Homework 3

STA 138 | Camden Possinger | Winter 2022

1. Suppose that, in an independent random sample of size 32, 9 cars are not in compliance with smog regulations.

Obtain the Wald, Clopper-Pearson, Agresti-Coull, and Wilson confidence intervals for the proportion of noncompliant cars in the population. Which is largest? Which is smallest?

```
n <- 32
x <- 9
pi <- x/n
conf_level <- 0.95

### Wald ###
sd <- sqrt(pi*(1-pi)/n)
wald_stat <- qnorm(c((1-conf_level)/2, 1-(1-conf_level)/2))

wald_interval <- pi + wald_stat*sd

### Clopper-Pearson ###
clopper_pearson_interval <- binom.test(x,n,pi)$conf.int

### Agresti-Coull ###
n_adj <- n + 4
x_adj <- x + 2
pi_adj <- x_adj/n_adj

sd_adj <- sqrt(pi_adj*(1-pi_adj)/n_adj)


agresti_coull_interval <- pi_adj + wald_stat*sd_adj

### Wilson ###
wilson_interval <- prop.test(x,n)$conf.int


intervals <- data.frame("Name" = c("Wald",
                                   "Clopper-Pearson",
                                   "Agresti-Coull",
                                   "Wilson"),

                        "Lower_Bound" = c(wald_interval[1],
                                          clopper_pearson_interval[1],
                                          agresti_coull_interval[1],
                                          wilson_interval[1]),

                        "Upper_Bound" = c(wald_interval[2],
```

```
                                        clopper_pearson_interval[2],
                                        agresti_coull_interval[2],
                                        wilson_interval[2])
)

intervals %>% kbl
```

Name	Lower_Bound	Upper_Bound
Wald	0.1254712	0.4370288
Clopper-Pearson	0.1374569	0.4674711
Agresti-Coull	0.1550818	0.4560293
Wilson	0.1439858	0.4697966

```
max_index <- (intervals$Upper_Bound - intervals$Lower_Bound) %>% which.max
min_index <- (intervals$Upper_Bound - intervals$Lower_Bound) %>% which.min

max_interval <- intervals$Name[max_index]
min_interval <- intervals$Name[min_index]
```

```
## [1] "Largest Interval: Clopper-Pearson"

## [1] "Smallest Interval: Agresti-Coull"
```

2. Repeat homework 2 problem 3. (e) using a likelihood ratio test instead of Pearson's test of goodness of fit. The study found 241 businesses in category I, 69 in category II, and 41 in category III. Evaluate the claim

that half of restaurants go out of business every three years. Use a significance level of 0.01, and Likelihood Ratio Test

```
lr_test <- function(n,x,pi_null){

  pi_hat <- x/n


  L1 <- dmultinom(x,n,pi_hat) %>% log
  L0 <- dmultinom(x,n,pi_null) %>% log

  LR_stat <- 2*(L1-L0)

  p_val <- 1-pchisq(LR_stat, length(x) - 1 )

  return(list("test_stat" = LR_stat,"p_value" = p_val))
}
```

```
n <-409
x <- c(241,69,41, 58)
pi_null <- rep(0.5,4)

test_results <- lr_test(n,x,pi_null)

data.frame( "Test_Stat" = test_results$test_stat,
           "P_Value" = test_results$p_value ) %>% kbl
```

Test_Stat	P_Value
218.2729	0

Here we have evidence to reject the claim that for each time period half of the restaurants go out of business.

3 Repeat homework 2 problem 4. (a) using a likelihood ratio test instead of Pearson's test of goodness of fit. Is there evidence at significance level 0.01 to conclude that the colors in this study differ in the attention

that they attract? Use a Pearson test of goodness of fit.

```r
n <- 2064
x <- c(306,338,432,348,331,309)
pi_null <- rep(344/2064, 6)

test_results <- lr_test(n,x,pi_null)


data.frame( "Test_Stat" = test_results$test_stat,
           "P_Value" = test_results$p_value ) %>% kbl
```

Test_Stat	P_Value
29.50425	1.85e-05

Here we have evidence to reject the null hypothesis that the colors in this study receive the same level of attention. We can conclude that at least two of these shirt colors differ in attention.