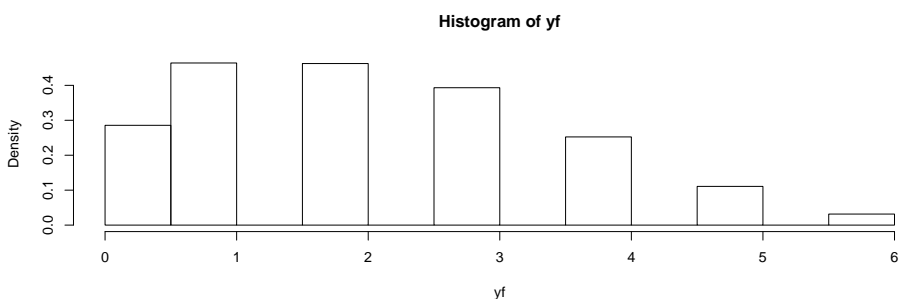
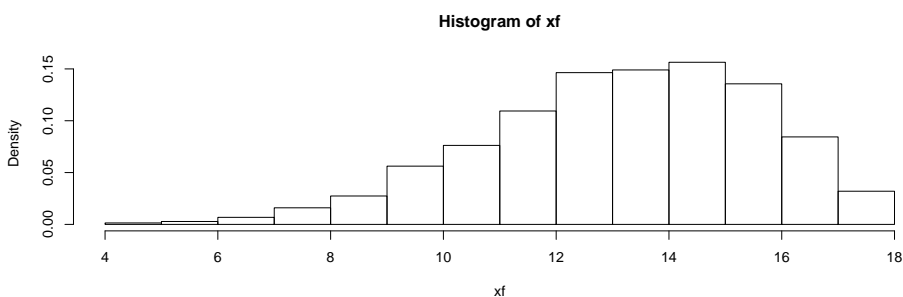


Example 2×2 table. Population 1: $m = 18, x = 14$.

Population 2: $n = 6, y = 2$. Beta(0.5,0.5) prior. Lets predict a future value for X and for Y

```
pist <- rbeta(5000,14.5,4.5)
  xf <- rbinom(5000,size=18,prob=pist)
rhst <- rbeta(5000,2.5,4.5)
yf <- rbinom(5000,size=6,prob=rhst)
par(mfrow=c(2,1))
hist(xf,prob=T)
hist(yf,prob=T)
mean(xf)
[1] 13.6744
mean(yf)
[1] 2.1608
```



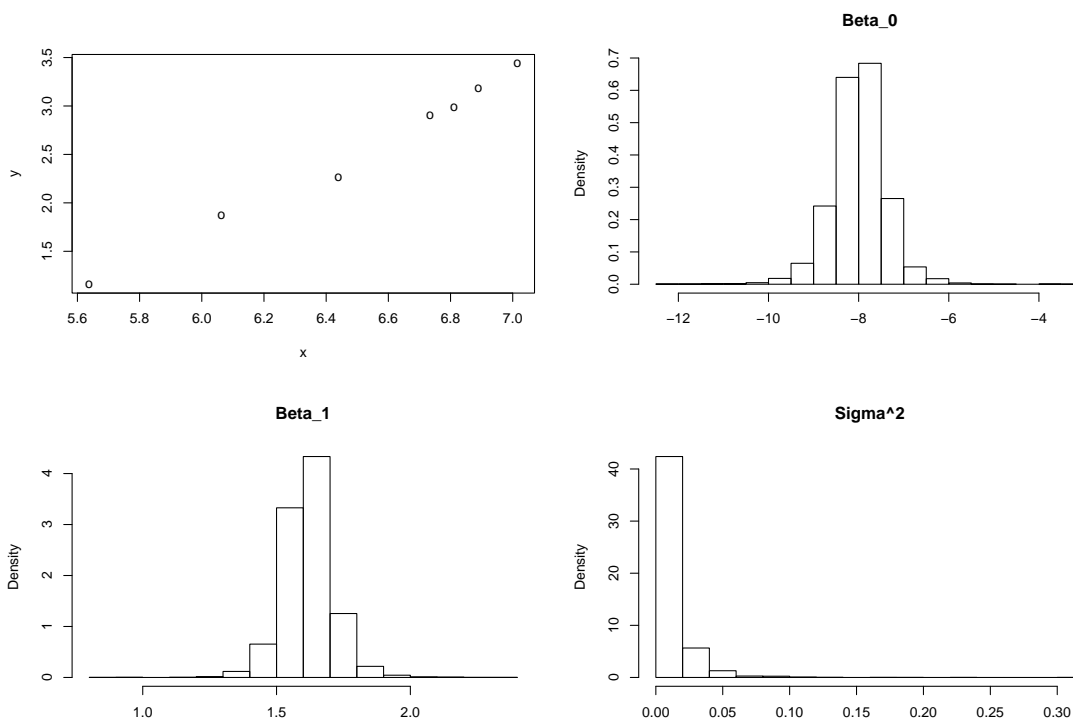
Example: Simple Linear Regression. Y is the logarithm of Body Mass (kg) and X is the logarithm of Metabolic rate (kcal/day) on a set of dogs.

Model: $Y = \beta_0 + \beta_1 X + \epsilon, \epsilon \sim N(0, \sigma^2)$

Non-informative prior: $p(\beta_0, \beta_1, \sigma^2) \propto 1/\sigma^2$

Figure shows:

- Scatter plot of Y vs X .
- Histogram of posterior samples for β_0 .
- Histogram of posterior samples for β_1 .
- Histogram of posterior samples for σ^2 .

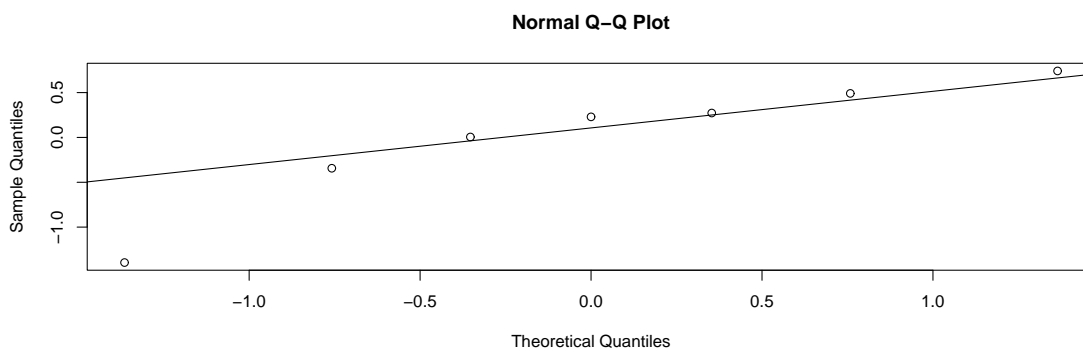
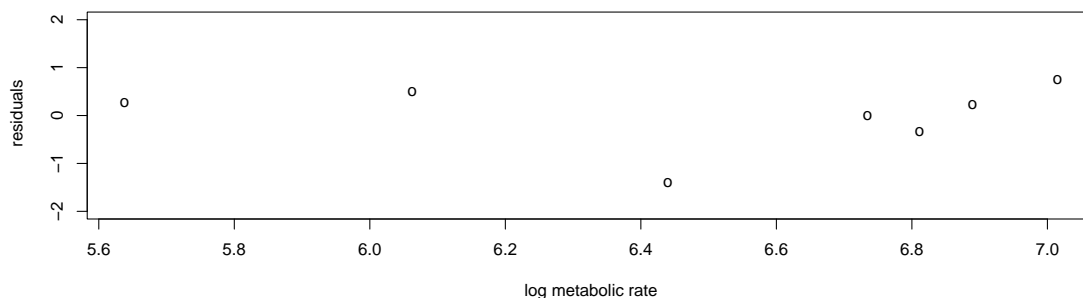


Summarizing the posterior simulations

	Posterior Quantiles				
	2.5%	25%	median	75%	97.5%
beta_0	-9.269	-8.337	-7.976	-7.641	-6.771
beta_1	1.431	1.564	1.615	1.670	1.811
sigma	0.055	0.078	0.095	0.122	0.221

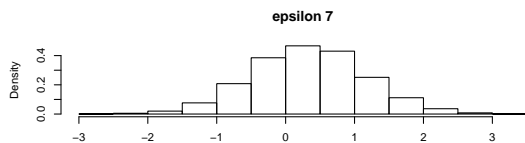
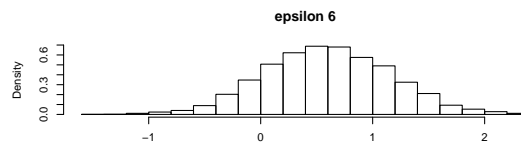
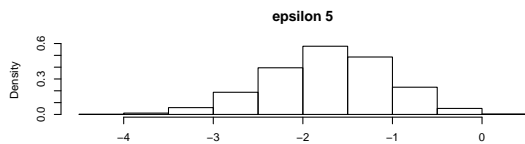
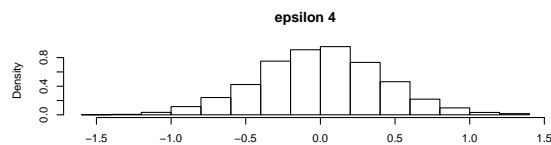
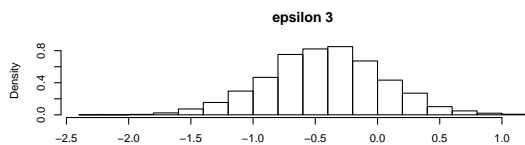
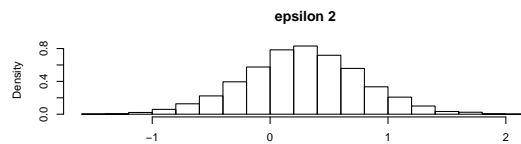
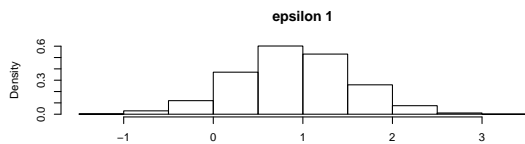
Residual Analysis: Following pictures consider the standardized residuals $(Y - X\hat{\beta})/\hat{s}$ where $\hat{\beta} = (X^t X)^{-1} X^t Y$

- Values of X vs. residuals
- Normal quantile plot of the residuals.

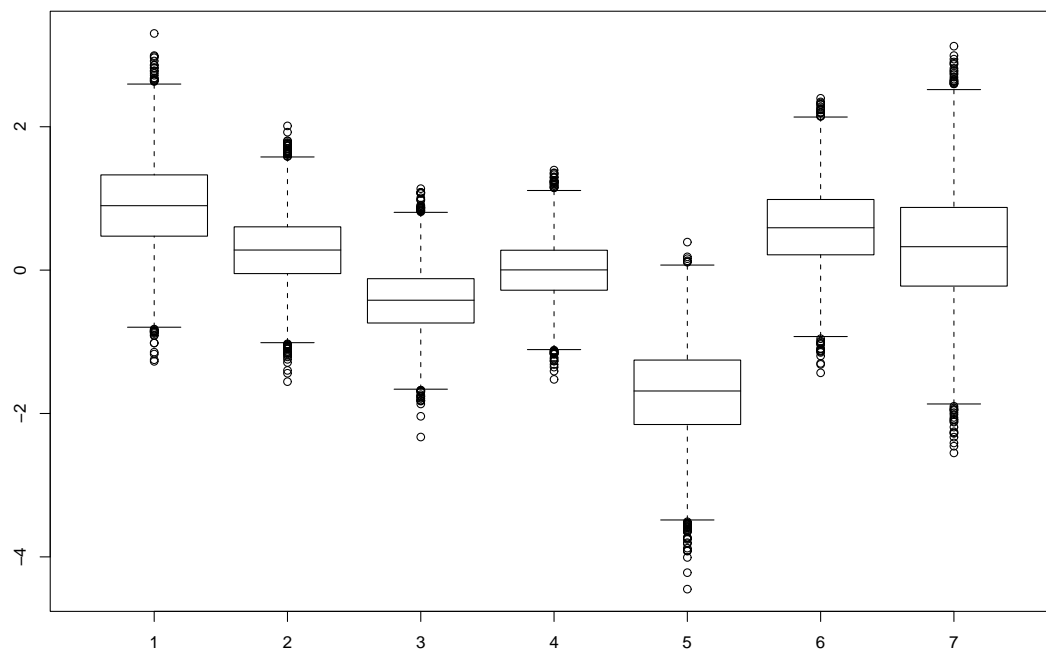


For each sampled value of (β, σ^2) from the posterior distribution, we may compute, $\epsilon^* = (Y - \beta X)/\sigma$ and produces samples for the errors in the model.

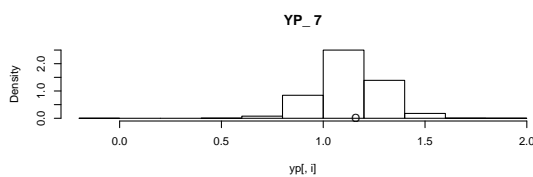
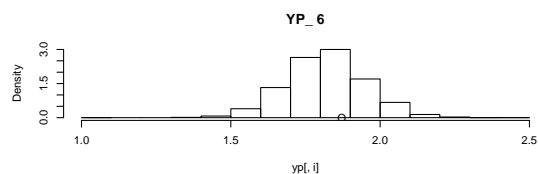
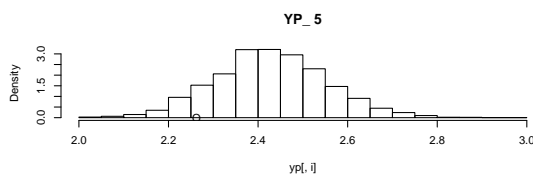
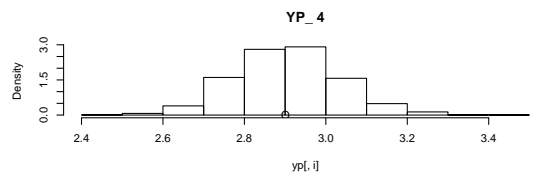
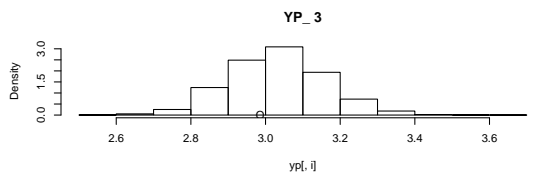
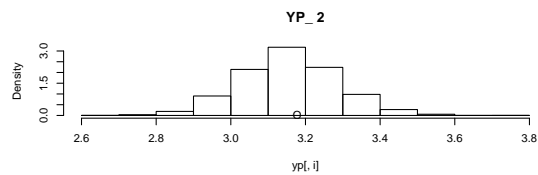
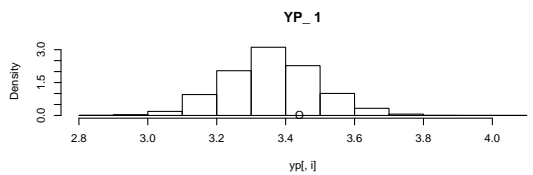
- Histograms of the sampled values for $\epsilon_1, \epsilon_2, \dots, \epsilon_7$.



- Boxplots for $\epsilon_1, \epsilon_2, \dots, \epsilon_7$.



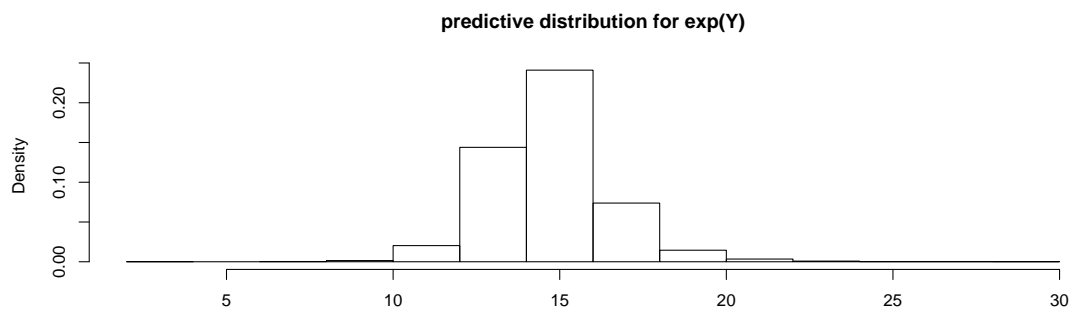
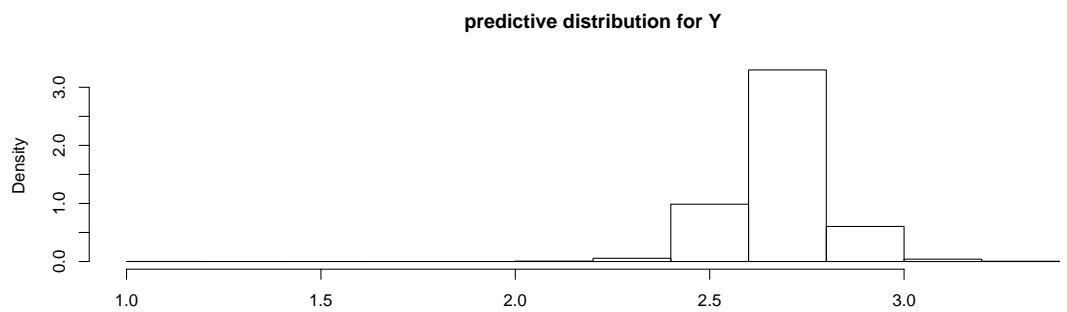
For predictive checks of the model, for each draw of (β, σ^2) , we may draw a new vector $Y^p \sim N(X\beta, \sigma^2)$. We may compare the samples of Y^p to the actual response values Y .



Predict value of Y corresponding to a new value of X ,
 $X = 6.6$ (or 735.0952 in kcal/day).

Figure shows:

- Histogram of values from predictive distribution for Y .
- Histogram of values from predictive distribution for $\exp(Y)$.



Splus/R code

```
# Example Pag. 265 of the book 'Bayesian Data Analysis'
# y=Body mass (kg)    x= metabolic rate (kcal/day)
y <- c(31.2,24,19.8,18.2,9.61,6.5,3.19)
y <- log(y)
n <- length(y)
x <- c(1113.2,981.8,908.2,840.8,626.2,429.5,280.9)
x <- log(x)
plot(x,y,pch="o")
# Preparing objects required for regression analysis
Xdes <- cbind(rep(1,n),x)
# the inverse of t(X)X and its Chol dec.
XX.t <- (t(Xdes)%*%Xdes)
XX.inv <- solve(t(Xdes)%*%Xdes)
XX.inv <- 0.5*(XX.inv + t(XX.inv))
XX.inv.cd <- t(chol(XX.inv))
# the MLE for ph
betahat <- XX.inv%*%t(Xdes)%*%y
resids <- (y - (Xdes)%*%betahat)
ssq <- sum (resids*resids)
#
```

```

# postreg is the function that
# samples values for beta and sigma^2
postreg <- function (m)
{
betahatmat <- matrix(betahat,2,m)
v <- rchisq(m,(n-2))
sg <-ssq/v
# generation of betas
norbi <- matrix(rnorm(2*m),ncol=m)
normv <- t(sqrt(sg)*t(norbi))
beta <- (XX.inv.cd%*%normv) + betahatmat
#result
beta <- t(beta)
return(beta,sg)
}
# obtaining 5000 samples from posterior
pst <- postreg(5000)
# summarzing the posterior distribution
  quantile(pst$beta[,1],probs=c(.025,.25,.5,.75,.975))
      2.5%      25%      50%      75%      97.5%
-9.269768 -8.337061 -7.976879 -7.641112 -6.771438
  quantile(pst$beta[,2],probs=c(.025,.25,.5,.75,.975))
      2.5%      25%      50%      75%      97.5%
1.431631 1.564013 1.615447 1.670739 1.811489
  quantile(sqrt(pst$sg),probs=c(.025,.25,.5,.75,.975))
      2.5%      25%      50%      75%      97.5%

```

```

0.05548276 0.07851021 0.09584138 0.12283993 0.22140462
# plotting the standardized residuals
plot(x,resids/sqrt(mean(pst$sg)),pch="o",
xlab="log metabolic rate",ylab=" residuals",
ylim=c(-2,2))
qqnorm(resids/sqrt(mean(pst$sg)))
qqline(resids/sqrt(mean(pst$sg)))
# computing sampled values for epsilons
eps <-matrix(NA,5000,7)
  for(i in 1:5000)
{ eps[i,] <- (y - Xdes%*%pst$beta[i,])/sqrt(pst$sg[i]) }
par(mfrow=c(4,2))
for(i in 1:7){hist(eps[,i],prob=T)}
  boxplot(eps[,1],eps[,2],eps[,3],eps[,4],eps[,5],
eps[,6],eps[,7])

```

```
# performing predictive checks
yp <- matrix(NA,5000,7)
for(i in 1:5000){yp[i,] <- Xdes%*%pst$beta[i,]
+ rnorm(1,mean=0,sd=sqrt(pst$sg)) }
par(mfrow=c(4,2))
for(i in 1:7){hist(yp[,i],prob=T); points(y[i],0,pch="0")}
# sampling the predictive distribution for
# a new data point
xf <- c(1,6.6)
# Corresponds to a body mass of 735.0952
yf <- rep(NA,5000)
for (i in 1:5000){
yf[i] <- rnorm(1,mean=sum(xf*pst$beta[i,]),sd=sqrt(pst$sg[i]))
}
```