

# 453 HW1

Max Tjen

2023-06-12

## Packages

```
library(binom)
library(PropCIs)
```

## Question 4

### Part A

To use one binomial distribution for the sample, we must assume that each woman's chlamydia status:

- is independent of the other observations
- has one of two possible outcomes (failure or success)
- has the same probability of success (positive result)

### Part B

```
binom.confint(x = 48, n = 750, conf.level = 0.95, methods = "agresti-coull")
```

```
##           method x    n mean      lower      upper
## 1 agresti-coull 48 750 0.064 0.04847048 0.0839731
```

Since our sample size has more than 40 observations, we will use the Agresti-Coull interval. By running the `binom.confint()` function, we get a 95% confidence interval of (0.048, 0.084). With this, we can determine that with 95% confidence, the true prevalence of chlamydia among asymptomatic pregnant women in Rotterdam, Netherlands is between 0.048 and 0.084.

## Question 5

```
table <- array(data = c(27, 17, 4, 702),
               dim = c(2, 2),
               dimnames = list(COBAS_AMPLICOR = c("Positive", "Negative"),
                               MagNA = c("Positive", "Negative")))

table
```

```
##                MagNA
## COBAS_AMPLICOR Positive Negative
##      Positive      27        4
##      Negative     17       702
```

```
# Agresti-Min procedure
```

```
diffpropci.mp(b = table[1,2], c = table[2,1], n = sum(table), conf.level = 0.95)
```

```
##
##
##
## data:
##
## 95 percent confidence interval:
##  0.005125038 0.029449430
## sample estimates:
## [1] 0.01728723
```

```
# McNemar's Test
```

```
mcnemar.test(x = table, correct = FALSE)
```

```
##
## McNemar's Chi-squared test
##
## data:  table
## McNemar's chi-squared = 8.0476, df = 1, p-value = 0.004556
```

The Agresti-Min confidence interval to estimate the difference in probabilities of positive test results is  $0.005 < \pi_{+1} - \pi_{1+} < 0.029$ . The interval is relatively small, with the important aspect being that 0 is not included within the interval. Because of this, along with the p-value of the McNemar's test being 0.005, we can sufficiently conclude that there is a difference in probabilities of positive test results between the COBAS system and MagNA kit.

## Question 6

### Part A

The product insert may give a Clopper-Pearson interval rather than other interval types because they want to show that the interval is exact. This means that the interval will likely be smaller than the other intervals as the others are approximations and don't provide as much accuracy and precision as to the testing results of the product. By providing an exact interval, the results may inspire more confidence among consumers.

### Part B

```
table <- array(data = c(190, 15, 7, 464),
               dim = c(2, 2),
               dimnames = list(True = c("+", "-"),
                                Assay = c("+", "-")))

table
```

```
##      Assay
## True   +   -
##      + 190   7
##      -   15 464
```

```
# sensitivity
binom.confint(x = 190, n = 190 + 7, conf.level = 0.95, methods = "exact")
```

```
## method x n mean lower upper
## 1 exact 190 197 0.964467 0.9281616 0.9855967
```

```
# specificity
binom.confint(x = 464, n = 464 + 15, conf.level = 0.95, methods = "exact")
```

```
## method x n mean lower upper
## 1 exact 464 479 0.9686848 0.9488757 0.9823693
```

For symptomatic males that provided swab specimens for chlamydia testing, the 95% Clopper-Pearson interval for sensitivity is (0.928, 0.986) and for specificity, it is (0.949, 0.982). This means that we can conclude with 95% confidence that the true sensitivity is between 0.928 and 0.986 and that the true specificity is between 0.949 and 0.982.

## Part C

```
# read data
aptima <- read.csv("/Users/mtjen/Desktop/453/Aptima_combo.csv")

result <- data.frame(matrix(ncol = 10))
colnames(result) <- c("Disease", "Gender", "Specimen", "Status",
                     "Sensitivity", "Sensitivity_lower", "Sensitivity_upper",
                     "Specificity", "Specificity_lower", "Specificity_upper")
result <- result[-1,]

for (rowIndex in 1:nrow(aptima)) {
  # combination details/description
  disease <- aptima[rowIndex, "Disease"]
  gender <- aptima[rowIndex, "Gender"]
  specimen <- aptima[rowIndex, "Specimen"]
  status <- aptima[rowIndex, "Symptoms_Status"]

  # get values
  sens_x <- as.numeric(aptima[rowIndex, "True_positive"])
  sens_n <- as.numeric(aptima[rowIndex, "True_positive"]) +
    as.numeric(aptima[rowIndex, "False_negative"])

  spec_x <- as.numeric(aptima[rowIndex, "True_negative"])
  spec_n <- as.numeric(aptima[rowIndex, "True_negative"]) +
    as.numeric(aptima[rowIndex, "False_positive"])

  # get confidence intervals
```

```

sens_int <- binom.confint(x = sens_x,
                          n = sens_n,
                          conf.level = 0.95,
                          methods = "exact")

spec_int <- binom.confint(x = spec_x,
                          n = spec_n,
                          conf.level = 0.95,
                          methods = "exact")

rowVals <- c(disease, gender, specimen, status,
             round(sens_int$mean, 3), round(sens_int$lower, 3),
             round(sens_int$upper, 3), round(spec_int$mean, 3),
             round(spec_int$lower, 3), round(spec_int$upper, 3))

result[nrow(result) + 1,] = rowVals
}

result

```

##	Disease	Gender	Specimen	Status	Sensitivity	Sensitivity_lower
## 1	Chlamydia	Male	Swab	Symptomatic	0.964	0.928
## 2	Chlamydia	Male	Swab	Asymptomatic	0.946	0.867
## 3	Chlamydia	Male	Urine	Symptomatic	0.985	0.957
## 4	Chlamydia	Male	Urine	Asymptomatic	0.963	0.894
## 5	Chlamydia	Female	Swab	Symptomatic	0.924	0.867
## 6	Chlamydia	Female	Swab	Asymptomatic	0.984	0.913
## 7	Chlamydia	Female	Urine	Symptomatic	0.938	0.885
## 8	Chlamydia	Female	Urine	Asymptomatic	0.968	0.888
## 9	Gonorrhea	Male	Swab	Symptomatic	0.99	0.972
## 10	Gonorrhea	Male	Swab	Asymptomatic	1	0.782
## 11	Gonorrhea	Male	Urine	Symptomatic	0.984	0.963
## 12	Gonorrhea	Male	Urine	Asymptomatic	1	0.753
## 13	Gonorrhea	Female	Swab	Symptomatic	1	0.962
## 14	Gonorrhea	Female	Swab	Asymptomatic	0.969	0.838
## 15	Gonorrhea	Female	Urine	Symptomatic	0.926	0.853
## 16	Gonorrhea	Female	Urine	Asymptomatic	0.875	0.71
##	Sensitivity_upper		Specificity	Specificity_lower	Specificity_upper	
## 1	0.986		0.969	0.949	0.982	
## 2	0.985		0.984	0.963	0.995	
## 3	0.997		0.984	0.968	0.993	
## 4	0.992		0.988	0.968	0.997	
## 5	0.961		0.967	0.951	0.979	
## 6	1		0.988	0.974	0.996	
## 7	0.971		0.988	0.977	0.995	
## 8	0.996		0.99	0.977	0.997	
## 9	0.998		0.988	0.972	0.996	
## 10	1		0.967	0.943	0.983	
## 11	0.995		0.998	0.987	1	
## 12	1		0.995	0.981	0.999	
## 13	1		0.981	0.969	0.989	
## 14	0.999		0.996	0.987	1	

## 15	0.97	0.991	0.982	0.996
## 16	0.965	0.995	0.985	0.999

## Question 7

### Part A

```
# 10 degrees Celsius
binom.confint(x = 0, n = 30, conf.level = 0.95, methods = "wilson")
```

```
##      method x  n mean lower      upper
## 1 wilson 0 30    0    0 0.1135134
```

```
# 15 degrees Celsius
binom.confint(x = 1, n = 30, conf.level = 0.95, methods = "wilson")
```

```
##      method x  n      mean      lower      upper
## 1 wilson 1 30 0.03333333 0.00590859 0.1667039
```

```
# 20 degrees Celsius
binom.confint(x = 25, n = 30, conf.level = 0.95, methods = "wilson")
```

```
##      method x  n      mean      lower      upper
## 1 wilson 25 30 0.83333333 0.6643565 0.9266346
```

Because the sample sizes are small for each group, we used a Wilson confidence interval. With this, we got the 95% confidence intervals for the probability that an egg hatches at each temperature is (0, 0.114) for 10 degrees Celsius, (0.006, 0.167) for 15 degrees Celsius, and (0.664, 0.927) for 20 degrees Celsius.

### Part B

With these confidence intervals, we can informally assess whether the probabilities could be the same at each temperature. The intervals for 10 and 15 degrees Celsius overlap and are relatively similar, which means that the true probability of an egg hatching is likely very similar between the two temperature settings. The 20 degree confidence interval doesn't overlap with any of the other intervals and is much higher, showing that it is the best temperature setting of the three in terms of having the best probability of an egg hatching.