Report

**Q1.** The summary statistics for the regression results are as follows. The estimated values of β0 and β1 are 68.20074 and 0.13511.

## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 68.20074 37.43707 1.822 0.0935 .   
## sqft 0.13511 0.01872 7.216 1.06e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 39.19 on 12 degrees of freedom  
## Multiple R-squared: 0.8127, Adjusted R-squared: 0.7971   
## F-statistic: 52.07 on 1 and 12 DF, p-value: 1.063e-05

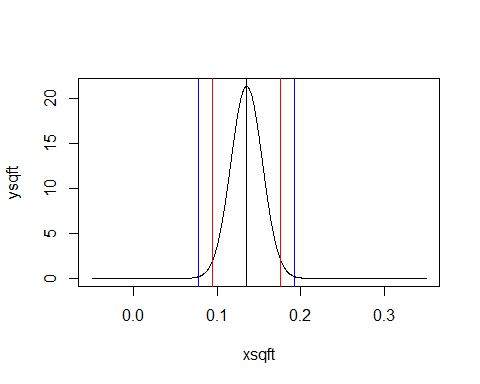
**Q2.** The coefficient estimate of β1 obtained from question 1 seems reasonable. The p-value is close zero, meaning that β1 is significantly different from zero.

**Q3.** Because the p-value for β1 is 1.06e-05, which is quite close to zero. Thus, it is statistically significant at all of the 10%, 5% and 1% level. The coefficient is 0.13511, meaning that every 1 square feet of living area increase would result in 0.13511(thousands of dollars) increase in sale price.

**Q4.** The 95% and 99% confidence intervals for the regression coefficient β1, and their corresponding plot could be shown as follows.

## 2.5 % 97.5 %  
## (Intercept) -13.36763120 149.7691150  
## sqft 0.09431191 0.1758987

## 0.5 % 99.5 %  
## (Intercept) -46.1522771 182.5537608  
## sqft 0.0779159 0.1922947



**Q5.** The model results are shown as follows. The difference of this model and the model in question 1 is that in this model, the coefficient β1 is 1000 times bigger than that in question 1. And the standard error has also become 1000 times than that in question 1.

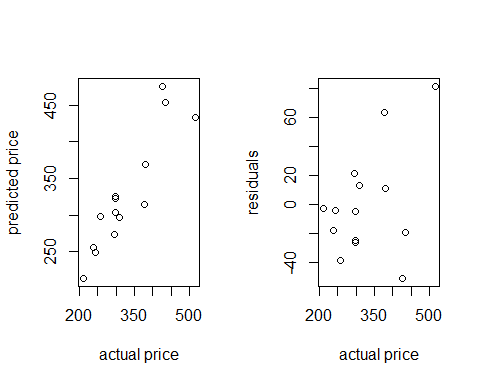
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 68.20 37.44 1.822 0.0935 .   
## s1000 135.11 18.72 7.216 1.06e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 39.19 on 12 degrees of freedom  
## Multiple R-squared: 0.8127, Adjusted R-squared: 0.7971   
## F-statistic: 52.07 on 1 and 12 DF, p-value: 1.063e-05

**Q6.** The model results are shown as follows. The only coefficient that is statistically significantly different from zero at 10%, 5%, and 1% levels is sqft. Coefficients of bedrms and baths both show no statistically significantly different from zero.

## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 106.66872 65.83994 1.620 0.136274   
## sqft 0.14359 0.03027 4.743 0.000788 \*\*\*  
## bedrms -15.13247 21.59533 -0.701 0.499452   
## baths -0.31736 30.16292 -0.011 0.991812   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 41.8 on 10 degrees of freedom  
## Multiple R-squared: 0.8224, Adjusted R-squared: 0.7691   
## F-statistic: 15.44 on 3 and 10 DF, p-value: 0.0004413

**Q7.** In model from question 6 and model from question 1, they both show that sqft has statistically significantly impact on price. Bedrooms and baths show no impact on price. These results match up with my theoretical and intuitive economic reasoning, because commonly, the larger the house, the more expensive the house will be.

**Q8.** The actual vs predicted values, and actual vs residuals are plotted as follows.



**Q9.** The mean of the residuals from this OLS regression is 1.903487e-15. The sum of the residuals from this OLS regression is 2.664535e-14. These values are expected. Because OLS requires that the variance should be constant with the sum of zero.

**Q10.** The coefficient on baths is not statistically significant because the p-value is relatively large. The interpretation of the coefficients on baths is that every 1 baths rooms increase would result in -0.31736(thousands of dollars) decrease of the price for the house.

**Appendix**

rm(list=ls())

setwd("D:/hw")

getwd()

hd<-read.csv("house.csv",header=T)

price<-hd$PRICE

sqft<-hd$SQFT

baths<-hd$BATHS

bedrms<-hd$BEDRMS

#Q1

reg<-lm(price~sqft)

summary(reg)

#Q4

confint(reg,level=0.95)

confint(reg,level=0.99)

xsqft<-seq(-0.05,0.35,length=1000)

ysqft<-dnorm(xsqft,mean=0.13511, sd=0.01872)

plot(xsqft, ysqft,type="l",lwd=1)

abline(v=0.13511)

abline(v=0.09431191,col="red")

abline(v=0.1758987,col="red")

abline(v=0.0779159,col="blue")

abline(v=0.1922947,col="blue")

#Q5

s1000<-sqft/1000

reg2<-lm(price~s1000)

summary(reg2)

#Q6

reg3<-lm(price~sqft+bedrms+baths)

summary(reg3)

#Q8

par(mfrow=c(1,2))

plot(price,fitted(reg3),xlab="actual price",ylab="predicted price")

plot(price,residuals(reg3),xlab="actual price",ylab="residuals")

par(mfrow=c(1,1))

#Q9

mean(residuals(reg3))

sum(residuals(reg3))