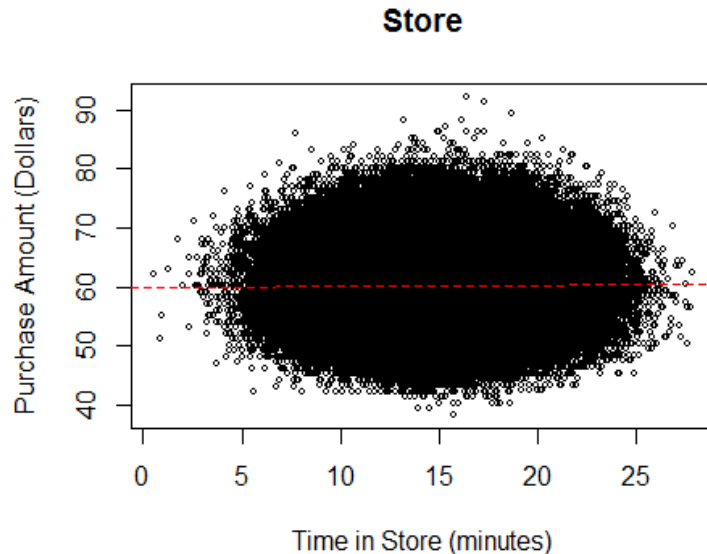


## Regression\_5\_MR-Class.R

```
#Exercise - Large n
# Data file
store <- read_excel("LargeNexample.xlsx", na="NA", col_names = TRUE)
attach(store)
# scatter plot
plot(Time, Amount, pch = 1, cex=.5, main = "Store", xlab = "Time in Store (minutes)", ylab = "Purchase Amount (Dollars)")
abline(lm(Amount~Time), lty=2, col="red")
```



```
# fit the model
linefit <- lm(Amount ~ Time)
summary(linefit)
## Call:
## lm(formula = Amount ~ Time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.983  -3.982   0.016   4.015  32.016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 59.975860   0.039500 1518.38  < 2e-16 ***
## Time         0.015470   0.002583    5.99  2.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.474 on 499998 degrees of freedom
## Multiple R-squared:  7.176e-05, Adjusted R-squared:  6.976e-05
## F-statistic: 35.88 on 1 and 499998 DF, p-value: 2.098e-09
detach(store)
```

```

# Exercise 7-1: House Values
# Data file
house <- read_excel("Values.xlsx", na="NA", col_names = TRUE)
attach(house)

# Parts a/b
# fit the model - 6 predictors
linefitH6 <- lm(AppraisedValue ~ LotSize + HouseSize + Age + Rooms + Baths +
Garage)
summary(linefitH6)

##
## Call:
## lm(formula = AppraisedValue ~ LotSize + HouseSize + Age + Rooms +
##     Baths + Garage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -103.793  -29.668    0.199   35.731  100.207
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   83.06381   68.78871   1.208  0.23950
## LotSize      292.18434   80.88338   3.612  0.00146 **
## HouseSize     0.10058    0.02796   3.597  0.00152 **
## Age          -1.25349    0.55150  -2.273  0.03269 *
## Rooms         10.68976    7.54185   1.417  0.16977
## Baths         6.27865   18.47597   0.340  0.73707
## Garage        15.97114   16.75384   0.953  0.35036
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.84 on 23 degrees of freedom
## Multiple R-squared:  0.8472, Adjusted R-squared:  0.8074
## F-statistic: 21.26 on 6 and 23 DF, p-value: 2.576e-08

# Part c
# confidence intervals for Beta-i
confint(linefitH6, level = .95)

##              2.5 %       97.5 %
## (Intercept) -59.23647111 225.3640888
## LotSize      124.86432506 459.5043570
## HouseSize     0.04273679  0.1584248
## Age          -2.39435315 -0.1126329
## Rooms        -4.91173374 26.2912606
## Baths        -31.94180328 44.4991111
## Garage       -18.68681193 50.6290935

```

```

# Part d
# * ANOVA analysis - sequential F tests
anova(linefitH6)

## Analysis of Variance Table
##
## Response: AppraisedValue
##          Df Sum Sq Mean Sq F value    Pr(>F)
## LotSize    1 158435   158435  56.7543 1.183e-07 ***
## HouseSize   1 166428   166428  59.6178 7.839e-08 ***
## Age         1  23836    23836   8.5387 0.007667 **
## Rooms       1   4192     4192   1.5016 0.232825
## Baths       1    671      671   0.2405 0.628486
## Garage      1   2537     2537   0.9087 0.350359
## Residuals  23  64206     2792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Parts e/f
# * fit the model with one predictor: Baths
linefitH1 <- lm(AppraisedValue ~ Baths)
summary(linefitH1)

##
## Call:
## lm(formula = AppraisedValue ~ Baths)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -204.22  -56.00  -21.05   46.49  316.86
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    217.71      60.06   3.625  0.00114 **
## Baths          86.07      28.42   3.029  0.00523 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 106.3 on 28 degrees of freedom
## Multiple R-squared:  0.2468, Adjusted R-squared:  0.2198
## F-statistic: 9.172 on 1 and 28 DF,  p-value: 0.005233

```

```

# Part g
# collinearity check: correlations among predictors
cor(house[,3:8])

##           LotSize HouseSize           Age           Rooms           Baths
## LotSize      1.00000000  0.4177096 -0.20264831  0.06618112  0.09200025
## HouseSize    0.41770962  1.00000000 -0.41040532  0.39957251  0.52131437
## Age          -0.20264831 -0.4104053  1.00000000 -0.00858975 -0.51086188
## Rooms         0.06618112  0.3995725 -0.00858975  1.00000000  0.13345696
## Baths         0.09200025  0.5213144 -0.51086188  0.13345696  1.00000000
## Garage        0.27196859  0.5713432 -0.53206415  0.02879164  0.49311367
##           Garage
## LotSize      0.27196859
## HouseSize    0.57134322
## Age          -0.53206415
## Rooms         0.02879164
## Baths         0.49311367
## Garage        1.00000000

# Part h
# fit the model - 4 predictors
linefitH4 <- lm(AppraisedValue ~ LotSize + HouseSize + Age + Rooms)
summary(linefitH4)

##
## Call:
## lm(formula = AppraisedValue ~ LotSize + HouseSize + Age + Rooms)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -105.587  -41.783    1.241   35.946   84.754
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 102.95667   59.78017   1.722  0.09737 .
## LotSize      287.35749   77.90157   3.689  0.00110 **
## HouseSize     0.11607    0.02309   5.027 3.48e-05 ***
## Age          -1.50801    0.47859  -3.151  0.00419 **
## Rooms         9.03443    7.24610   1.247  0.22403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.93 on 25 degrees of freedom
## Multiple R-squared:  0.8396, Adjusted R-squared:  0.8139
## F-statistic: 32.72 on 4 and 25 DF, p-value: 1.335e-09

```

```

# fit the model - 3 predictors
linefitH3 <- lm(AppraisedValue ~ LotSize + HouseSize + Age)
summary(linefitH3)

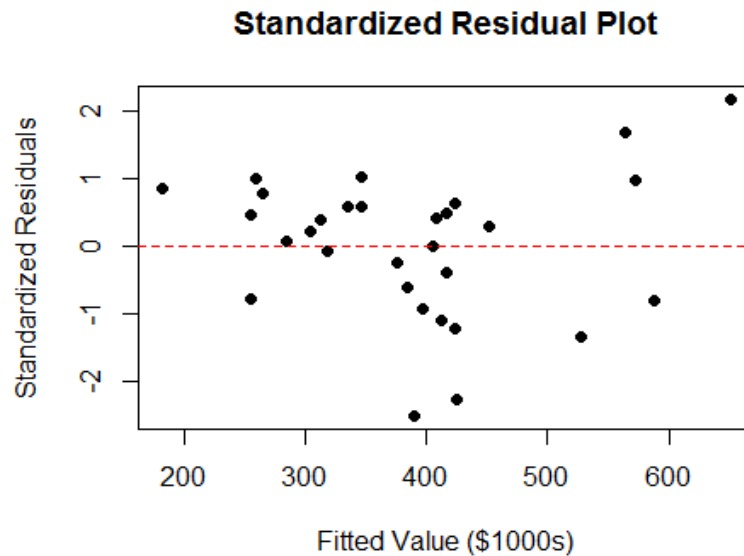
##
## Call:
## lm(formula = AppraisedValue ~ LotSize + HouseSize + Age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -117.80  -36.08   12.82   30.57   98.98
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  136.79404    53.82964   2.541  0.01735 *
## LotSize      276.08764    78.19613   3.531  0.00157 **
## HouseSize     0.12882     0.02092   6.157 1.65e-06 ***
## Age          -1.39893     0.47552  -2.942  0.00677 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.48 on 26 degrees of freedom
## Multiple R-squared:  0.8296, Adjusted R-squared:  0.81
## F-statistic: 42.2 on 3 and 26 DF, p-value: 3.911e-10

# fit the model - 2 predictors
linefitH2 <- lm(AppraisedValue ~ LotSize + HouseSize)
summary(linefitH2)

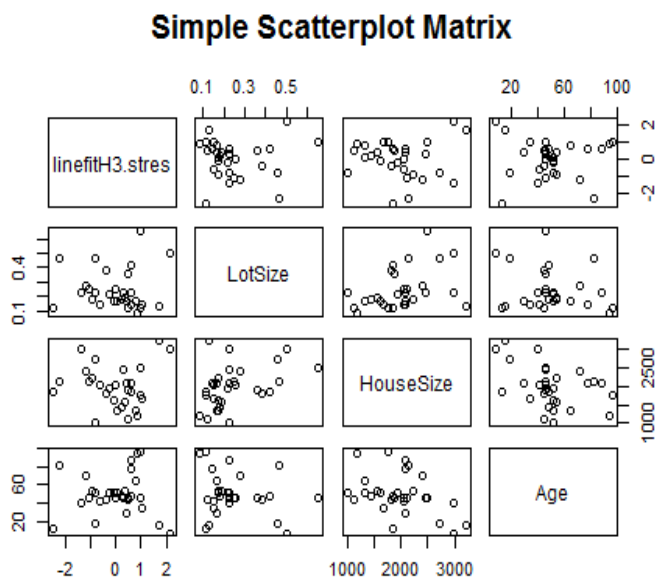
##
## Call:
## lm(formula = AppraisedValue ~ LotSize + HouseSize)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -154.626  -19.157    0.078   32.528  134.022
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.20904    41.27293   0.490  0.62834
## LotSize      284.75545    88.52728   3.217  0.00336 **
## HouseSize     0.15136     0.02206   6.862 2.26e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59.46 on 27 degrees of freedom
## Multiple R-squared:  0.7729, Adjusted R-squared:  0.7561
## F-statistic: 45.95 on 2 and 27 DF, p-value: 2.034e-09

```

```
# Part i
# Three predictor model - assumption checking
# standardized residual plot - on fitted values
linefitH3.stres <- rstandard(linefitH3)
plot(linefitH3$fitted.values, linefitH3.stres, pch = 16, main = "Standardized
Residual Plot", xlab = "Fitted Value ($1000s)", ylab = "Standardized Residual
s")
abline(0,0, lty=2, col="red")
```

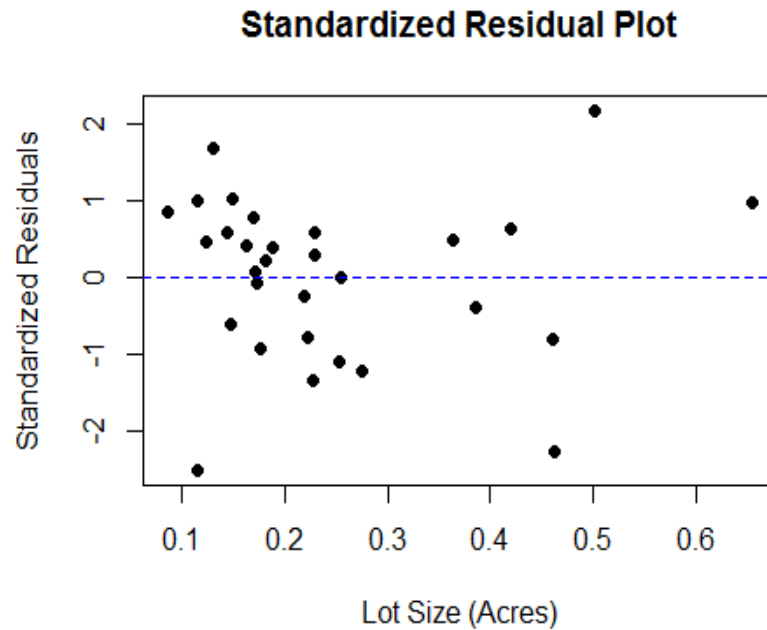


```
# Scatterplot Matrix
pairs(~ linefitH3.stres + LotSize + HouseSize + Age, main="Simple Scatterplot
Matrix")
```

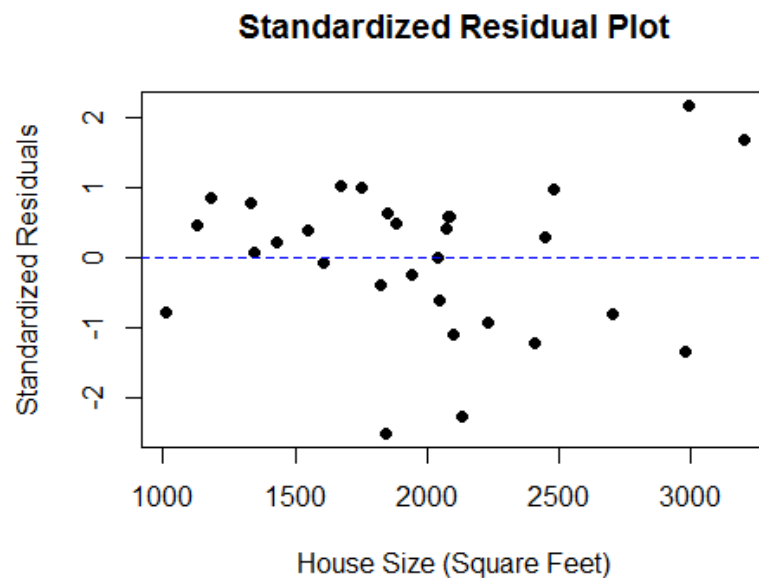


```
# Individual scatter plots against St Resids
# standardized residual plot - on Lot Size
plot(LotSize, linefitH3.stres, pch = 16, main = "Standardized Residual Plot",
```

```
xlab = "Lot Size (Acres)", ylab = "Standardized Residuals")
abline(0,0, lty=2, col="blue")
```

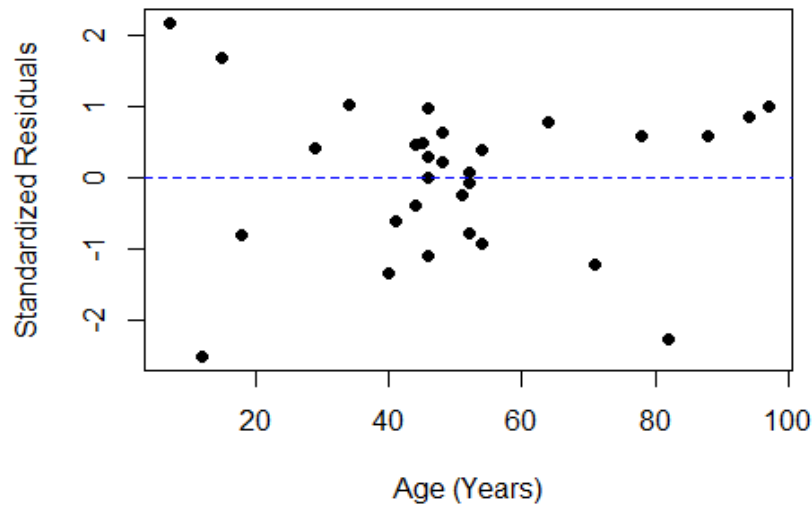


```
# standardized residual plot - on House Size
plot(HouseSize, linefitH3.stres, pch = 16, main = "Standardized Residual Plot",
xlab = "House Size (Square Feet)", ylab = "Standardized Residuals")
abline(0,0, lty=2, col="blue")
```



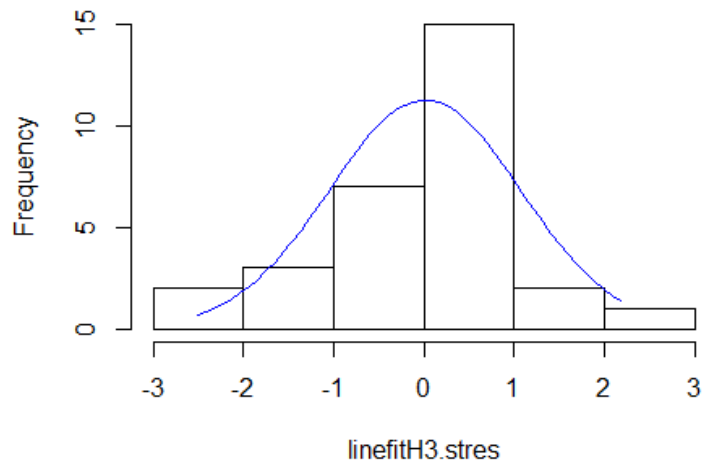
```
# standardized residual plot - on Age
plot(Age, linefitH3.stres, pch = 16, main = "Standardized Residual Plot", xlab = "Age (Years)",
ylab = "Standardized Residuals")
abline(0,0, lty=2, col="blue")
```

### Standardized Residual Plot



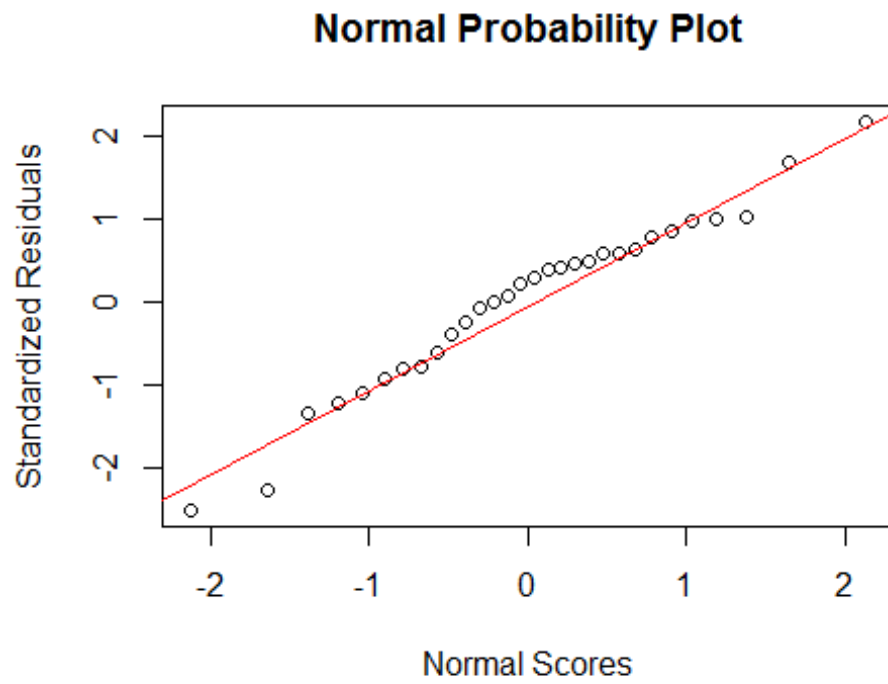
```
# Normality
# histogram with normal curve
h <- hist(linefitH3.stres)
x <- linefitH3.stres
xfit <- seq(min(x), max(x), length = 50)
yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue")
```

### Histogram of linefitH3.stres



```
# normal probability plot
qqnorm(linefitH3.stres, main = "Normal Probability Plot", xlab = "Normal Scores", ylab = "Standardized Residuals")
qqline(linefitH3.stres, col = "red")
```





```
# hypothesis test
shapiro.test(linefitH3.stres)

##
##  Shapiro-Wilk normality test
##
## data:  linefitH3.stres
## W = 0.96723, p-value = 0.4665

# clean up
detach(house)
```

```

# Exercise 7-2: Detergent
# Data file
detergent <- read_excel("Fresh.xlsx", na="NA", col_names = TRUE)
attach(detergent)

# Parts a/c
# fit the model - 3 predictors
linefitD3 <- lm(Demand ~ Price + IndPrice + AdvExp)
summary(linefitD3)

##
## Call:
## lm(formula = Demand ~ Price + IndPrice + AdvExp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41065 -0.11562 -0.00984  0.13466  0.51361
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.5891     2.4450   3.104 0.004567 **
## Price        -2.3577     0.6379  -3.696 0.001028 **
## IndPrice       1.6122     0.2954   5.459 1.01e-05 ***
## AdvExp         0.5012     0.1259   3.981 0.000491 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2347 on 26 degrees of freedom
## Multiple R-squared:  0.8936, Adjusted R-squared:  0.8813
## F-statistic: 72.8 on 3 and 26 DF, p-value: 8.883e-13

# Part b
# collinearity check: correlations among predictors
cor(detergent[,1:3])

##              Price    IndPrice    AdvExp
## Price      1.00000000 0.07836681 -0.468793
## IndPrice    0.07836681 1.00000000  0.604540
## AdvExp     -0.46879301 0.60454000  1.000000

```

```

# Part d
# fit the model - 2 predictors
linefitD2 <- lm(Demand ~ IndPrice + AdvExp)
summary(linefitD2)

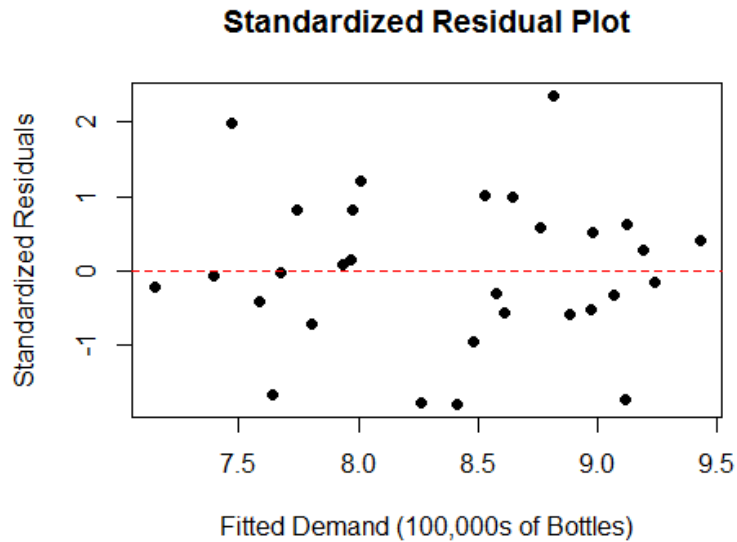
##
## Call:
## lm(formula = Demand ~ IndPrice + AdvExp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.68956 -0.14045  0.00676  0.14017  0.61142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.9524     0.9671  -0.985   0.33349
## IndPrice       1.0510     0.3070   3.423   0.00199 **
## AdvExp         0.8035     0.1159   6.931   1.9e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2844 on 27 degrees of freedom
## Multiple R-squared:  0.8377, Adjusted R-squared:  0.8257
## F-statistic: 69.69 on 2 and 27 DF,  p-value: 2.18e-11

# fit the model - 1 predictor
linefitD1 <- lm(Demand ~ AdvExp)
summary(linefitD1)

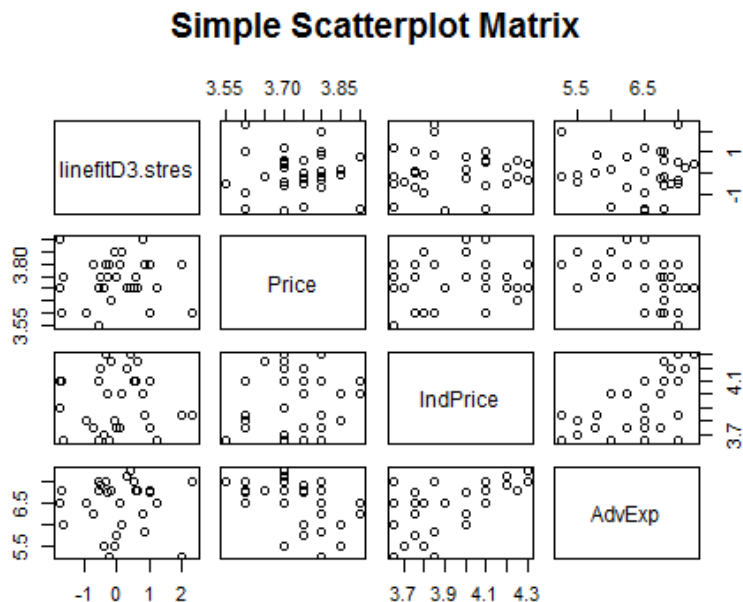
##
## Call:
## lm(formula = Demand ~ AdvExp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63964 -0.15886  0.01624  0.15343  0.74295
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.6490     0.7035   2.344   0.0264 *
## AdvExp        1.0434     0.1086   9.609 2.31e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3344 on 28 degrees of freedom
## Multiple R-squared:  0.7673, Adjusted R-squared:  0.759
## F-statistic: 92.32 on 1 and 28 DF,  p-value: 2.307e-10

```

```
# Part e
# Three predictor model - assumption checking
# standardized residual plot - on fitted values
linefitD3.stres <- rstandard(linefitD3)
plot(linefitD3$fitted.values, linefitD3.stres, pch = 16, main = "Standardized
Residual Plot", xlab = "Fitted Demand (100,000s of Bottles)", ylab = "Standar
dized Residuals")
abline(0,0, lty=2, col="red")
```

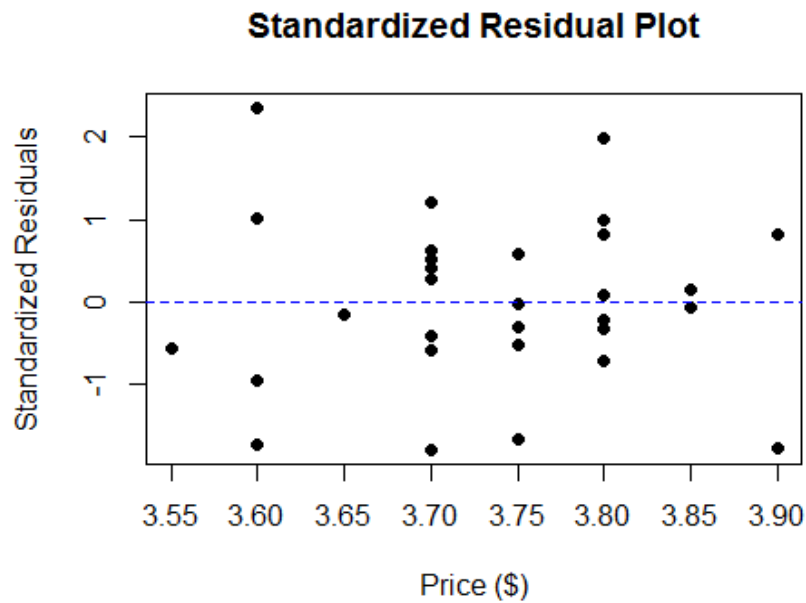


```
# Scatterplot Matrix
pairs(~ linefitD3.stres + Price + IndPrice + AdvExp, main="Simple Scatterplot
Matrix")
```

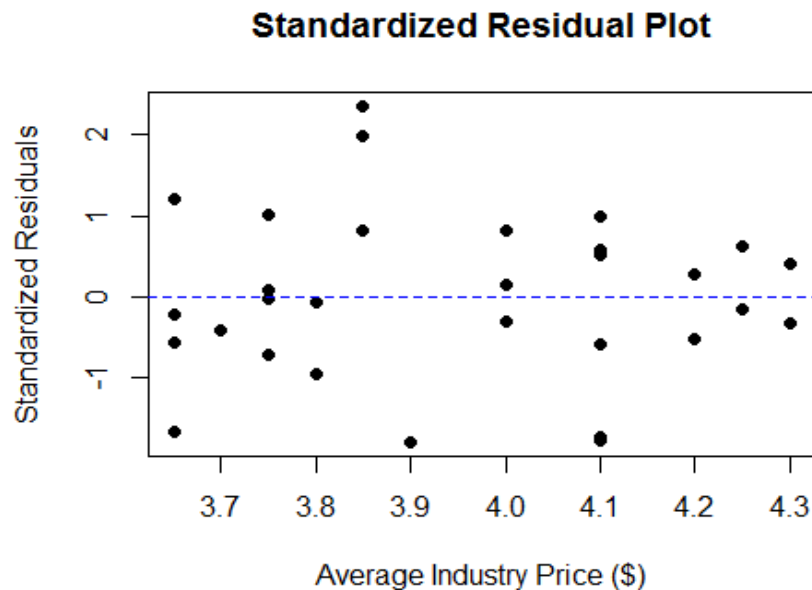


```
# individual scatter plots against St Resids
# standardized residual plot - on Price
```

```
plot(Price, linefitD3.stres, pch = 16, main = "Standardized Residual Plot", xlab = "Price ($)", ylab = "Standardized Residuals")
abline(0,0, lty=2, col="blue")
```

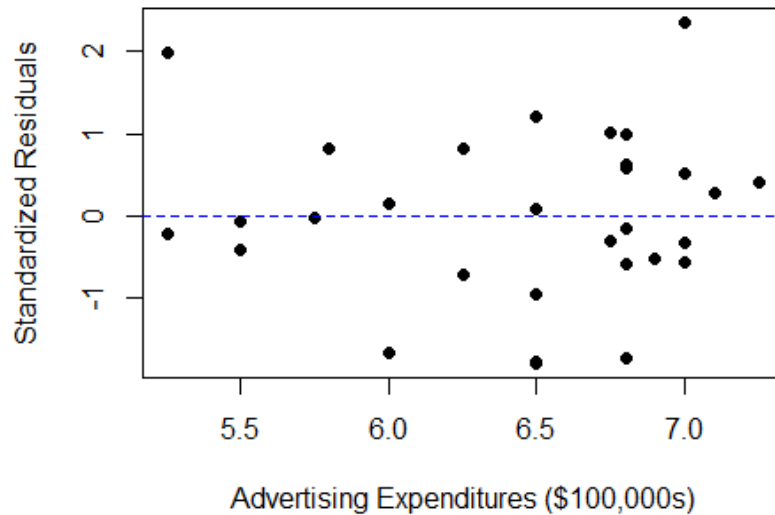


```
# standardized residual plot - on IndPrice
plot(IndPrice, linefitD3.stres, pch = 16, main = "Standardized Residual Plot", xlab = "Average Industry Price ($)", ylab = "Standardized Residuals")
abline(0,0, lty=2, col="blue")
```



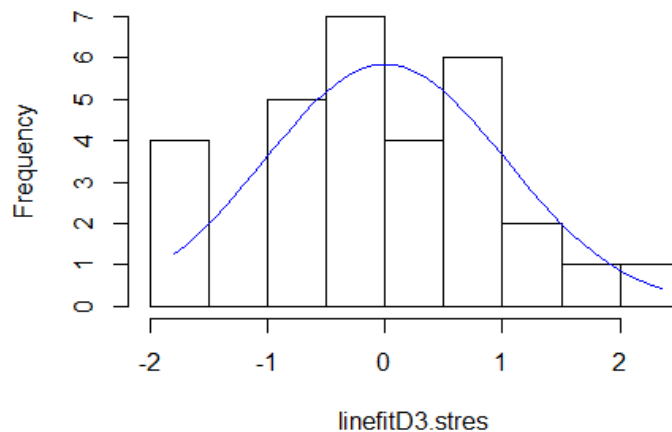
```
# standardized residual plot - on AdvExp
plot(AdvExp, linefitD3.stres, pch = 16, main = "Standardized Residual Plot", xlab = "Advertising Expenditures ($100,000s)", ylab = "Standardized Residuals")
abline(0,0, lty=2, col="blue")
```

### Standardized Residual Plot

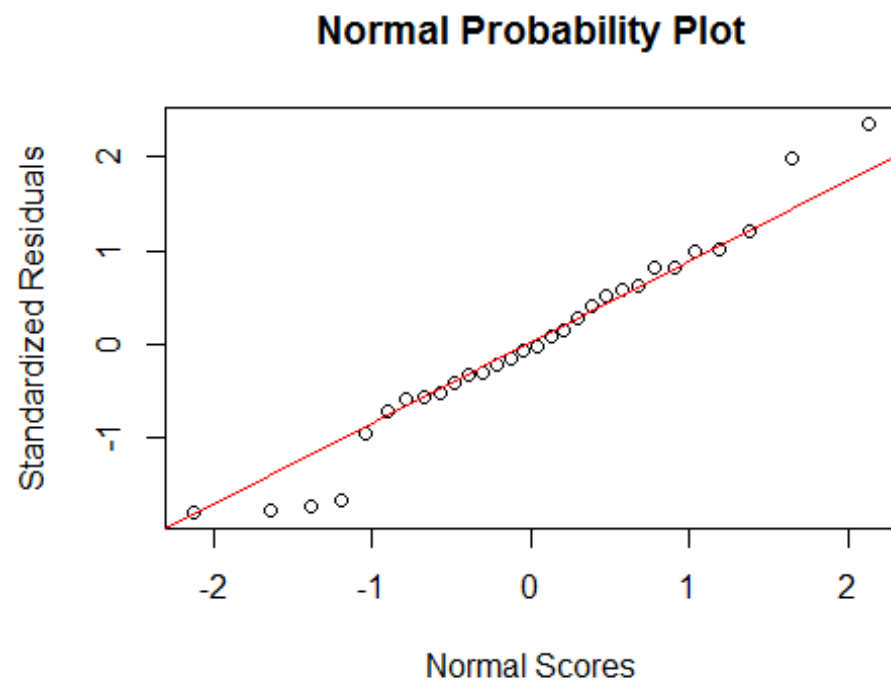


```
# Normality
# histogram with normal curve
h <- hist(linefitD3.stres)
x <- linefitD3.stres
xfit <- seq(min(x), max(x), length = 50)
yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue")
```

### Histogram of linefitD3.stres



```
# normal probability plot
qqnorm(linefitD3.stres, main = "Normal Probability Plot", xlab = "Normal Scores", ylab = "Standardized Residuals")
qqline(linefitD3.stres, col = "red")
```



```
# Hypothesis test
shapiro.test(linefitD3.stres)

##
##  Shapiro-Wilk normality test
##
## data:  linefitD3.stres
## W = 0.96993, p-value = 0.5371
```

```

# Part f
# fit the model - 2 modified predictors
linefitD2b <- lm(Demand ~ AdvExp + PriceDiff)
summary(linefitD2b)

##
## Call:
## lm(formula = Demand ~ AdvExp + PriceDiff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49779 -0.12031 -0.00867  0.11084  0.58106
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.4075     0.7223   6.102 1.62e-06 ***
## AdvExp        0.5635     0.1191   4.733 6.25e-05 ***
## PriceDiff     1.5883     0.2994   5.304 1.35e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2383 on 27 degrees of freedom
## Multiple R-squared:  0.886, Adjusted R-squared:  0.8776
## F-statistic: 105 on 2 and 27 DF, p-value: 1.845e-13

#clean up
detach(detergent)

```