**Part A:** Use the FHS dataset (frmgham2.csv) to complete the work below:

1. Read in the frmgham2.csv dataset. Create a new dataset (subset of the larger dataset) including only PERIOD=1.

```
## first set working directory
setwd("~/Dropbox/AAU materials/datasets/FHS")

## double check this
getwd()

## load the data
FHS <- read.csv("~/Dropbox/AAU materials/datasets/FHS/frmgham2.csv")

## look at the data
head(FHS)

## subset to PERIOD=1
FHS1 <- FHS[FHS$PERIOD==1,]
dim(FHS1)

## OR use the subset function
FHS1 <- subset(FHS, PERIOD==1)
dim(FHS1)
```

2. What is the average age & 95% CI? Be sure to interpret.

The average age in the FHS dataset (restricted to period 1) is 49.93 with a 95% CI of (49.67, 50.18).

We are 95% confident that the true mean age of the population is between 49.67 and 50.18.

```
t.test(FHS1$AGE)
```

3. There is a variable in the FHS dataset that defines people who have CVD at entry (Prevalent CVD).
    a. What is the age (95%CI) for those with prevalent CVD?

The mean age for those with prevalent CVD is 53.52 with a 95% CI of (53.04, 54.00).

```
## first subset the data
FHS1.CDV <- subset(FHS1, CVD==1)

## then run t.test to get 95% CI for the mean
t.test(FHS1.CDV$AGE)
```

b. What is the age (95% CI) for those without prevalent CVD?

The mean age for those with prevalent CVD is 48.66 with a 95% CI of (48.37, 48.95).

```
## first subset the data
FHS1.noCDV <- subset(FHS1, CVD==0)

## then run t.test to get 95% CI for the mean
t.test(FHS1.noCDV$AGE)
```

**Part B:** Use the FHS dataset (frmgham2.csv) for Period 1 to complete the work below:

1. What is the mean BMI in the FHS dataset? *Hint: Remember how to remove missing values!*

The mean BMI is 25.85

```
mean(FHS1$BMI,na.rm=TRUE)
```

2. We want to know whether the mean BMI is different than 25.
   a. What are the null and alternative hypotheses? Define your notation.

$$H_0: \mu = 25 \; vs. \; H_1: \mu \neq 25$$

where $\mu$ is the population (Framingham) mean BMI

   b. Now, answer this question using a hypothesis test. Be sure to state the: test statistic, p-value, and conclusion.

The test statistic is 25.85 with a p-value of <0.001. We conclude the BMI is significantly different that 25.

```
t.test(FHS1$BMI,mu=25,na.rm=TRUE)
```

   c. Answer this question based on the 95% CI. Justify your answer.

The 95% confidence interval is (25.73, 25.97), which does not cover 25.