Suppose that, as in lecture, we have a sample of n observations generated from the model

$$\begin{aligned} X_i &\overset{i.i.d.}{\sim} N(0,1), \\ u_i &\overset{i.i.d.}{\sim} N(0,1), \text{ independent of the } X_i, \\ Y_i \mid X_i, u_i &= \beta_0 + \beta_1 X_i + \epsilon_i, \text{ where} \\ \epsilon_i &= |X_i|\, u_i. \end{aligned}$$

In this assignment we will compare model-based, robust ("sandwich"), and bootstrap standard errors for estimates $\hat{\beta}_1$ of β_1 from linear regression models. We set $\beta_0 = 1$ and $\beta_1 = 2$.

1. For $n \in \{10, 100, 1000\}$, conduct a simulation that compares the three different standard error estimates (model-based, sandwich, bootstrap) of the true standard error of $\hat{\beta}$. In each simulation, you will need to replicate the following steps $B = 5000$ times:

   (a) Generate a sample of observations (X_i, Y_i) of size n according to the model above

   (b) Fit a linear regression model to the observed data and record the estimate $\hat{\beta}_1$

   (c) Compute model-based and robust standard errors for $\hat{\beta}_1$

   (d) Compute a bootstrap standard error for $\hat{\beta}_1$:

   - Draw 1000 bootstrap samples of size n from the observed sample (X, Y)
   - Compute $\hat{\beta}_1$ for each of these 1000 samples
   - Compute the standard deviation of these bootstrapped coefficient estimates

   You can use the function `doOne()` in the file `se_ex1.R` as a starting point for your simulations. You may do this on your own machine or on `cox`. Write your code so that the results are reproducible.

2. Conduct simulations using 3 additional values of n, of your choosing, this time using batch submission of jobs on `bayes`. Split each simulation into either 5 or 10 jobs (not 1 job and not 5000 individual jobs). Perform this batch submission either using a loop in your shell script, or by using a job array. **NB**: If `bayes` is full, `gosset` has an identical setup (i.e., use `qsub` to submit jobs) and is restricted to student use only. This is an older (i.e., slower) cluster, but often has available cores.

3. Present your results graphically and/or tabularly in a way that best illustrates your findings. Comment on what you see. Attach your code and any scripts in an Appendix.