Week 3 Classwork/Homework Assignment Answers

John Vinson

8/7/2017

For the assignment, you will submit a R Markdown (.rmd) file and the compiled document as both a pdf and html. Make sure that the compiled documents will display all the required code to get your results.

The assignment is due Friday September 1 at 8:00 PM.

1. Choose one of the following questions and discuss succently:

- What is a function? Compare and contrast mathematical functions with functions used in programming.
- Explain how functions can be represented. Use examples to aid in your explanations.
- What is mathematical model? How is this related to the topics that were covered this week?

Answers will vary.

2. A can in the shape of a cylinder is to be made to hold a volume of one liter (1000 cubic centimeters). The manufacturer needs the functions for both the volume and the surface area. Define these as functions of the radius (r) and height (h) of the can. Determine the possible values that h can take if the radius of the can must be between 0.5 and 10 cm. Which values of r and h (to three decimal places) minimize the surface area allowing the manufacturer to use the least amount of materials?

The formulas for the volume and surface area of cylinders are:

$$V = \pi r^2 h \tag{1}$$
$$SA = 2\pi rh + 2pir^2 \tag{2}$$

Since we are holding the volume of the cylinder at 1000 cubic cm, we can define our equations:

$$1000 = \pi r^2 h \tag{3}$$
$$SA = 2\pi rh + 2pir^2 \tag{4}$$

Our values are r are constrained such that it the smallest it can take is 0.5 cm and the biggest is 10 cm. So we determine the heights that these cans can take.

$$1000 = \pi 0.5^2 h \tag{5}$$
$$h_{0.5} = \frac{1000}{\pi 0.5^2} = 1273.24cm \tag{6}$$
$$h_{10} = \frac{1000}{\pi 10^2} \approx 3.183099cm \tag{7}$$

Therefore the height must be in the interval $[3.189099, 1273.24]$.

Now, we are interested in minimizing the surface area of the can. To address this, we will write a function that takes in a radius of the can and returns the surface area.

```
SA = function(r){
  h = 1000/(pi*r^2)
  return(2*pi*r*h+2*pi*r^2)
}
```

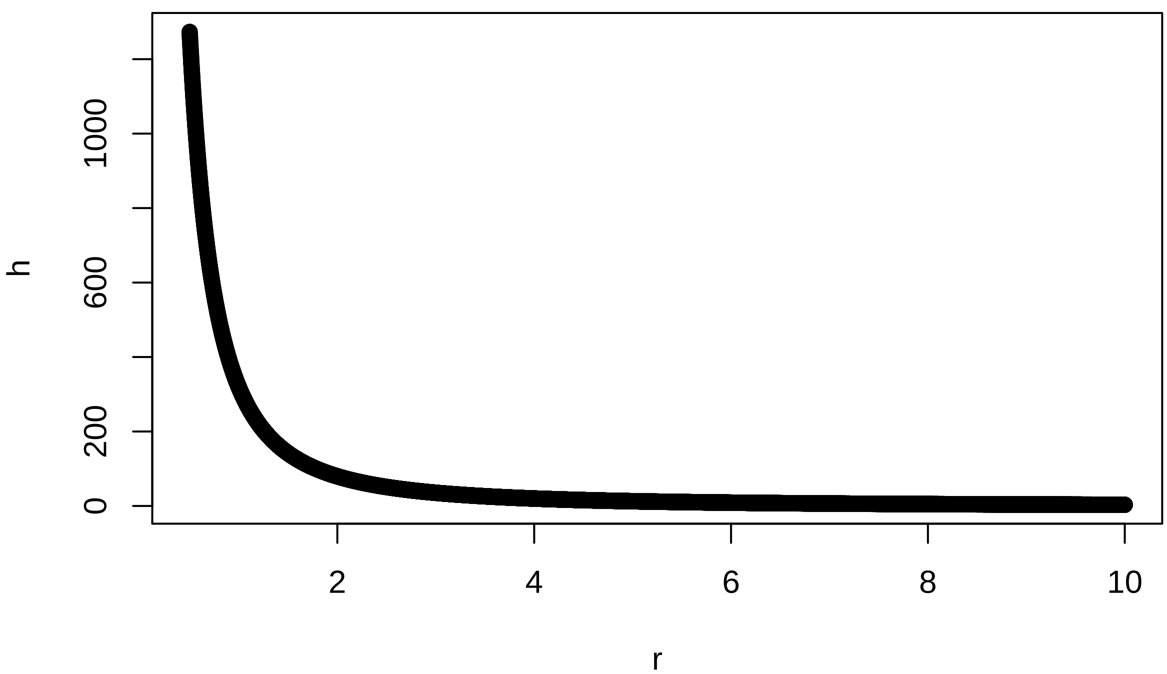
Using this function we will iterate over a sequence of radiuses from $[0.5, 10]$. Using these, we will approximate the radius that minimizes the surface area.

```
r.vec = seq(0.5, 10, 0.001)

SA.vec = SA(r.vec)

SA.min = min(SA.vec)

r.min = r.vec[which(SA.vec==SA.min)]

h.min = round(1000/(pi*r.min^2),3)
```

The diminsions of the can that minimizes the surface areas are radius of 5.419 cm and a height of 10.84 cm. The surface area of this can is 553.581 cm.

To visualize the shape of the function we can plot it.

```
plot(r.vec, 1000/(pi*r.vec^2), xlab="r", ylab="h")
```

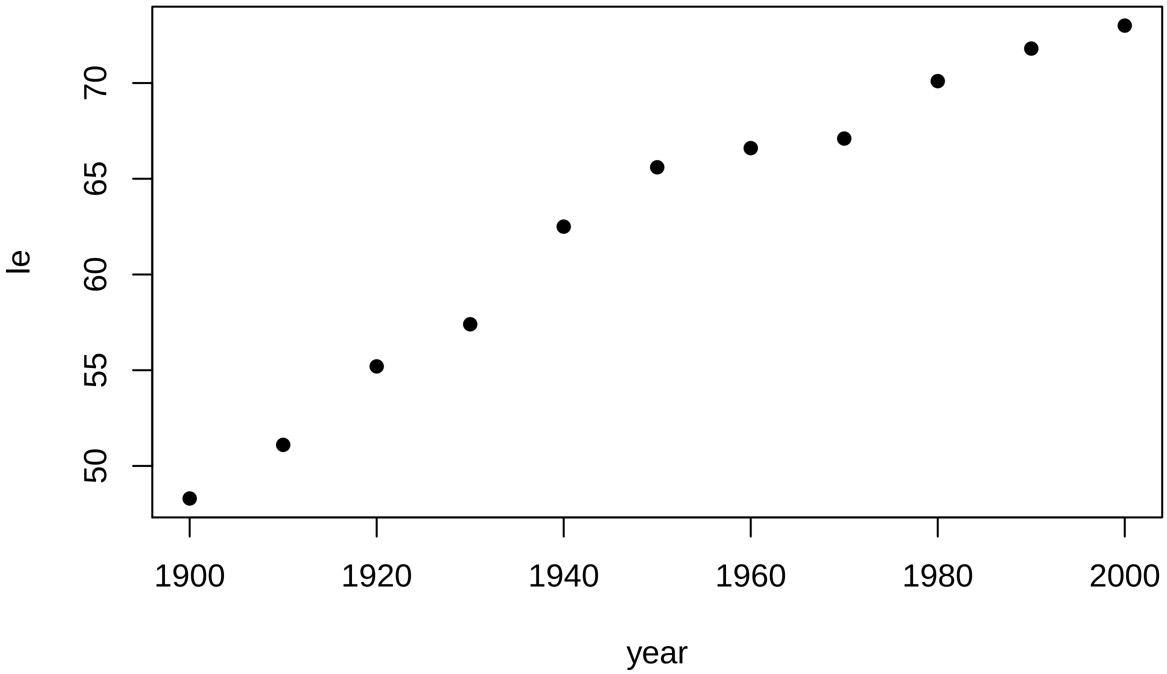


3. The following table includes the life expectancy (at birth) of males from 1900 to 2000.

year	life expectancy	year	life expectancy
1900	48.3	1960	66.6
1910	51.1	1970	67.1
1920	55.2	1980	70.1
1930	57.4	1990	71.8
1940	62.5	2000	73
1950	65.6		

Plot the data as a scatter plot and determine the appropiate model (function) to fit the data. Using this model, predict the life span of males born in 2010.

```
year = seq(1900, 2000, 10)
le = c(48.3,51.1,55.2,57.4,62.5, 65.6, 66.6,67.1,70.1,71.8,73.0)

plot(year, le, pch=16)
```



From this plot, the best fitting model is a linear model. One way that we can estimate the parameters of this model is to estimate the slope using the first and last points in the dataset.

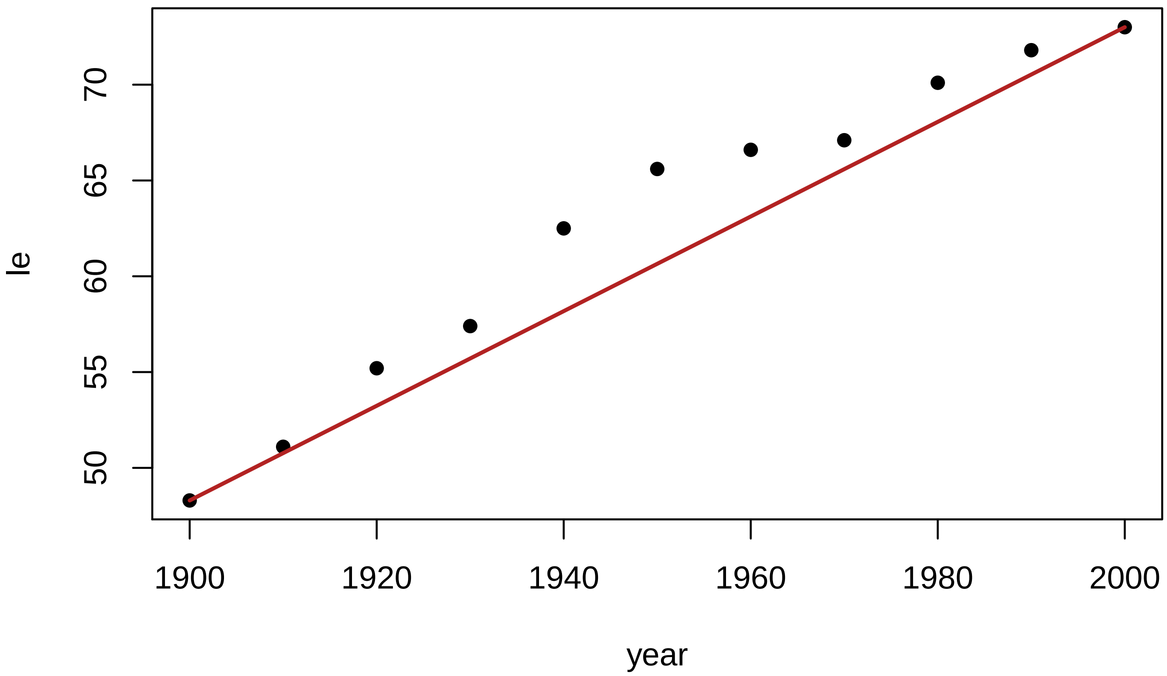
$m = \frac{73.0-48.3}{2000-1900} = 0.247$

Now, using a point from our data we can estimate the intercept.

$b = y - mx = 48.3 - 0.247(1900) = -421$

This results in a linear model, $le = 0.247 * year - 421$.

```
plot(year, le, pch=16)

le.lm = function(year, m, b){return(m*year+b)}

lines(year, le.lm(year, m=0.247, b=-421), col="firebrick", lwd=2.0)
```



Using this parameterization, the model results in many of the points above the line. This means on average our model is estimating a life expectancy lower than the actual value. To provide a better estimate, we can use a least squares method, which minimizes the difference between the estimated life expectancy and the actual data. To do this, we can write a function that given a set or parameters will return the a measure of the difference between the model and the data (sum of square difference).

$SSD(m, b) = \sum_{i}^{year}(le_i - m * i + b)^2$

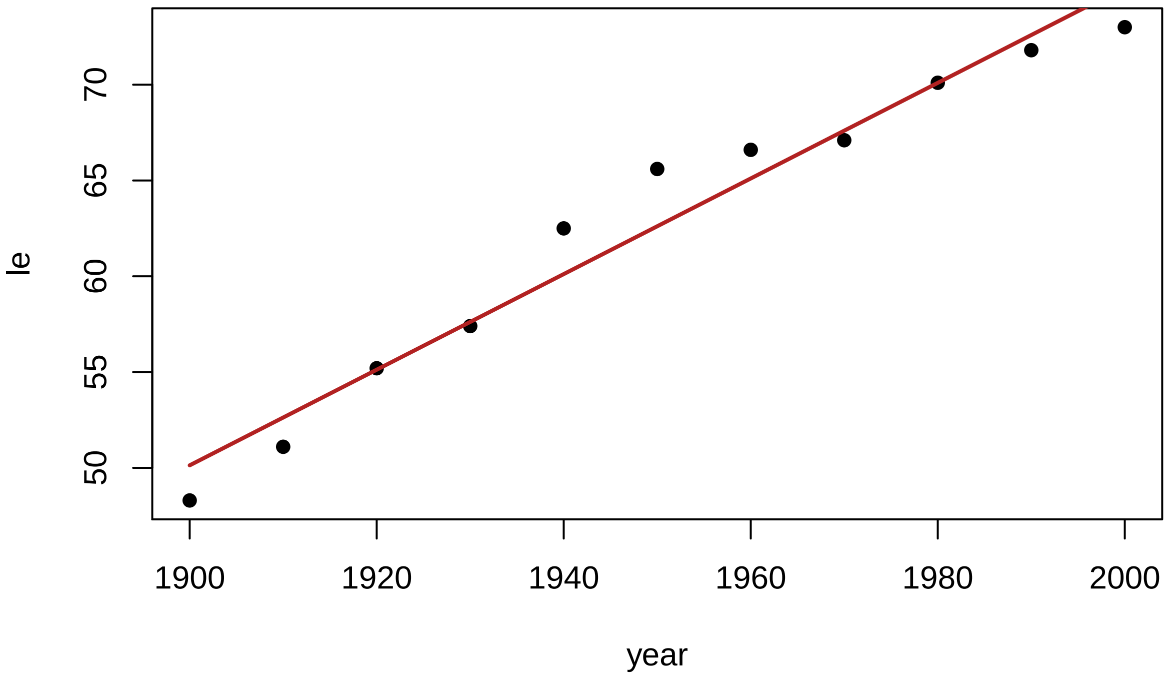
```
SSD = function(parms){
  out = 0

  for(i in 1:length(year)){
    out = out + (le[i] - (parms[1]*year[i] + parms[2]))^2
  }
  return(out)
}
```

Now, we can use the optim function to determine the parameter values that minimize the SSD. To use this function, we must provide initial values (guesses) that the optimization method (here, we use Nelder-Mead) will start at. Our initial values will be the estimated values from above.

Alternatively, we could do a grid search. However, this may result in inaccurate results.

```
best = optim(c(0.247,-421), fn=SSD, method="Nelder-Mead")
```

This provide estimates (rounded to 4/3 d.p.) of m=0.2495 and b=-423.997.



In R, there is a built-in function *lm* which will perform a linear regression and provide estimates for the parameters of a linear model.

The question asks for the life expectancy in 2010. Using our first model (the first and last point method), we would expect males in 2010 to have a life expectancy of 69.47 years. Using the model using the least squares method, we would expect males to live ~77.6 years in 2010.

4. Scientists have modeled the average surface temperature (T; in $^\circ C$) as a linear function:

$$T = 0.02t + 8.50 \tag{8}$$

where, t represents the number of years since 1900. What does the slope and T-intercept represent? What is the predicted temperature in 2100? In what year will average surface temperture exceed $10^\circ C$? Is a linear model an appropiate model for investigating the surface temperature? Explain and discuss possible alternatives.

Here, slope represents the average increase in surface temperature per year since 1900. Since 1900, each year the surface temperature has risen by $0.02^\circ C$ on average. The intercept is the average surface temperature in 1900.

In 2100, the average surface temperature is predicted to be

$$\begin{aligned} T =& 0.02t + 8.50 && (9)\\ T =& 0.02(110) + 8.50 && (10)\\ T =& 10.7^\circ C && (11) \end{aligned}$$

To determine which year the average surface temperature will be greater than $10^\circ C$, we need need to solve the inequality:

$$\begin{aligned}10 \geq& 0.02t + 8.50 \quad (12)\\ 1.5 \geq& 0.02t \quad (13)\\ t \geq& 75 \quad (14)\end{aligned}$$

After 1975, the average surface temperature will be greater than $10°C$ assuming this linear model.

Discussion of the assumptions of a linear model and the impact it will have on analyses. Discussion of other possible models and their assumptions.

5. The number of individuals in a bacteria population at time t ($N(t)$) with an initial population size of 100 individuals is described by the following function:

$$N(t) = 100 * 2^{t/3}. \quad (15)$$

What is the doubling time (i.e. the time it takes to double the number of individuals) of the population? Find the inverse of the function and describe what this new function represents. Determine how long it will take for the population to reach 50000.

To determine the doulbing time, we need to set $N(t) = 200$ individuals and solve to t.

$$\begin{aligned}N(t) =& 100 * 2^{t/3} \quad (16)\\ 200 =& 100 * 2^{t/3} \quad (17)\\ \log_2 2 =& (\log_2 2)^{t/3} \quad (18)\\ 1 =& t/3 \quad (19)\\ t =& 3 \quad (20)\end{aligned}$$

To find the inverse we solve for t.

$$\begin{aligned}N(t) =& 100 * 2^{t/3} \quad (21)\\ \frac{N(t)}{100} =& (2)^{t/3} \quad (22)\\ \log_2(\frac{N(t)}{100}) =& \log_2(2)^{t/3} \quad (23)\\ \log_2(\frac{N(t)}{100}) =& t/3 \quad (24)\\ t =& 3\log_2(\frac{N(t)}{100}) \quad (25)\\ & \quad (26)\end{aligned}$$

This is a function that takes an population size ($N(t)$)and outputs the time (t) it takes the population to reach that size.

```r
3*log2(50000/100)
```

```
## [1] 26.89735
```