Bayes, Homework 2

1. Expected Value of σ^2

Our conjugate prior for σ is

$$\sigma^2 \sim \frac{\nu\,\lambda}{\chi^2_\nu}$$

Check that

$$E(\sigma^2) = \frac{\nu\,\lambda}{\nu - 2}.$$

That is, if $X \sim \chi^2_\nu$, then $E(1/X) = 1/(\nu - 2)$.

2. The American Returns

We looked at the Canadian returns in the data set conret.csv by looking at the posteriors $\mu|\sigma, y$ and $\sigma|\mu, y$.

Do the same thing for the American returns (usa = read.csv("conret.csv")$usa).

Are the American returns different from the Canadian returns? How do μ and σ compare? Typically people like a big μ and a small σ.

What is the predictive distribution of the sum of the next 12 American returns (given the data and σ)?

3. The House Price Data

The file midcity.csv on the webpage house data on sold houses.

Each observation corresponds to a recently sold house.
We are interested in developing a model to predict $Y = Price$ using the variables: "Nbhd" "SqFt" "Brick" "Bedrooms" "Bathrooms".

Let's compare using just the variable SqFt with using all the variables.

Below are the two regressions.

```
hd = read.csv("midcity.csv")
hd = hd[,c(2,4:8)] #drop unused variables
hd$Nbhd = as.factor(hd$Nbhd) #Nbhd is categorical
hd$Price = hd$Price/1000 #rescale
hd$SqFt = hd$SqFt/1000 #rescale

print(summary(hd))
```

```
##  Nbhd        SqFt       Brick       Bedrooms       Bathrooms
##  1:44   Min.   :1.450   No :86   Min.   :2.000   Min.   :2.000
##  2:45   1st Qu.:1.880   Yes:42   1st Qu.:3.000   1st Qu.:2.000
##  3:39   Median :2.000            Median :3.000   Median :2.000
##         Mean   :2.001            Mean   :3.023   Mean   :2.445
```

```
##         3rd Qu.:2.140            3rd Qu.:3.000   3rd Qu.:3.000
##         Max.   :2.590            Max.   :5.000   Max.   :4.000
##      Price
##  Min.   : 69.1
##  1st Qu.:111.3
##  Median :126.0
##  Mean   :130.4
##  3rd Qu.:148.2
##  Max.   :211.2
```

```
lmS = lm(Price~SqFt,hd) #Price on SqFt
print(summary(lmS))
```

```
##
## Call:
## lm(formula = Price ~ SqFt, data = hd)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -46.59 -16.64  -1.61  15.12  54.83
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -10.091     18.966  -0.532    0.596
## SqFt          70.226      9.426   7.450  1.3e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.48 on 126 degrees of freedom
## Multiple R-squared:  0.3058, Adjusted R-squared:  0.3003
## F-statistic:  55.5 on 1 and 126 DF,  p-value: 1.302e-11
```

```
lmA = lm(Price~.,hd) #Price on all
print(summary(lmA))
```

```
##
## Call:
## lm(formula = Price ~ ., data = hd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.008  -7.323  -0.119   7.819  33.392
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.919     10.474   1.711  0.08967 .
## Nbhd2          4.866      2.722   1.788  0.07633 .
## Nbhd3         34.084      3.169  10.755  < 2e-16 ***
## SqFt          35.930      6.404   5.610 1.30e-07 ***
## BrickYes      18.508      2.396   7.723 3.65e-12 ***
## Bedrooms       1.902      1.902   1.000  0.31933
## Bathrooms      6.827      2.563   2.664  0.00878 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.15 on 121 degrees of freedom
## Multiple R-squared:  0.805,  Adjusted R-squared:  0.7954
## F-statistic: 83.27 on 6 and 121 DF,  p-value: < 2.2e-16
```

Recall the our regression model (with normal errors) is

$$Y_i = x_i' \beta + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma^2), \; iid.$$

Using just SqFt, the estimate of σ from the standard regression output is 22.48.

Using all the variables, the estimate of σ is 12.15.

Let's get a rough gauge of our uncertainty about σ by plugging in the estimates of β so that we can pretend we observe the errors:

$$Y_i - x_i' \hat{\beta} \approx \epsilon_i \sim N(0, \sigma^2)$$

For both regression (using just SqFt and using all the variables), get the posterior of σ. How do they compare? Is σ smaller with all the variables.