
sex	1.02	2.77	[1.98, 4.52]
Z_Age:sex	-0.27	0.76	[0.53, 1.12]
Random			
τ_{00}	1.22	3.40	[1.44, 8.68]
τ_{11}	0.54	1.72	[1.06, 3.24]
τ_{22}	0.47	1.60	[1.02, 4.09]
τ_{33}	0.21	1.23	[1.01, 2.61]
R^2_m	0.13	1.14	[,]
R^2_c	0.43	1.54	[,]

Part A

Is there an interaction between age and sex?
No, there is no interactionAge and sex (b = -0.27, OR = 0.76, 95% CI = [0.53, 1.12]). Sex does not moderate the relationship between age and feeling unsafe.

Part B

What is the simplest Level 2 variance model that can be used? Test the following simplifications (the intercept is retained in all random effects specifications):

Part i

Eliminate sex:Z_Age from the random effects.

```
fit3i <- glmer(unsafe ~ Z_Age*sex + (Z_Age + sex | street),
               data = dat, family = binomial("logit"), control=cl1)
tab3i <- table_fun(fit3i)
```

Part ii

Eliminate Z_Age and sex:Z_Age from the random effects.

```
fit3ii <- glmer(unsafe ~ Z_Age*sex + (sex | street),
                data = dat, family = binomial("logit"))
tab3ii <- table_fun(fit3ii)
```

Part iii

Eliminate sex and sex:Z_Age from the random effects.

```
fit3iii <- glmer(unsafe ~ Z_Age*sex + (Z_Age | street),
                 data = dat, family = binomial("logit"))
tab3iii <- table_fun(fit3iii)
```

Part iv

Eliminate sex, Z_Age, and sex:Z_Age from the random effects.

```r
fit3iv <- glmer(unsafe ~ Z_Age*sex + (1 | street),
                data = dat, family = binomial("logit"))
tab3iv <- table_fun(fit3iv)
```

```r
tab3i %>% mutate(Model = "3i") %>%
  full_join(tab3ii %>% mutate(Model = "3ii")) %>%
  full_join(tab3ii %>% mutate(Model = "3iii")) %>%
  full_join(tab3ii %>% mutate(Model = "3iv")) %>%
  mutate(b = sprintf("\\makecell{%s \\\\ {%s}}", OR, CI)) %>%
  select(type:b, Model) %>%
  spread(key = Model, value = b) %>%
  select(-type) %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c("", rep("OR [CI]", 4))) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1, 4) %>%
  group_rows("Model", 5, 6) %>%
  group_rows("Random", 7,9)
```

	OR [CI]	OR [CI]	OR [CI]	OR [CI]
Fixed				
Intercept	0.56 [0.41, 0.79]	0.58 [0.45, 0.81]	0.58 [0.45, 0.81]	0.58 [0.45, 0.81]
sex	2.89 [1.96, 3.93]	2.63 [1.97, 3.83]	2.63 [1.97, 3.83]	2.63 [1.97, 3.83]
Z_Age	2.26 [1.67, 3.03]	2.08 [1.65, 2.76]	2.08 [1.65, 2.76]	2.08 [1.65, 2.76]
Z_Age:sex	0.80 [0.54, 1.16]	0.82 [0.60, 1.19]	0.82 [0.60, 1.19]	0.82 [0.60, 1.19]
R^2_c	1.54 [,]	1.43 [,]	1.43 [,]	1.43 [,]
R^2_m	1.14 [,]	1.13 [,]	1.13 [,]	1.13 [,]
Random				
τ_{00}	3.22 [1.57, 7.85]	3.08 [1.53, 6.45]	3.08 [1.53, 6.45]	3.08 [1.53, 6.45]
τ_{11}	1.48 [1.06, 2.14]	1.40 [1.00, 2.61]	1.40 [1.00, 2.61]	1.40 [1.00, 2.61]
τ_{22}	1.43 [1.01, 2.75]			

Compared to the original model, identify what you consider to be the simplest model using the likelihood ratio test as well as AIC.

```r
anova(fit3, fit3i)
```

```
## Data: dat
## Models:
## fit3i: unsafe ~ Z_Age * sex + (Z_Age + sex | street)
## fit3: unsafe ~ Z_Age * sex + (Z_Age * sex | street)
```

```
##       Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## fit3i 10 1228.3 1277.4 -604.16   1208.3
## fit3  14 1233.7 1302.4 -602.83   1205.7 2.665      4     0.6154
```

```r
anova(fit3, fit3ii)
```

```
## Data: dat
## Models:
## fit3ii: unsafe ~ Z_Age * sex + (sex | street)
## fit3: unsafe ~ Z_Age * sex + (Z_Age * sex | street)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit3ii  7 1233.3 1267.7 -609.66   1219.3
## fit3   14 1233.7 1302.4 -602.83   1205.7 13.658      7    0.05761 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit3, fit3iii)
```

```
## Data: dat
## Models:
## fit3iii: unsafe ~ Z_Age * sex + (Z_Age | street)
## fit3: unsafe ~ Z_Age * sex + (Z_Age * sex | street)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit3iii  7 1223.4 1257.8 -604.70   1209.4
## fit3    14 1233.7 1302.4 -602.83   1205.7 3.7368      7     0.8095
```

```r
anova(fit3, fit3iv)
```

```
## Data: dat
## Models:
## fit3iv: unsafe ~ Z_Age * sex + (1 | street)
## fit3: unsafe ~ Z_Age * sex + (Z_Age * sex | street)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit3iv  5 1230.4 1254.9 -610.18   1220.4
## fit3   14 1233.7 1302.4 -602.83   1205.7 14.704      9     0.0994 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 4

Now add Z_Crowded to the simplest model from Question 3 and include all two-way interactions and the three-way interaction.

```r
fit4 <- glmer(unsafe ~ Z_Age*sex*Z_Crowded + (sex*Z_Crowded | street),
              data = dat, family = binomial("logit"), control=cl1)
tab4 <- table_fun(fit4)

tab4 %>%
  select(-type) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("", "b", "OR", "CI")) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,8) %>%
  group_rows("Random", 9,12) %>%
  group_rows("Model", 13,14)
```

	b	OR	CI
Fixed			
Intercept	-0.57	0.57	[0.39, 0.80]
Z_Age	0.69	2.00	[1.58, 2.69]
sex	1.00	2.71	[1.94, 4.61]
Z_Crowded	-0.53	0.59	[0.40, 0.85]
Z_Age:sex	-0.16	0.85	[0.57, 1.20]
Z_Age:Z_Crowded	-0.33	0.72	[0.53, 0.90]
sex:Z_Crowded	-0.26	0.77	[0.48, 1.25]
Z_Age:sex:Z_Crowded	0.43	1.54	[1.13, 2.20]
Random			
τ_{00}	0.89	2.44	[1.17, 6.13]
τ_{11}	0.10	1.10	[1.00, 3.54]
τ_{22}	0.22	1.24	[1.00, 3.92]
τ_{33}	0.52	1.69	[1.01, 5.76]
R^2_m	0.21	1.24	[,]
R^2_c	0.37	1.45	[,]

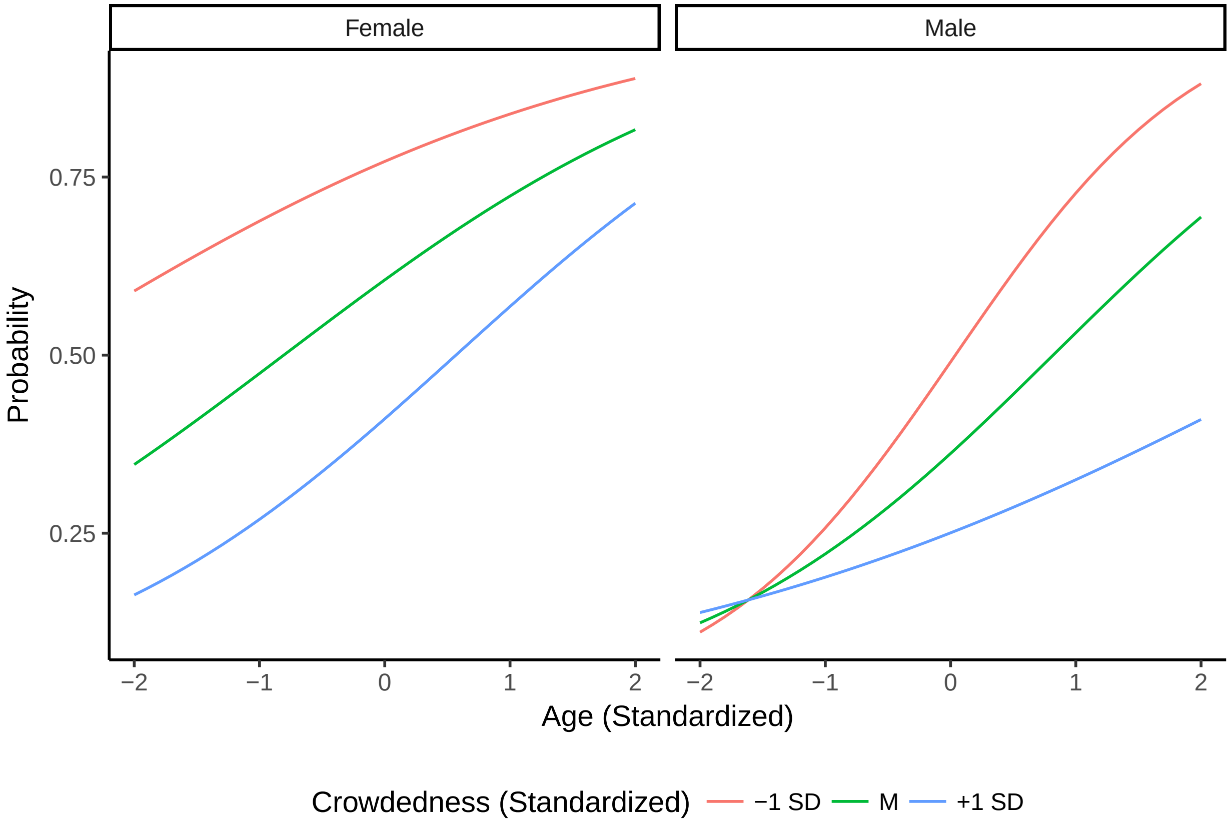
Part A

What is the highest order effect that is significant?
The three way interaction between Age, sex, and crowdedness is significant.

Part B

Plot the highest order significant effect with probability of feeling unsafe as the outcome.

```r
crossing(
  Z_Age = seq(-2,2,.1),
  Z_Crowded = seq(-1,1,1),
  sex = c(0,1)
) %>%
  mutate(pred = predict(fit4, newdata = ., re.form = NA),
         ci = predict(fit4, newdata =., interval = "conf", re.form = NA),
         OR = exp(pred),
         p = OR/(1+OR),
         Z_Crowded = mapvalues(Z_Crowded, c(-1,0,1), c("-1 SD", "M", "+1 SD")),
         Z_Crowded = factor(Z_Crowded, levels = c("-1 SD", "M", "+1 SD")),
         sex = mapvalues(sex, c(0,1), c("Male", "Female"))) %>%
  ggplot(aes(x = Z_Age, y = p, color = Z_Crowded)) +
  geom_line() +
  facet_grid(~sex) +
  theme_classic() +
  labs(x = "Age (Standardized)", y = "Probability", color = "Crowdedness (Standardized)") +
  theme(legend.position = "bottom")
```



Part C

Provide an interpretation for the plotted effect.

For young men, the probability of feeling unsafe dones't vary across different levels of crowdedness. As they age, however, men become much more likely to feel unsafe on uncrowded streets. Across different ages, women are more likely to feel unsafe on less crowded streets. Older women in general are also more likely to feel unsafe.