Solutions

Question 1

```
set.seed(1)
y <- rnorm(5,2,1)
sims <- 10000
means <- numeric(sims)
for (i in 1:sims) {
  temp <- sample(y,length(y),replace=TRUE)
  means[i] <- mean(temp)
}
means<-sort(means)
c(means[sims*0.025],means[sims*0.975])
qnorm(c(0.025,0.975), mean(y), 1/sqrt(length(y)))
```

Exact confidence interval is [1.25, 3.00]. Bootstrap interval is [1.44, 2.90]

For $n = 50$, exact confidence interval is [1.82, 2.38]. Bootstrap interval is [1.87, 2.32]

For $n = 200$, exact confidence interval is [1.90, 2.16]. Bootstrap interval is [1.90, 2.17]

We see the bootstrap interval becoming very close to the true interval for moderate sample sizes.

Question 2

The MLE point estimate is. $60/100 = 0.6$. For the confidence interval:

```
y <- c(rep(1,60),rep(0,40))
sims <- 10000
means <- numeric(sims)
for (i in 1:sims) {
  temp <- sample(y,length(y),replace=TRUE)
  means[i] <- mean(temp)
}
means<-sort(means)
```

```
c(means[sims*0.025],means[sims*0.975])
```

Giving an interval of [0.51, 0.69]. For $n = 1000$ this becomes. [0.57,0.63]

Question 3

```
sims <- 10000
meandiffs <- numeric(sims)
for (i in 1:sims) {
  y1temp <- sample(y1,length(y1),replace=TRUE)
  y2temp <- sample(y2,length(y2),replace=TRUE)
  meandiffs[i] <- mean(y1temp)-mean(y2temp)
}
meandiffs<-sort(meandiffs)
c(meandiffs[sims*0.025],meandiffs[sims*0.975])
```

Confidence interval is $[1.50, 17.11]$. This doesnt include 0, so the difference in means is significant at the 0.95% level

Question 4

```
n <- dim(cd4)[1]
index <- 1:n
sims <- 10000
beta0s <- numeric(sims)
beta1s <- numeric(sims)

for (i in 1:sims) {
  index_star <- sample(index, size=n, replace=TRUE)
  cd4_star <- cd4[index_star, ]
  mod <- lm(oneyear~baseline,data=cd4_star)
  beta0s[i] <- mod$coef[1]
  beta1s[i] <- mod$coef[2]
}
beta0s  <- sort(beta0s)
beta1s  <- sort(beta1s)
```

Confidence intervals are:

```
> c(beta0s[sims*0.025],beta0s[sims*0.975])
[1] -0.5653221  1.8733597
> c(beta1s[sims*0.025],beta1s[sims*0.975])
[1] 0.6949909 1.4341808
```