Report_3

Your Name

11 November 2019

Instructions:

1. Update the read.csv command so that R knows where to find the data on your computer.
2. Set the random number seed (see below) to a unique value.
3. Replace all the italic text with your own discussion (this is a focus of this report).
4. After you have completed the blood pressure vs pulse rate section, repeat the analysis for blood pressure and age (copy and paste your code, change the variable names, then update the discussion).

Introduction

The NHANES data set consists of four data files: adult.csv, youth.csv, lab.csv and exam.csv. The adult file contains information about subjects who were over 17 years old. The youth file contains information about younger subjects. The lab and exam files contain additional data about both adults and youth.

Reading the data into R

For this report, we only need to read the exam.csv file into R.

```
exam = read.csv("../nhanes/exam.csv", header=TRUE)
```

The size of this data set is

```
dim(exam)
```

```
## [1] 31311  2368
```

There are 31311 rows and 2368 columns in the exam file.

Setting a Random Number Seed.

Change the 10 in the following command to some other positive whole number. Pick a number that is different from that of other students. This value is used to randomly select a subject from the data set.

```
set.seed(10)
```

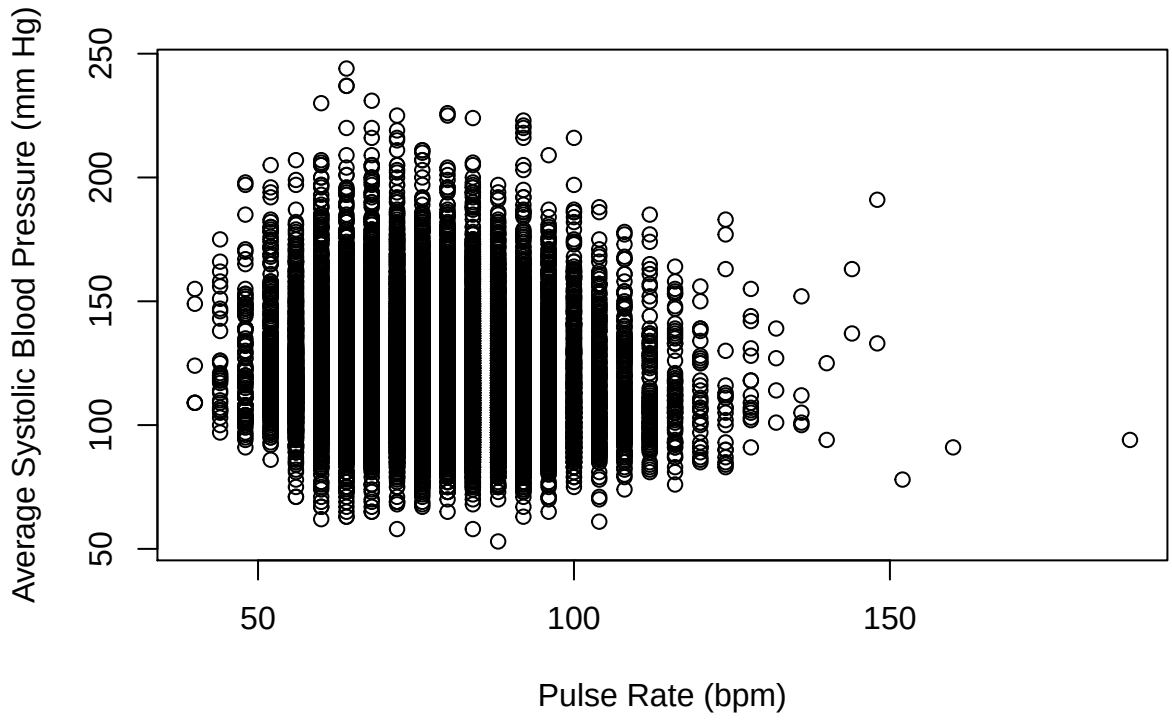
Blood Pressure and Pulse Rate

Explain what the variables PEP6DR and PEPMNK1R mean.

```
x1 <- data.frame(SEQN=exam$SEQN, PEP6DR=exam$PEP6DR, PEPMNK1R=exam$PEPMNK1R)
x2 <- na.omit(x1)
x3 <- x2[(x2$PEP6DR != 888) & (x2$PEPMNK1R != 888),  ]
```

Plot the pulse rate and average systolic blood pressure.

```
plot(x3$PEP6DR, x3$PEPMNK1R, xlab="Pulse Rate (bpm)",
     ylab="Average Systolic Blood Pressure (mm Hg)")
```



Comment on any features you notice about the graph.

Regression Line

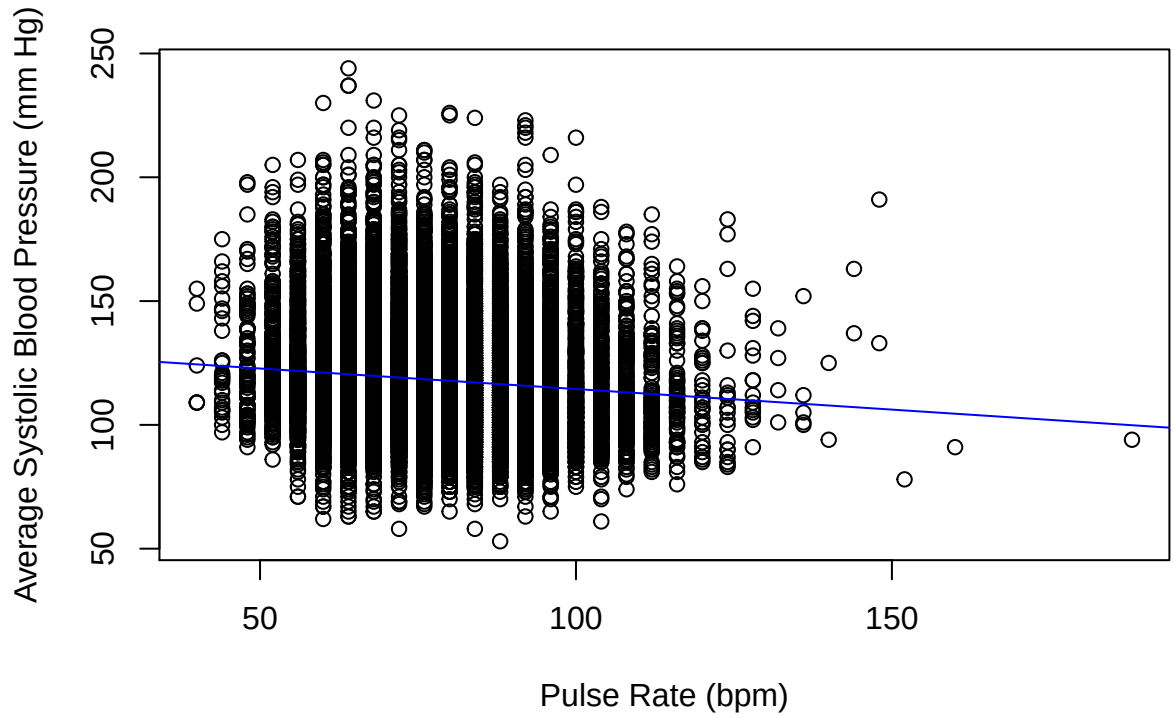
The correlation between the blood pressure and pulse rates is.

```
r = cor(x3$PEP6DR, x3$PEPMNK1R)
r
```

```
## [1] -0.1000979
```

Now compute the regression line and add it to the plot.

```
model = lm(x3$PEPMNK1R ~ x3$PEP6DR)
plot(x3$PEP6DR, x3$PEPMNK1R, xlab="Pulse Rate (bpm)",
     ylab="Average Systolic Blood Pressure (mm Hg)")
abline(model, col='blue')
```




```
summary(model)
```

```
##
## Call:
## lm(formula = x3$PEPMNK1R ~ x3$PEP6DR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.466 -14.474  -3.795  10.877 123.549
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 131.07848    0.83791  156.44   <2e-16 ***
## x3$PEP6DR    -0.16605    0.01084  -15.32   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 21.05 on 23201 degrees of freedom
## Multiple R-squared:  0.01002,    Adjusted R-squared:  0.009977
## F-statistic: 234.8 on 1 and 23201 DF,  p-value: < 2.2e-16
```

The slope of the regression line is

```
slope = as.vector(model$coefficients[2])
slope
```

```
## [1] -0.1660492
```

The y-intercept is

```
y.int = as.vector(model$coefficients[1])
y.int
```

```
## [1] 131.0785
```

The RMS error for the regression line is

```
rms = sd(x3$PEPMNK1R)*sqrt(1 - r * r)
rms
```

```
## [1] 21.05257
```

Comment on the magnitude and sign of the correlation. Do you think an increase in pulse rate causes an increase in blood pressure? Decrease? What else might account for the value of the correlation?

Study a Randomly-Selected Subject.

We randomly select a subject for which the pulse rate and average systolic blood pressure measurements are available (not blank).

```
row.number    = sample(1:dim(x3)[1], 1)
subject_id    = x3$SEQN[row.number]
subject_pulse = x3$PEP6DR[row.number]
subject_bp    = x3$PEPMNK1R[row.number]
```

The value of the HSSEX variable is

```
exam[exam$SEQN == subject_id,]$HSSEX
```

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

Comment on the meaning of this value.

The age of the subject is

```
exam[exam$SEQN == subject_id,]$HSAGEIR
```

```
## [1] 40
```

The value of DMPCREGN is

```
exam[exam$SEQN == subject_id,]$DMPCREGN
```

```
## [1] 3
```

Comment on the meaning of this value.

The value of DMARACER is

```
exam[exam$SEQN == subject_id,]$DMARACER
```

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

Comment on the meaning of this value.

We randomly selected patient 32708. This subject's pulse rate was 84 and average systolic blood pressure was 146.

Compared to all subjects, the z-score for this pulse rate is

```
(subject_pulse - mean(x3$PEP6DR))/sd(x3$PEP6DR)
```

```
## [1] 0.6063244
```

Comment on this z-score

Compared to all subjects, the z-score for this blood pressure measurement is

```
(subject_bp - mean(x3$PEPMNK1R))/sd(x3$PEPMNK1R)
```

```
## [1] 1.303734
```

Comment on this z-score

The blood pressure predicted by the regression line is

```
predicted_bp = slope * subject_pulse + y.int
predicted_bp
```

```
## [1] 117.1303
```

Compared to subjects with the same pulse rate, the z-score of this subject's pulse rate is

```
(subject_bp - predicted_bp) / rms
```

```
## [1] 1.371313
```

Comment on this z-score

Blood Pressure and Age

Explain what the variables PEPMNK1R and HSAGEIR mean. Repeat the analysis that you did for PEP6DR and PEPMNK1R using HSAGEIR and PEPMNK1R.

Regression Line

Repeat the regression line analysis that you did for blood pressure and pulse.

Study a Randomly-Selected Subject

Repeat the analysis that you did for blood pressure and pulse.