Name: ___________________________________

Computing in Statistics, STAT:5400
Midterm 2, Fall 2017

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

A large-sample, normal-theory confidence interval for a population proportion can be computed as follows:

$$\begin{aligned} \hat{p} &= \frac{y}{n} \\ se(\hat{p}) &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ c.i. &= \hat{p} \pm z^* se(\hat{p}) \end{aligned}$$

where n is the sample size, y is the number of successes in the sample, and z^* is a standard normal quantile determined by the confidence level desired.

The plus-four method of confidence interval is calculated in the same way, after adding 2 to the number of successes and 2 to the number of failures in the sample. That is:

$$\begin{aligned} \tilde{p} &= \frac{y+2}{n+4} \\ se(\tilde{p}) &= \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \\ c.i. &= \tilde{p} \pm z^* se(\tilde{p}) \end{aligned}$$

The `prop.test` function, with the Yates correction argument set to FALSE, computes large-sample normal-theory confidence intervals. With appropriate adjustment to the to the numeric arguments, it can also compute plus-four method intervals.

Conduct a simulation study to compare the true coverage probabilities of large-sample normal-theory intervals versus plus-four method intervals, when

- the nominal coverage is 90%
- the sample size is 25
- the true population proportion is 0.16.

Use a large enough number of replicate datasets so that the standard error of your estimated coverage probabilities will be no greater than 0.0025.

Time how long it takes to run your simulation study.

Submit:

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

2 Root-finding

The function

$$f(x) = x^4 - 2x^3 + 1$$

has two real roots in the interval [0.5, 2.0] .

Use the `uniroot` function in such a way that you are able to find both roots.

Submit:

- Your R code.
- R output containing results for each root.
- Any other R output that helped you do this.

3 Relational database structures

A journal editor wishes to store information about the papers published in her journal and the authors of those papers. Below are the attributes that she wishes to store. Develop a relational database structure, in third normal form, for these data. For each table, indicate any primary and/or foreign keys. Note that the same person could be an author of more than one paper.

```
Paper title
Paper volume number
Paper issue number
Paper starting page number
Paper author name 1
Paper author affiliation 1
Paper author email 1
Paper author name 2
Paper author affiliation 2
Paper author email 2
Paper author name 3
Paper author affiliation 3
Paper author email 3
Paper author name 4
Paper author affiliation 4
Paper author email4
Paper author name 5
Paper author affiliation 5
Paper author email 5
```