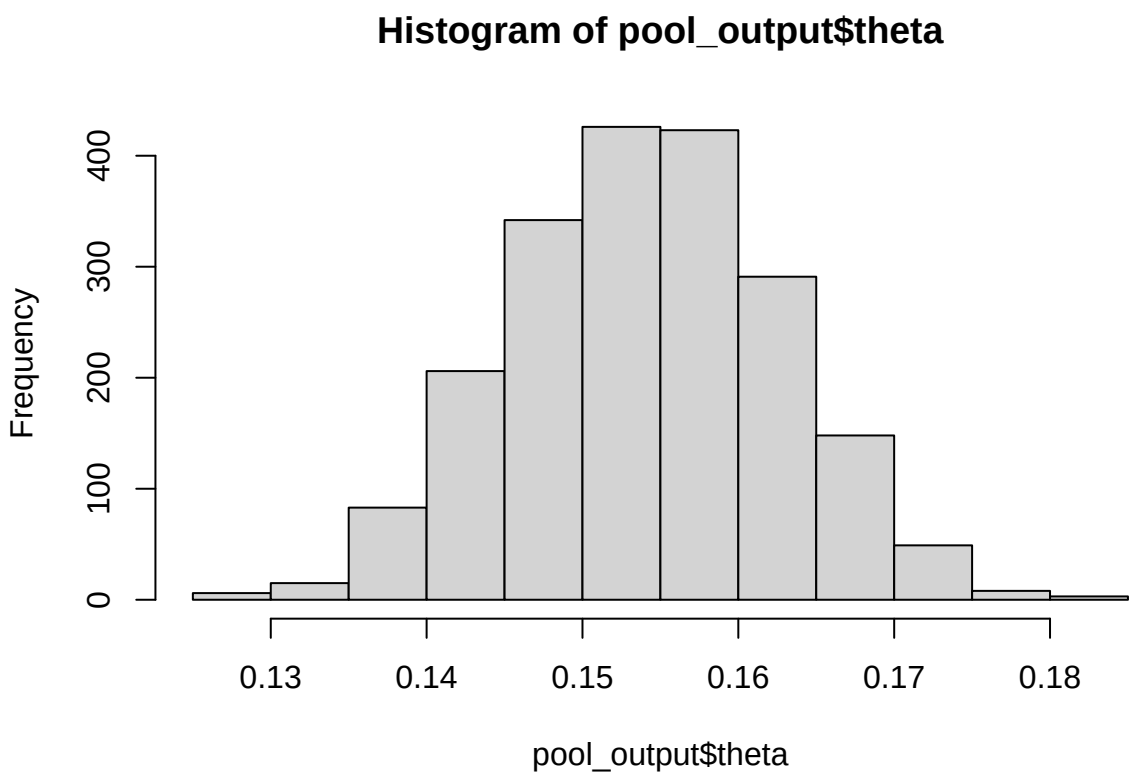
STA 602 Lab

Ryan Tang

12 September, 2022

Exercise 1

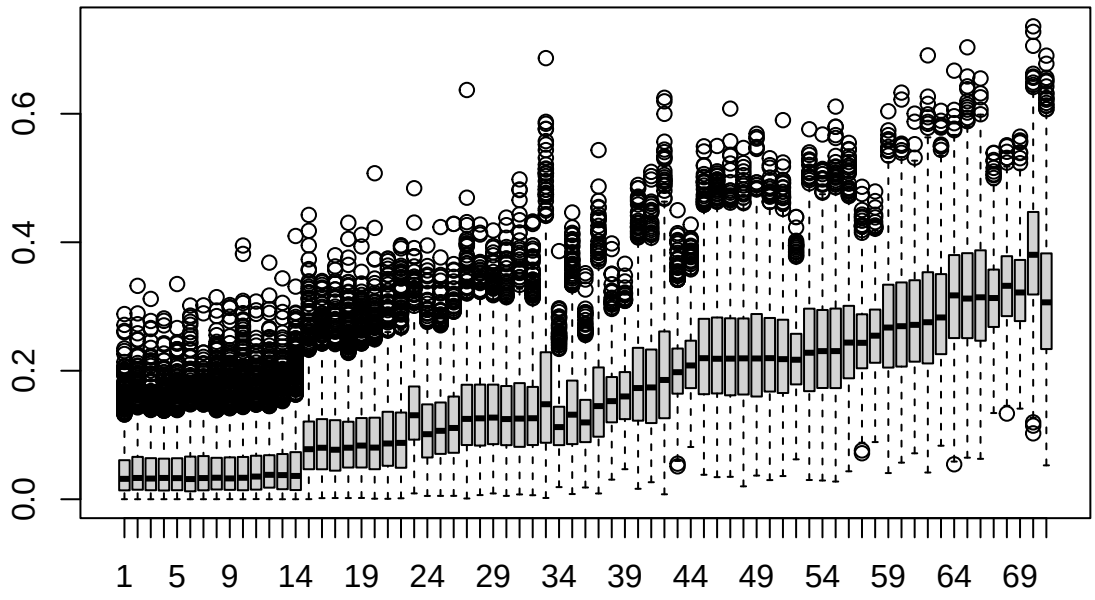
```
hist(pool_output$theta)
```



Exercise 2

So we estimated θ_i for each trial in the sample using the $p(\theta_i) = Beta(1, 1)$ uniform priors and likelihood function $p(y_i|\theta_i) = Binom(N_i, k_i)$, and plotted the posterior distributions $p(\theta_i|y_i)$ for each trial in a Boxplot, one for each trial. The points here are just outliners in terms of boxplot's definition, which is outside the 75% or 25% percentiles.

```
boxplot(nopool_output$theta)
```



Exercise 3

The only pertinent difference between the two STAN models is that `pool` has one parameter and `noPool` has `n` parameters — same as the number of trials. In other words, we assume all trials are iid from the `pool` model but not identically distributed in the `nopool` model.

Exercise 4

We can intuitively see that based on the Binom likelihood, given a trial with N draws, a represents the number of rats with tumor and b represents the number rats do not have tumor.

Exercise 5

From out trail samples, the chance of obtaining tumor is low. And *Beta* distribution's is given by the below equation.

$$E[\theta|Y] = \frac{a}{a+b}$$

Hence, we are consistently seeing samples with a much lower than b.

```
total_rats <- sum(N)
total_a <- sum(y)
total_b <- total_rats - total_a
print(paste("Total rats =", total_rats))
```

```
## [1] "Total rats = 1739"
```

```r
print(paste("Total rats with Tumor = a =", total_a))
```

```
## [1] "Total rats with Tumor = a = 267"
```

```r
print(paste("Total rats without Tumor = b =", total_b))
```

```
## [1] "Total rats without Tumor = b = 1472"
```

Exercise 6

The batch with b set less or equal to 25 did pretty well. The higher it is, the stronger your belief that the tumor is rare represented in terms of sample size. Hence, it needs more data to "convince" your prior belief.

Exercise 7

The MLE results is for the $\hat{\theta}_i = \frac{y_i}{N_i}$ and $\hat{\theta} = \frac{\sum y_i}{\sum N_i}$. They are just the sample means. The differences between approaches 1 and 2 are because we used a Uniform prior that represented us do not hold a strong belief of the true tumor rate (effective sample size 2). Hence, the posteriors are heavily affected by the sample mean instead of the prior.
