BAYESIAN STATISTICS

CLASS 3

Sampling distribution is Lognormal – there are two parameters for a Lognormal: μ, σ^2

$\mu \sim$ Uniform(0,80)
$\sigma^2 \sim$ Uniform(0,80)

STAN

```
data{
  int <lower=0> n ;
  real y[n];
}

parameters {
  real mu;
  real <lower=0> sigma;
}

model {
  mu ~uniform(0,80);
  sigma~uniform(0,80);
  for (i in 1:n)
   y[i] ~ lognormal(mu,sigma);
}
```

R Code

```
library(lmtest)
library(rstan)
lognorm.dat<-list(y=unemployment[,1],n=length(unemployment[,1]))
log.stan=stan(file='Q:\\My
Drive\\Bayesian\\Code\\lognormal_model.stan',data=lognorm.dat,seed=10678)
log.extract=extract(log.stan)
new.mu=log.extract$mu
new.sigma=log.extract$sigma
mean(new.mu)
mean(new.sigma)
mean(new.sigma/new.mu)
```

```
mu.sigma=cbind(new.mu,new.sigma)
head(mu.sigma)
```

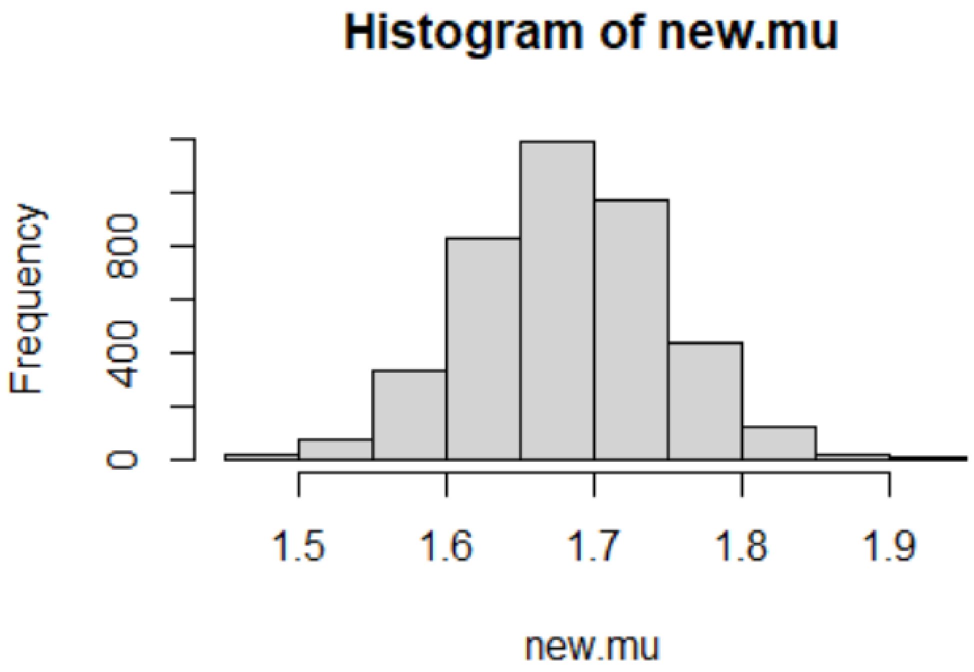
```
       new.mu   new.sigma
[1,] 1.780556 0.5457839
[2,] 1.663109 0.6653600
[3,] 1.646789 0.5786628
[4,] 1.606961 0.5906967
[5,] 1.698649 0.5329611
[6,] 1.765465 0.6531658
```

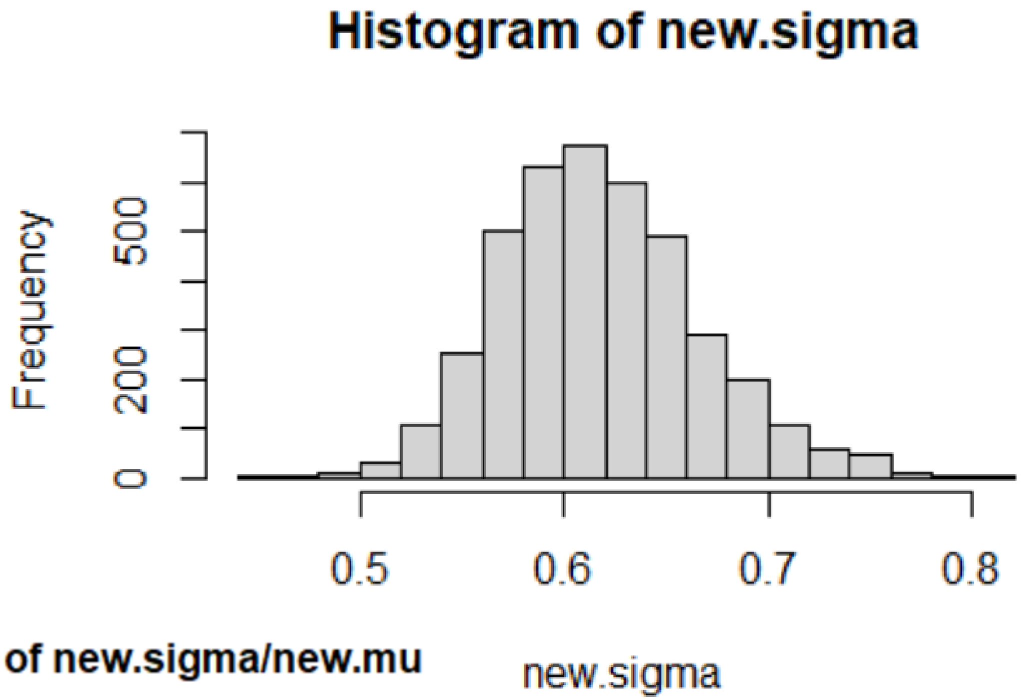
Posterior Joint distribution:
$P(\mu,\sigma|Y)$

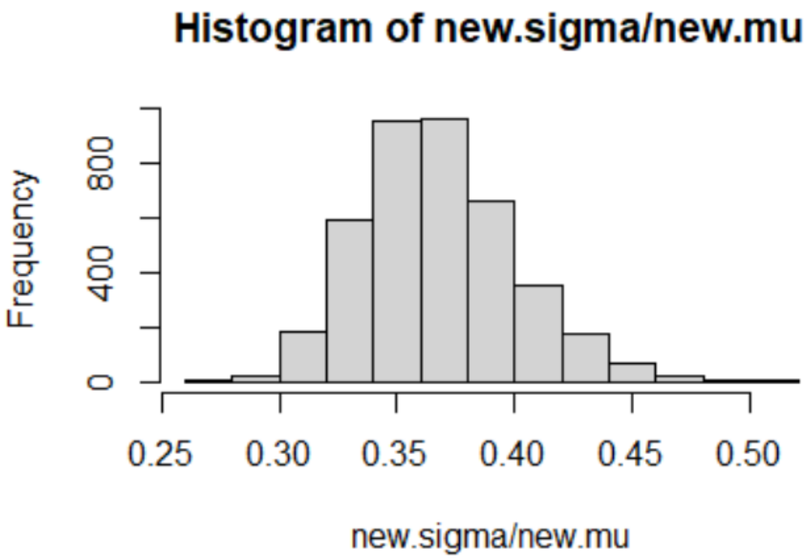
Posterior Marginal distribution of mean:
$P(\mu|Y)$

Posterior Marginal distribution of std dev:
$P(\sigma|Y)$

Together, these are the posterior joint distribution, however, you can just use one column for the posterior marginal distribution







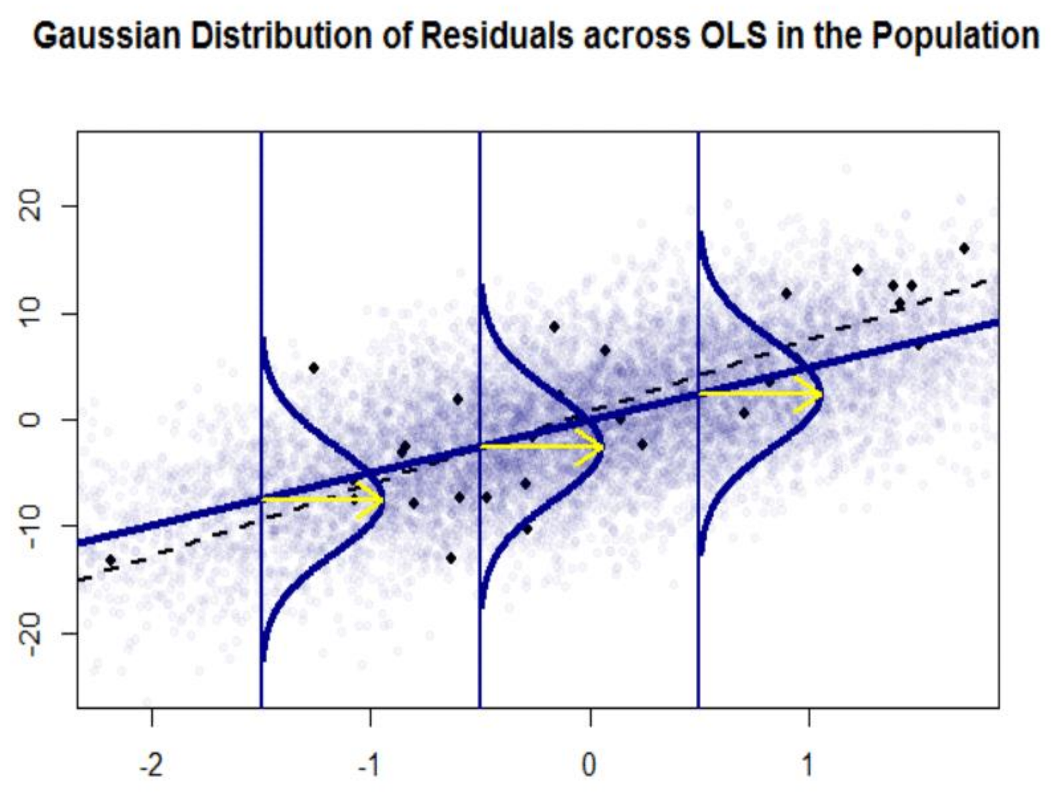
NOTE: SIGMA is the **standard deviation** (not variance)

```
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.226   3.858   5.233   6.478   7.527  22.981 
> summary(new.mu)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.455   1.638   1.681   1.682   1.727   1.934 
> summary(new.sigma)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.4535  0.5836  0.6141  0.6177  0.6478  0.8078 
```

```
> summary(new.mu)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.455   1.638   1.681   1.682   1.727   1.934 
> summary(new.sigma)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.4535  0.5836  0.6141  0.6177  0.6478  0.8078 
> sd(log(y))
[1] 0.610505
> mean(log(y))
[1] 1.682152
```

```
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.226   3.858   5.233   6.478   7.527  22.981
> summary(exp(new.mu+0.5*new.sigma^2))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.049   6.207   6.500   6.530   6.820   8.607
> summary(sqrt((exp(2*new.mu+new.sigma^2)*(exp(new.sigma^2)-1))))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.978   4.038   4.380   4.473   4.827   7.552
> sd(y)
[1] 4.338271
```

Linear Regression assumes Normal distribution with a mean of $X\beta$ and constant standard deviation



'Liv.Area','Base.Area','Garage.Area','Porch.Area','Age

Recall Ames Housing data….trying to predict Y (sales price of house)

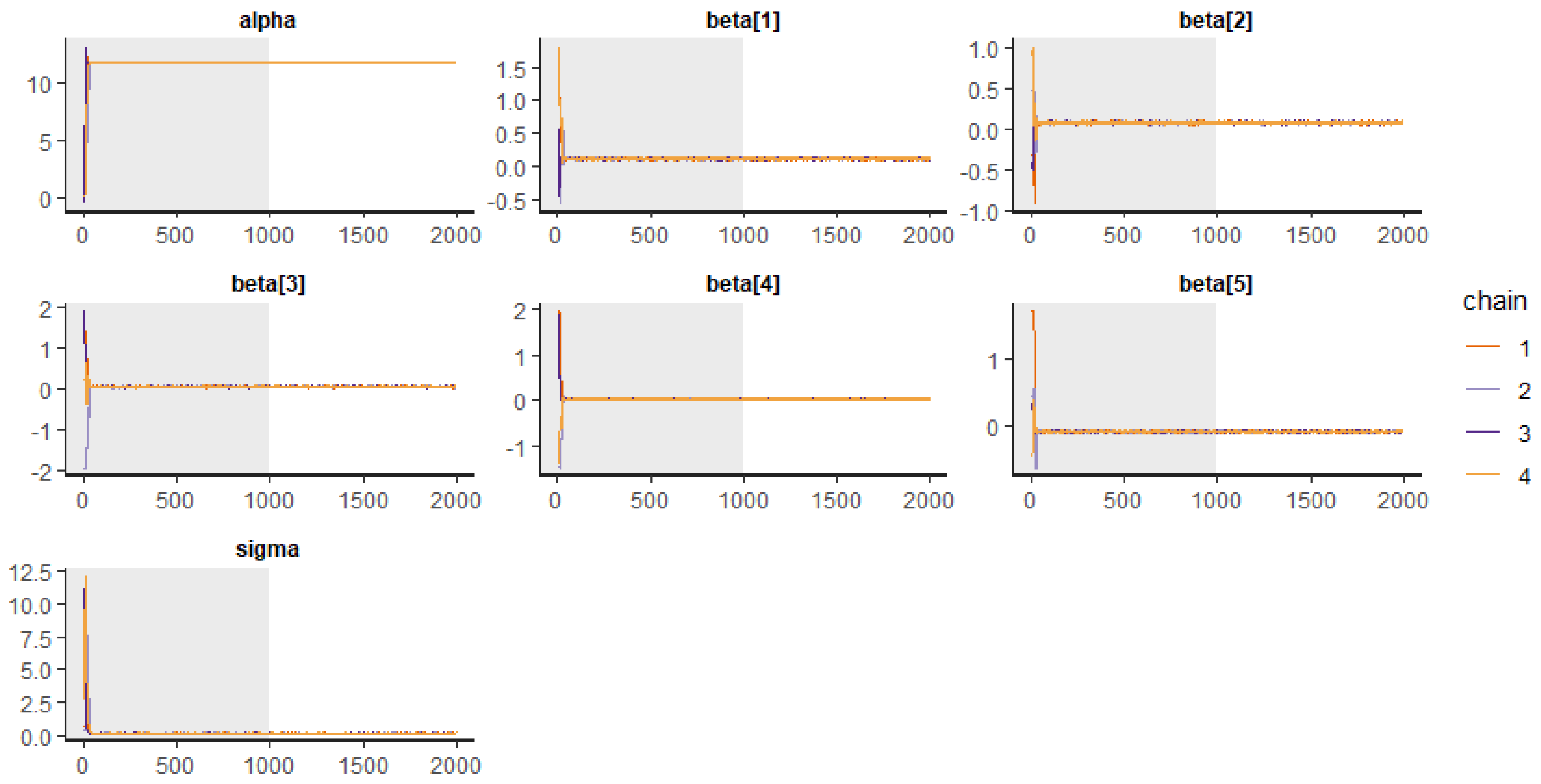
We will use the following predictor variables: X1 = Living Area, X2 = Basement Area, X3 = Garage Area, X4 = Porch Area and X5 = Age of house

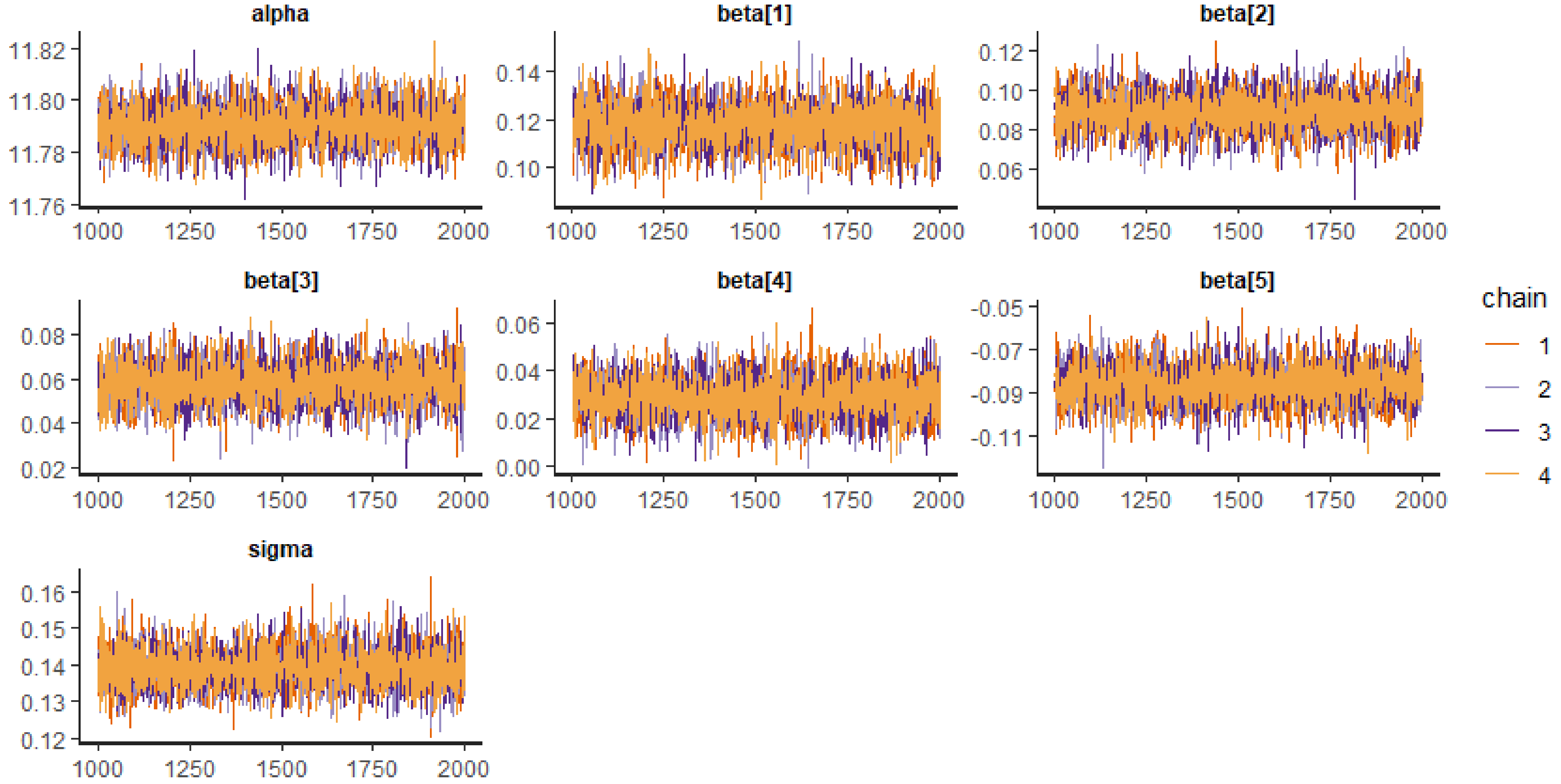
$$\text{Log}(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 5 x_5 + \varepsilon$$

Regression in stan

```
data{
  int <lower=0> n;
  vector[n] y;
  matrix[n,5] x;
}
parameters{
  real alpha;
  vector[5] beta;
  real<lower=0> sigma;
}
model {
y ~normal(alpha + x*beta, sigma);
}
```

```
library(sas7bdat)
ameshousing<-read.sas7bdat('Q:\\My Drive\\Summer II - Statistics Bootcamp\\Bootcamp
Data\\ameshousing3.sas7bdat')
x=cbind(ameshousing$Gr_Liv_Area, ameshousing$Basement_Area,
ameshousing$Garage_Area, ameshousing$Deck_Porch_Area, ameshousing$Age_Sold)
x=as.data.frame(scale(x))
colnames(x)=c('Liv.Area','Base.Area','Garage.Area','Porch.Area','Age')
regress.dat=list(n=nrow(x),x=x,y=ameshousing$Log_Price)
regress.stan=stan(file='Q:\\My
Drive\\Bayesian\\Code\\ameshousing.stan',data=regress.dat,seed=93457)
output.stan=extract(regress.stan)
traceplot(regress.stan,inc_warmup=T)
```


alpha
beta[1]
beta[2]
beta[3]
beta[4]
beta[5]
sigma
chain
1
2
3
4



alpha
beta[1]
beta[2]
beta[3]
beta[4]
beta[5]
sigma
chain
1
2
3
4

```
print(regress.stan,  probs=c(.025,.975))
```

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
alpha	11.79	0.00	0.01	11.78	11.81	6588	1
beta[1]	0.12	0.00	0.01	0.10	0.14	5907	1
beta[2]	0.09	0.00	0.01	0.07	0.11	5643	1
beta[3]	0.06	0.00	0.01	0.04	0.08	5779	1
beta[4]	0.03	0.00	0.01	0.01	0.05	6864	1
beta[5]	-0.09	0.00	0.01	-0.10	-0.07	5316	1
sigma	0.14	0.00	0.01	0.13	0.15	5456	1
lp__	440.42	0.05	1.89	435.92	443.12	1744	1