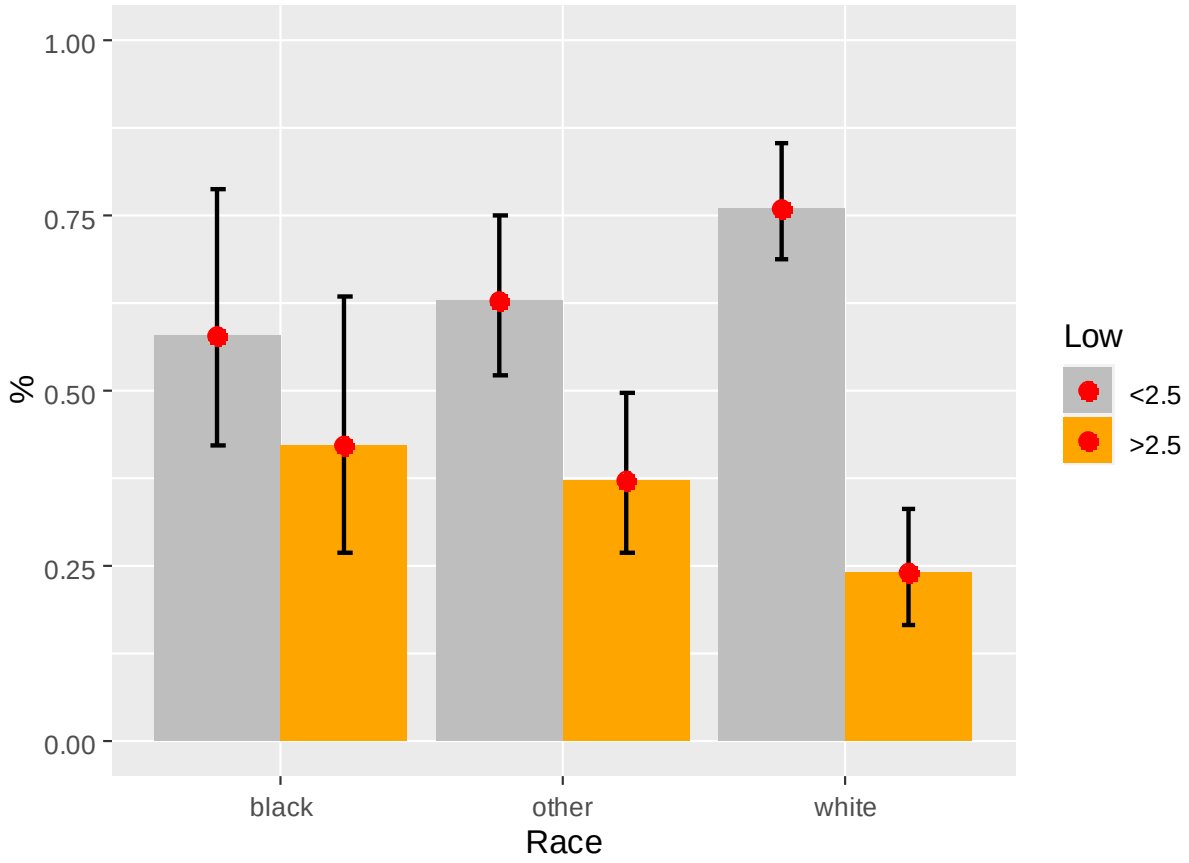
Bar plots with percentages

Summary

Use **compareGroups** for obtaining a table of percentage

Plot percentages of a table as a bar plot

Estimate simultaneous CI for proportions



Load libraries and data

We shall use the database **birthwt** available in the **UsingR** package

We need the packages:

```
library(tidyverse)
library(DescTools)
library(questionr)
library(plyr)
library(formattable)
library(gmodels)
```

We shall use the data **birthwt**

head(birthwt)

	low	age	lwt	race	smoke	ptl	ht	ui	ftv	bwt
85	0	19	182	2	0	0	0	1	0	2523
86	0	33	155	3	0	0	0	0	3	2551
87	0	20	105	1	1	0	0	0	1	2557
88	0	21	108	1	1	0	0	1	2	2594
89	0	18	107	1	1	0	0	1	0	2600
91	0	21	124	3	0	0	0	0	0	2622

Define variables as factors

```
birthwt$race<- factor(birthwt$race,labels=c('white','black','other'))
birthwt$low<- factor(birthwt$low,
                     labels =c('>2.5','<2.5'))
birthwt$ht<-factor(birthwt$ht,labels=c('No','Yes'))
head(birthwt)
```

```
    low age lwt  race smoke ptl ht ui ftv  bwt
85 <2.5  19 182 black     0   0 No  1   0 2523
86 <2.5  33 155 other     0   0 No  0   3 2551
87 <2.5  20 105 white     1   0 No  0   1 2557
88 <2.5  21 108 white     1   0 No  1   2 2594
89 <2.5  18 107 white     1   0 No  1   0 2600
91 <2.5  21 124 other     0   0 No  0   0 2622
```

Use **compareGroups** for obtaining a table of frequencies

```
createTable(compareGroups(race~low,birthwt))
```

race Will appear as columns and **low** as rows.

	white N=96	black N=26	other N=67	p.overall
low:				0.082
>2.5	73 (76.0%)	15 (57.7%)	42 (62.7%)	
<2.5	23 (24.0%)	11 (42.3%)	25 (37.3%)	

Percentages are computed within columns. In this case We obtain the percentage of births with low and normal weight.

Confidence intervals

```
createTable(compareGroups(race~low,birthwt),show.ci=TRUE)
```

This option for createTable compute CI for the probabilities

	white N=96	black N=26	other N=67	p.overall
low:				0.082
>2.5	76.0% [66.3%;84.2%]	57.7% [36.9%;76.6%]	62.7% [50.0%;74.2%]	
<2.5	24.0% [15.8%;33.7%]	42.3% [23.4%;63.1%]	37.3% [25.8%;50.0%]	

Prepare percentage table for plotting

```
t <- birthwt %>% group_by(race,low) %>%
  summarise(n=n()) %>% mutate(freq=n/sum(n))
t
```

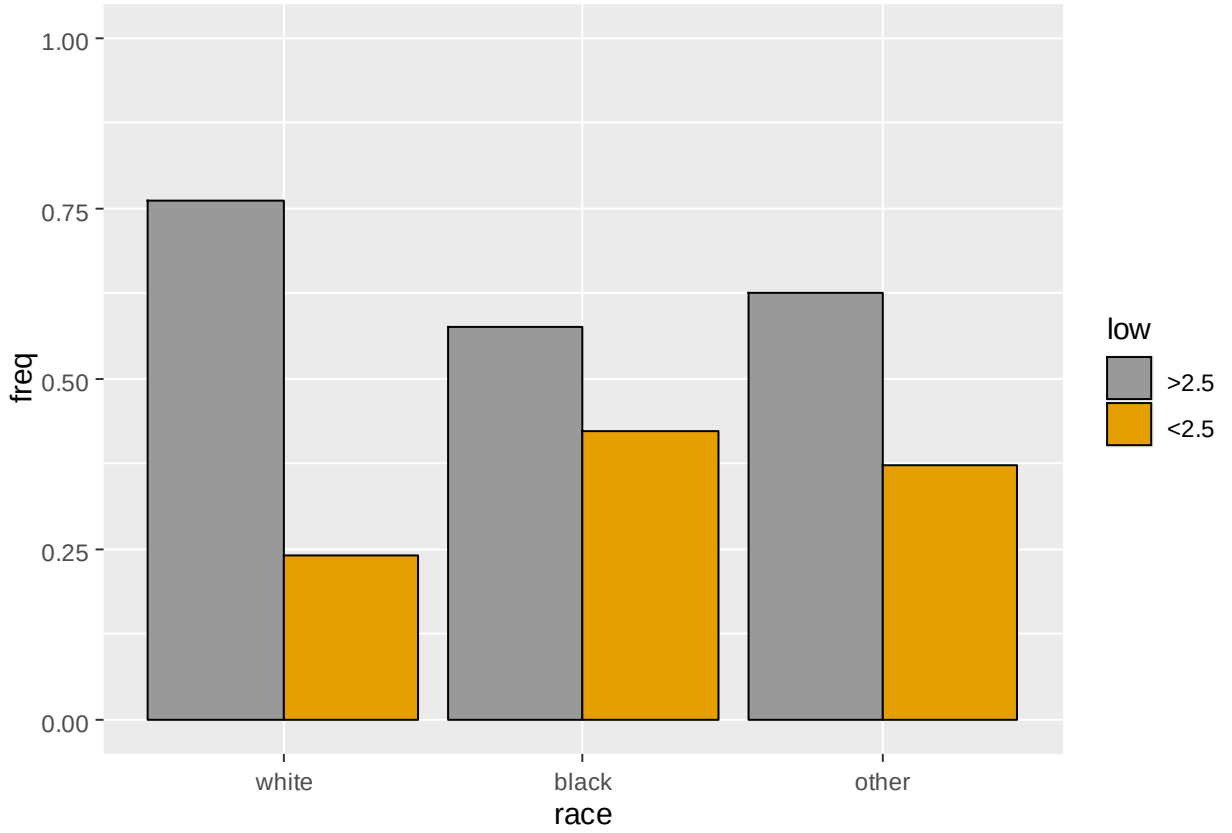
```
# A tibble: 6 x 4
# Groups:   race [3]
  race  low       n  freq
  <fct> <fct> <int> <dbl>
1 white >2.5     73 0.760
2 white <2.5     23 0.240
3 black >2.5     15 0.577
4 black <2.5     11 0.423
5 other >2.5     42 0.627
6 other <2.5     25 0.373
```

- First compute the percentages. This can be done by grouping by the variables and then calculate (summarise) the number of cases for each cell and calculate the frequencies.
- Frequencies are computed over the observations in each race.
- In the group_by the first variable are the groups and the second one are the events within groups.
- For instance, the proportion of births with low weight in black mothers is 0.423

Bar plot

```
ggplot(t,aes(x=race,fill=low,freq))+
   geom_bar(stat="identity",
            position=position_dodge(),
            color="black")+
  scale_fill_manual(values=c("#999999", "#E69F00"))+
  ylim(0,1)
```

- If you want to plot the computed proportions, the aesthetic in ggplot should include the x (race) the groups (fill=low), and the variable that has the proportions (freq)
- Then geom_bar should indicate the stat='identity'.
- Manual colors can be indicated by scale_fill_manual



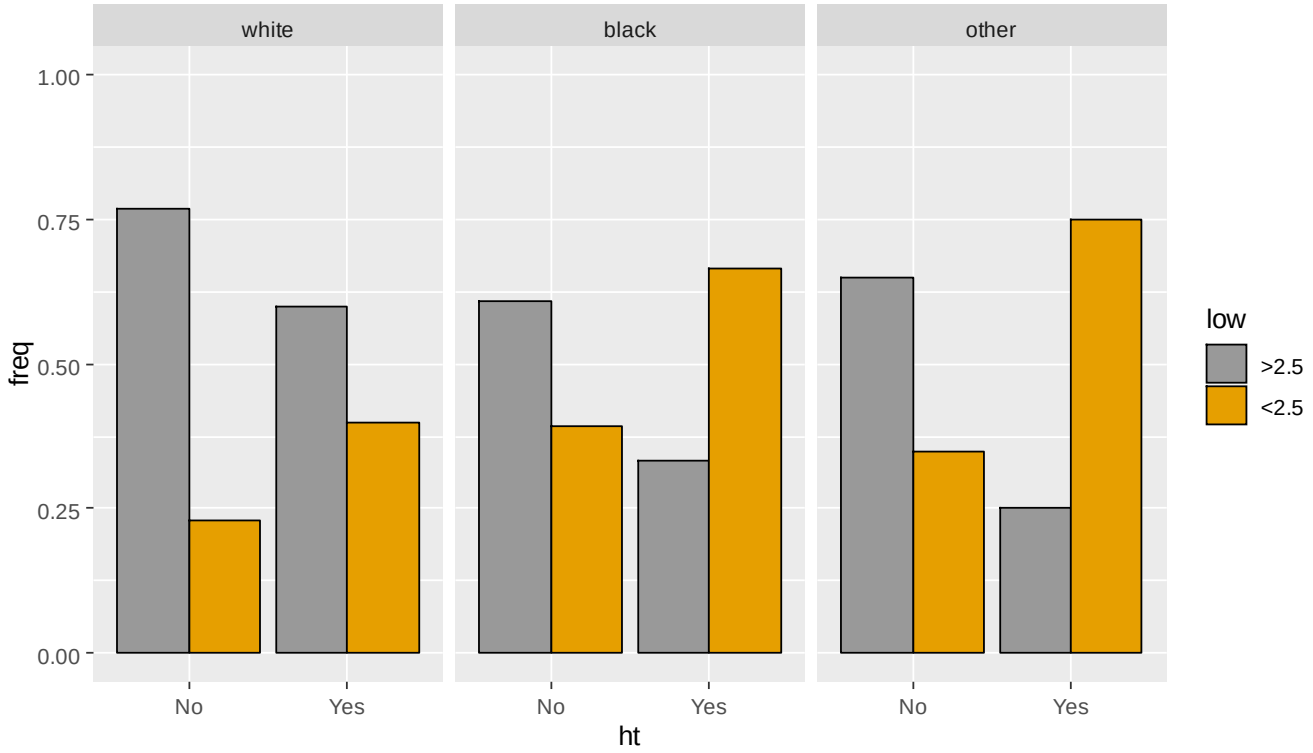
Proportion of low within ht by race

```
res <-  compareGroups(ht~low,birthwt)
t1<-createTable(update(res,subset=race=='white'))
t2<-createTable(update(res,subset=race=='black'))
t3<-createTable(update(res,subset=race=='other'))

cbind('White'=t1,'Black'=t2,'Other'=t3)
```

	White			Black			Other		
	No N=91	Yes N=5	p.overall	No N=23	Yes N=3	p.overall	No N=63	Yes N=4	p.overall
low:			0.590			0.556			0.143
>2.5	70 (76.9%)	3 (60.0%)		14 (60.9%)	1 (33.3%)		41 (65.1%)	1 (25.0%)	
<2.5	21 (23.1%)	2 (40.0%)		9 (39.1%)	2 (66.7%)		22 (34.9%)	3 (75.0%)	

Bar plots




```
t <- birthwt %>% group_by(race,ht,low) %>%
  summarise(n=n()) %>% mutate(freq=n/sum(n))
t

ggplot(t,aes(x=ht,fill=low,freq))+
  geom_bar(stat="identity",
           position=position_dodge(),
           color="black")+
  scale_fill_manual(values=c("#999999", "#E69F00"))+
  ylim(0,1)+
  facet_wrap(.~race)
```

- You can group by three variables. Proportions of low within ht groups are computed for each race.
- Then a barplot for each race can be obtained using facet_wrap

Simultaneous CI for proportions

Suppose we have a sample of blood types (A,B,AB,O) in a population. The results are:

A: 39
B: 11
AB: 4
O: 48

We can obtain CI for the population probabilities by using the **MultinomCI** within the **DescTools** package

```
> MultinomCI(c(39,11,4,48)) %>% round(2)
      est lwr.ci upr.ci
[1,] 0.38   0.28   0.48
[2,] 0.11   0.01   0.21
[3,] 0.04   0.00   0.14
[4,] 0.47   0.37   0.57
```

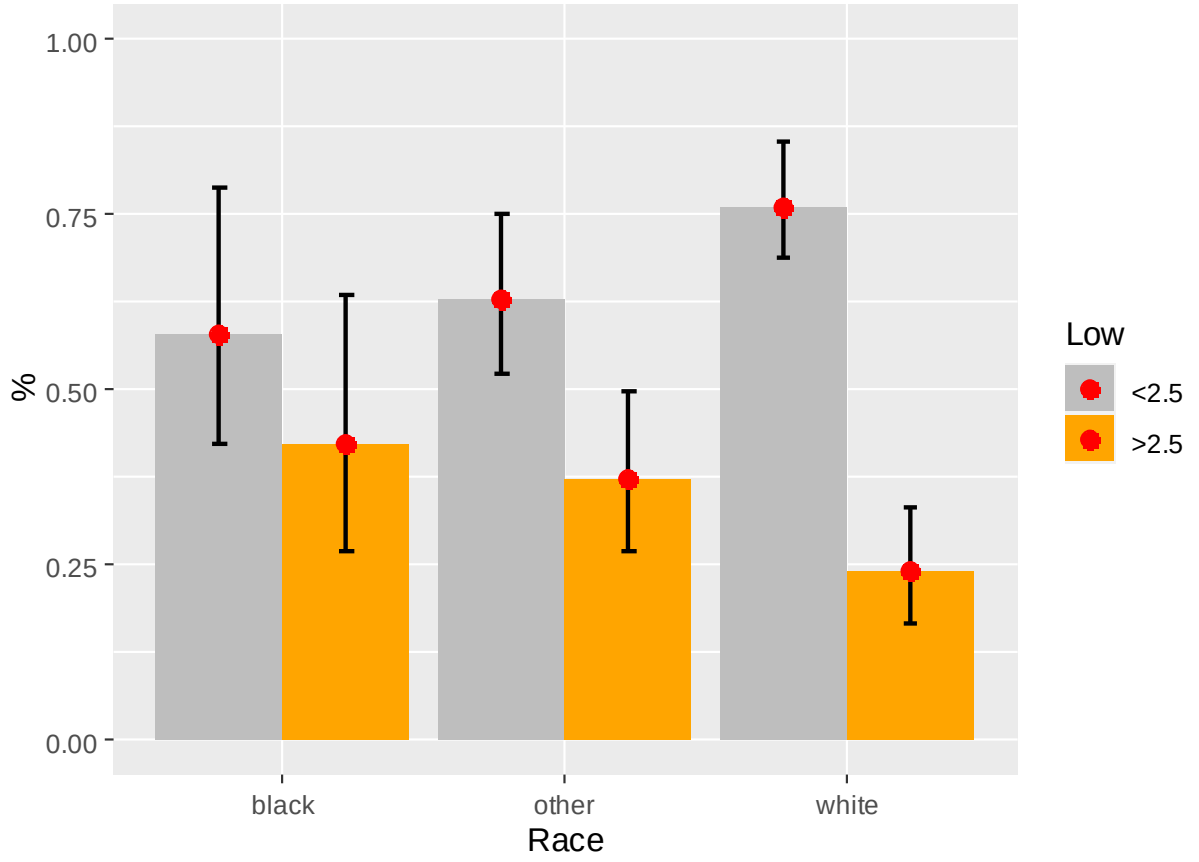
It is not correct to use the CI for a proportion repeated for each group!! For instance for the B group:

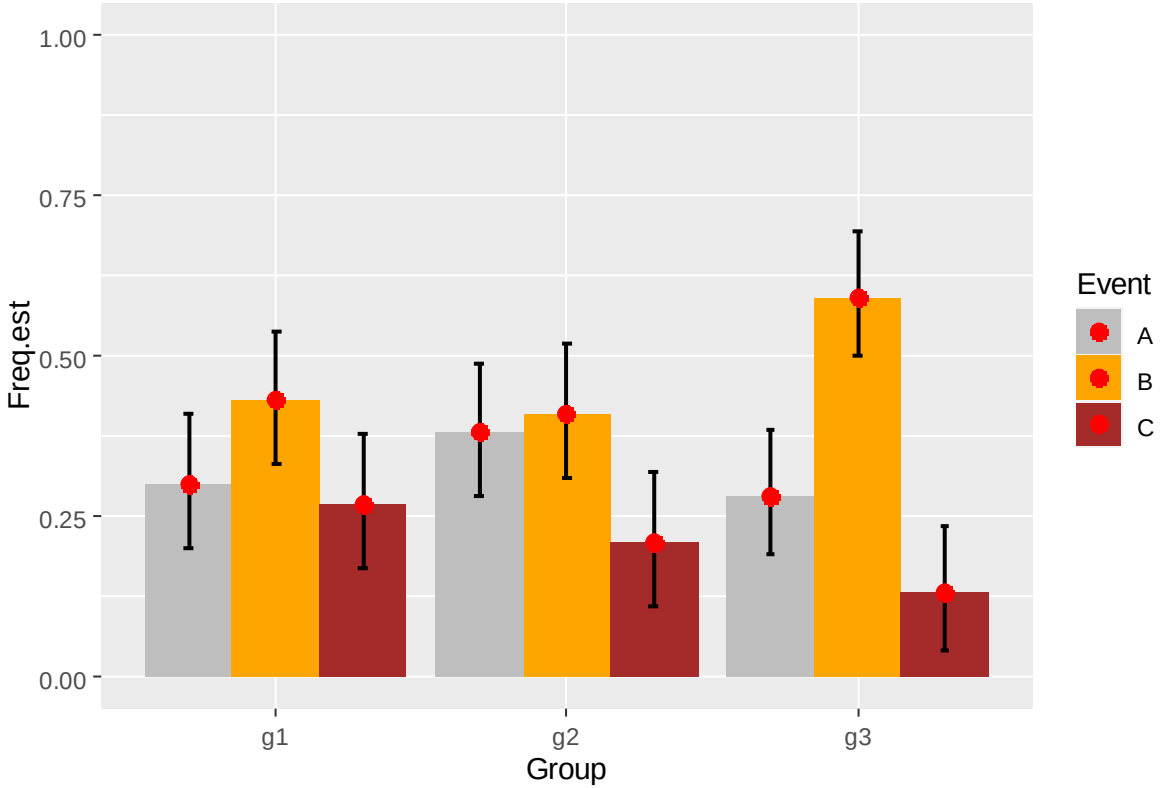
```
> binom.exact(11,sum(c(39,11,4,48))) %>% round(2)
   x   n proportion lower upper conf.level
1 11 102       0.11  0.06  0.18       0.95
```

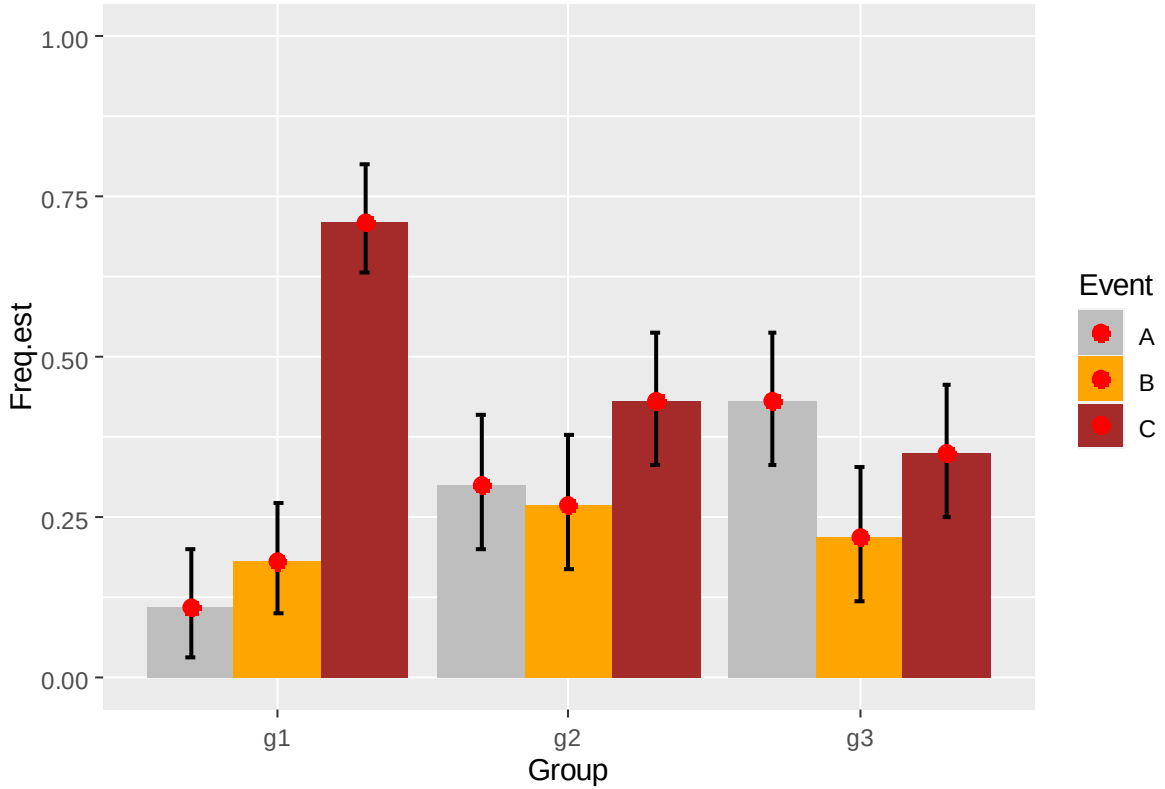
CI for proportions

	white N=96	black N=26	other N=67	p.overall
low:				0.082
<2.5	73 (76.0%)	15 (57.7%)	42 (62.7%)	
>2.5	23 (24.0%)	11 (42.3%)	25 (37.3%)	

- CI from proportions should be computed as simultaneous CI using MultinomCI

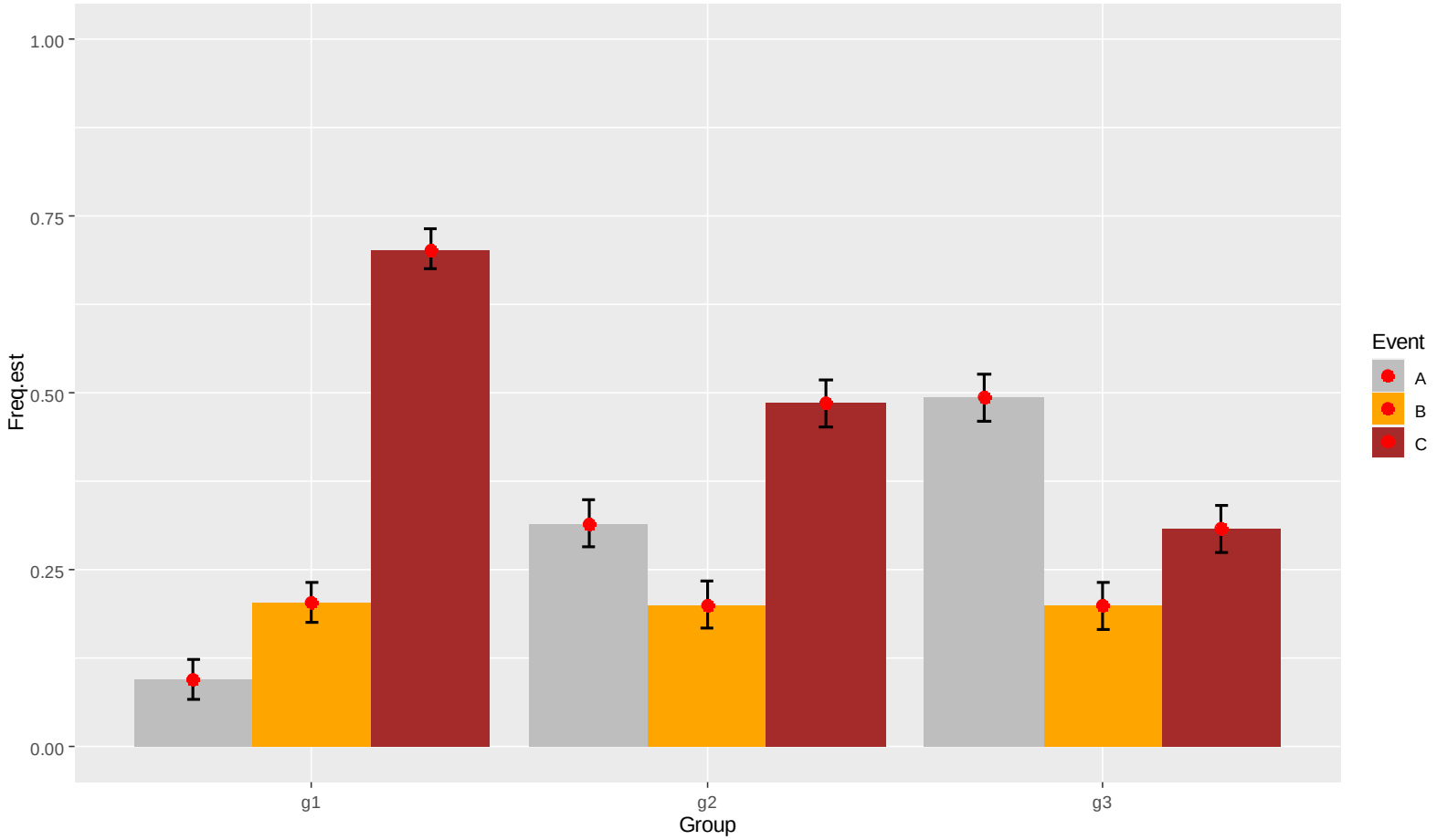



1.00
0.75
0.50
0.25
0.00
Freq.est
g1
g2
g3
Group
Event
A
B
C


1.00
0.75
0.50
0.25
0.00
Freq.est
g1
g2
g3
Group
Event
A
B
C

CI and adjusted residuals in contingency tables

	A	B	C	Row Total
g1	94	204	702	1000
	9.400%	20.400%	70.200%	33.333%
	-17.455	0.290	15.776	
g2	315	200	485	1000
	31.500%	20.000%	48.500%	33.333%
	1.211	-0.097	-1.033	
g3	493	199	308	1000
	49.300%	19.900%	30.800%	33.333%
	16.245	-0.193	-14.743	
Column Total	902	603	1495	3000



CI and adjusted residuals in contingency tables

	A	B	C	Row Total
g1	10	23	67	100
	10.000%	23.000%	67.000%	33.445%
	-4.491	0.174	3.800	
g2	23	20	52	95
	24.211%	21.053%	54.737%	31.773%
	-0.504	-0.384	0.763	
g3	45	24	35	104
	43.269%	23.077%	33.654%	34.783%
	4.941	0.203	-4.511	
Column Total	78	67	154	299