BIOS668 Homework #2

Sara O'Brien

2/7/2023

Honor code: On my honor, I have neither given nor received unauthorized aid on this assignment. *Sara O'Brien*

Question 1 (2') *Testing by hand calculation*

The claim that the rate of symptomatic COVID-19 illness is the same across the mRNA-1273 and placebo groups can be tested using a Chi-square test under the null hypothesis H0: p1 = p2. A by-hand calculation Chi-square test yields a p-value of **1.06e-35**, which is significantly below our alpha-level of 0.05. We have sufficient evidence to reject the null hypothesis that rates of infection are equal across the vaccination and placebo groups. This is evidence against the claim that mRNA-1273 has no effect.

Testing by simulation

We can also test this hypothesis by simulation to determine the probability of observing a X-squared value larger than the X-squared value from our by-hand calculation under the null hypothesis. We obtain a **probability of 0. This is further evidence against our null** that there is no difference in rates of infection for the mRNA-1273 and placebo groups because the simulation found the simulated X-squared values to be less than actual X-squared value, 155.5458, indicating a greater probability of a significant difference between treatment groups in the actual data from the Phase 3 trial of mRNA-1273. This is additional evidence against the claim that mRNA-1273 has no effect.

```
# Q1. Create data table
ModernaP3 <- as.table(rbind(c(11, 15210-11), c(185, 15210-185)))
dimnames(ModernaP3) <- list(Group=c("mRNA-1273", "Placebo"), Outcome=c('Infected',"Not-Infected"))

ModernaP3
```

```
##             Outcome
## Group        Infected Not-Infected
##   mRNA-1273        11        15199
##   Placebo         185        15025
```

```
## Q1 (2'). Test H0: p1=p2 by hand
n1 <- 11; n2 <- 185
Obs_n <- c(n1, 15210-n1, n2, 15210-n2)

# Expected number of infected cases from the 15210 people in mRNA-1273
E_n1 <- 15210*(196/sum(c(n1, 15210-n1, n2, 15210-n2)))

# Expected number of infected cases from the 15210 people in placebo group
E_n2 <- 15210*(196/sum(c(n1, 15219-n1, n2, 15210-n2)))

stat.obs <- sum(  (c(n1, 15210-n1, n2, 15210-n2) - c(E_n1, 15210-E_n1, E_n2, 15210-E_n2))**2/c(E_n1, 152

stat.obs
```

```
## [1] 155.5458
```

```r
pchisq(stat.obs, df=1, lower.tail=FALSE) # Prob(x(1) > Stat.obs)
```

```
## [1] 1.063961e-35
```

```r
# Other methods to test H0:
chisq.test(ModernaP3) # Also yields significant p-value
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  ModernaP3
## X-squared = 153.69, df = 1, p-value < 2.2e-16
```

```r
fisher.test(ModernaP3) # Also yields significant p-value
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  ModernaP3
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.02881611 0.10771565
## sample estimates:
## odds ratio
## 0.05877327
```

```r
# Q1 (2'). Test H0 by simulation

nrep <- 5000
set.seed(730317945) # Set seed to PID

Y_mat <- matrix(NA, ncol=4, nrow=nrep)
Y.stats <- rep(NA, nrep)

for (ii in 1:nrep)
{
    # Simulate infected cases
  n.temp <- sample.int(15210+15210, 196, replace=FALSE)
  # Number of infected cases from the 15210 people in mRNA-1273 group
  n1 <- sum(n.temp <= 15210)
  # Number of infected cases from the 15210 people in placebo group
  n2 <- sum(n.temp > 15210)

  Y_mat[ii, ] <- c(n1, 15210-n1, n2, 15210-n2)
  # Expected number of infected cases from the 15210 people in mRNA-1273 group
  E_n1 <- 15210*(196/sum(c(n1, 15210-n1, n2, 15210-n2)))
  # Expected number of infected cases from the 15210 people in placebo group
  E_n2 <- 15210*(196/sum(c(n1, 15210-n1, n2, 15210-n2)))

  # Chi-square test-statistics
  Y.stats[ii] <- sum( (c(n1, 15210-n1, n2, 15210-n2) - c(E_n1, 15210-E_n1, E_n2, 15210-E_n2))**2/c(E_n1
}

# Probability of observing stat.obs or larger under H0
```

```
sum(Y.stats > stat.obs)/nrep
```

```
## [1] 0
```

Question 2.1 (2') Using Sample 2, the claim that a woman uses about 20,000 words per day can be tested with a two-tailed test under the null hypothesis H0: mu = 20000. The test yields a p-value of **9.57e-9**, which is significantly below the alpha level of 0.05. We have sufficient evidence to reject the null hypothesis that a woman uses about 20,000 words per day.

```
# Q2.1: one-arm, two-tailed test

# Pr(Y >= 20000-14297 | H0)
pnorm(14297-20000, mean=0, sd=6441/sqrt(42), lower.tail=TRUE) +
  # Pr(Y < -(20000-14297) | H0)
  pnorm(-(14297-20000), mean=0, sd=6441/sqrt(42), lower.tail=FALSE)
```

```
## [1] 9.569527e-09
```

Question 2.2 (2') The claim that women are more talkative than men can be tested using a one-tailed test under the null hypothesis H0: mu1 > mu2. The test yields a p-value of **0.45**, which is above the alpha level of 0.05. We do not have sufficient evidence to reject the null hypothesis that women speak more than men using the data from Sample 2.

```
# Q2.2: two arm, one-tailed, HA Y1 - Y2 > 0

# Pr(Y1 - Y2 >= 14297-14060 | H0)
pnorm(14297-14060, mean=0, sd=sqrt(((6441**2)/42)+((9065**2)/37)), lower.tail=FALSE)
```

```
## [1] 0.4473706
```

```
# Don't need to add Pr(Y1 - Y2 < 14297-14060 | H0) since we are using one-tailed test
# pnorm(-(14297-14060), mean=0, sd=sqrt(((6441**2)/42)+((9065**2)/37)), lower.tail=TRUE)
```