Question 2 Assignment 1

Dhanraj Patel

2a

```r
# set the number of simulations -- the number of samples we take
m <- 10000

# size of each sample
n <- 30

t1 <- rep(NA, m)
t2 <- rep(NA, m)
t3 <- rep(NA, m)
SD <- rep(NA, m) # need to calc bin width


for (i in 1:m) {
  my_sample <- rpois(30 , lambda = 2)

  t1[i] <- (length(which(my_sample == 0)))/30
  t2[i] <- exp(-mean((my_sample)))
  t3[i] <- exp(-var((my_sample)))
  SD[i] <- sd(my_sample) # need to calc bin width

}


# calculate the bias
bias_t1 = mean(t1) - exp(-2)
bias_t2 = mean(t2) - exp(-2)
bias_t3 = mean(t3) - exp(-2)
print("the bias for t1 is:")
```

```
## [1] "the bias for t1 is:"
```

```r
bias_t1
```

```
## [1] 0.0004313834
```

```r
print("the bias for t2 is:")
```

```
## [1] "the bias for t2 is:"
```

```r
bias_t2
```

```
## [1] 0.005120881
```

```r
print("the bias for t3 is:")
```

```
## [1] "the bias for t3 is:"
```

```r
bias_t3
```

```
## [1] 0.0224709
```

2b

```r
# calculate the variance
var_t1 = var(t1)
var_t2 = var(t2)
var_t3 = var(t3)

print("the variance for t1 is:")
```

```
## [1] "the variance for t1 is:"
```

```r
var_t1
```

```
## [1] 0.003812905
```

```r
print("the variance for t2 is:")
```

```
## [1] "the variance for t2 is:"
```

```r
var_t2
```

```
## [1] 0.001309401
```

```r
print("the variance for t3 is:")
```

```
## [1] "the variance for t3 is:"
```

```r
var_t3
```

```
## [1] 0.006598075
```

2c

```r
bias_t1
```

```
## [1] 0.0004313834
```

```r
bias_t2
```

```
## [1] 0.005120881
```

```r
bias_t3
```

```
## [1] 0.0224709
```

```r
var_t1
```

```
## [1] 0.003812905
```

```r
var_t2
```

```
## [1] 0.001309401
```

```
var_t3
```

```
## [1] 0.006598075
```

```
# MSE(T) = VAR(T) + bias^2

MSE_T1 <- var_t1 + (bias_t1 ** 2)
MSE_T2 <- var_t2 + (bias_t2 ** 2)
MSE_T3 <- var_t3 + (bias_t3 ** 2)




print("the MSE for t1 is:")
```

```
## [1] "the MSE for t1 is:"
```

```
MSE_T1
```

```
## [1] 0.003813091
```

```
print("the MSE for t2 is:")
```

```
## [1] "the MSE for t2 is:"
```

```
MSE_T2
```

```
## [1] 0.001335624
```

```
print("the MSE for t3 is:")
```

```
## [1] "the MSE for t3 is:"
```

```
MSE_T3
```

```
## [1] 0.007103017
```

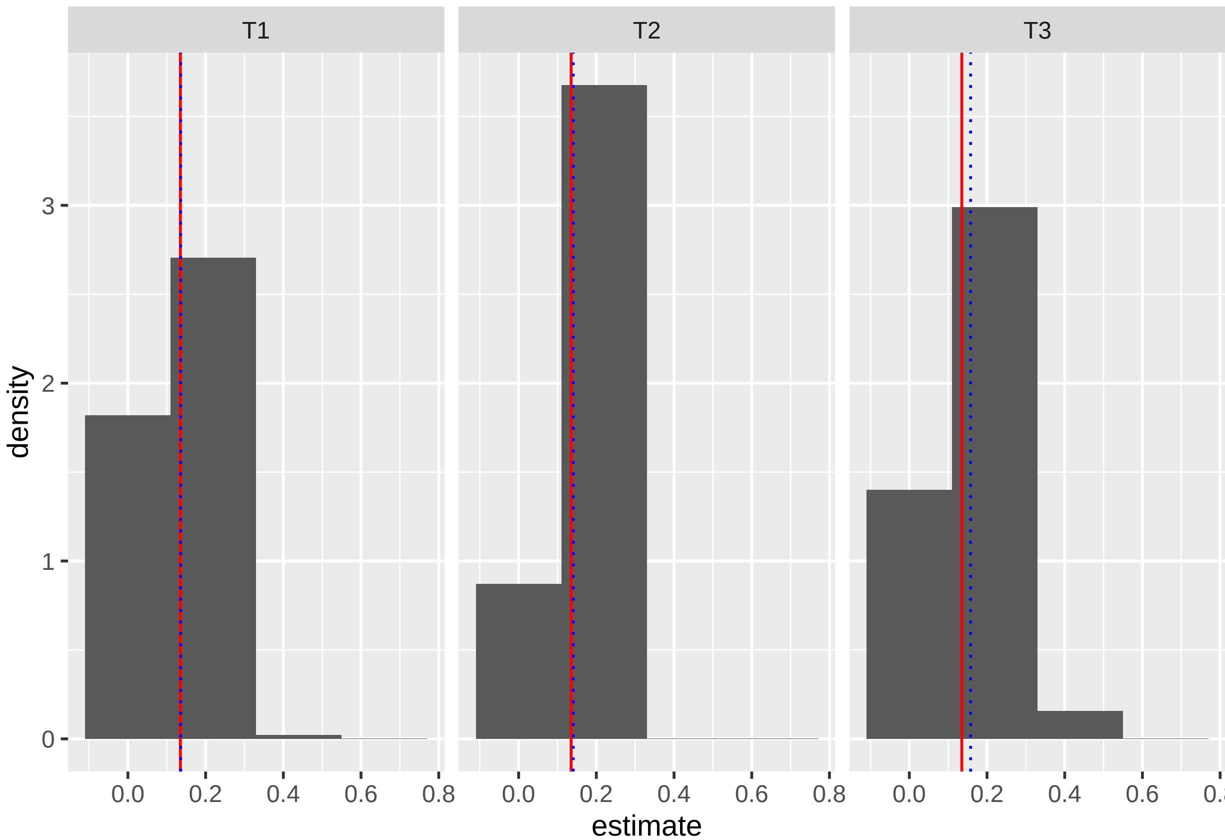
2d

```
# plot the sampling distribution
to_plot <- data.frame(t1 = t1, t2 = t2, t3 = t3)
to_plot <- data.frame(estimate = c(t1, t2, t3), estimator = c(rep("T1", m), rep("T2", m), rep("T3", m))

estimator_group <- to_plot %>%
  group_by(estimator) %>%
  summarize(estimate = mean(estimate))

final_plot <- to_plot %>%
  ggplot(aes(x = estimate)) + geom_histogram(aes(y=..density..), binwidth = .22) +
  geom_vline(xintercept = 0.135, color = "red") +
  facet_wrap(~estimator)+
  geom_vline(data = estimator_group, aes(xintercept = estimate), colour = "blue", linetype = "dotted")

final_plot
```



2e

In order to assess the best estimator you need to look at 2 different factors, the biasedness and the consistency. Biasedness is a measure of how accurate the estimators and consistency is the measure of how precise the estimator is. The first measure to look at is biasedness and since out of all the estimators T1 is the only one that is unbaised, we can conclude that it is the best estimator.

Best estimator = T1

2f

There are some cases where a baised estimate is preferable to an unbaised one. One such case is when you want to minimize the mean square error (MSE). In many cases, a small increase in bais can minimize the variance enough to decrease the MSE value. Other examples of when a biased estimator is preferable to a unbaised one is when Managing risk and Efficient testing.