

HW2Sol

Problem 1

(a)

```
url<-"https://people.stat.sc.edu/hoyen/STAT540/Data/dataP1HW2.txt"
mydata<- read.table(url, head=T)
head(mydata)
```

```
##      X      Y
## 1 35.0 13.7
## 2 41.8 16.7
## 3 42.6 16.4
## 4 33.2  9.2
## 5 47.2 13.8
## 6 53.7 17.8
```

(b)

The means of X and Y are 41.83, 14.54. The variances of X and Y are 67.51, 10.64

```
attach(mydata)
c(mean(X), var(X))
```

```
## [1] 41.82667 67.51067
```

```
c(mean(Y), var(Y))
```

```
## [1] 14.54000 10.63829
```

(c)

The Pearson correlation between X and Y is 0.26.

```
r=cor(X, Y)
r
```

```
## [1] 0.2567613
```

(d)

The results from lm are shown as below.

```
help(lm)
fit<- lm(Y~X, data=mydata)
summary(fit)
```

```
##
## Call:
## lm(formula = Y ~ X, data = mydata)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.461 -1.695 -0.570  1.915  6.242
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.2768     4.5301   2.269   0.041 *
## X              0.1019     0.1064   0.958   0.356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.271 on 13 degrees of freedom
## Multiple R-squared:  0.06593,    Adjusted R-squared:  -0.005925
## F-statistic: 0.9175 on 1 and 13 DF,  p-value: 0.3556
```

(e)

```
fit$coefficients
```

```
## (Intercept)          X
##  10.2768323    0.1019246
```

(f)

```
readme="This is a simple linear regression problem."
readme
```

```
## [1] "This is a simple linear regression problem."
```

(g)

Create a list named listP1HW2 as below.

```
listP1HW2<- list(mydata, fit$coefficients, r, readme)
listP1HW2
```

```
## [[1]]
##      X      Y
## 1  35.0  13.7
## 2  41.8  16.7
## 3  42.6  16.4
## 4  33.2   9.2
## 5  47.2  13.8
## 6  53.7  17.8
## 7  34.0  19.6
## 8  34.3  13.3
## 9  42.0  20.8
## 10 58.8  15.7
## 11 47.3  11.1
## 12 37.0  12.6
## 13 46.8  12.0
## 14 44.6  14.1
## 15 29.1  11.3
##
## [[2]]
```

```
## (Intercept)          X
## 10.2768323    0.1019246
##
## [[3]]
## [1] 0.2567613
##
## [[4]]
## [1] "This is a simple linear regression problem."
names(listP1HW2)<- c("data", "coef", "corr", "readme")
listP1HW2
```

```
## $data
##      X      Y
## 1  35.0 13.7
## 2  41.8 16.7
## 3  42.6 16.4
## 4  33.2  9.2
## 5  47.2 13.8
## 6  53.7 17.8
## 7  34.0 19.6
## 8  34.3 13.3
## 9  42.0 20.8
## 10 58.8 15.7
## 11 47.3 11.1
## 12 37.0 12.6
## 13 46.8 12.0
## 14 44.6 14.1
## 15 29.1 11.3
##
## $coef
## (Intercept)          X
## 10.2768323    0.1019246
##
## $corr
## [1] 0.2567613
##
## $readme
## [1] "This is a simple linear regression problem."
listP1HW2$Data$X
## NULL
```

Problem 2

Using the formula

```
n<-100
total<-n*(n+1)/2
total

## [1] 5050
```

Alternative approach

```
n<-100  
sum(1:n)
```

```
## [1] 5050
```

Problem 3

(a)

```
library(dslabs)  
data(murders)  
str(murders)  
  
## 'data.frame': 51 obs. of 5 variables:  
## $ state : chr "Alabama" "Alaska" "Arizona" "Arkansas" ...  
## $ abb : chr "AL" "AK" "AZ" "AR" ...  
## $ region : Factor w/ 4 levels "Northeast","South",...: 2 4 4 2 4 4 1 2 2 2 ...  
## $ population: num 4779736 710231 6392017 2915918 37253956 ...  
## $ total : num 135 19 232 93 1257 ...
```

(b)

```
a <- murders$abb  
class(a)  
  
## [1] "character"
```

(c)

```
b <- murders[["abb"]]  
class(b)  
  
## [1] "character"  
  
identical(a,b)  
  
## [1] TRUE
```

(d)

```
index<-match(c("AK", "MI", "IA"), murders$abb)  
murders$state[index]  
  
## [1] "Alaska" "Michigan" "Iowa"
```

(e)

The states are shown as below:

```
murder_rate<-murders$total/murders$population*100000  
low<-murder_rate < 1  
murders[low,]
```

##	state	abb	region	population	total
## 12	Hawaii	HI	West	1360301	7
## 13	Idaho	ID	West	1567582	12
## 16	Iowa	IA	North Central	3046355	21
## 20	Maine	ME	Northeast	1328361	11
## 24	Minnesota	MN	North Central	5303925	53
## 30	New Hampshire	NH	Northeast	1316470	5
## 35	North Dakota	ND	North Central	672591	4
## 38	Oregon	OR	West	3831074	36
## 42	South Dakota	SD	North Central	814180	8
## 45	Utah	UT	West	2763885	22
## 46	Vermont	VT	Northeast	625741	2
## 51	Wyoming	WY	West	563626	5

(f)

The states are shown as below:

```
south<-murders[murders$state=="South",]
south2<-south[order(south$population),]
south2
```

```
## [1] state      abb      region      population total
## <0 rows> (or 0-length row.names)
```

(g)

The actual abbreviations are shown below:

```
ind<- c("MA", "ME", "MI", "MO", "MU") %in% murders$abb
c("MA", "ME", "MI", "MO", "MU")[ind]
```

```
## [1] "MA" "ME" "MI" "MO"
```