CSI-674 Jericho McLeod
Homework 10 G00986513

Problem 1:

The precision, Rho, is from a gamma distribution with alpha 14.5 and beta 0.00472.

Given Rho, the intercept, Eta, is a normal distribution with mean 32 and precision 31*Rho.

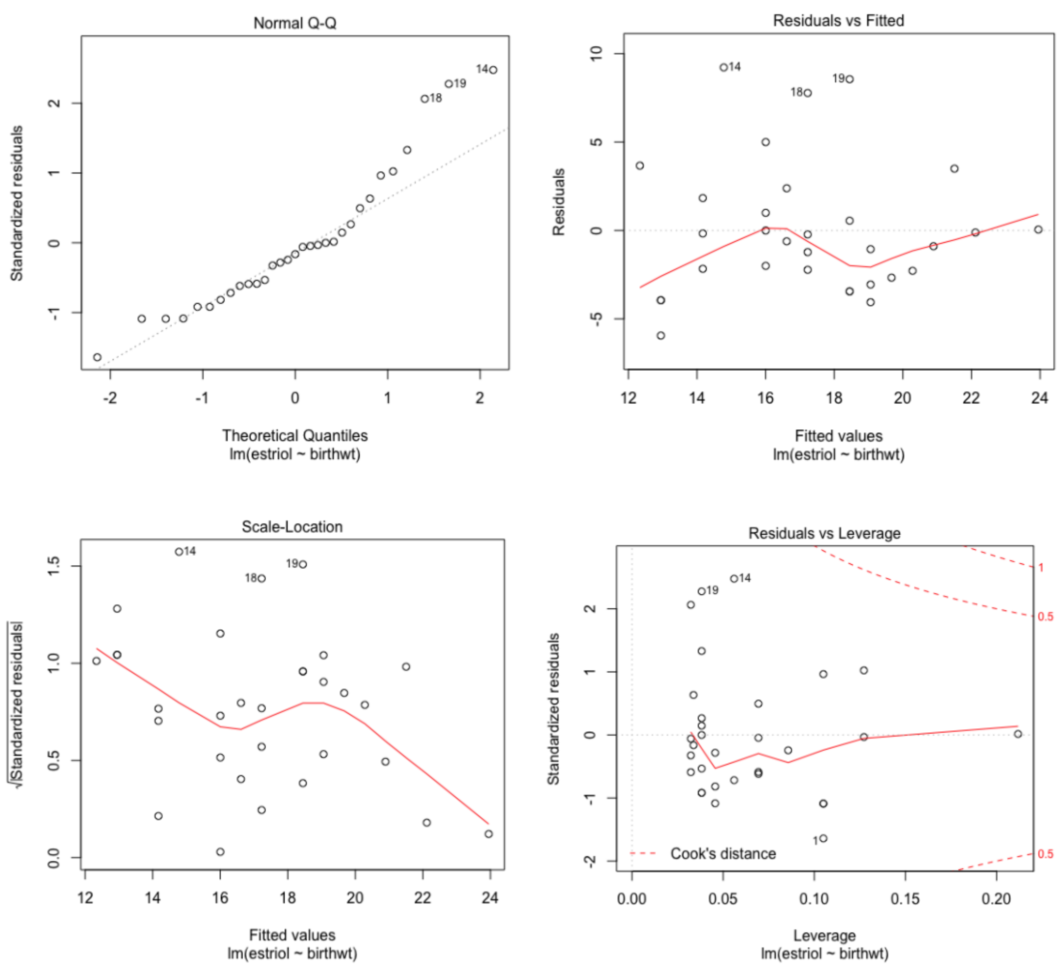
Given Rho, the slope, Beta, is also a normal distribution with mean 0.6082 and precision 677.419*Rho.

R Code used for Problem 1:

```
library(data.table)
data <- fread('http://www.biostat.umn.edu/~lynn/iid/estriol.dat')
head(data)

x <- data$estriol
y <- data$birthwt
n <- length(x)

xbar <- mean(x)
ybar <- mean(y)

b <- sum( (x-xbar) * (y-ybar)) / sum( (x-xbar)^2 )
a <- ybar - b*xbar

sxx <- sum( (x-xbar)^2 )
syy <- sum( (y-ybar)^2 )
sxy <- sum( (x-xbar) * (y-ybar) )
see <- sum( (y-ybar-b*(x-xbar))^2)

#posteriors
rho_alpha <- (n-2)/2
rho_beta  <- 2/see

eta_center <- ybar
eta_spread <- (n*(n-2)/see)^-0.5
eta_degf   <- n-2

beta_center <- b
beta_spread <- (see*(n-2)/see)^-0.5
beta_degf   <- n-2

s2<- see/(n-2)
s <- sqrt(s2)
```

Problem 2:

I do not think the assumptions for normal linear regression are met. The QQ plot shows a concerning fluctuation of residuals for higher quantiles. The fitted residuals show a somewhat non-linear pattern that appears to be more of a third order polynomial than linear. The scale location plot shows that the data is not homoskedactic. The residuals versus leverage looks reasonable; there are some influential data points but they tend to have low leverage.



R Code for Problem 2:

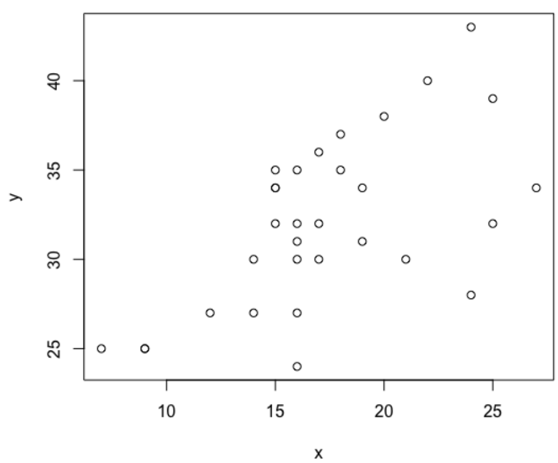
```
fitted <- lm(formula= estriol ~ birthwt,data=data)
plot(fitted)
```

Jericho McLeod
G00986513

Problem 3:

The predictive distribution for birthweight given milligrams of estriol is a nonstandard T distribution with center 33.079, spread 3.891, and 29 degrees of freedom. A 90% credible interval for this distribution is from 26.468 to 39.690.

A quick comparison of the actual data plotted shows that this appears to be a reasonable range given the data; meaning there are no grossly misstated values.



R Code used for Problem 3:

```
xnew = 19
pred_center <- xnew*b+a
pred_spread <- sqrt( (((xnew-xbar)^2/sxx)+(1/n)+1) * (see/(n-2)))
pred_df     <- n-2
pred_ci <- qt(c(0.05,0.95),df=pred_df)*pred_spread
pred_ci <- pred_ci + pred_center
pred_ci
plot(x,y)
```

Problem 4:

The Monte Carlo simulation draws 4 values: a rho, beta, and eta for the values of the linear regression, and then the expected Y value from a normal distribution where the mean depends on eta and beta and the standard deviation depends on rho.

The result is 26.766 to 39.691, which is very nearly the same as the exact predictive distribution. Increasing the Monte Carlo simulation length to 1,000,000 draws makes the credible interval be within 0.01% of the exact answer. This indicates that the exact marginal distribution for the values does in fact give a regression, and the Monte Carlo verifies it. Given the ease of doing either, it is worthwhile to just calculate the exact values in this case, though there are likely situations where the complexity is such that it is simpler to use the Monte Carlo method.

Code used for Problem 4:

```
numSim <- 1000000
rho_k  <- rgamma(numSim,shape=((n-2)/2),scale=(2/see))
eta_k  <- rnorm(numSim,mean=ybar,sd=sqrt(1/(n*rho_k)))
beta_k <- rnorm(numSim,mean=b,sd=sqrt(1/(sxx*rho_k)))
y_sim  <- rnorm(numSim,mean=(eta_k+beta_k*(19-xbar)),sd=sqrt(1/rho_k))
quantile(y_sim,c(0.05,0.95))
density(y_sim)
plot(density(y_sim))
```