Problem Section 6

KEY

Monday Feb 12 2024

Learning Outcomes

The problems are designed to build conceptual understanding and problem-solving skills. The emphasis is on learning to find, evaluate and build confidence. The specific tasks include:

- Implement a non-parametric bootstrap scheme

- Calculate bootstrap confidence intervals

- Back up and support work with relevant explanations

Exercises

1. (Non-parametric bootstrap) Consider the following set of 20 numbers which were independently drawn from some distribution with mean μ_0 and variance σ_0^2.

```
obs_df <- tibble(
  x = c(3.56, 0.69, 0.10, 1.84, 3.93,
        1.25, 0.18, 1.13, 0.27, 1.21,
        0.50, 0.67, 0.01, 0.61, 0.82,
        1.70, 0.39, 0.11, 1.20, 0.72)
)
```

In this problem we are interested in estimating σ_0^2 using the method of moments estimator

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

a. Calculate the observed value of the estimator $\hat{\sigma}_0^2$ for this sample and save it in a variable called `obs_sigma2hat`. Report its value using inline code rounded to four digits.

```
#calculate obs_sigma2hat for the observed sample
nsamp <- nrow(obs_df)

obs_sigma2hat <- (nsamp-1)*var(obs_df$x)/nsamp
```

The observed value of the estimator $\hat{\sigma}_0^2$ is 1.0668.

Now we wish to construct the bootstrapped sampling distribution of $\hat{\sigma}_0^2$ using non-parametric bootstrap to resample from the original data.

b. Construct the bootstrapped sampling distribution of $\hat{\sigma}_0^2$. Follow these steps.

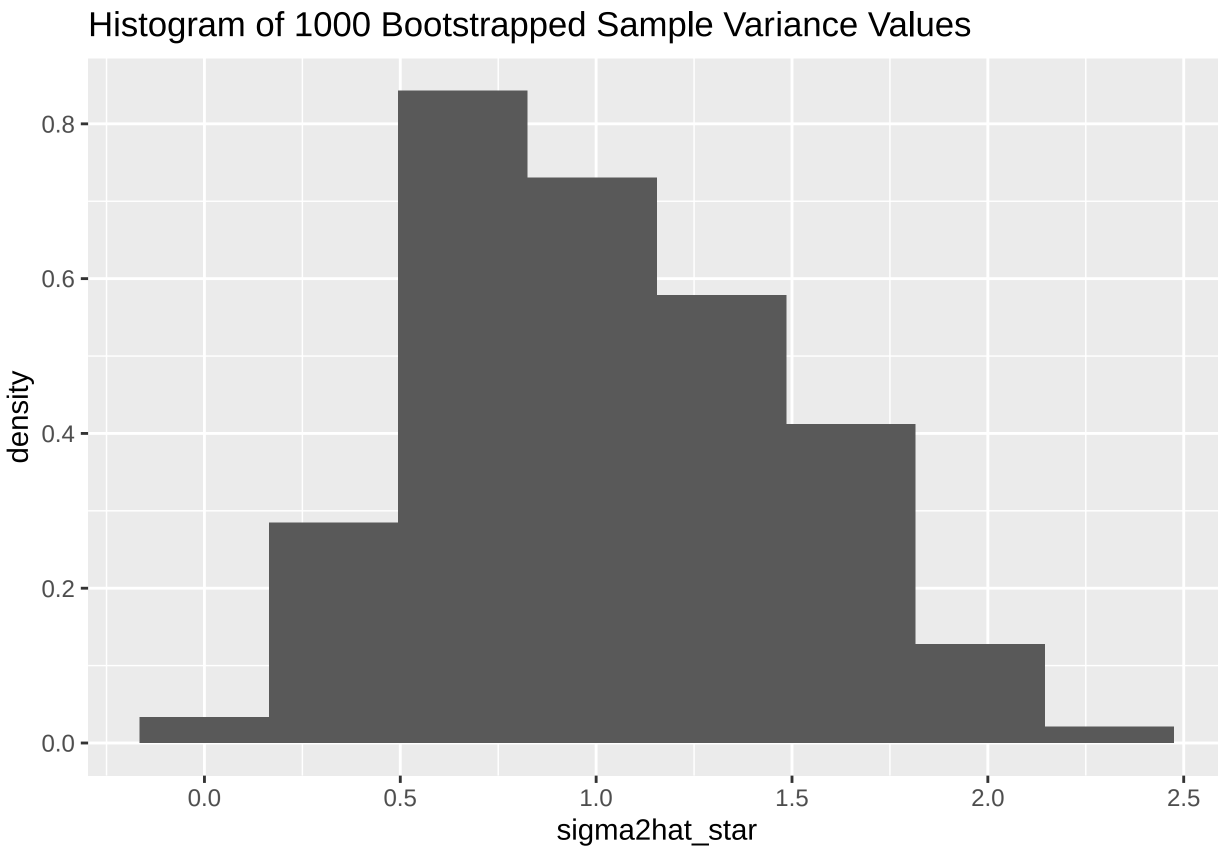
- Take B random samples of size 20 each from the observed sample data with replacement and calculate the value of the bootstrapped estimate - call it sigma2hat_star - for each resample. Save the estimates in a data frame called `boot_df`.

- Make a histogram of the bootstrapped estimates. Examine the shape. Calculate the mean and standard deviation of the bootstrapped estimates
- If the bootstrap results indicate evidence of bias in the estimator, adjust your point estimate from part a accordingly to calculate a bias-corrected version
- Calculate a 90% bootstrap confidence interval. You need to decide whether to use the standard method or the percentile based method.
- Report your point estimate and confidence interval in a complete sentence using inline code.

```
set.seed(1415151)
B = 1000       #number of replications

#1. Take B random samples of size 20 each from your data with replacement
#and calculate the value of sigma2hat for the resample.
#2. Make a histogram of the bootstrapped values of s2.
#3. Examine the shape, calculate the mean and standard deviation of the bootstrapped values.
#4. Calculate a 95% bootstrap confidence interval for sigma2 using the standard bootstrap...
#...confidence interval and also the simple percentile method

#Step 1:

boot_df <- tibble(
  #calculate variance of each resample created using replicate
  sigma2hat_star = replicate(n = B,
                             expr = var(sample(x = obs_df$x,
                                               size = nsamp,
                                        replace=TRUE))*(nsamp-1)/nsamp) )

#Our DF will have a total of B variances. Look at the first five rows
boot_df %>% slice_head(n=5)
```

```
## # A tibble: 5 x 1
##   sigma2hat_star
##            <dbl>
## 1     0.80763275
## 2     1.290506
## 3     0.214096
## 4     1.8197988
## 5     1.055916
```

```
#Step 2: Create a histogram

ggplot(data=boot_df,
       mapping = aes(x=sigma2hat_star,
                     y=after_stat(density)) ) +
  geom_histogram(binwidth = .33)+
  labs(title="Histogram of 1000 Bootstrapped Sample Variance Values")
```




```r
#Step 3: Summarise the bootstrap samples

boot_df_summary <- boot_df %>%
                        summarize(bootstrap_mean = mean(sigma2hat_star),
                                  bootstrap_se = sd(sigma2hat_star))
boot_df_summary
```

```
## # A tibble: 1 x 2
##   bootstrap_mean bootstrap_se
##            <dbl>        <dbl>
## 1      1.0294272   0.45588455
```

The bootstrapped sampling distribution of the estimator is skewed to the right. We see that the average of the bootstrap estimates, 1.0294, is slightly lower than the sample value we saw before from the raw sample.

We could do a bias correction here to our observed estimate. This is shown below

```r
bias <- boot_df_summary$bootstrap_mean - obs_sigma2hat

bias_corrected_estimate <- obs_sigma2hat - bias      #bias corrected estimator of sigma20
```

The bias-corrected estimate of σ_0^2 is 1.1041.

To calculate a confidence interval we can consider two different methods. In the first method we will construct the **standard bootstrap** confidence interval. That is we will have that our 95% CI will be:

$$[(\hat{\sigma}_0^2 - bias) - 1.96 * se_{bootstrap}, (\hat{\sigma}_0^2 - bias) + 1.96 * se_{bootstrap}]$$

```r
#standard bootstrap confidence interval

std_ci <- bias_corrected_estimate +
           c(-1,1)*qnorm(p = 0.975)*boot_df_summary$bootstrap_se
```

The standard bootstrap confidence interval is 0.2106, 1.9976.

We can also consider the **simple percentile** method for constructing our confidence interval. In this case we will simply take the 2.5 and 97.5 percentiles of our bootstrap samples. This yields:

```r
percentile_ci <- boot_df %>%
                      summarize(lower = quantile(sigma2hat_star,.025),
                                upper = quantile(sigma2hat_star, .975))
```

The simple percentile confidence interval is 0.2068 - 1.9301

In this case we would prefer the simple percentile method, as from looking at our histogram, the distribution is not very symmetric.

To summarise: the estimate for σ_0^2, the variance of the population from which our data is generated is 1.1041. The 95% bootstrap confidence interval for σ_0^2 is 0.2068 - 1.9301