Inference for categorical data Solution

MATH224 - Intro to Stat

Exercise 1 (2 Points)

```
yrbss %>%
  count(text_while_driving_30d)
```

```
## # A tibble: 9 x 2
##   text_while_driving_30d     n
##   <chr>                  <int>
## 1 0                       4792
## 2 1-2                      925
## 3 10-19                    373
## 4 20-29                    298
## 5 3-5                      493
## 6 30                       827
## 7 6-9                      311
## 8 did not drive           4646
## 9 <NA>                     918
```

Exercise 2 (3 Points)

For Section 001

```
yrbss %>%
  select(helmet_12m, text_while_driving_30d)%>%
  na.omit()%>%
  filter(text_while_driving_30d != "did not drive",
         helmet_12m == "never")%>%
  count(text_while_driving_30d)%>%
  mutate(prop = n/sum(n))
```

```
## # A tibble: 7 x 3
##   text_while_driving_30d     n   prop
##   <chr>                  <int>  <dbl>
## 1 0                       2566 0.585
## 2 1-2                      515 0.117
## 3 10-19                    207 0.0472
## 4 20-29                    180 0.0410
## 5 3-5                      281 0.0641
## 6 30                       463 0.106
## 7 6-9                      175 0.0399
```

For Section 007

```r
no_helmet <- yrbss %>%
  filter(helmet_12m == "never")

no_helmet <- no_helmet %>%
  mutate(text_ind = ifelse(text_while_driving_30d == "30", "yes", "no"))%>%
  filter(!is.na(text_ind))

no_helmet%>%
  count(text_ind)%>%
  mutate(p = n/sum(n))
```

```
## # A tibble: 2 x 3
##   text_ind     n      p
##   <chr>    <int>  <dbl>
## 1 no        6040 0.929
## 2 yes        463 0.0712
```

Exercise 3 (3 Points)

```r
no_helmet <- yrbss %>%
  filter(helmet_12m == "never")

no_helmet <- no_helmet %>%
  mutate(text_ind = ifelse(text_while_driving_30d == "30", "yes", "no"))%>%
  filter(!is.na(text_ind))

prop_test(no_helmet,
          text_ind ~ NULL,
          success = "yes",
          z = TRUE,
          conf_int = TRUE,
          conf_level = 0.95, correct = FALSE) # 2 Point
```

```
## # A tibble: 1 x 5
##   statistic p_value alternative lower_ci upper_ci
##       <dbl>   <dbl> <chr>          <dbl>    <dbl>
## 1     -69.2       0 two.sided     0.0652   0.0777
```

```r
(0.07770443 - 0.06519769)/2 # 1 Point
```

```
## [1] 0.00625337
```

Exercise 4 (3 Points)

```r
dat = yrbss%>%
  select(gender)%>%
  na.omit() # 1 Point

prop_test(dat,
          gender ~ NULL,
          success = "male",
          z = TRUE,
          conf_int = TRUE,
          conf_level = 0.95, correct = FALSE) # 1 Point
```

```
## # A tibble: 1 x 5
##   statistic p_value alternative lower_ci upper_ci
##       <dbl>   <dbl> <chr>          <dbl>    <dbl>
## 1      2.82 0.00474 two.sided      0.504    0.521
```

```r
(0.5205266 - 0.5037094)/2 # 1 Point
```

```
## [1] 0.0084086
```

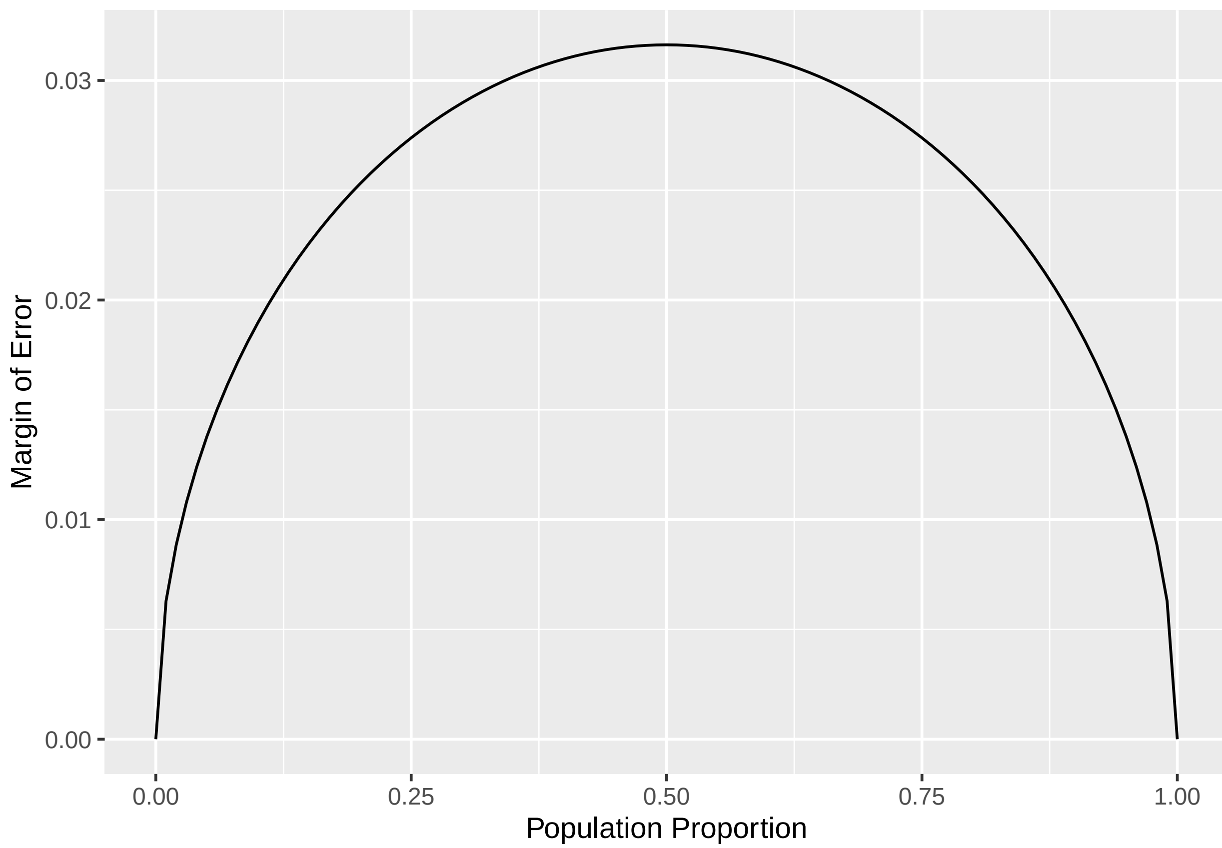
Exercise 5 (3 Points)

2 Point When p = 0 & 1, ME = 0, but as p increases ME increases. ME maximizes when p = 0.5, after that it starts decreasing as p increases.

```r
n <- 1000 #0.25

p <- seq(from = 0, to = 1, by = 0.01) # 0.25 Point
me <- 2 * sqrt(p * (1 - p)/n) # 0.25 Point

dd <- data.frame(p = p, me = me)
ggplot(data = dd, aes(x = p, y = me)) +
  geom_line() +
  labs(x = "Population Proportion", y = "Margin of Error") # 0.25 Point
```



Exercise 6 (3 Points)

2 Points We would reject the Null Hypothesis H_0 because p-values is less than 0.05

```
prop_test(no_helmet,
          text_ind~NULL,
          success = "yes",
          p = 0.08,
          z = TRUE) # 1 Point
```

```
## # A tibble: 1 x 3
##   statistic p_value alternative
##       <dbl>   <dbl> <chr>
## 1     -2.62 0.00889 two.sided
```

Exercise 7 (3 Points)

1 Point for question Hypothesis testing on the gender variable to see if the male proportion is 50% or not

```
prop_test(dat,
          gender~NULL,
          success = "male",
          p = 0.50,
          z = TRUE) # 1 Point
```

```
## # A tibble: 1 x 3
##   statistic p_value alternative
##       <dbl>   <dbl> <chr>
## 1      2.82 0.00474 two.sided
```

1 Point for inference Since the p-value is less than 0.05, we will reject the null hypothesis of $H_0 : p = 0.5$