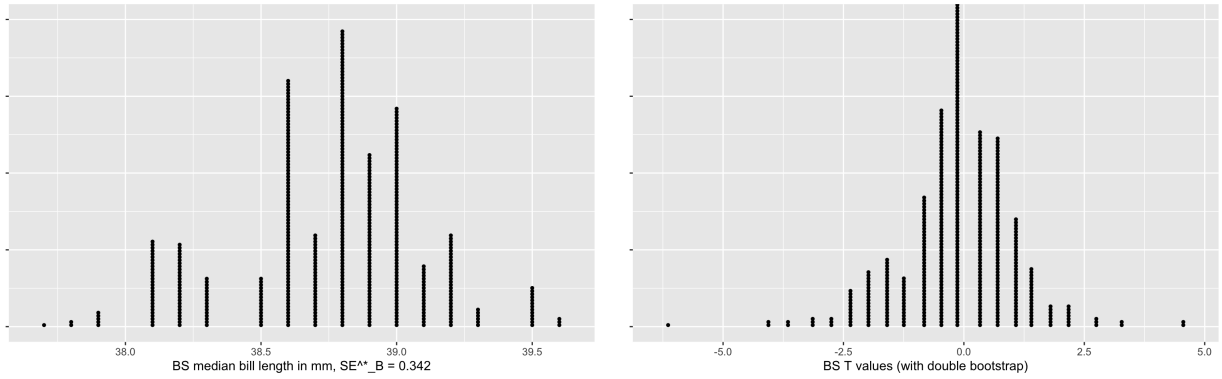
Math 152, Fall 2022
Jo Hardin
WU # 13
Tuesday 10/25/22

Your name: _____

Names of people you worked with: _____

Task: Consider a dataset describing 151 Adelie penguins collected at Palmer Station, Antarctica LTER. The dataset has been bootstrapped 500 times, and each time the median bill length (in mm) has been computed. Additionally, a separate SE was computed for each of the 500 BS samples (by performing a double bootstrap with 50 resamples each). Below are plots for both the distribution of the bootstrapped median as well as for the bootstrapped median normalized by subtracting the observed median (38.8 mm) and dividing by the individual bootstrapped SEs ($\hat{SE}^*(b)$).



Find the following three intervals:

1. 90% BS SE confidence interval for θ (You might need to know that `qnorm(.95, 0, 1) = 1.645`.)
2. 90% BS-t confidence interval for θ
3. 90% BS percentile confidence interval for θ

Solution:

1. $$38.8 \pm 1.645 \cdot 0.342$$

   A 90% BS SE interval would be from 38.24 mm to 39.36 mm.

2. The quantiles are given by the T distribution. They are (approximately):

   $$\hat{q}^*_{0.05} = -2.11 \qquad \hat{q}^*_{0.95} = 1.51$$

   Resulting in an interval given by:

   $$(\hat{\theta} - \hat{q}^*_{0.95}\hat{SE}^*_B, \hat{\theta} - \hat{q}^*_{0.05}\hat{SE}^*_B)$$

   A 90% BS-t interval would be from 38.28 mm to 39.02 mm.

3. A 90% percentile interval would be from 38.1 mm to 39.2 mm. Interval calculated by determining the bounds on roughly 90% of the bootstrapped medians.

Which is better? Which is "right"? We can't know for sure. The intervals are not substantially different, so they are probably all reasonably correct or close to correct (where *correct* means that the process captures θ in 90% of datasets).

That said, the bootstrap distribution of the medians does not seem particularly Normal, so we probably aren't going to report the first interval. We may want to think about whether a transformation of the median might lead to a normal bootstrap sampling distribution (in order to use the percentile interval). For now, the BS-t interval seems safest.