---

title: "Correlation Using R"

output: html\_document

---

#1. Read in a csv file and assign it to a data frame object#

```{r}

health <- read.csv("cdc.csv")

str(health)

```

#2. Take a look a the variables of interest#

```{r}

library("lattice") # load lattice

histogram(~health$weight)

histogram(~health$height)

```

#3. Plot the two variables in a scatterplot#

```{r}

# Women and Men

xyplot(health$height ~ health$weight)

# Add a linear fit line

xyplot(height ~ weight, data = health,

main = "Weight and Height",

type = c("p", "r"))

# Subset by groups (a factor)

xyplot(height ~ weight, data = health,

main = "Weight and Height",

type = c("p", "r"),

groups = factor(gender, labels = c("Women", "Men")))

# Add group labels

xyplot(height ~ weight, data = health,

main = "Weight and Height",

type = c("p", "r"),

groups = factor(gender, labels = c("Women", "Men")),

auto.key = list(columns = 2))

# Subset by a continuous variable

xyplot(height ~ weight, data = health,

main = "Weight and Height",

groups = age)

# Plot separate plots for each level of a factor

xyplot(height ~ weight | genhlth, data = health,

main = "Weight and Height",

type = c("p", "r"))

```

#4. Calculate independent variance estimates and assign them to objects#

```{r}

x <- var(health$weight)

y <- var(health$height)

```

#5. Calculate the covariance estimate and assign to objects#

```{r}

xycov <- cov(health$weight, health$height)

```

#6. Calculate the linear correlation using Pearson's \*r\* and assign to objects#

```{r}

rxy <- xycov / sqrt(x \* y)

rxy # take a look at the correlation value

r <- cor(health$weight, health$height)

r # take a look at the correlation value

```

#7. Create vector objects and combine them into a Data Frame object#

```{r}

failures <- c(5, 10, 3)

sleepless <- c(2, 3, 1)

failures; sleepless

df <- data.frame(failures, sleepless)

```

#7. Calculate Pearson's \*r\* and assign to an object#

```{r}

rsleep <- cor(df$failures, df$sleepless)

rsleep

# take a look at a scatterplot

xyplot(failures ~ sleepless, data = df, type = c("p", "r"))

```