1 Rats

As a small example, the survival times of 9 rats were 10, 27, 30, 40, 46, 51, 52, 104, and 146 days. Because of the skewness in the data, consider estimating the population median survival time θ through the sample median.

Let $n = 9$ and $\tilde{\theta}$ be the sample median.

1.1 bootstrap-t method: median

Compute a 95% CI for θ using the bootstrap-t method. Use $B_1 = 1000$ first-level bootstrap samples and $B_2 = 50$ second level bootstrap samples (to estimate the standard error). Interpret the CI.

We are 95% confident that the true value of the median is between 20.46346 and 78.24754.

1.2 bootstrap percentile CI: median

Compute a 95% CI for θ using the bootstrap percentile CI with B = 1000 bootstrap samples. Interpret the CI.

We are 95% confident that the true value of the median is between 27 and 53.3.

1.3 bootstrap percentile CI: mean

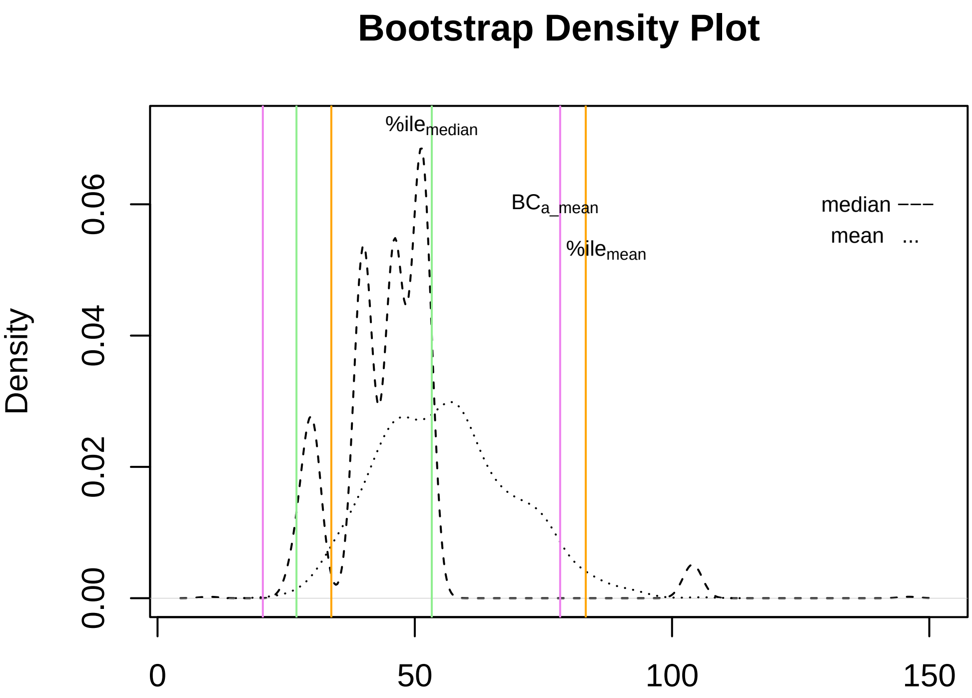
Compute a 95% confidence interval for the mean time between failures θ using the basic bootstrap method with B = 1000 bootstrap samples. Interpret the CI.

We are 95% confident that the true value of the mean is between 33.775 and 83.22778.

1.4 density estimate

Plot a density estimate of the data. In R, you can do this through the density function. Compare the results in parts (a), (b), and (c). If there is a difference in the results, does it have to do with the shape of the data?

First, the percentile CI's focuse on areas in which the data has a higher concentration. Because the data is skewed to the right and median is robust to outliers, percentile CI based on it can be seen to the left of the percentile CI base on the mean. The mean is pulled to the right due to some extremely high values (104, 146). It could be observed that the percentile CI's are narrower than that of the BC_a and that the BC_a appears to consider areas of large concentration from both the mean and the median percentile CI's.



2 Spatial test

Consider the spatial test data from Table 14.1 of Efron and Tibshirani (1993) shown below. From the table's description, it is clear that the measurements A and B are paired. Suppose the data consist of a random sample from an unknown joint distribution of A and B. Whenever ratios are scientifically or statistically preferred to differences, we gain stability by considering the logarithm of the ratios. Let $\theta_1 = \log E\left(\frac{A_i}{B_i}\right)$, $\theta_2 = E\left(\log \frac{A_i}{B_i}\right)$ for all i. Exclude observation #14 because the logarithm of its ratio is undefined. Use 2000 bootstrap samples.

```
[1] "The file exists; loading ../../../problems/ps_02/datasets/spatial_test_data.RData"

    CODE FILENAME: ../R/s01_i01_load_data.R

    data <- "../../../problems/ps_02/datasets/spatial_test_data.RData"

    if (file.exists(data)) {
      print(paste(c("The file exists; loading", data), collapse = ' '))
      load(data)
    } else {
      paste(c("The file does not exist; creating, loading and saving", data),
            collapse = ' ')
      spatial_test_data <- data.frame(
        'i' = 1:25,
        'A' = c(48, 36, 20, 29, 42, 42, 20, 42, 22, 41, 45, 14, 6,
                33, 28, 34, 4, 32, 24, 47, 41, 24, 26, 30, 41),
        'B' = c(42, 33, 16, 39, 38, 36, 15, 33, 20, 43, 34, 22, 7,
                34, 29, 41, 13, 38, 25, 27, 41, 28, 14, 28, 40)
```

```
    )
    n1 <- dim(spatial_test_data)[1]
    seed <- 7

    save(spatial_test_data, seed, n1, file=data)
  }

  rm(data)
```

2.1 Bootstrap percentile CI for θ_1

Compute a bootstrap percentile confidence interval for θ_1. Interpret the CI.

```
     2.5%     97.5%
-0.068671  0.164869
```

2.2 BC_a CI for θ_1

Compute a BCa confidence interval for θ_1. Interpret the CI.

```
1.677644% 91.46218%
 -0.07823   0.12771
```

-0.06867057 0.16486927

2.3 BC_a CI for θ_2

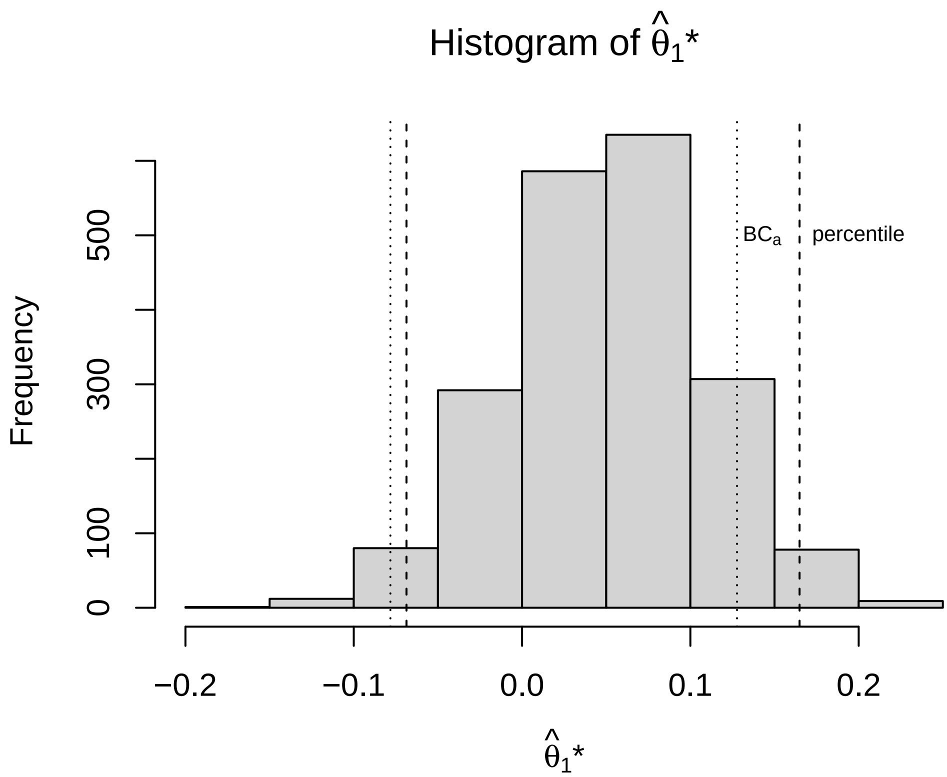
Compute a BC_a confidence interval for θ_2. Interpret the CI.

```
0.00004380274%      47.78721%
    -0.2629559      0.0019218
```

2.4 Bootstrap percentile vs BC_a CI for θ_1

Compare your CIs in 2.1 and 2.2. How different are the two CIs?

The CI based on the BC_a is narrower than that of the bootstrap percentile CI. Also, the endpoints of the former is shifted to the left.



3 References

4 Appendix

4.1 Code to read data for items 1 & 2

```
CODE FILENAME: ../R/s01_i01_load_data.R

data <- "../../../problems/ps_02/datasets/spatial_test_data.RData"

if (file.exists(data)) {
  print(paste(c("The file exists; loading", data), collapse = ' '))
  load(data)
} else {
  paste(c("The file does not exist; creating, loading and saving", data),
        collapse = ' ')
  spatial_test_data <- data.frame(
    'i' = 1:25,
    'A' = c(48, 36, 20, 29, 42, 42, 20, 42, 22, 41, 45, 14, 6,
            33, 28, 34, 4, 32, 24, 47, 41, 24, 26, 30, 41),
    'B' = c(42, 33, 16, 39, 38, 36, 15, 33, 20, 43, 34, 22, 7,
            34, 29, 41, 13, 38, 25, 27, 41, 28, 14, 28, 40)
  )
  n1 <- dim(spatial_test_data)[1]
  seed <- 7
```

```
  save(spatial_test_data, seed, n1, file=data)
}

rm(data)
```

4.2 Code to read data for item 3

```
CODE FILENAME: ../R/s01_i01_load_data.R

data <- "../../../problems/ps_02/datasets/spatial_test_data.RData"

if (file.exists(data)) {
  print(paste(c("The file exists; loading", data), collapse = ' '))
  load(data)
} else {
  paste(c("The file does not exist; creating, loading and saving", data),
        collapse = ' ')
  spatial_test_data <- data.frame(
    'i' = 1:25,
    'A' = c(48, 36, 20, 29, 42, 42, 20, 42, 22, 41, 45, 14, 6,
            33, 28, 34, 4, 32, 24, 47, 41, 24, 26, 30, 41),
    'B' = c(42, 33, 16, 39, 38, 36, 15, 33, 20, 43, 34, 22, 7,
            34, 29, 41, 13, 38, 25, 27, 41, 28, 14, 28, 40)
  )
  n1 <- dim(spatial_test_data)[1]
  seed <- 7

  save(spatial_test_data, seed, n1, file=data)
}

rm(data)
```