

Name: _____

Computing in Statistics, 22S:166
Midterm 2, Fall 2014

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. You don't have to type any other text except where needed to answer a question.

1 R and Bootstrap

1. Use an appropriate R function to compute the theoretical 0.25 quantile of the exponential distribution with parameter $\frac{1}{6}$. Show your code and output.
2. Carry out the following steps in R.
 - (a) Generate a vector of 20 random values from the exponential density with parameter $\frac{1}{6}$.
 - (b) With the `boot` package, use the nonparametric bootstrap with 1000 bootstrap replications to obtain a bootstrap "unbiased estimate" of the 0.25 quantile of the population from which your sample was drawn. Your statistic should be the sample 0.25 quantile. Your code should calculate the "unbiased estimate" directly, not as a separate step after the output from the `boot` function has been displayed. Show your code and output.

2 Simulation studies

Now carry out a simulation study to determine whether the bootstrap unbiased estimate of the 0.25 quantile is truly unbiased when the population of interest is exponential with parameter $\frac{1}{6}$ and the data is a sample of size 20. Use 1000 bootstrap samples for each bootstrap run. Develop and test your code using a very small number S of replicate datasets, because it is time-consuming to run bootstrap on a large number of datasets. Do your final run with $S = 500$ replicate datasets.

Your code should display:

- The amount of time required to run the simulation study
- The estimate of the bias in the bootstrap unbiased estimates
- The standard error of your estimate of bias

3 SAS

You do not need to run SAS for this problem. Just answer the questions below.

The datastep below was used to read in a dataset.

```
data cells ;
infile '/mnt/nfs/netapp1/homepage/kcowles/Datasets/cell.dat' ;
input type $ diam ;
run ;
```

Additional SAS code produced the following table:

	Diameter		
	Mean	StdDev	N
Cell type			
lymph	6.95	1.60	40
tumor	17.92	2.97	50
All	13.04	6.00	90

1. The name of the SAS procedure used to create the table is _____.
2. Which `table` statement would have been needed in the procedure to produce the table?
 - (a) `table type, diam * (mean stddev n*f=3.0) ;`
 - (b) `table type all, diam * (mean stddev n)*f=3.0 ;`
 - (c) `table type all, diam * (mean stddev n) ;`
 - (d) `table diam * (mean stddev n*f=3.0) , type all ;`
 - (e) none of the above
3. In the table, the phrases `Diameter` and `Cell type` appear instead of the variable names `diam` and `cell`. Which kind of statement accomplished this?
 - (a) `format`
 - (b) `keylabel`
 - (c) `label`