Math 327, Chapter 3 Homework on Property Values

David Oniani

September 18, 2019

```
load ("./Appendix_B_data.Rdata")
```

```
## Warning: namespace 'emmeans' is not available and has been replaced
## by .GlobalEnv when processing object '.Last.ref_grid'
```

The table in Appendix B.4 has data on property values, as measured by sale price, as well as nine predictor variables as listed here:

y = Sale price ($1000)

x1 = taxes: local, school, county ($1000)

x2 = number of baths (#)

x3 = Lot Size (1000 sq ft)

x4 = Living Space (1000 sq ft)

x5 = Number of garage stalls

x6 = Number of rooms

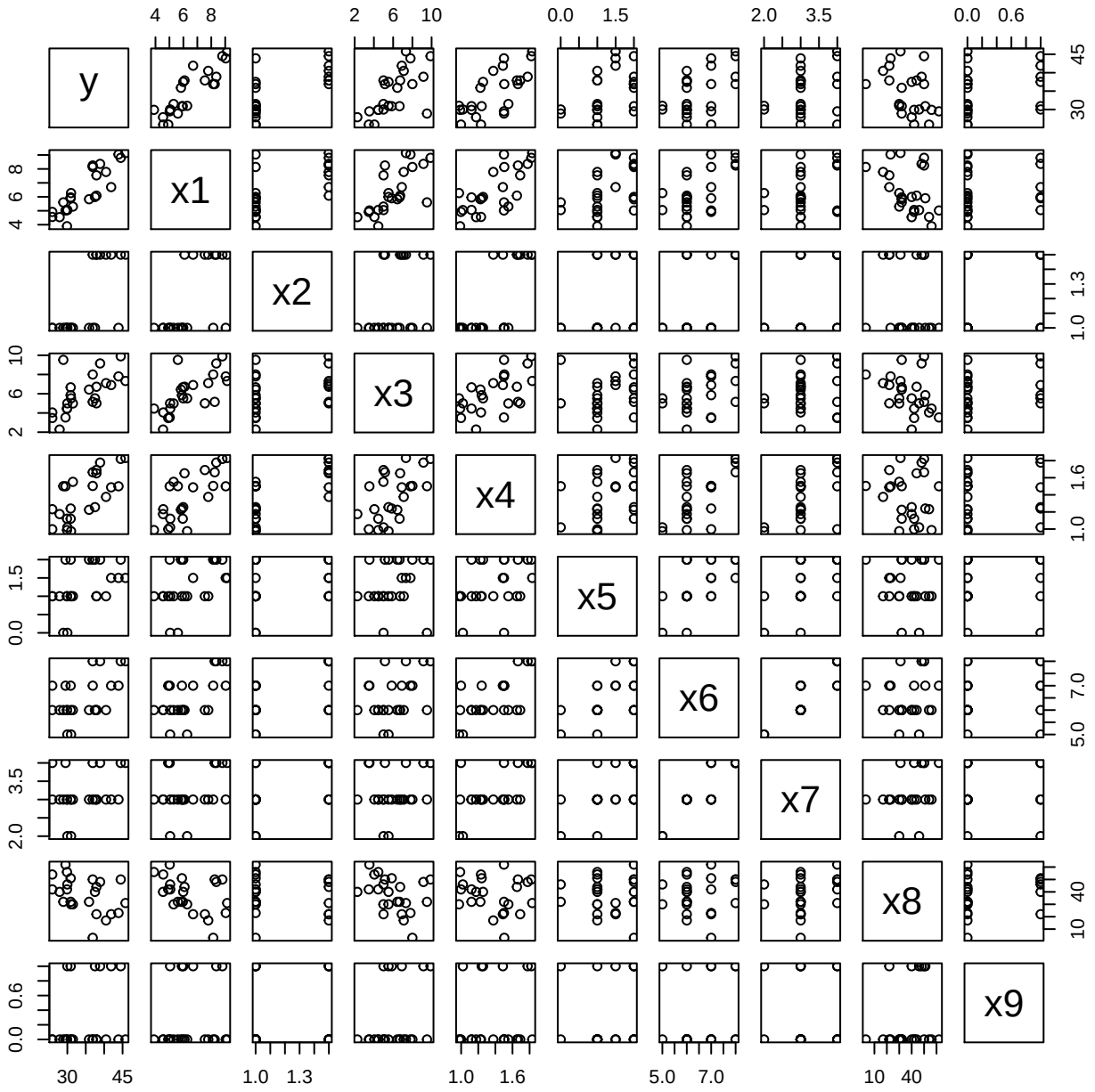
x7 = Number of bedrooms

x8 = Age of the home (years)

x9 = # fireplaces

First, produce the scatter plot matrix of all variables

```
plot (dataB.4)
```

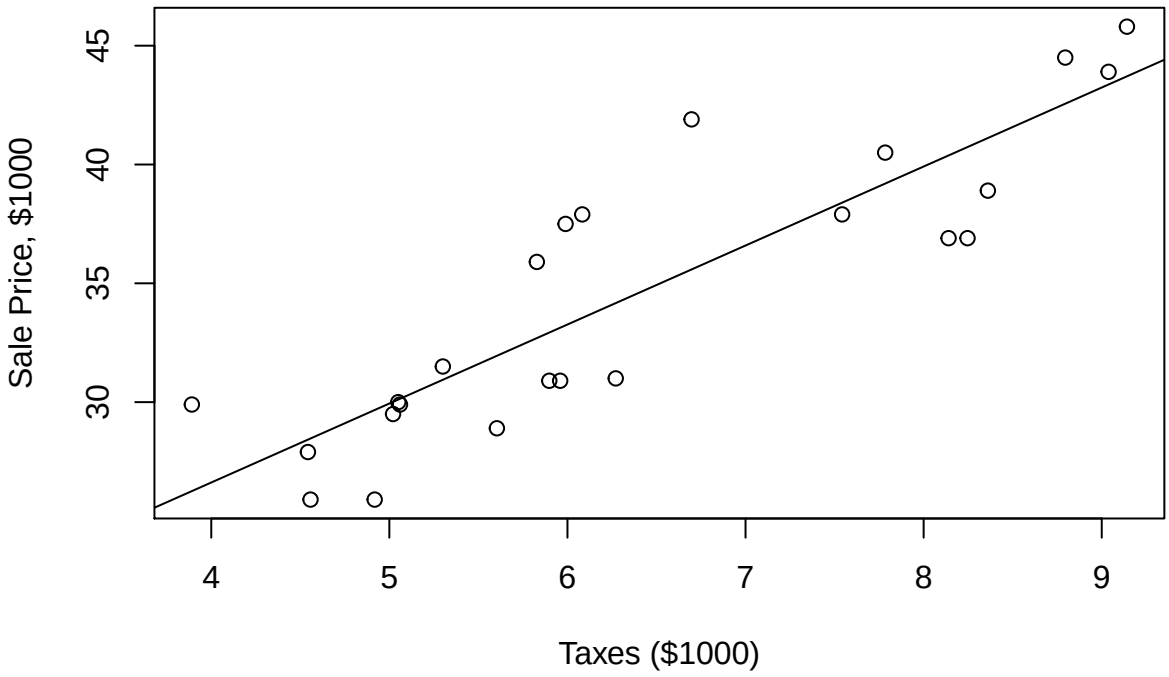


Fit Sale Price vs. Taxes

```
plot (y ~ x1, data=dataB.4, ylab="Sale Price, $1000", xlab="Taxes ($1000)")

# Fit sale price vs taxes

fitB4.1 = lm (y ~ x1, data=dataB.4)
abline (fitB4.1)
```




```
summary (fitB4.1)
```

```
##
## Call:
## lm(formula = y ~ x1, data = dataB.4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8343 -2.3157 -0.3669  1.9787  6.3168
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.3202     2.5717   5.179 3.42e-05 ***
## x1            3.3244     0.3903   8.518 2.05e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.961 on 22 degrees of freedom
## Multiple R-squared:  0.7673, Adjusted R-squared:  0.7568
## F-statistic: 72.56 on 1 and 22 DF,  p-value: 2.051e-08
```

```
confint (fitB4.1)
```

```
##                 2.5 %    97.5 %
## (Intercept) 7.986755 18.653604
## x1          2.514988  4.133754
```

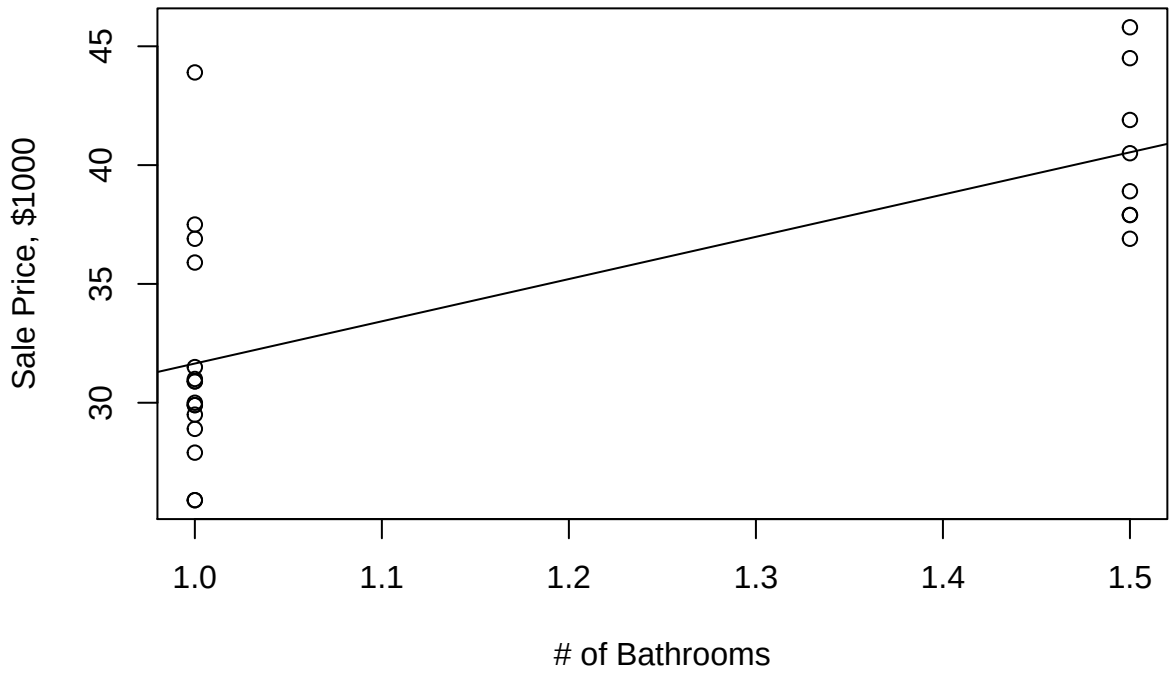
Q1: Summarize the fit of sale price vs taxes. At a minimum, include interpretation of the slope, the coefficient of determination, and the residual standard error.

The mean response changes between 2.515 and 4.134 Sale Price per Taxes with 95% confidence. The coefficient of determination (R^2) is 0.7673 which means that 76.73% of variation in Sales Price is explained by Taxes. The residual standard error is 2.961 which tells us that, on average, our predictions are 2.961 (thousand) dollars off from the real value.

Fit Sale Price vs # of bathrooms

```
plot (y ~ x2, data=dataB.4, ylab="Sale Price, $1000", xlab="# of Bathrooms")

fitB4.2 = lm (y ~ x2, data=dataB.4)
abline (fitB4.2)
```




```
summary (fitB4.2)
```

```
##
## Call:
## lm(formula = y ~ x2, data = dataB.4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.750 -2.638 -1.194  2.013 12.250
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.875      4.438   3.127  0.00491 **
## x2            17.775      3.728   4.767 9.27e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.305 on 22 degrees of freedom
## Multiple R-squared:  0.5081, Adjusted R-squared:  0.4858
## F-statistic: 22.73 on 1 and 22 DF,  p-value: 9.268e-05
```

```
confint (fitB4.2)
```

```
##                 2.5 %   97.5 %
## (Intercept)  4.671615 23.07839
## x2          10.042610 25.50739
```

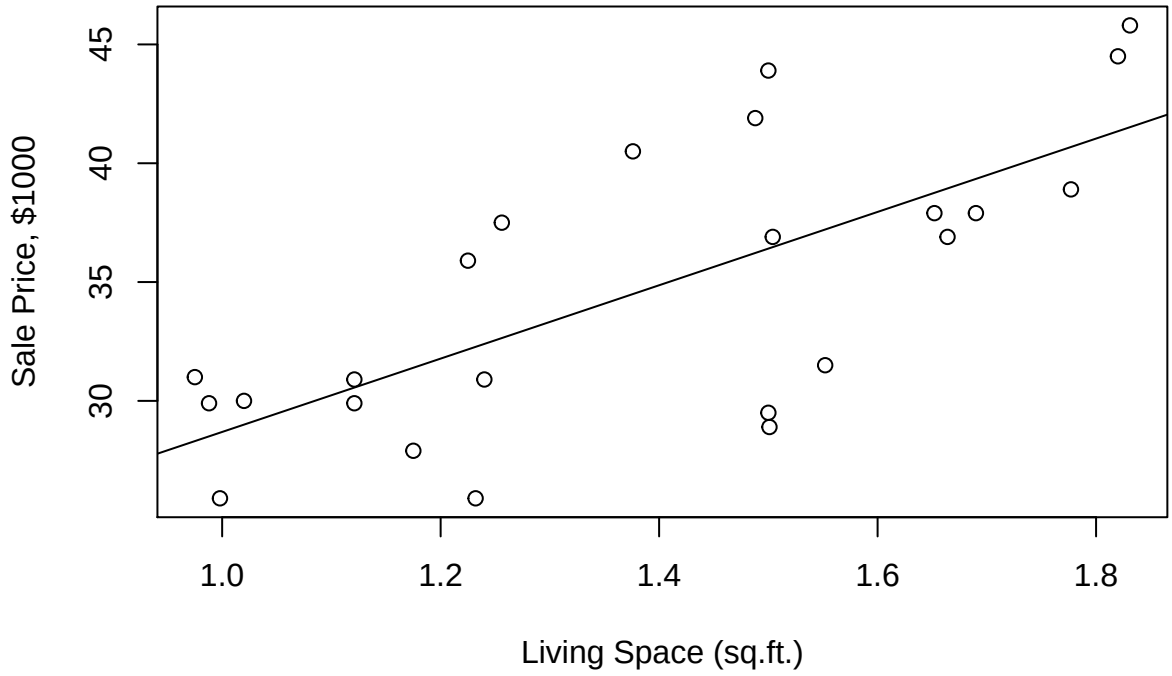
Q2: Summarize the fit of sale price vs. number of bathrooms. At a minimum, include interpretation of the slope, the coefficient of determination, and the residual standard error.

The mean response changes between 10.043 and 25.507 Sale Price per # of Bathrooms with 95% confidence. The coefficient of determination (R^2) is 0.5081 which means that 50.81% of variation in Sales Price is explained by the number of Bathrooms. The residual standard error is 4.305 which tells us that, on average, our predictions are 4.305 (thousand) dollars off from the real value.

Fit Sale Price vs. Living Space

```
plot (y ~ x4, data=dataB.4, ylab="Sale Price, $1000", xlab="Living Space (sq.ft.)")

fitB4.3 = lm (y ~ x4, data=dataB.4)
abline (fitB4.3)
```




```
summary (fitB4.3)
```

```
##
## Call:
## lm(formula = y ~ x4, data = dataB.4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5246 -2.2206 -0.1601  3.2979  7.4908
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.260      4.599   2.883 0.008629 **
## x4            15.433      3.262   4.731 0.000101 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.322 on 22 degrees of freedom
## Multiple R-squared:  0.5043, Adjusted R-squared:  0.4818
## F-statistic: 22.38 on 1 and 22 DF,  p-value: 0.0001012
```

```
confint (fitB4.3)
```

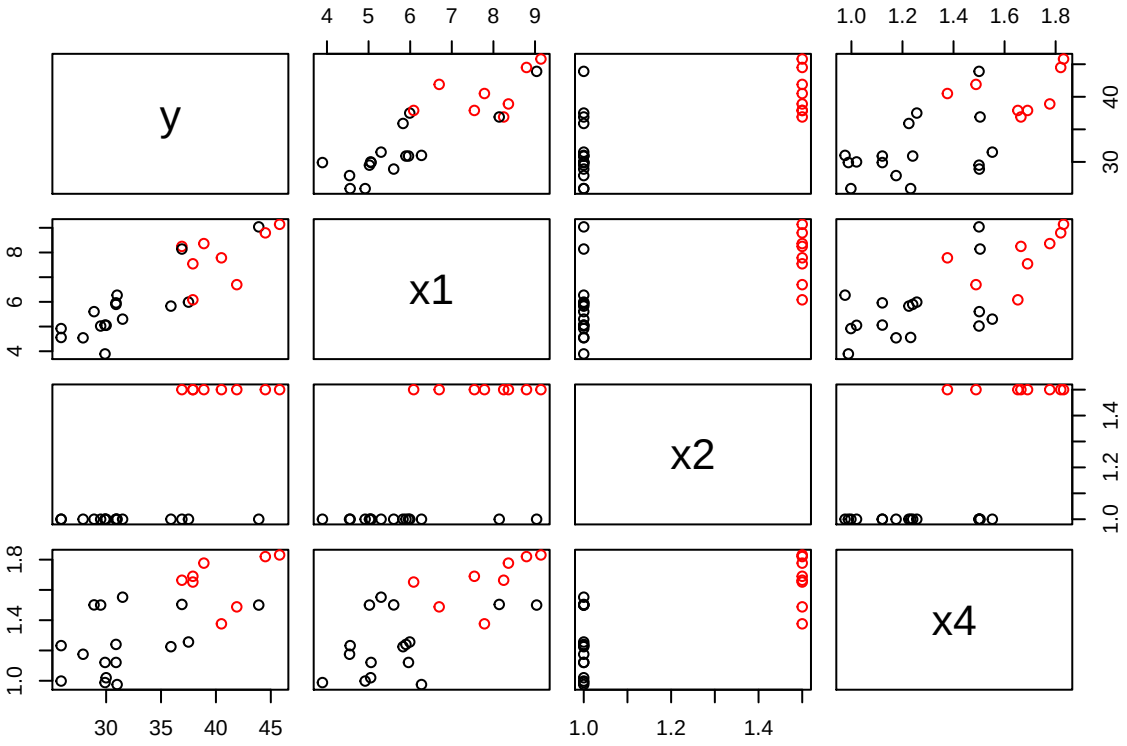
```
##                2.5 %   97.5 %
## (Intercept) 3.722526 22.79678
## x4          8.667971 22.19804
```

Q3: Summarize the fit of sale price vs. living space. At a minimum, include interpretation of the slope, the coefficient of determination, and the residual standard error.

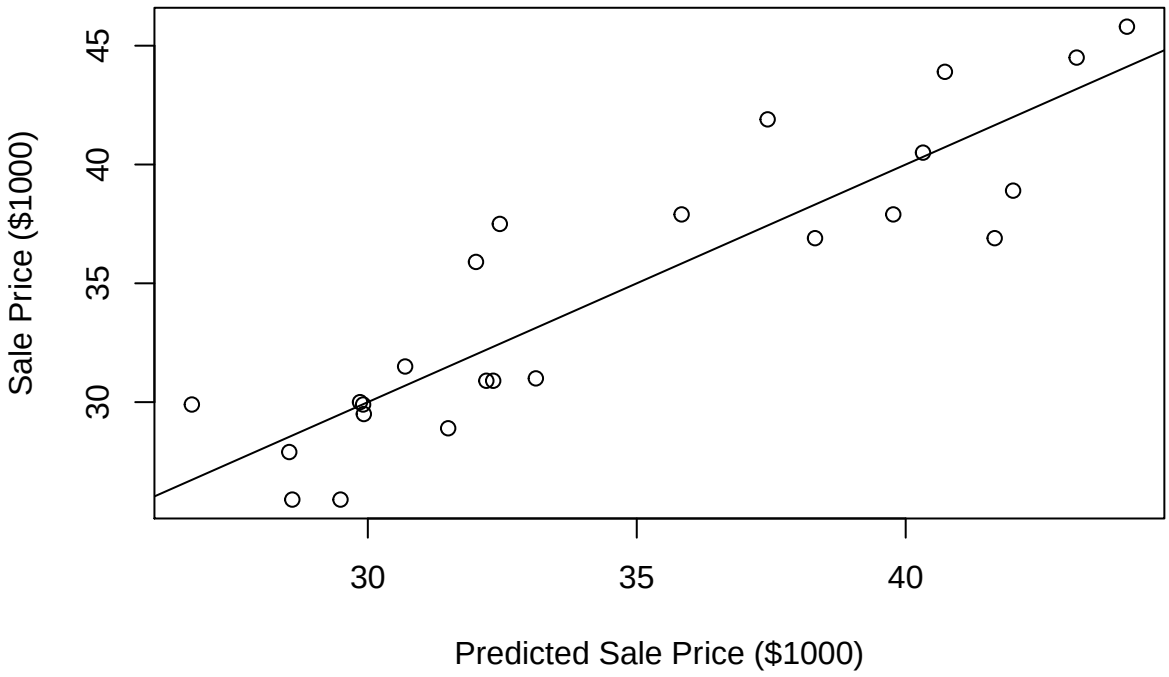
The mean response changes between 8.668 and 22.198 Sale Price per Living Space (in sq.ft.) with 95% confidence. The coefficient of determination (R^2) is 0.5043 which means that 50.43% of variation in Sales Price is explained by the Living Space (in sq.ft.). The residual standard error is 4.322 which tells us that, on average, our predictions are 4.322 (thousand) dollars off from the real value.

Fit Sale Price vs Taxes, number of bathrooms and living space

```
# Scatterplot matrix of Y, X1, X2, and X4, which are in columns
# 1, 2, 3, and 5 of the data frame.
plot (dataB.4[,c(1,2,3,5)], col=-0.5+2*dataB.4$x2)
```




```
fitB4.4 = lm (y ~ x1 + x2 + x4, data=dataB.4)
# A useful plot for any regression model is y vs y_hat
plot (dataB.4$y ~ fitB4.4$fitted.values,
      ylab="Sale Price ($1000)", xlab="Predicted Sale Price ($1000)")
abline (0,1)
```




```
summary (fitB4.4)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x4, data = dataB.4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7543 -1.9327 -0.2189  1.7815  5.0460
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.9491     3.2179   3.092 0.005751 **
## x1            2.6888     0.5736   4.688 0.000141 ***
## x2            6.0085     3.7354   1.609 0.123393
## x4            0.3122     3.6391   0.086 0.932494
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.861 on 20 degrees of freedom
## Multiple R-squared:  0.8026, Adjusted R-squared:  0.773
## F-statistic: 27.11 on 3 and 20 DF,  p-value: 3.011e-07
```

```
confint (fitB4.4)
```

```
##                 2.5 %    97.5 %
## (Intercept)  3.236710 16.661414
## x1           1.492297  3.885320
## x2          -1.783442 13.800477
## x4          -7.278965  7.903297
```

Q4: Summarize the multiple regression model above. At a minimum, include interpretation of each predictor's parameter estimate, the adjusted coefficient of determination, and the residual standard error.

The mean response changes between 1.492 and 3.885 Sales Prices per taxes (local, school, county in $1000) (x1), for any 1-unit increase in the predictor with 95% confidence, holding all other predictors fixed.

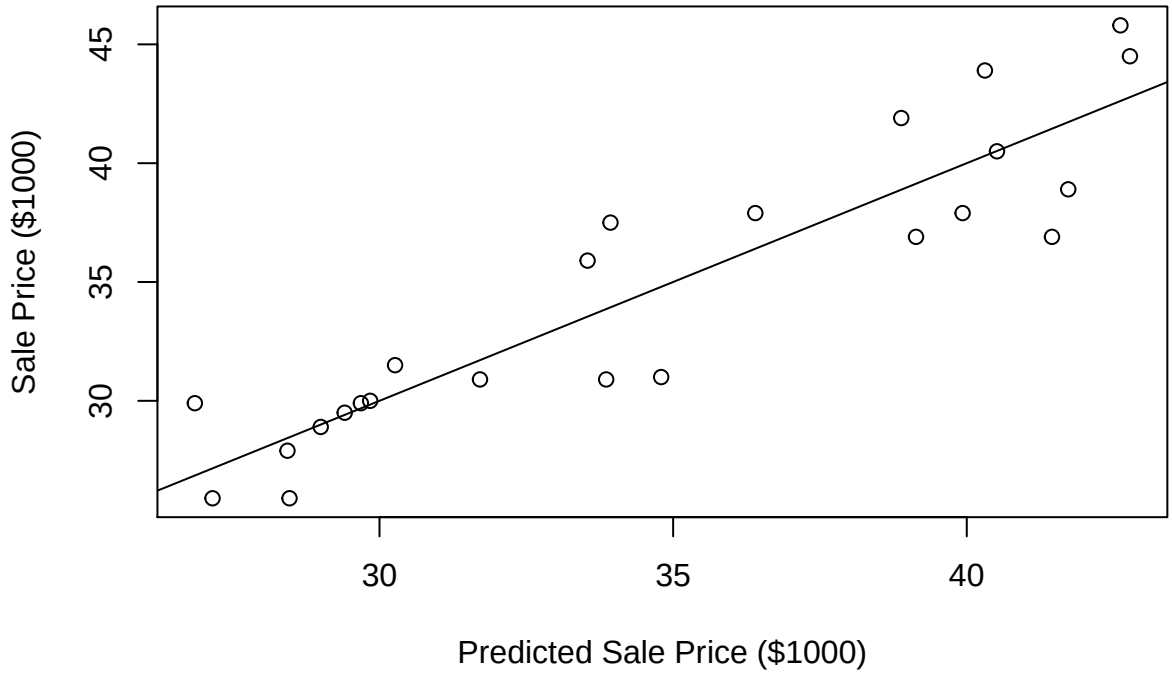
The mean response changes between -1.783 and 13.800 Sales Prices per number of baths (#) (x2), for any 1-unit increase in the predictor with 95% confidence, holding all other predictors fixed.

The mean response changes between -7.279 and 7.903 Sales Prices per Living Space (1000 sq ft) (x4), for any 1-unit increase in the predictor with 95% confidence, holding all other predictors fixed.

The coefficient of determination (R^2) is 0.8026 which means that 80.26% of variation in Sales Price is explained by the Predicted Sale Price (in $1000). In this case, however, since we have multiple predictor variables, we prefer using Adjusted R-squared value, which tells us that 77.3% of a variation in Sale Price is explained by Predicted Sale Price (in $1000). The residual standard error is 2.861 which tells us that, on average, our predictions are 2.861 (in thousands) dollars off from the real value.

```
# Fit x1, x2, x5, and x7

fitB4.5 = lm (y ~ x1 + x2 + x5 + x7, data=dataB.4)
plot (dataB.4$y ~ fitB4.5$fitted.values,
      ylab="Sale Price ($1000)", xlab="Predicted Sale Price ($1000)")
abline (0,1)
```




```
summary (fitB4.5)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x5 + x7, data = dataB.4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5510 -2.0807  0.0391  1.8805  3.5912
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.5091     3.6331   3.718 0.001458 **
```

```
## x1            2.4192     0.5169   4.680 0.000163 ***
## x2            8.4802     3.2943   2.574 0.018577 *
## x5            2.0006     1.2104   1.653 0.114793
## x7           -2.1823     1.2762  -1.710 0.103557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.681 on 19 degrees of freedom
## Multiple R-squared:  0.8352, Adjusted R-squared:  0.8005
## F-statistic: 24.08 on 4 and 19 DF,  p-value: 3.249e-07
```

```
confint (fitB4.5)
```

```
##                  2.5 %     97.5 %
## (Intercept)  5.9048562 21.1132565
## x1           1.3373124  3.5010054
## x2           1.5852744 15.3752250
## x5          -0.5327977  4.5339692
## x7          -4.8533911  0.4888895
```

Q5: Summarize the multiple regression model above. At a minimum, include interpretation of each predictor's parameter estimate, the adjusted coefficient of determination, and the residual standard error.

The mean response changes between 1.337 and 3.501 Sales Prices per taxes (local, school, county in $1000) (x1), for any 1-unit increase in the predictor with 95% confidence, holding all other predictors fixed.

The mean response changes between 1.585 and 15.375 Sales Prices per number of baths (#) (x2), for any 1-unit increase in the predictor with 95% confidence, holding all other predictors fixed.

The mean response changes between -0.533 and 4.534 Sales Prices per Number of garage stalls (x5), for any 1-unit increase in the predictor with 95% confidence, holding all other predictors fixed.

The mean response changes between -4.853 and 0.489 Sales Prices per Number of bedrooms (x7), for any 1-unit increase in the predictor with 95% confidence, holding all other predictors fixed.

The coefficient of determination (R^2) is 0.8352 which means that 83.52% of variation in Sales Price is explained by the Predicted Sale Price (in $1000). In this case, however, since we have multiple predictor variables, we prefer using Adjusted R-squared value, which tells us that 80.05% of a variation in Sale Price is explained by Predicted Sale Price (in $1000). The residual standard error is 2.681 which tells us that, on average, our predictions are 2.681 (in thousands) dollars off from the real value.

Q6: Compare models fitB4.4 and fitB4.5. Based on the information provided in these analyses, which one would be better for predicting sale price of properties?

It seems like fitB4.5 would be a better model for predicting sale price of properties. The reason is that the coefficient of determination has a higher value of $R^2 = 0.8352$ as opposed to $R^2 = 0.8026$ in fitB4.4. More importantly, for Adjusted R-squared values, we have 0.8005 for fitB4.5 and 77.3% for fitB4.4 which, once again, bolsters the argument that fitB4.5 is a better model. Besides, the residual standard error is smaller in fitB4.5 which has a value of 2.681 as opposed to 2.861 in fitB4.4. This means that, on average, fitB4.5 will be $2.861 - 2.681 = 0.18$ thousands of dollars better at predicting the sale price.