HW2

Rabia Suleiman

8/22/2019

```
knitr::opts_chunk$set(echo = TRUE)
```

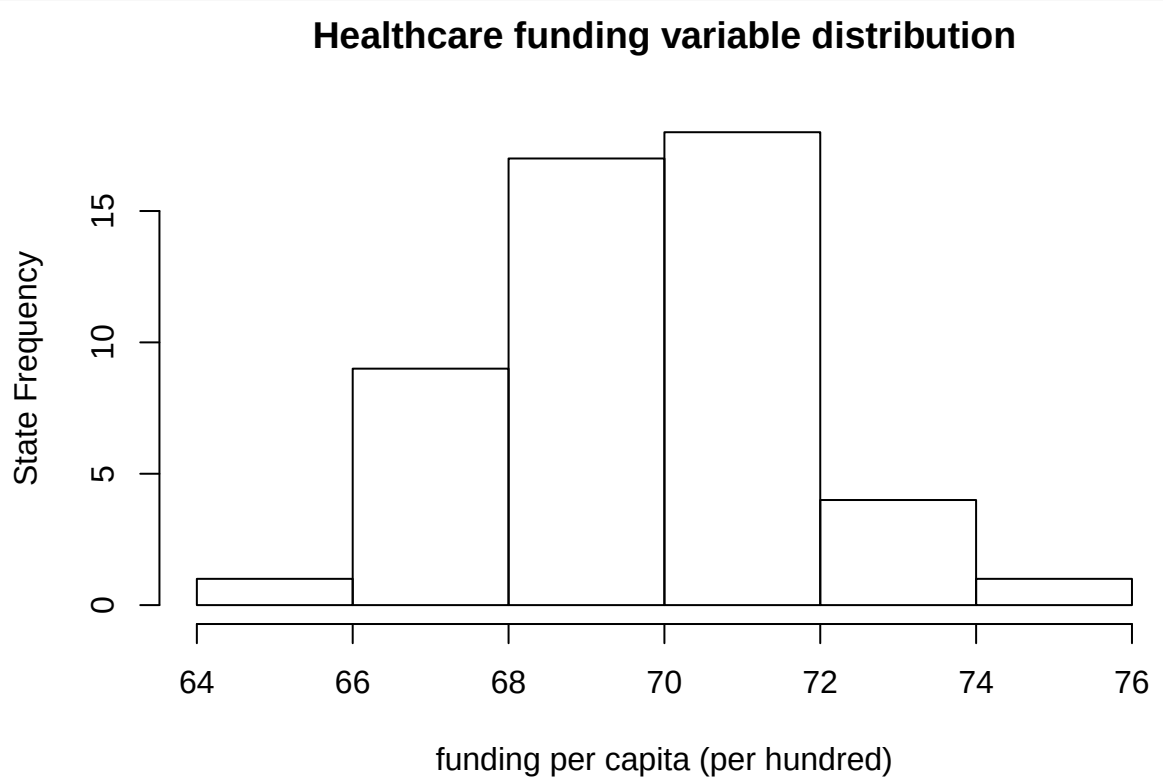
Question

There are many variables that can differences in life expectancy like sex, location (urban/rural, major regions/provinces) and socio-economic characteristics (e.g. education, wealth quintile). Average life expectancy varies by state in the United States. This question seeks to determine the impact of healthcare spending per capita of each state on the life expectancy of individuals residing there. We expect that the more the state invest in healthcare, the greater the life expectancy (As access to healthcare provides preventative care as education programs.) We also control for socioeconomic status and ses.

Health Care Investemnt State level= health care by state follows a normal distribution with the majority of states allocating between $6350 and $7700 per year, (9 states allocated less while 8 allocated more)

Unit of Analysis = 5000 Individuals in the United States

```
rm(list=ls())
 N = 50
health<-rnorm(n=N, mean=70, sd=2)
hist((health), xlab= "funding per capita (per hundred)", ylab= "State Frequency", main="Healthcare fundi
```




```
summary (health)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   64.69   68.36   69.79   69.85   71.35   74.55
```

Gender

The sex of an indivudal plays a significant role in life expectancy. A binomial distribution with.5 probability was used to assign gender.

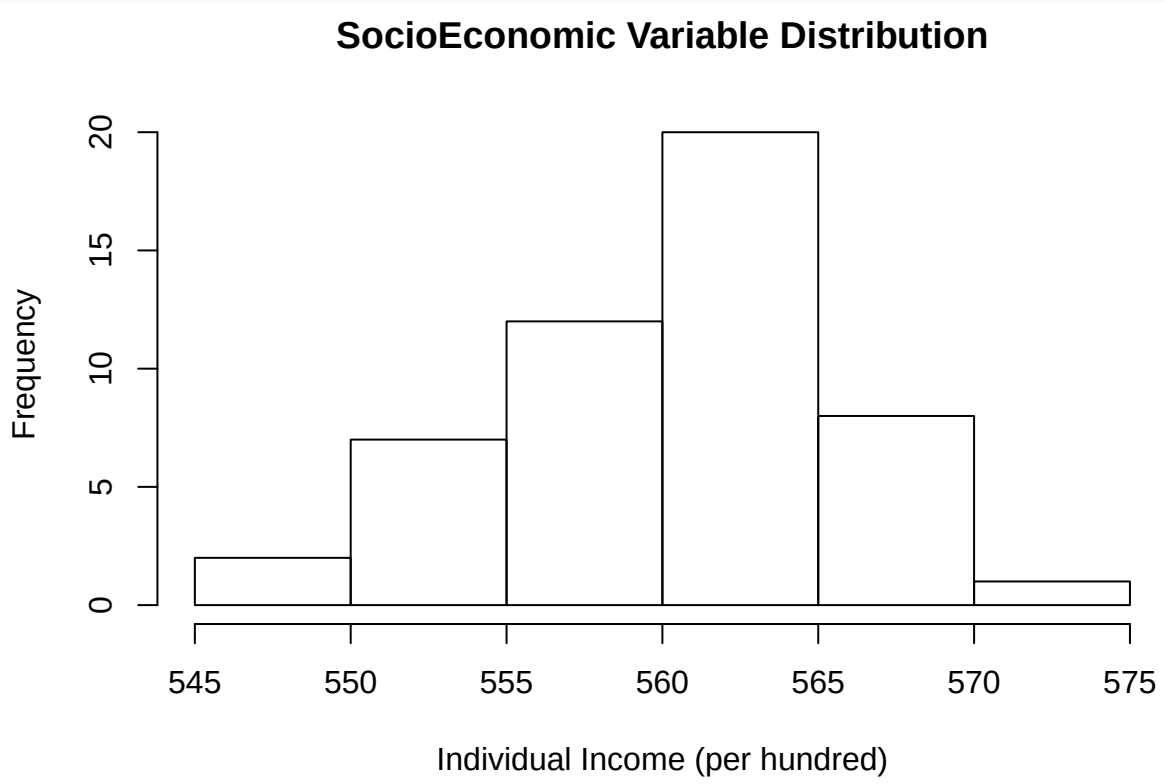
```
sex=rbinom(n=N, size=1, prob=0.5) #1 is female, 0 is male
summary(sex)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    1.00    0.58    1.00    1.00
```

Socioeconomic Status

The average American is $56,000, we allow for a normal distribution (although this a hypothetical situaltion in reality the distrubtion is somewhat skewed to the poor but its not quite an exponential distribution)

```
ses2<-rnorm(n=N, mean=560, sd=5)
hist((ses2), xlab= "Individual Income (per hundred)", ylab= " Frequency", main="SocioEconomic Variable
```




```
summary (ses2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   545.4   557.2   561.3   560.3   564.1   570.6
```

```
mean_ses<-560
```

life expectancy

```
 a = 48
 b1 = 0.35
 b2 = 0.35
 b3 = 0.025
```

```r
error = rnorm(n=N, mean=0, sd=0.1) # N(0,1)
expectancy<-a+ b1*health + b2*sex+ b3*ses2 +error
hist((expectancy) , xlab= "years", ylab= "Frequency", main="Life Expectancy Histogram")
```


```r
regression1 <- lm(expectancy ~ health + sex+ ses2)
summary(regression1)
```

```
##
## Call:
## lm(formula = expectancy ~ health + sex + ses2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17404 -0.05776 -0.00300  0.05642  0.23075
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 50.083804   1.484586  33.736  < 2e-16 ***
## health       0.356521   0.006723  53.033  < 2e-16 ***
## sex          0.364360   0.028351  12.852  < 2e-16 ***
## ses2         0.020451   0.002500   8.181  1.6e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09421 on 46 degrees of freedom
## Multiple R-squared:  0.9864, Adjusted R-squared:  0.9855
## F-statistic:  1114 on 3 and 46 DF,  p-value: < 2.2e-16
```

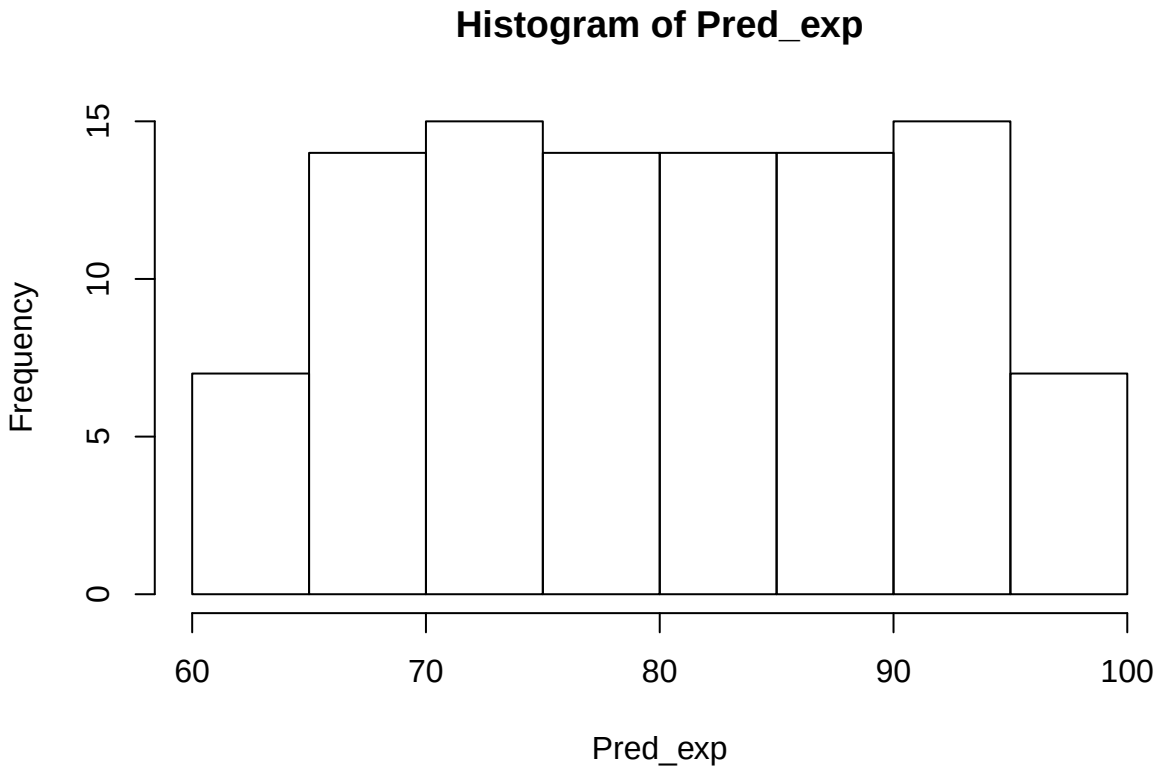
```r
health_c<- seq(from=1, to=100, by=1)
```

```
Pred_exp<- NA

for(i in 1:100){
Pred_exp[i] = 47.44 + (0.35*health_c[i]) + (0.351*1) + (.026*mean_ses)

}
```

```
 head(Pred_exp)
```

```
## [1] 62.701 63.051 63.401 63.751 64.101 64.451
```

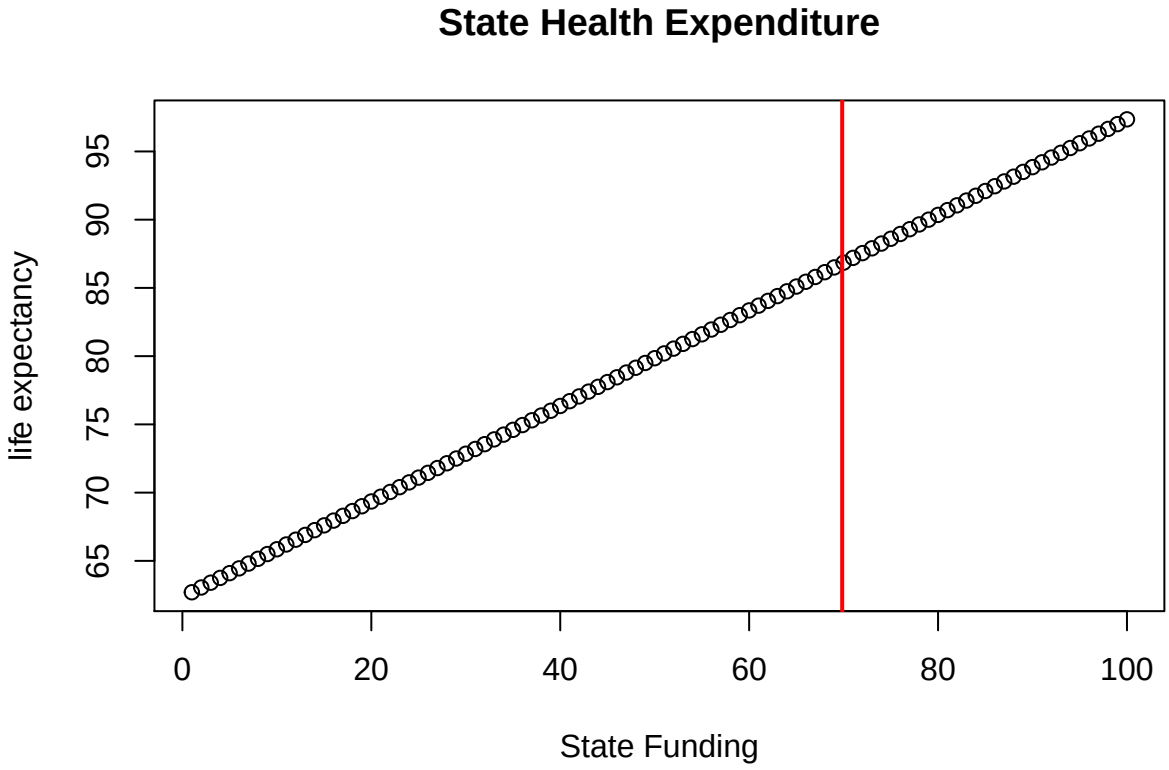
```
 hist(Pred_exp)
```




```
 summary(Pred_exp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   62.70   71.36   80.03   80.03   88.69   97.35
```

```
plot((Pred_exp ~ health_c),  xlab="State Funding", ylab="life expectancy",
main="State Health Expenditure")
abline(v=mean(health), col="red", lwd=2)
```


State Health Expenditure
life expectancy
65
70
75
80
85
90
95
0
20
40
60
80
100
State Funding