

Exercise 2B

**Problem 1.**

AGE

	Males	Females
Mean	50.46	48.58
St. Dev.	18.44	18.54
Min	20	20
25 <sup>th</sup> %	35	32
Median	50	47
75 <sup>th</sup> %	66	63.75
Max	85	85

**Problem 2.**

BMI

	Males	Females
Mean	28.61	29.03
St. Dev.	6.26	7.19
Min	14.20	13.36
25 <sup>th</sup> %	24.56	23.88
Median	27.66	27.90
75 <sup>th</sup> %	31.49	32.64
Max	130.21	72.28

24 missing for males  
32 missing for females

**Problem 3.**

		BMI Category		
	Underweight ( $<18.5$ )	Normal ( $\geq 18.5$ & $<25$ )	Overweight ( $\geq 25$ & $<30$ )	Obese ( $\geq 30$ )
Total cholesterol, mean (SD)	<b>188 (35.67)</b>	<b>193.4 (41.007)</b>	<b>201.8 (42.13)</b>	<b>199.2 (43.51)</b>
Smoking status, N (%)				
Never	<b>32 (42.67%)</b>	<b>662 (48.60%)</b>	<b>856 (53.50%)</b>	<b>909 (53.91%)</b>
Former	<b>8 (10.67%)</b>	<b>308 (22.61%)</b>	<b>426 (26.62%)</b>	<b>467 (27.70%)</b>
Current	<b>35 (46.67%)</b>	<b>392 (28.78%)</b>	<b>318 (19.88%)</b>	<b>310 (18.39%)</b>

## Code.

```
>setwd("~/Desktop/Hw2") #setting working directory
>
>library(sas7bdat) #calling package
>options(max.print=999999) #setting parameters for file input
>read.sas7bdat("bpa.sas7bdat") #reading in SAS file
>
> bpa <- read.sas7bdat("bpa.sas7bdat") #setting file to a variable
> save(bpa, file="bpa.rda") #saving for future use
>
> #Problem 1
>
>#AGE DISTRIBUTION FOR MALES
>
>bpa_males <- subset(bpa, gender==1, select=c(age)) #subsetting age by male gender
>summary(bpa_males) #obtaining five-number summary
age
Min. :20.00
1st Qu.:35.00
Median :50.00
Mean :50.46
3rd Qu.:66.00
Max. :85.00
>
> sapply(bpa_males,sd) #obtaining standard deviation
age
18.44326
>
>#AGE DISTRIBUTION FOR FEMALES
>
> bpa_females <- subset(bpa, gender==2, select=c(age)) #subsetting age by female gender
> summary(bpa_females) #obtaining five-number summary
age
Min. :20.00
1st Qu.:32.00
Median :47.00
Mean :48.58
3rd Qu.:63.75
Max. :85.00
>
> sapply(bpa_females,sd) #obtaining standard deviation
age
18.5475
>
```

```

>
>#PROBLEM 2
>
>#BMI DISTRIBUTION FOR MALES
> bmi_males <- subset (bpa, gender==1, select=c(bmi)) #subsetting bmi by male gender
> summary(bmi_males) #obtaining five-number summary
      bmi
Min.   : 14.20
1st Qu.: 24.56
Median : 27.66
Mean    : 28.61
3rd Qu.: 31.49
Max.    :130.21
NA's    :24
>
> sapply(bmi_males, sd, na.rm=TRUE) #obtaining standard deviation
      bmi
6.263154
>
>#BMI DISTRIBUTION FOR FEMALES
> bmi_females <- subset (bpa, gender ==2, select=c(bmi)) #subsetting bmi by female gender
> summary(bmi_females) #obtaining five-number summary
      bmi
Min.   :13.36
1st Qu.:23.88
Median :27.90
Mean    :29.03
3rd Qu.:32.64
Max.    :72.28
NA's    :32
>
> sapply(bmi_females, sd, na.rm=TRUE)
      bmi
7.193834
>
>
>#PROBLEM 3
>
>#Setting BMI categories
> bmicat[bpa$bmi < 18.5] <- "Underweight"
> bmicat[bpa$bmi >= 18.5 & bpa$bmi <25] <- "Normal Weight"
> bmicat[bpa$bmi >=25 & bpa$bmi <30] <- "Overweight"
> bmicat[bpa$bmi >=30] <- "Obese"
>
>#Subsetting cholesterol by BMI categories
> underweight_chol <- subset (bpa, bmicat== "Underweight", select=c(chol))

```

```

> normal_weight_chol <- subset (bpa, bmicat== "Normal Weight", select=c(chol))
> overweight_chol <- subset (bpa, bmicat== "Overweight", select=c(chol))
> obese_chol <- subset (bpa, bmicat== "Obese", select=c(chol))
>
>#Cholesterol Distribution for Underweight Category
> summary (underweight_chol)
  chol
Min.  : 95
1st Qu.:162
Median :189
Mean   :188
3rd Qu.:215
Max.   :288
NA's   :6
>
>> sapply (underweight_chol, sd, na.rm=TRUE)
  chol
35.67418
>
>#Cholesterol Distribution for Normal Weight Category
> summary(normal_weight_chol)
  chol
Min.  : 87.0
1st Qu.:166.0
Median :191.0
Mean   :193.4
3rd Qu.:217.0
Max.   :431.0
NA's   :74
> sapply (normal_weight_chol, sd, na.rm=TRUE)
  chol
41.00746
>
>#Cholesterol Distribution for Overweight Category
> summary(overweight_chol)
  chol
Min.  : 98.0
1st Qu.:172.0
Median :200.0
Mean   :201.8
3rd Qu.:227.0
Max.   :386.0
NA's   :75
>
> sapply(overweight_chol, sd, na.rm=TRUE)
  chol

```

42.13073

>

>#Cholesterol Distribution for Obese Category

> summary(obese\_chol)

chol

Min. : 83.0

1st Qu.:170.0

Median :196.0

Mean :199.2

3rd Qu.:223.0

Max. :615.0

NA's :83

> sapply (obese\_chol, sd, na.rm=TRUE)

chol

43.50734

>

>#Subsetting smoking status by BMI categories

>

> smoke\_underweight <- subset (bpa, bmicat== "Underweight", select=c(SMK))

> smoke\_normal\_weight <- subset (bpa, bmicat== "Normal Weight", select=c(SMK))

> smoke\_overweight <- subset (bpa, bmicat== "Overweight", select=c(SMK))

> smoke\_obese <- subset (bpa, bmicat== "Obese", select=c(SMK))

>

>#Finding N and % by smoking status of BMI categories

>

>#Package source code

> tblFun <- function(x){

+ tbl <- table(x)

+ res <- cbind(tbl,round(prop.table(tbl)\*100,2))

+ colnames(res) <- c('Count','Percentage')

+ res

+ }

>

> do.call(rbind,lapply(smoke\_underweight,tblFun))

Count Percentage

0 32 42.67

1 8 10.67

2 35 46.67

>

> do.call(rbind,lapply(smoke\_normal\_weight,tblFun))

Count Percentage

0 662 48.60

1 308 22.61

2 392 28.78

>

> do.call(rbind,lapply(smoke\_overweight,tblFun))

```
      Count Percentage
0  856    53.50
1  426    26.62
2  318    19.88
>
> do.call(rbind,lapply(smoke_obese,tblFun))
      Count Percentage
0  909    53.91
1  467    27.70
2  310    18.39
>
>
>#End of code
```