S631 HW1

Erik Parker

August 25, 2017

Assume the data is the entire population of interest, S = {set of all UN members}, and $X : S \rightarrow R$ (a function which assigns to each country a corresponding female life expectancy in years, rounded to the nearest integer).

---

1. Is X a random variable? Why or why not.

X is a random variable, because it is a function which maps our set of possible objects (here S, the set of all UN members) to a set of labels represented by all real numbers (R). Specifically, we can use this function to find the average life expectancy in years of females in any randomly selected UN country, and the CDF can be evaluated to find the probability of a country's women having a lifespan at least as great as a particular value from R for all real numbers.

2. Write the mathematical expression for the following sets:

a) Countries with female life expectancy less than or equal to 80 years.

F(80) = P(X <= 80) = P({sES: X(s) <= 80})

b) Countries with female life expectancy equal to 75.

f(75) = P(X = 75) = P({sES: X(s) = 75})

c) Countries with female life expectancy between 65 and 70 years.

f(65 < X < 70) = P(65 < X < 70) = P({sES: 65 < X(s) < 70})

3. Write down the probabilities of the previous expressions using only CDFs.

a)

F(80) = P(X <= 80)

b)

f(75) = F(75) - F(74) = P(X <= 75) - P(X <= 74)

c)

Given that P(X=x) = 0 for all R,
f(65 < X < 70) = f(65 < X <= 69) = F(69) - F(65)

4. Obtain the previously defined probabilities using the *UN11* data.

```
library(alr4)

life <- UN11$lifeExpF

life <- round(life,0)
```

F(80) = P(X <= 80)

```
length(life[life<=80]) / length(life)
```

```
## [1] 0.7889447
```

So, the probability that a country in this dataset has a female life expectancy of at least 80 years is 0.7889, or 79%.

f(75) = F(75) - F(74)

```
(length(life[life<=75]) - length(life[life<=74])) / length(life)
```

```
## [1] 0.06030151
```

So, the probability that a country in this dataset has a female life expectancy of 74 years is 0.06, or 6%.

f(65 < X < 70) = F(69) - F(65)

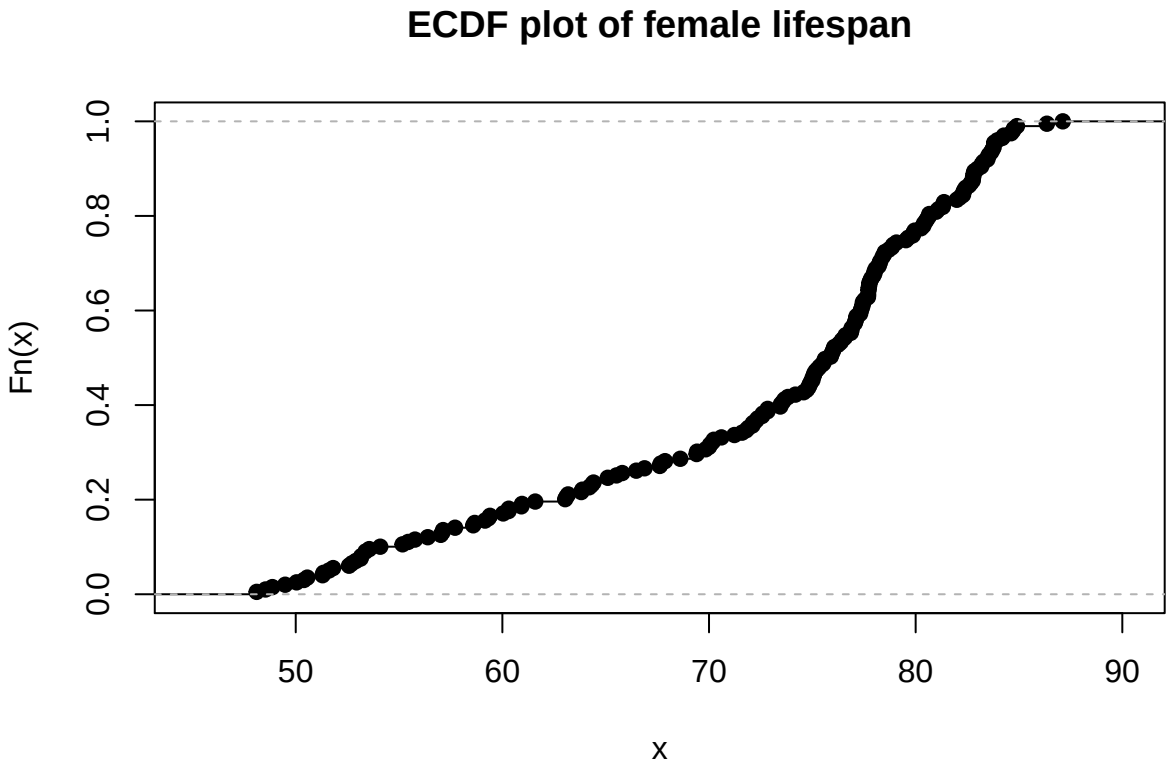
```
(length(life[life<=69]) - length(life[life<=65])) / length(life)
```

```
## [1] 0.05527638
```

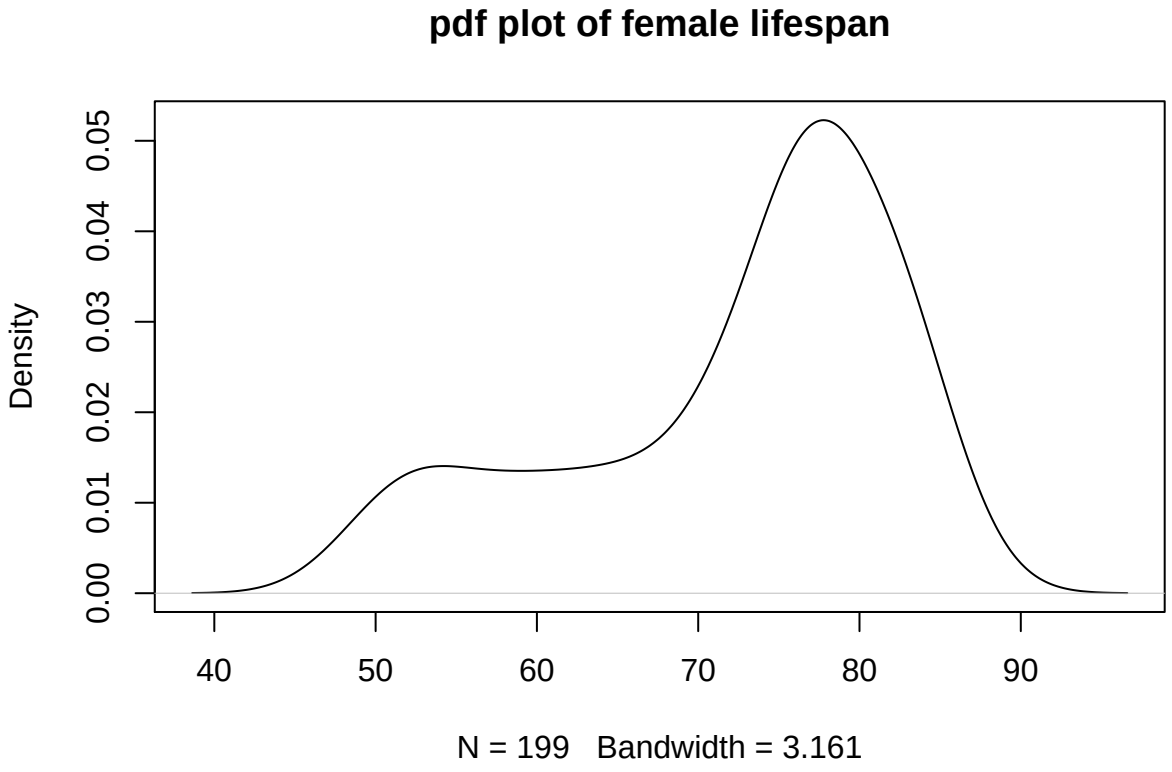
So, the probability that a country in this dataset has a female life expectancy between 70 and 65 years is equal to the probability of seeing a life expectancy of 66 + 67 + 68 + 69 years, which is 0.055, or 5.5%.

5. Draw the ECDF and pdf for "LifeExpF" without rounding. Does the variable seem to follow a normal distribution? Explain.

```
plot(ecdf(UN11$lifeExpF), main = "ECDF plot of female lifespan")
```




```
plot(density(UN11$lifeExpF), main = "pdf plot of female lifespan")
```



The variable "LifeExpF", representing the average expected life span for females in the 199 UN countries in this dataset, does not appear to follow a normal distribution. The distribution is weighted heavily to the right, as the majority of countries show lifespans that are higher than the

mean - thus creating a left-skewed, non-normal distribution.