Name: ___________________________________

Computing in Statistics, STAT:5400
Midterm 2, Fall 2018

You must work in the Linux environment. Submit your answers in the ICON drop box as an .Rnw file and .pdf file produced using Sweave, with your name as author. If you can't get your .Rnw file to compile, submit it anyway and include your R output in a separate text file.

In your document, have a named section for each problem, and, where needed, a numbered list of answers to multipart questions. You don't have to type any other text except where needed to answer a question.

1 Simulation study

Carry out two simulation studies to estimate the bias of the sample .95 quantile as an estimator of the population .95 quantile when the population distribution is poisson with parameter 6. (Note that the R `quantile` function computes sample quantiles, and the `qpois` function produces the true theoretical quantile.) Do one simulation study for each of two sample sizes – $n = 10$ and $n = 50$. In both cases, set your number of replicate datasets S such that the standard error of your estimate of bias will be no larger than 0.01.

Time how long it takes to run each simulation study.

Submit:

- Your R code.
- Your R output containing the results of the simulation studies.
- R output showing how long it took to run each simulation study.
- A brief paragraph interpreting the results.

```
> # find out the true populaton value
>
> trueq <- qpois(0.95,6)
> # function to carry out simulation study
>
> simstudy <- function(S,n, truth)
+ {
+         mydat <- matrix(rpois(S * n, 6), nrow = S)
+         sampquants <- apply(mydat,1,quantile, 0.95)
+         variance = var(sampquants)
```

```
+         list( bias = mean(sampquants) - truth, variance = variance,
+              se = sqrt(variance/S) )
+ }
> set.seed(24)
> # pilot study to estimate standard error
>
> S <- 500
> n <- 10
> pilot.out <- simstudy(S,n,trueq)
> S <- ceiling(pilot.out$variance / 0.01^2)
> # carry out the sim studies
>
> system.time(n10out <- simstudy(S, n, trueq))

   user  system elapsed
  1.310   0.003   1.317

> n <- 50
> system.time(n50out <- simstudy(S, n, trueq))

   user  system elapsed
  1.395   0.002   1.403

> results <- rbind(cbind(n10out$bias, n50out$bias), cbind(n10out$se, n50out$se))
> colnames(results) <- cbind("n = 10", "n = 50")
> rownames(results) <- cbind("bias", "standard error")
> print(results)

                   n = 10      n = 50
bias           -0.6965284 0.024727578
standard error  0.0095673 0.005610331
```

The number of replicate datasets required to keep the standard error below 0.01 was 21474. The bias of the sample quantile was substantial and negative for sample size $n = 10$ at -0.6965. When the sample size was $n = 50$, the bias was positive and small at 0.02473, but still significantly different from zero since the standard error was only 0.00561.

2 The bootstrap

Here is a dataset of size 10.

```
4 8 5 5 6 6 3 5 6 7
```

2.1 Nonparametric bootstrap

Now perform a nonparametric bootstrap analysis using this dataset to assess the bias in the .95 sample quantile as an estimator of the .95 quantile of the population from which this sample was drawn.

Sutmit:

- Your R code.
- Your R output containing the numeric results of the nonparametric bootstrap

```
> library(boot)
> mydat <- c(4, 8, 5, 5, 6, 6, 3, 5, 6, 7)
> myquantfunc <- function(x,ind)
+         quantile(x[ind], 0.95)
> bootout <- boot(mydat, myquantfunc, 3000)
> print(bootout)

ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = mydat, statistic = myquantfunc, R = 3000)


Bootstrap Statistics :
    original     bias    std. error
t1*     7.55 -0.3130333   0.6555342

> bias <- mean(bootout$t) - bootout$t0
> print(paste("Estimated bias is", bias))

[1] "Estimated bias is -0.313033333333333"
```

2.2 Parametric bootstrap

Now perform a parametric bootstrap analysis for the same purpose. Assume that the population is poisson.

Submit:

- Your R code.
- Your R output containing the numeric results of the parametric bootstrap

```
> myrandgen <- function(x,d)
+    rpois(length(x), d$parm)
> myfunc2 <- function(x) quantile(x,0.95)
> bootout2 <- boot(mydat,myfunc2, 3000, sim="parametric", ran.gen = myrandgen,
+                  mle = list(parm = mean(mydat) ))
> print(bootout2)

PARAMETRIC BOOTSTRAP


Call:
boot(data = mydat, statistic = myfunc2, R = 3000, sim = "parametric",
    ran.gen = myrandgen, mle = list(parm = mean(mydat)))


Bootstrap Statistics :
    original  bias    std. error
t1*     7.55 1.14215    1.366179

> bias2 <- mean(bootout2$t) - bootout2$t0
> print(paste("Estimated bias is", bias2))

[1] "Estimated bias is 1.14215"
```