PS2 - Q3

Ben Juarez

Part a

```
birth = importdata("birth.txt");
bwt = birth(:,1);
bwt = bwt * 0.0283495;
bwt = bwt(bwt<999);
n = 100;
N = length(bwt);

pop_mean = mean(bwt);
disp("Population mean weight:");disp(pop_mean);
```

```
Population mean weight:
    3.3899
```

```
X = datasample(bwt, n);
sample_mean = mean(X);
disp("Sample mean weight:");disp(sample_mean);
```

```
Sample mean weight:
    3.4300
```

```
se = sqrt(((std(bwt)^2)/n) * (1 - ((n - 1)/(N - 1))));
disp("Standard error of sample mean:");disp(se);
```

```
Standard error of sample mean:
    0.0496
```

Part b

```
B = 10^3;
P = repmat(X, round(N/n), 1);
sample_means = zeros();
for i = 1:B
    sample_means(i) = mean(datasample(P,n));
end
sum = 0;
avg_sample_means = mean(sample_means);
for i = 1:B
    sum = sum + (sample_means(i) - avg_sample_means)^2;
end
se_hat = sqrt(1/B * sum);
disp("Bootstrap estimate of standard error:");disp(se_hat);
```

```
Bootstrap estimate of standard error:
    0.0519
```

Part c

```
k = floor(N/n);
```

```
r = rem(N, n);
p = (1 - (r/n))*(1-(r/(N-1)));
P1 = repmat(X, k, 1);
P2 = repmat(X, k+1, 1);
means = zeros();
for i = 1:B
    if rand() < p
        means(i) = mean(datasample(P1, n));
    else
        means(i) = mean(datasample(P2, n));
    end
end
sum = 0;
avg_sample_means = mean(means);
for i = 1:B
    sum = sum + (means(i) - avg_sample_means)^2;
end
se_hat = sqrt(1/B * sum);
disp("Bootstrap estimate of standard error:");disp(se_hat);
```

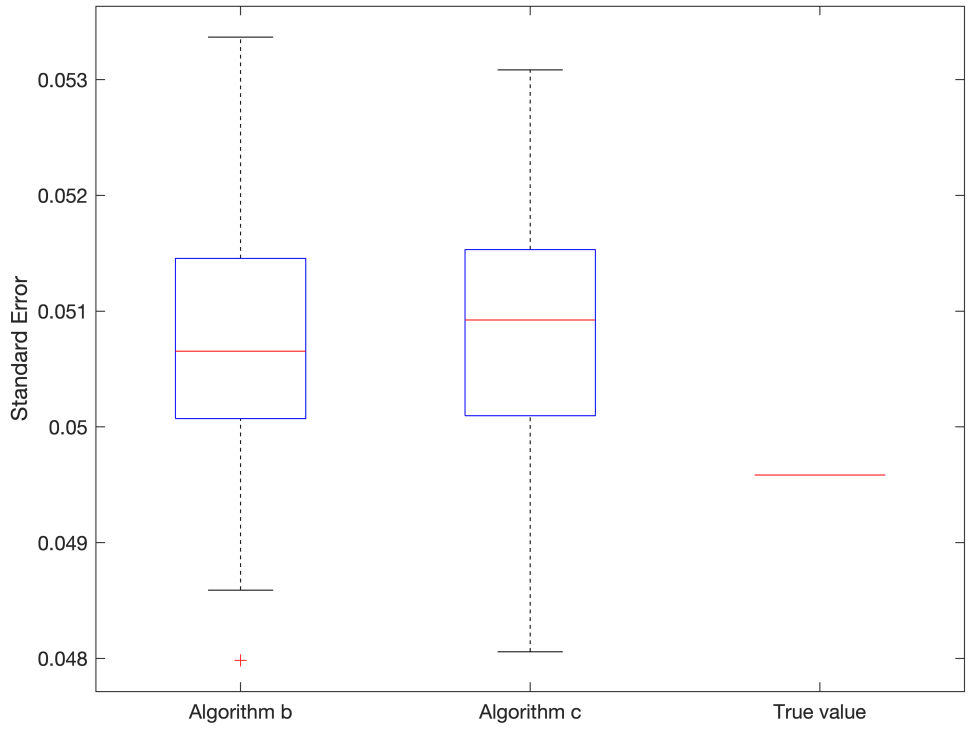
```
Bootstrap estimate of standard error:
    0.0507
```

Part d

```
part_b = zeros();
for z = 1:100
    P = repmat(X, round(N/n), 1);
    sample_means = zeros();
    for i = 1:B
        sample_means(i) = mean(datasample(P,n));
    end
    sum = 0;
    avg_sample_means = mean(sample_means);
    for i = 1:B
        sum = sum + (sample_means(i) - avg_sample_means)^2;
    end
    part_b(z) = sqrt(1/B * sum);
end
part_c = zeros();
for z = 1:100
    k = floor(N/n);
    r = rem(N, n);
    p = (1 - r/n)*(1-r/(N-1));
    P1 = repmat(X, k, 1);
    P2 = repmat(X, k+1, 1);
    means = zeros();
    for i = 1:B
        if rand() < p
            means(i) = mean(datasample(P1, n));
        else
            means(i) = mean(datasample(P2, n));
        end
    end
    sum = 0;
```

```
    avg_sample_means = mean(means);
    for i = 1:B
        sum = sum + (means(i) - avg_sample_means)^2;
    end
    part_c(z) = sqrt(1/B * sum);
end

figure;
boxplot([transpose(part_b), transpose(part_c), repmat(se, 100, 1)], ...
    'Labels', {'Algorithm b', 'Algorithm c', 'True value'});
ylabel("Standard Error");
```



Examining the box plot, we can see that both algorithms have fairly similar distributions. Additionally, the true value rarely falls within the IQR of either distribution. For most iterations, the true value falls within the largest/smallest observations (excluding outliers).