HW4

Kaylee Vo

July 17, 2025

Make sure you put your name in the author and date above.

20 points total for this section.

```
#You know the drill!
#Please set your HSS where you have the NHANES.RData and this template saved, as your working directory
#If you continue to have trouble, you could just click on the "NHANES.RData" file name under "Files" at

load("NHANES.RData")
```

Study about waist vs. hip circumference among 30+ year old males"

First, set up your study dataset:

```
#You know the drill here too!
#[1 pt] Create a new dataset called NHANES_2 according to the flowchart in Canvas - 1 line:
NHANES_2 = subset(
  NHANES,
  age>= 30 &
    sex == "Male" &
    !is.na(waist_circ) &
    !is.na(hip_circ)
)
```

```
#You know the drill here too!
#[1 pt] Check the age range - first, show that the original NHANES includes ages 0-80; then, show that
range(NHANES$age)
```

```
## [1]  0 80
```

```
range(NHANES_2$age)
```

```
## [1] 30 80
```

```
#[1 pt] Check the sex - first, show that the original NHANES includes females and males; then, show tha
table(NHANES$sex, exclude=F)
```

```
##
##   Male Female
##   4946   5054
```

```r
table(NHANES_2$sex, exclude=F)
```

```
##
##   Male Female
##   3050      0
```

```r
#[1 pt] Check for missing data for waist_circ - first, show the number of missing data for  in the orig
sum(is.na(NHANES$waist_circ))
```

```
## [1] 641
```

```r
sum(is.na(NHANES_2$waist_circ))
```

```
## [1] 0
```

```r
#[1 pt] Check for missing data for hip_circ - first, show the number of missing data for  in the origin
sum(is.na(NHANES$hip_circ))
```

```
## [1] 1720
```

```r
sum(is.na(NHANES_2$hip_circ))
```

```
## [1] 0
```

Evaluate the linear relationship between x=waist_circ (continuous) and y=hip_circ (continuous)

```r
#[5pts] Plot the scatterplot and the linear regression line - 2 lines:
plot(
  NHANES_2$waist_circ,
  NHANES_2$hip_circ,
  main="Regression of Waist Circ on Hip Circ",
  xlab="Waist Circumference",
  ylab="Hip Circumference"
)
abline(lm(hip_circ ~ waist_circ, data=NHANES_2), col = "red")
```

Regression of Waist Circ on Hip Circ
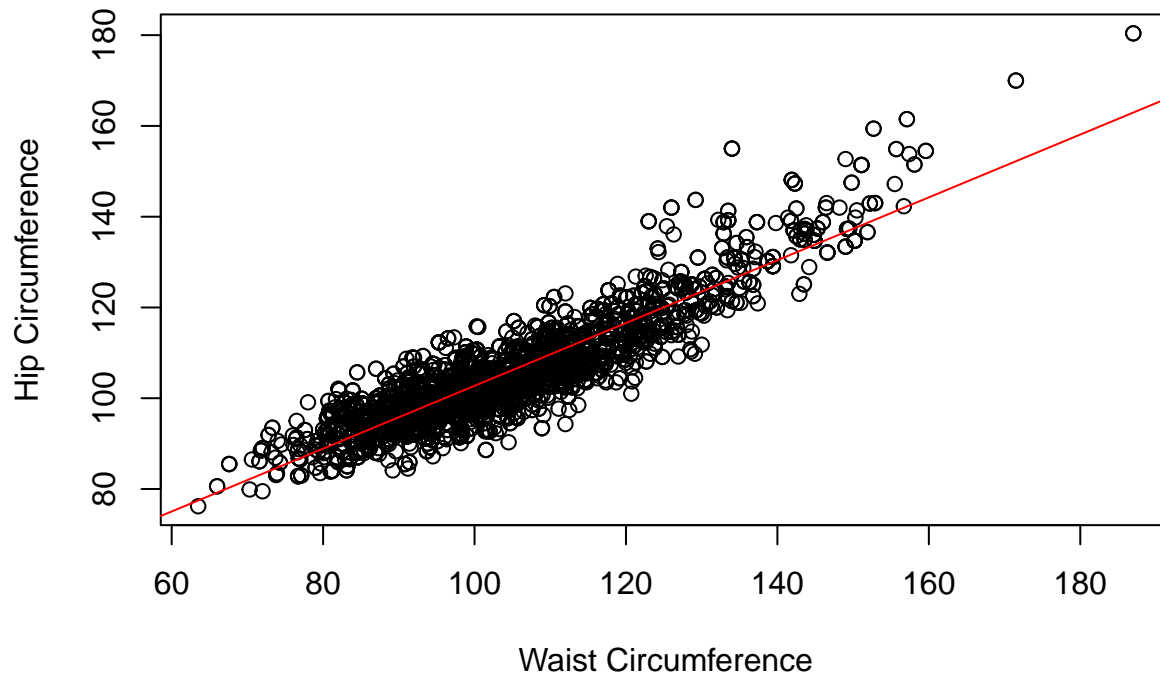

```r
#[5pts] Perform lm, save the results as a new object called "linreg" and evaluate the lm results with t
linreg = lm(hip_circ ~ waist_circ, data=NHANES_2)
summary(linreg)
```

```
##
## Call:
## lm(formula = hip_circ ~ waist_circ, data = NHANES_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.7353  -3.7741  -0.4235   3.4887  28.7382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.518465   0.648418   51.69   <2e-16 ***
## waist_circ   0.692115   0.006129  112.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.382 on 3048 degrees of freedom
## Multiple R-squared:  0.8071, Adjusted R-squared:  0.807
## F-statistic: 1.275e+04 on 1 and 3048 DF,  p-value: < 2.2e-16
```

```r
#[5pts] Plot the residual plots - 1 line:
#For your discussion, you can focus on the first two plots (1) Residuals vs. Fitted plot and (2) Q-Q pl
plot(linreg)
```

Residuals vs Fitted

lm(hip_circ ~ waist_circ)

Q−Q Residuals

lm(hip_circ ~ waist_circ)

4

Scale−Location

√|Standardized residuals|

90122

Fitted values
lm(hip_circ ~ waist_circ)

Residuals vs Leverage

Standardized residuals

7059

3004

Cook's distance

Leverage
lm(hip_circ ~ waist_circ)