

Question 1

Jin Barai

```
library(carData)
data(TitanicSurvival)
Titanic = na.omit(TitanicSurvival)
Titanic = Titanic[Titanic$sex == "female", ]
Titanic$survived1 = as.numeric(Titanic$survived == "yes")
head(Titanic)
```

```
##               survived    sex age passengerClass survived1
## Allen, Miss. Elisabeth Walton      yes female   29         1st         1
## Allison, Miss. Helen Loraine       no female    2         1st         0
## Allison, Mrs. Hudson J C (Bessi    no female   25         1st         0
## Andrews, Miss. Kornelia Theodos   yes female   63         1st         1
## Appleton, Mrs. Edward Dale (Cha    yes female   53         1st         1
## Astor, Mrs. John Jacob (Madelei    yes female   18         1st         1
```

```
TitanicSample <- c(57, 349, 147, 21, 310, 293, 62, 96, 304, 31, 258, 5,
381, 295, 180, 354, 346, 245, 88, 1, 383, 188, 333, 216, 351)
```

```
popSize <- function(pop) {nrow(as.data.frame(pop))}
sampSize <- function(samp) {popSize(samp)}

createInclusionProbFn <- function(pop, sampSize) {
  N <- popSize(pop)
  n <- sampSize
  function(u) { n/N }
}

createJointInclusionProbFn <- function(pop, sampSize) {
  N <- popSize(pop)
  n <- sampSize
  function(u,v) {
    ## Note that the answer depends on whether u and v
    ## are the same or different
    if (u == v) {n/N} else {(n * (n-1)) / (N * (N-1))}
  }
}

createHTestimator <- function(pi_u_fn) {
  function(samp, variateFn) {
    Reduce('+',
      Map(function(u) {variateFn(u) / pi_u_fn(u)}, samp),
      init = 0
    )
  }
}
```

```

    }
  }

createHTVarianceEstimator <- function(pop, pi_u_fn, pi_uv_fn) {
  function(samp, variateFn) {
    Reduce('+',
      Map(function(u) {
        pi_u <- pi_u_fn(u)
        y_u <- variateFn(u)
        Reduce('+',
          Map(function(v) {
            pi_v <- pi_u_fn(v)
            pi_uv <- pi_uv_fn(u, v)
            y_v <- variateFn(v)
            Delta_uv <- pi_uv - pi_u * pi_v
            result <- (Delta_uv * y_u * y_v)
            result <- result / (pi_uv * pi_u * pi_v)
            result
          },
            samp),
          init = 0)
        },
      samp
    ),
    init = 0)
  }
}

```

(a)

```

n = 25
N = nrow(Titanic)
inclusionProb <- createInclusionProbFn(1:N, sampSize = n)
inclusionJointProb <- createJointInclusionProbFn(1:N, sampSize = n)

titanicHTestimator <- createHTestimator(inclusionProb)
HTVarianceEstimator <- createHTVarianceEstimator(1:N,
  pi_u_fn = inclusionProb,
  pi_uv_fn = inclusionJointProb)

createVariateFnAvgAge <- function(popData, variate1, N=1, y=NULL) {
  function(u) { popData[u, variate1] / N }
}
TitanicAvgAge <- createVariateFnAvgAge(Titanic, "age", N=N)

```

The HT estimate of the average age is

```
titanicHTestimator(TitanicSample, TitanicAvgAge)
```

```
## [1] 28.47
```

An estimate of the variance or the standard error is (square root of the variance estimate)

```
sqrt(HTVarianceEstimator(TitanicSample, TitanicAvgAge))
```

```
## [1] 3.076505
```

(b)

```
createvariateFnNy <- function(popData, variate1, N=1, y=NULL) {  
  function (u) { (popData[u, variate1] <= y )/N}  
}
```

```
propTitanicAge25 =createvariateFnNy(Titanic, "age", N=N, y=25)
```

```
pair <- round( c( titanicHTestimator(TitanicSample, propTitanicAge25),  
  sqrt(HTVarianceEstimator(TitanicSample, propTitanicAge25)) ), 3)
```

The HT estimate of the proportion of age less than or equal to 25 is

```
pair[1]
```

```
## [1] 0.44
```

The standard error of the above estimate

```
pair[2]
```

```
## [1] 0.098
```

(c)

The HT estimate of the proportion of age less than or equal to 45 is on the left and the standard error is on the right:

```
propTitanicAge45 =createvariateFnNy(Titanic, "age", N=N, y=45)
```

```
round( c( titanicHTestimator(TitanicSample, propTitanicAge45),  
  sqrt(HTVarianceEstimator(TitanicSample, propTitanicAge45)) ), 3)
```

```
## [1] 0.880 0.064
```

(d)

```
yseq = c(0, sort(Titanic$age[TitanicSample]), 80)
```

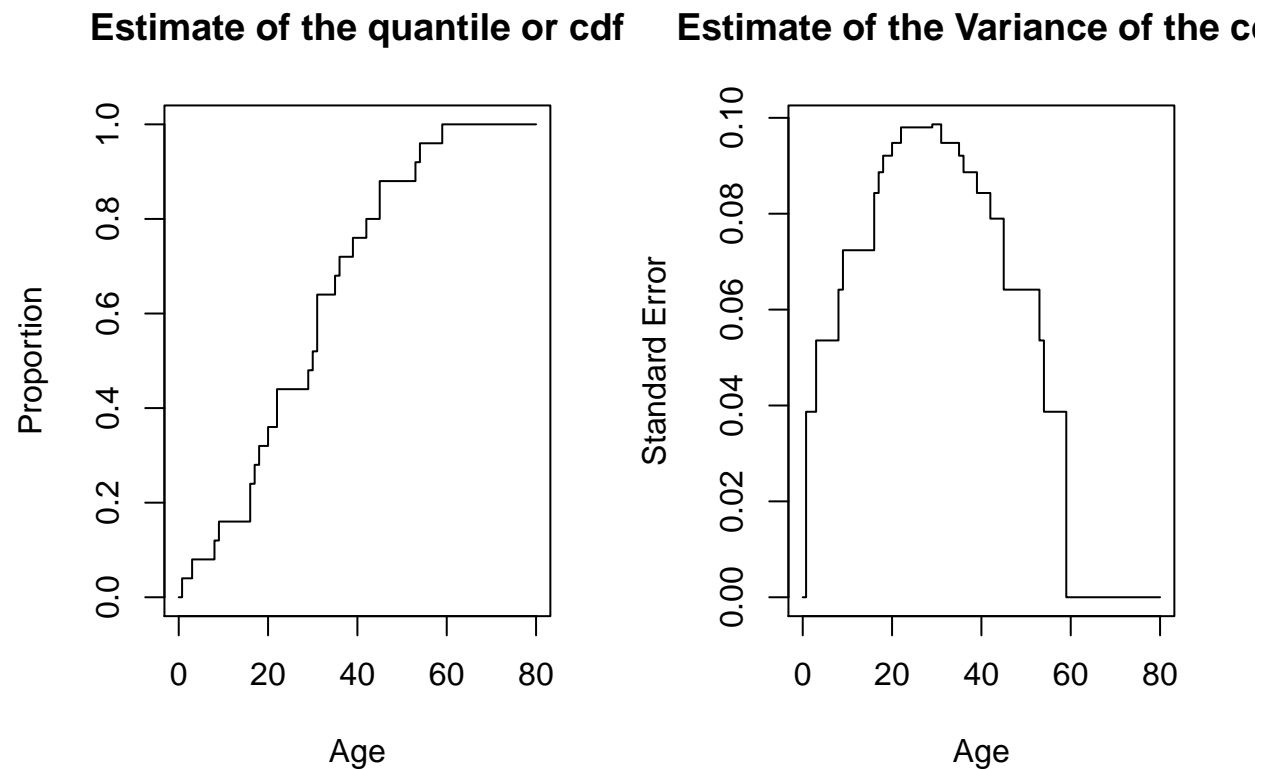
```
cdf.estimate.sd = sapply(yseq, function(y) {  
  propAge <- createvariateFnNy(Titanic, "age", N=N, y=y)  
  
  c( titanicHTestimator(TitanicSample, propAge),  
    sqrt( round(HTVarianceEstimator(TitanicSample, propAge), 14) ) )  
})
```

```

} )

par(mfrow=c(1,2) )
plot(yseq, cdf.estimate.sd[1,], type='s', ylab="Proportion",
     xlab="Age",
     main="Estimate of the quantile or cdf")
plot(yseq, cdf.estimate.sd[2,], type='s', ylab="Standard Error",
     xlab="Age",
     main="Estimate of the Variance of the cdf")

```



(e)

```

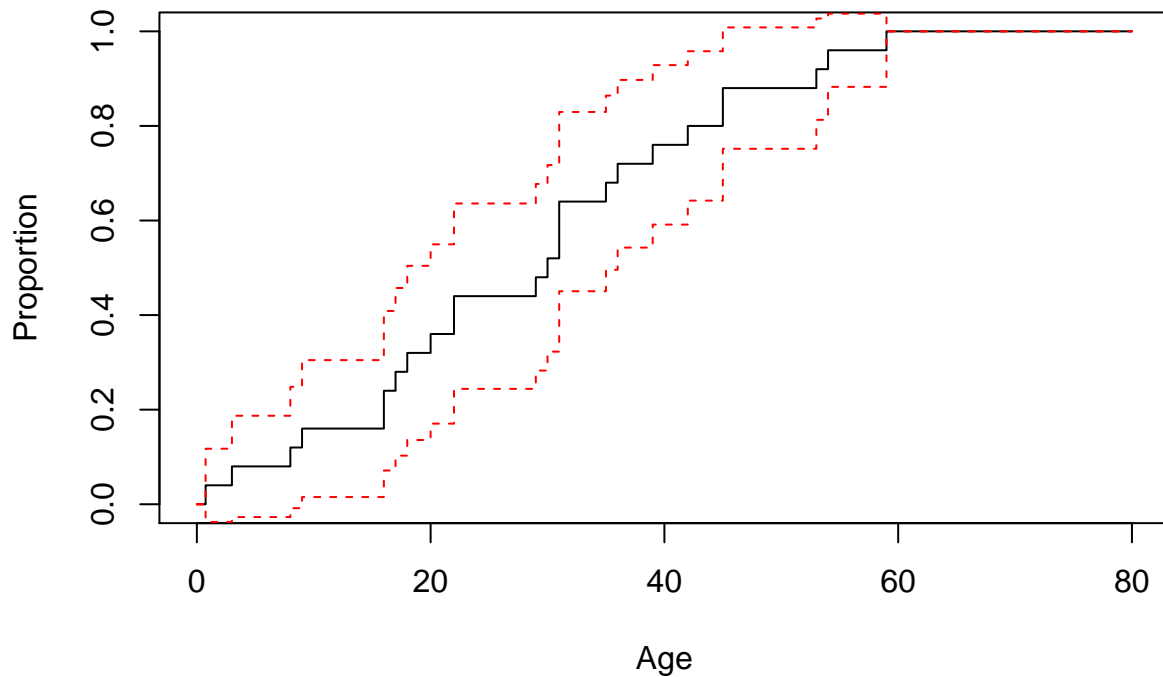
plot(yseq, cdf.estimate.sd[1,], type='s', ylim=c(0,1),
     ylab="Proportion", xlab="Age",
     main="Estimate of the quantile or cdf")

cdf.lower = cdf.estimate.sd[1,] - 2*cdf.estimate.sd[2,]
cdf.upper = cdf.estimate.sd[1,] + 2*cdf.estimate.sd[2,]

lines(yseq, cdf.lower, type='s', col=2, lty=2)
lines(yseq, cdf.upper, type='s', col=2, lty=2)

```

Estimate of the quantile or cdf



(f)

The HT estimate of the proportion of females who survived is on the left and standard error is on the right:

```
createvariateFnNyBool <- function(popData, variate1, N=1, y=NULL) {
  function (u) { (popData[u, variate1] == y )/N}
}

propTitanicSurvived <- createvariateFnNyBool(Titanic, variate1="survived1", y=1, N=N)
round( c( titanicHTestimator(TitanicSample, propTitanicSurvived),
  sqrt(HTVarianceEstimator(TitanicSample, propTitanicSurvived)) ), 3)
```

```
## [1] 0.680 0.092
```

(g)

The HT estimate of the proportion of females in 1st class is on the left and standard error is on the right:

```
propTitanicFirstClass<- createvariateFnNyBool(Titanic, variate1 = "passengerClass",y="1st", N=N)
round( c( titanicHTestimator(TitanicSample, propTitanicFirstClass),
  sqrt(HTVarianceEstimator(TitanicSample, propTitanicFirstClass)) ), 3)
```

```
## [1] 0.320 0.092
```