Question 3.

Table of Contents

Introduction:

Statistical values provide lots of useful information, for example, mean, median, mode, even skewedness and kurtosis. In this study, I will use the median and kurtosis estimates to fit a model as accurately as possible to the data in lifetime.txt. Using bootstrap methods, nonparametric and parametric, I will compare the bootstrapped values to the theoretical and estimated values we calculate repeatedly sampling with replacement.

Analysis and results:

The first thing we want to do is analyze the data in lifetime.txt and find the expected value with the appropriate confidence interval. The expected value that was calculated from the data[1] which is 68.81. I found this using the mean(X) method in R. Deriving the variance using the $\mathrm{V(X)} = \frac{1}{\mathrm{n}}\sum(\mathrm{X} - \mathrm{E(x)})^2$ formula we find 4.67 as the variance. Now we can find the 95% confidence interval for the expected value which is (64.56958, 73.04438).

I found the sample median and kurtosis values first[2] so that I could compare to the nonparametric bootstrapped values. The sample median was 48.42 and the sample kurtosis was 8.17. Next, I applied the nonparametric bootstrap procedure to estimate the median and kurtosis[3] of the data. I found the kurtosis value by using $\widehat{K_4} = \frac{\frac{1}{n}\sum(X_i - \bar{X})^4}{s^4}$ formula in the bootstrap. This process gave me the following estimates and 95% confidence interval for the median and kurtosis, respectively: 47.97574 and 8.041159 with (43.87595, 52.95716) and (5.685464, 10.67128).

Now let's verify that the data follows an exponential distribution. I will conduct the following test: $H_O: X \sim \exp(\lambda)\, vs. H_A: X \nsim \exp(\lambda)\ with\ \propto = 0.05$. First I must try to find a value for λ. I used a library[4] called "fitdistrplus" to fit the data to a distribution. A summary[5] of this method provided me with the estimate: $\hat{\lambda} = 0.01453341$. The next step then was to test this estimate using $H_O: X \sim \exp(0.01453341)\, vs. H_A: X \nsim \exp(0.01453341)\ with\ \propto = 0.05$. This test was done using the One-sample Kolmogorov-Smirnov test and produced a p-value = 0.9085. Since this p-value is much larger than $\propto$, and very close to 1, I fail to reject the null hypothesis and can say, with 95% confidence, that $X \sim \exp(0.01453341)$.

[1] Appendix # 3. A
[2] Appendix # 3. B
[3] Appendix # 3. B
[4] Appendix # 3. C
[5] Appendix # 3. C

Now that we have a distribution to fit to the data, we can apply the parametric bootstrap procedure to estimate the median and kurtosis values. I generated Y~ $\exp(0.01453341)$[6]. Now, I applied the bootstrap by randomly sampling through Y, sampling with replacement and calculating the median and kurtosis with the same method and formula as before, med(X) and $\widehat{K_4} = \frac{\frac{1}{n}\sum(X_i-\bar{X})^4}{s^4}$. I found that the bootstrapped estimates and 95% confidence interval of the median and kurtosis were, respectively: 47.81113 and 9.201635 with (47.31662, 56.14983) and (6.64921, 11.66061).

The final step is to compare these values to the theoretical values. Given: $f(x) = \frac{1}{0.01453341} e^{-0.01453341x}$ $and\ F(x) = 1 - e^{-0.01453341x}$. Let's find the median and kurtosis; $median = P(X \le x) = 0.5 \Rightarrow 1 - e^{-\lambda x} = 0.5 \Rightarrow x = \frac{\ln 2}{\lambda} = \frac{\ln 2}{0.01453341} = 47.69336175 \therefore$ $median \approx 47.69$. Now for kurtosis: $k_4 = \frac{E(x-E(x))^4}{\left(E(x-E(x))^2\right)^2}$, $first\ let's\ expand\ E\left(x-E(x)\right)^4$:

$$E\left(x-E(x)\right)^4 = E(x^4 + 4x^3E(x) + 6x^2E(x)^2 + 4xE(x) + E(x)^4$$

$$= E(x^4) - 4\frac{1}{\lambda}E(x^3) + 6\frac{1}{\lambda^2}E(x^2) - 4\frac{1}{\lambda^3}E(x) + \frac{1}{\lambda^4}$$

$$= 4!\frac{1}{\lambda^4} - 4\cdot 3!\frac{1}{\lambda^3}\frac{1}{\lambda} + 6\cdot 2!\frac{1}{\lambda^2}\frac{1}{\lambda^2} - 4\frac{1}{\lambda}\frac{1}{\lambda^3} + \frac{1}{\lambda^4} = 12\frac{1}{\lambda^4} - 4\frac{1}{\lambda^4} + \frac{1}{\lambda^4} = 9\frac{1}{\lambda^4}$$

$$so, k_4 = \frac{E\left(x-E(x)\right)^4}{\left(E\left(x-E(x)\right)^2\right)^2} = {9\frac{1}{\lambda^4}}\Big/{V(x)^2} = {9\frac{1}{\lambda^4}}\Big/{(\lambda^2)^2} = {\frac{9}{\lambda^4}}\Big/{\lambda^4} = 9$$

Now we can compare the theoretical estimates to the bootstrapped estimates of $\exp(0.01453341)$ as well as the estimates from the data.

	Theoretical, $\exp(\mathbf{0.01453341})$	**Sampled from $\exp(\mathbf{0.01453341})$**	**Bootstrapped estimate exp(0.01453341)**	**Bootstrapped, $\exp(\mathbf{0.01453341})$, CI**
Median	47.69336	51.46687	47.81113	(46.6395, 53.81261)
Kurtosis	9	9.192094	9.201635	(7.108722, 11.27547)

	Sampled from lifetime.txt	**Bootstrapped estimate from lifetime.txt**	**Bootstrapped lifetime.txt CI**
Median	48.41656	47.97574	(43.87595, 52.95716)
Kurtosis	8.17837	8.041159	(5.685464, 10.67128)

[6] Appendix # 3. D

From this table we can see that the sampled $exponential(\hat{\lambda})$ and theoretical medians are not too far off and neither are the kurtosis values. The difference between them could simply be due to the randomization in the sampling, because if we compare the theoretical values to the 95% confidence interval, we can observe that all the estimates are contained in each of the confidence intervals. To go further, I compared the sampled data estimates to that of the bootstrapped estimates and to the previous table as well and even with these comparisons we can see that all the estimates are close and not too far off from the theoretical estimate.

Conclusion

To conclude, if we look at the histograms[7] of the sample and the distribution of $\exp(0.01453341)$ we can see that they are very alike and have similar properties. The confidence intervals are also very similar for the two median CI's and the kurtosis CI's. The difference can be accounted for by the randomization of the sampling. These bootstrap procedures helped us find the best estimate for the median and kurtosis and also to make sure that the estimate we got for λ was a good estimate to fit the data.

Appendix

 # Course: STA312
 # Final Project: # 3.A
 # Last Name: MANSOOR, First Name: SARAH
 # St. #: 1004183251

```
> data_proj <- stack(lifetime)
> X <- data_proj$values
> expected_value <- mean(X)
> expected_value
[1] 68.80698
> n <- 1000
> variance_proj <- mean ((X - expected_value) ^2) / n
> variance_proj
[1] 4.673983
> conf_proj <- c (expected_value - 1.96 * sqrt(variance_proj), expected_value + 1.96 *
sqrt(variance_proj))
> conf_proj
[1] 64.56958 73.04438
```

 # Course: STA312
 # Final Project: # 3.B
 # Last Name: MANSOOR, First Name: SARAH
 # St. #: 1004183251

```
> n <- 1000
```

[7] Appendix # Visuals # Visual 2

```
> median(X) # sample median
[1] 48.41656
> kurtosis_s = sum((X-mean(X)) ^4)/n/ var(X)^2
> kurtosis_s # sample kurtosis
[1] 8.17837
> B <- 200
> bootstrap.statM <- NULL
> bootstrap.statK <- NULL
> for (b in 1:B){
+   i <- sample(1:n, size = n, replace = TRUE)
+   X.boot <- X[i]
+   M <- median(X.boot)
+   fourth.moment <- sum((X.boot - mean(X.boot))^4)/n
+   K <- fourth.moment/var(X.boot)^2
+   bootstrap.statM <- c(bootstrap.statM, M)
+   bootstrap.statK <- c(bootstrap.statK, K)
+ }
> M <- mean(bootstrap.statM) # bootstrap estimate of median for data
> K <- mean(bootstrap.statK) # bootstrap estimate of kurtosis for data
> M
[1] 47.97574
> K
[1] 8.041159
> # 95% CI for median and kurtosis
> alpha <- 0.05
> # Normal Bootstrap CI for median and kurtosis
> lowerM <- median(X) - qnorm(1 - alpha/2)*sd(bootstrap.statM)
> upperM <- median(X) + qnorm(1 - alpha/2)*sd(bootstrap.statM)
> lowerM
[1] 43.87595
> upperM
[1] 52.95716
> lowerK <- kurtosis_s - qnorm(1 - alpha/2)*sd(bootstrap.statK)
> upperK <- kurtosis_s + qnorm(1 - alpha/2)*sd(bootstrap.statK)
> lowerK
[1] 5.685464
> upperK
[1] 10.67128
```

 # Course: STA312
 # Final Project: # 3.C
 # Last Name: MANSOOR, First Name: SARAH
 # St. #: 1004183251

```
> library("fitdistrplus")
> fw <- fitdist(X, 'exp')
```

```
> summary(fw)
> summary(fw)
Fitting of the distribution ' exp ' by maximum likelihood
Parameters :
       estimate     Std. Error
rate  0.01453341   0.0004573987
Loglikelihood:  -5231.305   AIC:  10464.61   BIC:  10469.52
> ks.test(X, 'pexp', rate = 0.01453341)

One-sample Kolmogorov-Smirnov test

data:  X
D = 0.017823, p-value = 0.9085
alternative hypothesis: two-sided
```

 # Course: STA312
 # Final Project: # 3.D
 # Last Name: MANSOOR, First Name: SARAH
 # St. #: 1004183251

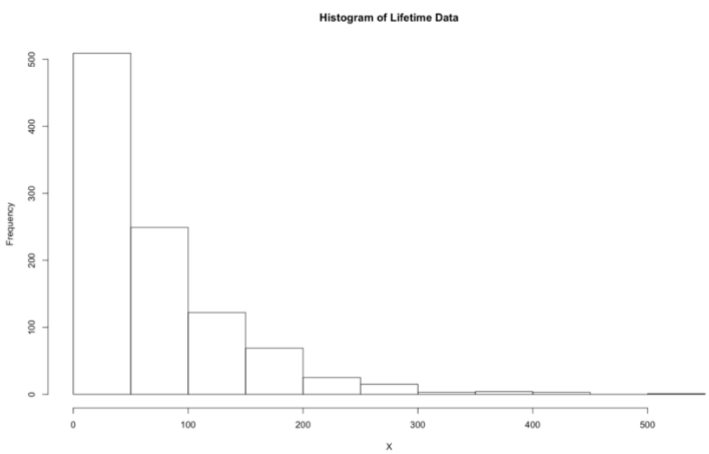
```
> n <- 1000
> Y <- rexp(n, rate = 0.01453341)
> median(Y) # sample exp estimate of median
[1] 51.46687
> kurtosis_t = (sum((Y-mean(Y))^4) / n) / var(Y)^2
> kurtosis_t # sample exp estimate of kurtosis
[1] 9.192094

> n <- 1000
> Y <- rexp(n, rate = 0.01453341)
> B <- 200
> bootstrap.statMT <- NULL
> bootstrap.statKT <- NULL
> for (i in 1:B){
+   boot.s <- NULL
+   boot.s <- sample(Y, replace = TRUE)
+   M_T <- median(boot.s)
+   fourth.moment_T <- sum((boot.s - mean(boot.s))^4)/n
+   K_T <- fourth.moment_T/var(boot.s)^2
+   bootstrap.statMT <- c(bootstrap.statMT, M_T)
+   bootstrap.statKT <- c(bootstrap.statKT, K_T)
+ }
> M_T <- mean(bootstrap.statMT) # bootstrap estimate of median exp
> K_T <- mean(bootstrap.statKT) # bootstrap estimate of kurtosis exp
> M_T
[1] 47.81113
```

```
> K_T
[1] 9.201635
> # 95% CI for median and kurtosis
> alpha <- 0.05
> # Normal Bootstrap CI for median and kurtosis
> lowerM_T <- median(Y) - qnorm(1 - alpha/2)*sd(bootstrap.statMT)
> upperM_T <- median(Y) + qnorm(1 - alpha/2)*sd(bootstrap.statMT)
> lowerM_T # lower exp bootstrapped median
[1] 46.6395
> upperM_T # upper exp bootstrapped median
[1] 53.81261
> lowerK_T <- kurtosis_t - qnorm(1 - alpha/2)*sd(bootstrap.statKT)
> upperK_T <- kurtosis_t + qnorm(1 - alpha/2)*sd(bootstrap.statKT)
> lowerK_T # lower exp bootstrapped kurtosis
[1] 7.108722
> upperK_T # upper exp bootstrapped kurtosis
[1] 11.27547
```

 # Visuals

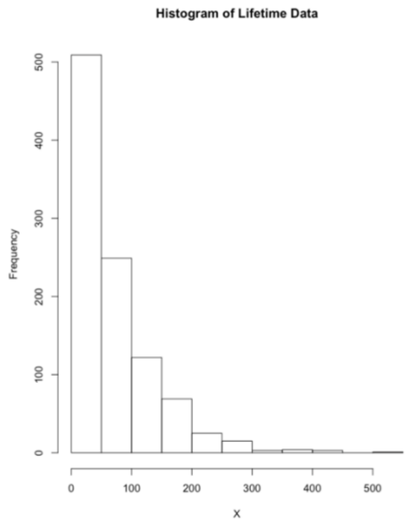
 # Visual 1

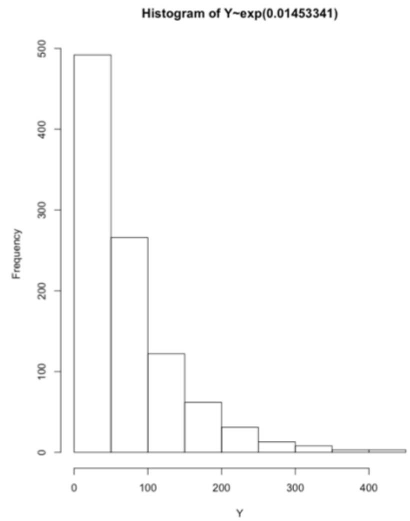
```
> hist(X, main = 'Histogram of Lifetime Data')
```



 # Visual 2

```
> par(mfrow = c(1,2))
> hist(Y, main = 'Histogram of Y~exp(0.01453341)')
```


Histogram of Lifetime Data
Frequency
X
0
100
200
300
400
500


Histogram of Y~exp(0.01453341)
Frequency
Y
0
100
200
300
400
500