

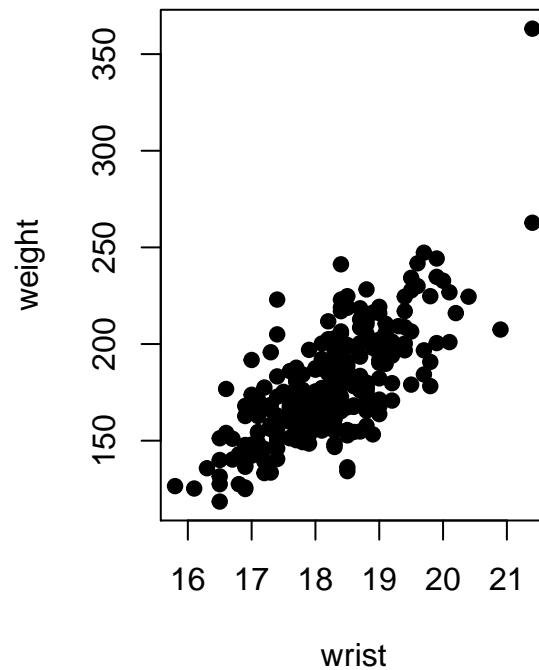
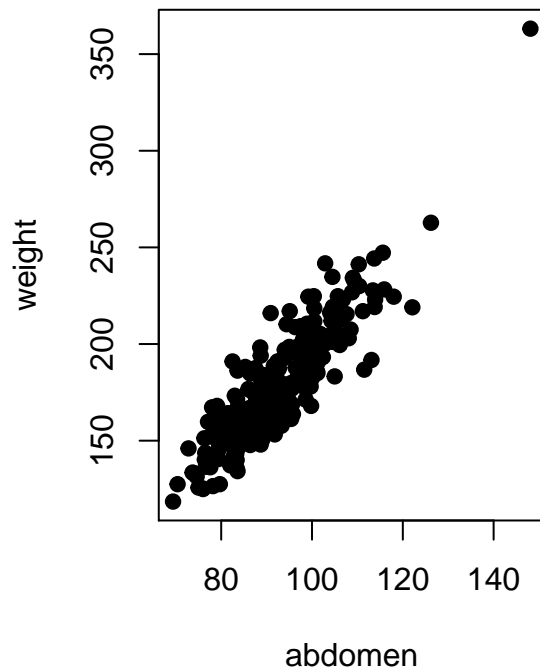
## lecture\_5

### R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#####  
##### READING IN BODYFAT DATA #####  
#####  
  
data <- read.table("data/bodyfat (1).csv",header=T,sep=",")  
attach(data)  
data[1:5,]  
  
## density bodyfat age weight height neck chest abdomen hip thigh knee ankle  
## 1 1.0708 12.3 23 154.25 67.75 36.2 93.1 85.2 94.5 59.0 37.3 21.9  
## 2 1.0853 6.1 22 173.25 72.25 38.5 93.6 83.0 98.7 58.7 37.3 23.4  
## 3 1.0414 25.3 22 154.00 66.25 34.0 95.8 87.9 99.2 59.6 38.9 24.0  
## 4 1.0751 10.4 26 184.75 72.25 37.4 101.8 86.4 101.2 60.1 37.3 22.8  
## 5 1.0340 28.7 24 184.25 71.25 34.4 97.3 100.0 101.9 63.2 42.2 24.0  
## biceps forearm wrist  
## 1 32.0 27.4 17.1  
## 2 30.5 28.9 18.2  
## 3 28.8 25.2 16.6  
## 4 32.4 29.4 18.2  
## 5 32.2 27.7 17.7  
  
par(mfrow=c(1,2))  
plot(abdomen,weight,pch=19)  
plot(wrist,weight,pch=19)
```

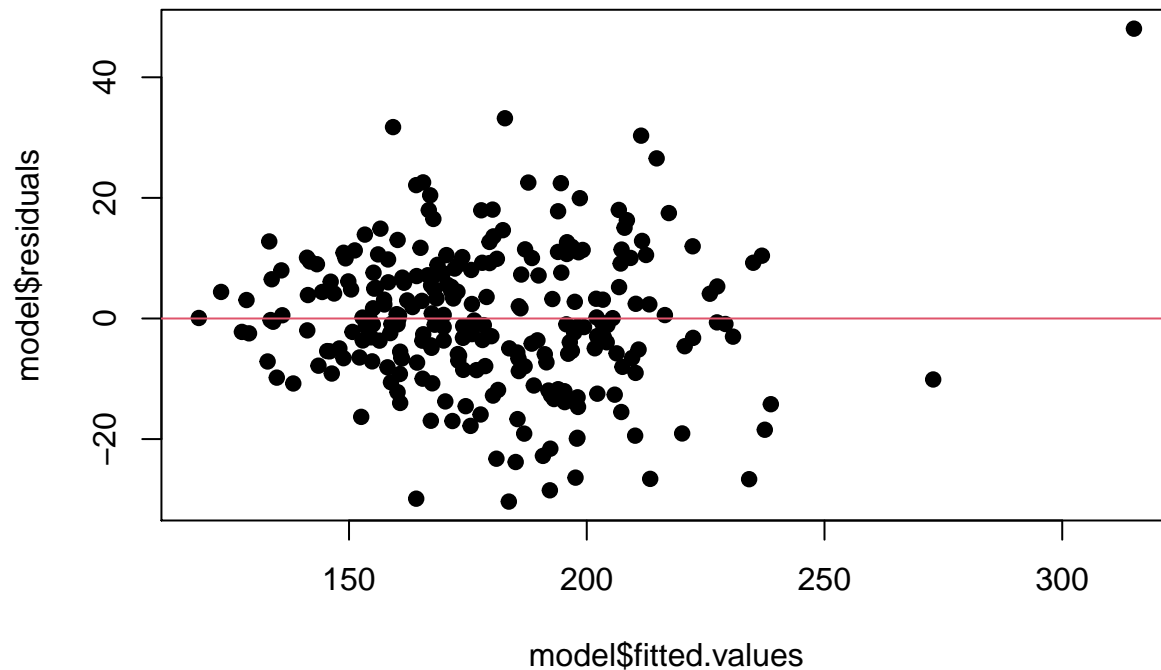


```
#####
##### MLE FIT OF REGRESSION MODEL #####
#####
```

```
model <- lm(weight~abdomen+wrist)
summary(model)
```

```
##
## Call:
## lm(formula = weight ~ abdomen + wrist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30.362  -6.730  -0.593   7.598  48.059
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -166.69048    14.90795  -11.181  <2e-16 ***
## abdomen      1.92816     0.08783   21.954  <2e-16 ***
## wrist        9.16918     1.01440    9.039  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.77 on 249 degrees of freedom
## Multiple R-squared:  0.8408, Adjusted R-squared:  0.8395
## F-statistic: 657.4 on 2 and 249 DF, p-value: < 2.2e-16

par(mfrow=c(1,1))
plot(model$fitted.values,model$residuals,pch=19)
abline(h=0,col=2)
```



```
#####
### SAMPLING FROM POSTERIOR DISTRIBUTION ###
#####
```

```
beta.hat <- model$coef
n <- length(weight)
p <- length(beta.hat)
s2 <- (n-p)*summary(model)$sigma^2
V.beta <- summary(model)$cov.unscaled

numsamp <- 1000
beta.samp <- matrix(NA,nrow=numsamp,ncol=p)
sigsq.samp <- rep(NA,numsamp)
for (i in 1:numsamp){
  temp <- rgamma(1,shape=(n-p)/2,rate=s2/2)
  cursigsq <- 1/temp
  curvarbeta <- cursigsq*V.beta
  curvarbeta.chol <- t(chol(curvarbeta))
  z <- rnorm(p,0,1)
  curbeta <- beta.hat+curvarbeta.chol%*%z
  sigsq.samp[i] <- cursigsq
  beta.samp[i,] <- curbeta
}
```

```
#####
### SUMMARIZING POSTERIOR DISTRIBUTIONS ###
#####
```

```
## posterior means
postmean.beta <- apply(beta.samp,2,mean)
postmean.sigsq <- mean(sigsq.samp)
```

```

postmean.beta

## [1] -166.751516    1.933480    9.144747
postmean.sigsq

## [1] 139.6963

## posterior correlation between variables
cor(cbind(beta.samp,sigsq.samp))

##                                sigsq.samp
##          1.00000000  0.20923380 -0.89685488  0.02105217
##          0.20923380  1.00000000 -0.61836064  0.04557855
##          -0.89685488 -0.61836064  1.00000000 -0.03470097
## sigsq.samp  0.02105217  0.04557855 -0.03470097  1.00000000

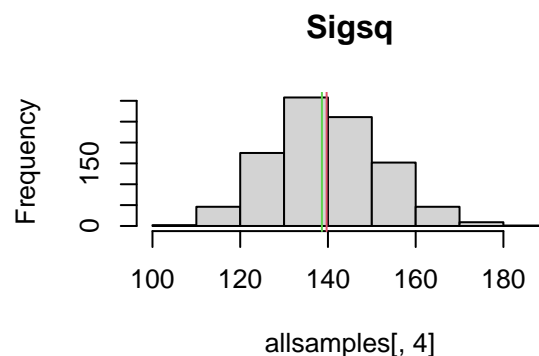
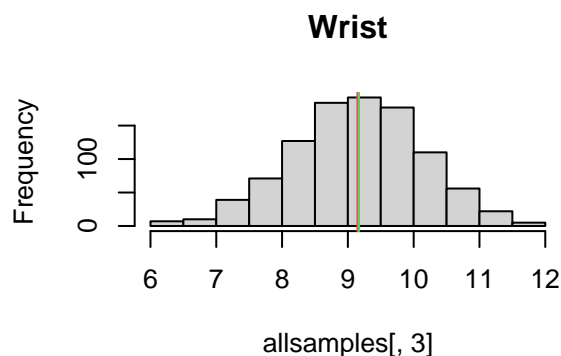
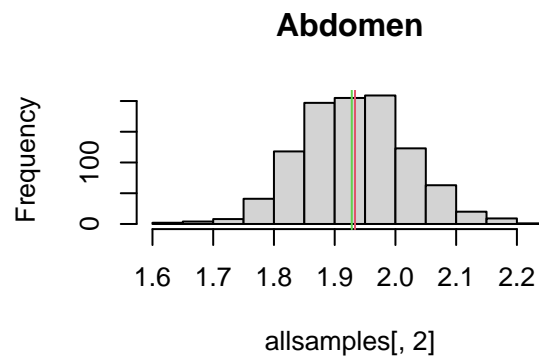
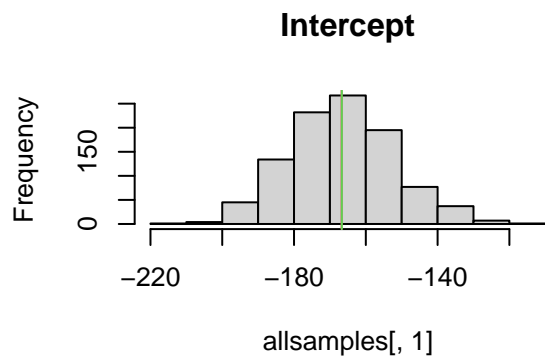
## 95% posterior intervals
allsamples <- cbind(beta.samp,sigsq.samp)
allsamples.sort <- apply(allsamples,2,sort)
allsamples.sort[25,]

##                                sigsq.samp
## -193.624906    1.766581    7.165422  117.788463
allsamples.sort[975,]

##                                sigsq.samp
## -136.783427    2.105086   11.028066  164.287332

## posterior histograms
par(mfrow=c(2,2))
hist(allsamples[,1],main="Intercept")
abline(v=postmean.beta[1],col=2)
abline(v=beta.hat[1],col=3)
hist(allsamples[,2],main="Abdomen")
abline(v=postmean.beta[2],col=2)
abline(v=beta.hat[2],col=3)
hist(allsamples[,3],main="Wrist")
abline(v=postmean.beta[3],col=2)
abline(v=beta.hat[3],col=3)
hist(allsamples[,4],main="Sigsq")
abline(v=postmean.sigsq,col=2)
abline(v=summary(model)$sigma^2,col=3)

```



```
#####
#### POSTERIOR PREDICTIVE DISTRIBUTION ####
#### FOR A NEW COVARIATE VECTOR Xstar #####
#####

Xstar <- c(1,130,17) # new person with abdomen = 130 and wrist = 17
Xstar <- t(Xstar)    # making it a row vector

## use posterior samples from before:

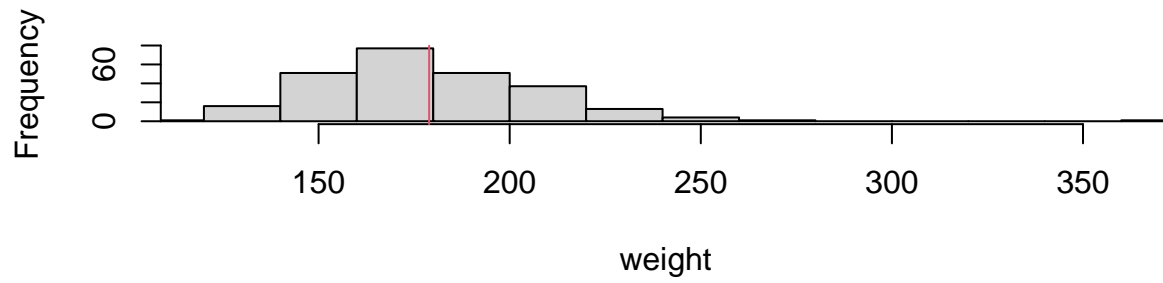
ystar.samp <- rep(NA,numsamp)
for (i in 1:numsamp){
  xstarbeta <- Xstar%*%t(t(beta.samp[i,]))
  ystar.samp[i] <- rnorm(1,mean=xstarbeta,sd=sqrt(sigsq.samp[i]))
}

ystar.postmean <- mean(ystar.samp)
ystar.postmean

## [1] 240.3006

par(mfrow=c(2,1))
xmin <- min(weight,ystar.postmean)
xmax <- max(weight,ystar.postmean)
hist(weight,main="Dataset Weights",xlim=c(xmin,xmax))
abline(v=mean(weight),col=2)
hist(ystar.samp,main="Predicted Weight of Person with abdomen = 130 and wrist = 17",xlim=c(xmin,xmax))
abline(v=ystar.postmean,col=2)
```

### Dataset Weights



### Predicted Weight of Person with abdomen = 130 and wrist = 17

