Project\_1\_Linear Regression Models

Sanjeev Tiwari

October 28, 2017

grades<-read.csv(file.choose()) # REading the Grades data set within a variable  
summary(grades) # Displaying a summary of dataset

## Sr\_No id lastname firstname   
## Min. : 1 Min. :106484 LANGFORD : 3 ERIC : 3   
## 1st Qu.: 27 1st Qu.:337908 HUANG : 2 DAWN : 2   
## Median : 53 Median :574170 JONES : 2 JACKIE : 2   
## Mean : 53 Mean :571367 ROBINSON : 2 JIM : 2   
## 3rd Qu.: 79 3rd Qu.:807963 VALENZUELA: 2 ROBERT : 2   
## Max. :105 Max. :988808 AHGHEL : 1 STACY : 2   
## (Other) :93 (Other):92   
## gender ethnicity year lowup section   
## Min. :1.00 Min. :1.000 Min. :1.000 Min. :1.00 Min. :1   
## 1st Qu.:1.00 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:2.00 1st Qu.:1   
## Median :1.00 Median :4.000 Median :3.000 Median :2.00 Median :2   
## Mean :1.39 Mean :3.352 Mean :2.943 Mean :1.79 Mean :2   
## 3rd Qu.:2.00 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:2.00 3rd Qu.:3   
## Max. :2.00 Max. :5.000 Max. :4.000 Max. :2.00 Max. :3   
##   
## gpa extrc review quiz1   
## Min. :1.140 Min. :1.00 Min. :1.000 Min. : 0.000   
## 1st Qu.:2.300 1st Qu.:1.00 1st Qu.:1.000 1st Qu.: 6.000   
## Median :2.720 Median :1.00 Median :2.000 Median : 8.000   
## Mean :2.779 Mean :1.21 Mean :1.667 Mean : 7.467   
## 3rd Qu.:3.490 3rd Qu.:1.00 3rd Qu.:2.000 3rd Qu.:10.000   
## Max. :4.000 Max. :2.00 Max. :2.000 Max. :10.000   
##   
## quiz2 quiz3 quiz4 quiz5   
## Min. : 3.000 Min. : 0.000 Min. : 2.0 Min. : 2.000   
## 1st Qu.: 7.000 1st Qu.: 6.000 1st Qu.: 6.0 1st Qu.: 7.000   
## Median : 8.000 Median : 9.000 Median : 8.0 Median : 8.000   
## Mean : 7.981 Mean : 7.981 Mean : 7.8 Mean : 7.867   
## 3rd Qu.: 9.000 3rd Qu.:10.000 3rd Qu.:10.0 3rd Qu.: 9.000   
## Max. :10.000 Max. :10.000 Max. :10.0 Max. :10.000   
##   
## final total percent grade passfail  
## Min. :40.00 Min. : 51.0 Min. :41.00 A:23 F: 6   
## 1st Qu.:57.00 1st Qu.: 92.0 1st Qu.:74.00 B:32 O: 1   
## Median :62.00 Median :103.0 Median :82.00 C:35 P:98   
## Mean :61.48 Mean :100.6 Mean :80.34 D: 9   
## 3rd Qu.:68.00 3rd Qu.:111.0 3rd Qu.:87.00 F: 6   
## Max. :75.00 Max. :124.0 Max. :99.00   
##

library('psych')  
describe(grades) # Display the extended summary of hte dataset including hte skewness and kurtosis

## vars n mean sd median trimmed mad  
## Sr\_No 1 105 53.00 30.45 53.00 53.00 38.55  
## id 2 105 571366.67 277404.13 574170.00 575307.45 350282.04  
## lastname\* 3 105 50.26 28.28 49.00 50.32 35.58  
## firstname\* 4 105 49.15 28.18 48.00 49.07 35.58  
## gender 5 105 1.39 0.49 1.00 1.36 0.00  
## ethnicity 6 105 3.35 1.06 4.00 3.38 1.48  
## year 7 105 2.94 0.69 3.00 2.96 0.00  
## lowup 8 105 1.79 0.41 2.00 1.86 0.00  
## section 9 105 2.00 0.80 2.00 2.00 1.48  
## gpa 10 105 2.78 0.76 2.72 2.80 0.76  
## extrc 11 105 1.21 0.41 1.00 1.14 0.00  
## review 12 105 1.67 0.47 2.00 1.71 0.00  
## quiz1 13 105 7.47 2.48 8.00 7.76 2.97  
## quiz2 14 105 7.98 1.62 8.00 8.12 1.48  
## quiz3 15 105 7.98 2.31 9.00 8.34 1.48  
## quiz4 16 105 7.80 2.28 8.00 8.11 2.97  
## quiz5 17 105 7.87 1.77 8.00 8.02 1.48  
## final 18 105 61.48 7.94 62.00 61.74 8.90  
## total 19 105 100.57 15.30 103.00 101.80 13.34  
## percent 20 105 80.34 12.14 82.00 81.29 10.38  
## grade\* 21 105 2.46 1.10 2.00 2.38 1.48  
## passfail\* 22 105 2.88 0.47 3.00 3.00 0.00  
## min max range skew kurtosis se  
## Sr\_No 1.00 105 104.00 0.00 -1.23 2.97  
## id 106484.00 988808 882324.00 -0.09 -1.33 27071.87  
## lastname\* 1.00 99 98.00 0.00 -1.19 2.76  
## firstname\* 1.00 98 97.00 0.04 -1.24 2.75  
## gender 1.00 2 1.00 0.44 -1.82 0.05  
## ethnicity 1.00 5 4.00 -0.44 -0.63 0.10  
## year 1.00 4 3.00 -0.45 0.40 0.07  
## lowup 1.00 2 1.00 -1.41 -0.02 0.04  
## section 1.00 3 2.00 0.00 -1.44 0.08  
## gpa 1.14 4 2.86 -0.05 -0.87 0.07  
## extrc 1.00 2 1.00 1.41 -0.02 0.04  
## review 1.00 2 1.00 -0.70 -1.53 0.05  
## quiz1 0.00 10 10.00 -0.83 0.04 0.24  
## quiz2 3.00 10 7.00 -0.64 -0.35 0.16  
## quiz3 0.00 10 10.00 -1.10 0.59 0.23  
## quiz4 2.00 10 8.00 -0.89 -0.09 0.22  
## quiz5 2.00 10 8.00 -0.69 0.16 0.17  
## final 40.00 75 35.00 -0.33 -0.42 0.78  
## total 51.00 124 73.00 -0.81 0.77 1.49  
## percent 41.00 99 58.00 -0.81 0.78 1.18  
## grade\* 1.00 5 4.00 0.45 -0.34 0.11  
## passfail\* 1.00 3 2.00 -3.58 11.11 0.05

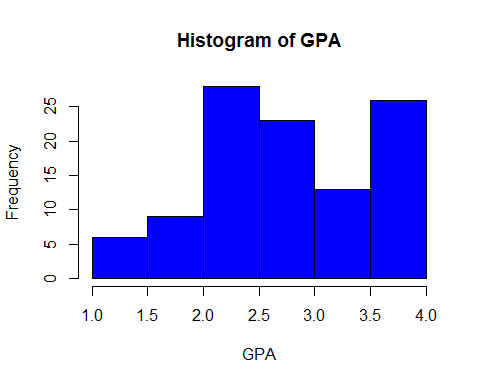
describe(grades$gpa)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 105 2.78 0.76 2.72 2.8 0.76 1.14 4 2.86 -0.05 -0.87  
## se  
## X1 0.07

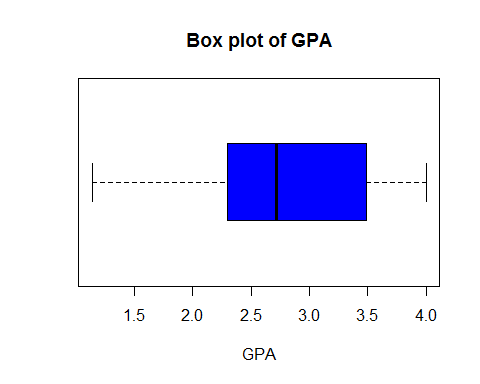
stem(grades$gpa)

##   
## The decimal point is 1 digit(s) to the left of the |  
##   
## 10 | 48  
## 12 | 443  
## 14 | 0  
## 16 | 1677  
## 18 | 04515  
## 20 | 23919  
## 22 | 1225780123445788  
## 24 | 03356671446667  
## 26 | 1623477  
## 28 | 004480266  
## 30 | 1256379  
## 32 | 85  
## 34 | 2599334778  
## 36 | 460  
## 38 | 4400000000000058  
## 40 | 0

hist(grades$gpa, main = "Histogram of GPA", xlab = "GPA", ylab = "Frequency", col = "Blue")



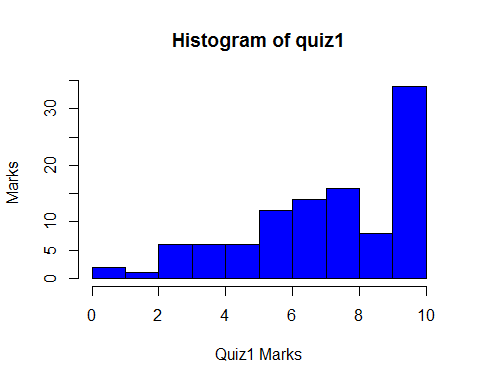
boxplot(grades$gpa, main = "Box plot of GPA", xlab = "GPA", col = "Blue", horizontal = T)



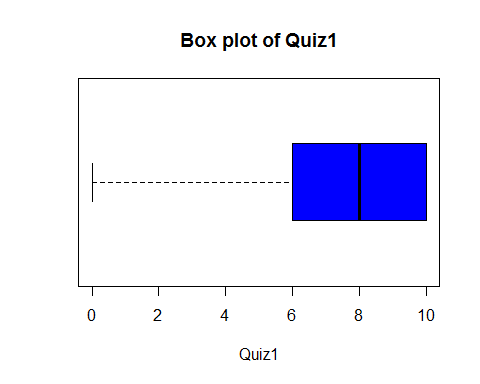
describe(grades$quiz1)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 105 7.47 2.48 8 7.76 2.97 0 10 10 -0.83 0.04  
## se  
## X1 0.24

hist(grades$quiz1, main = "Histogram of quiz1", xlab = "Quiz1 Marks", ylab = "Marks", col = "Blue")



boxplot(grades$quiz1, main = "Box plot of Quiz1", xlab = "Quiz1", col = "Blue", horizontal = T)



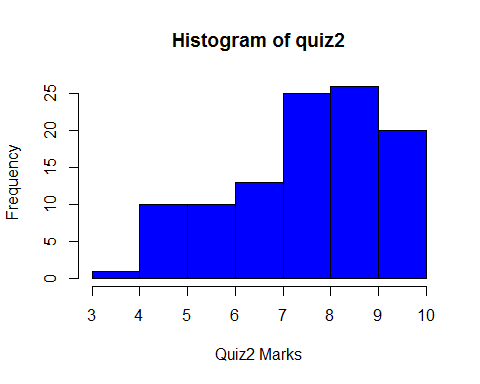
stem(grades$quiz1)

##   
## The decimal point is at the |  
##   
## 0 | 00  
## 1 |   
## 2 | 0  
## 3 | 000000  
## 4 | 000000  
## 5 | 000000  
## 6 | 000000000000  
## 7 | 00000000000000  
## 8 | 0000000000000000  
## 9 | 00000000  
## 10 | 0000000000000000000000000000000000

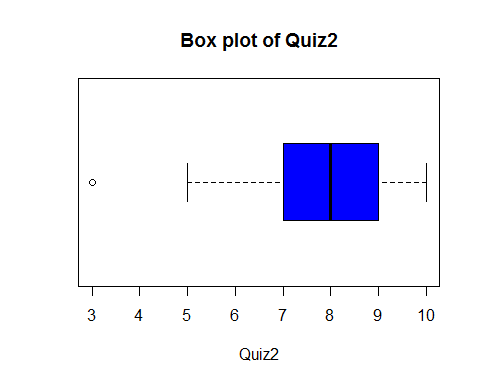
describe(grades$quiz2)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 105 7.98 1.62 8 8.12 1.48 3 10 7 -0.64 -0.35  
## se  
## X1 0.16

hist(grades$quiz2, main = "Histogram of quiz2", xlab = "Quiz2 Marks", ylab = "Frequency", col = "Blue")



boxplot(grades$quiz2, main = "Box plot of Quiz2", xlab = "Quiz2", col = "Blue", horizontal = T)



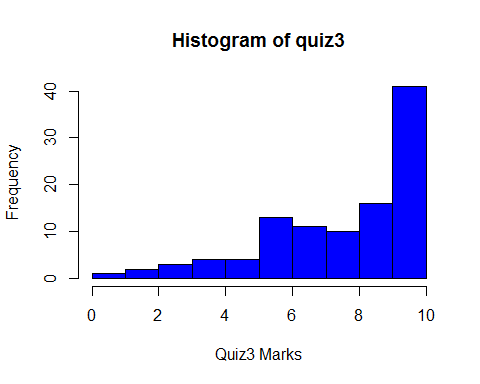
stem(grades$quiz2)

##   
## The decimal point is at the |  
##   
## 3 | 0  
## 3 |   
## 4 |   
## 4 |   
## 5 | 0000000000  
## 5 |   
## 6 | 0000000000  
## 6 |   
## 7 | 0000000000000  
## 7 |   
## 8 | 0000000000000000000000000  
## 8 |   
## 9 | 00000000000000000000000000  
## 9 |   
## 10 | 00000000000000000000

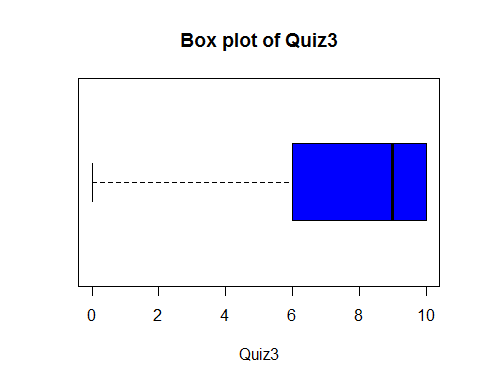
describe(grades$quiz3)

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 105 7.98 2.31 9 8.34 1.48 0 10 10 -1.1 0.59 0.23

hist(grades$quiz3, main = "Histogram of quiz3", xlab = "Quiz3 Marks", ylab = "Frequency", col = "Blue")



boxplot(grades$quiz3, main = "Box plot of Quiz3", xlab = "Quiz3", col = "Blue", horizontal = T)



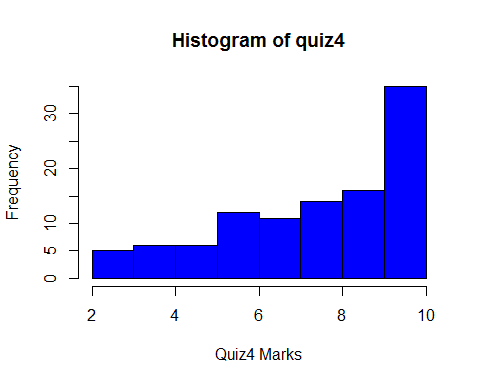
stem(grades$quiz3)

##   
## The decimal point is at the |  
##   
## 0 | 0  
## 1 |   
## 2 | 00  
## 3 | 000  
## 4 | 0000  
## 5 | 0000  
## 6 | 0000000000000  
## 7 | 00000000000  
## 8 | 0000000000  
## 9 | 0000000000000000  
## 10 | 00000000000000000000000000000000000000000

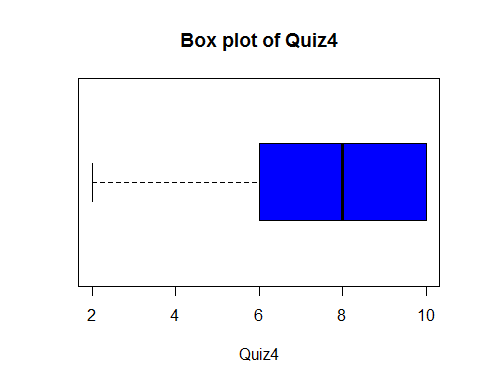
describe(grades$quiz4)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 105 7.8 2.28 8 8.11 2.97 2 10 8 -0.89 -0.09  
## se  
## X1 0.22

hist(grades$quiz4, main = "Histogram of quiz4", xlab = "Quiz4 Marks", ylab = "Frequency", col = "Blue")



boxplot(grades$quiz4, main = "Box plot of Quiz4", xlab = "Quiz4", col = "Blue", horizontal = T)



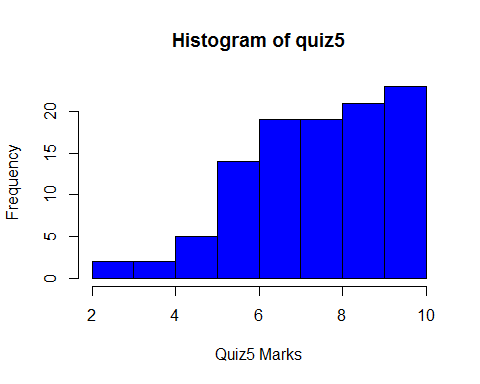
stem(grades$quiz4)

##   
## The decimal point is at the |  
##   
## 2 | 00000  
## 2 |   
## 3 |   
## 3 |   
## 4 | 000000  
## 4 |   
## 5 | 000000  
## 5 |   
## 6 | 000000000000  
## 6 |   
## 7 | 00000000000  
## 7 |   
## 8 | 00000000000000  
## 8 |   
## 9 | 0000000000000000  
## 9 |   
## 10 | 00000000000000000000000000000000000

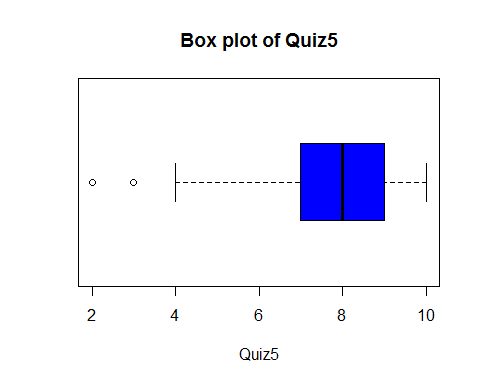
describe(grades$quiz5)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 105 7.87 1.77 8 8.02 1.48 2 10 8 -0.69 0.16  
## se  
## X1 0.17

hist(grades$quiz5, main = "Histogram of quiz5", xlab = "Quiz5 Marks", ylab = "Frequency", col = "Blue")



boxplot(grades$quiz5, main = "Box plot of Quiz5", xlab = "Quiz5", col = "Blue", horizontal = T)



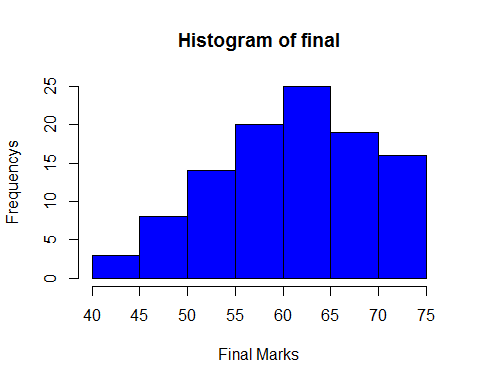
stem(grades$quiz5)

##   
## The decimal point is at the |  
##   
## 2 | 0  
## 2 |   
## 3 | 0  
## 3 |   
## 4 | 00  
## 4 |   
## 5 | 00000  
## 5 |   
## 6 | 00000000000000  
## 6 |   
## 7 | 0000000000000000000  
## 7 |   
## 8 | 0000000000000000000  
## 8 |   
## 9 | 000000000000000000000  
## 9 |   
## 10 | 00000000000000000000000

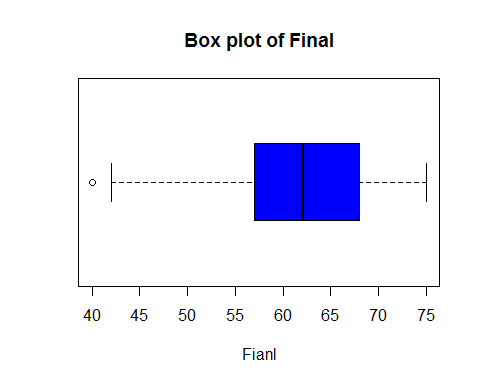
describe(grades$final)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 105 61.48 7.94 62 61.74 8.9 40 75 35 -0.33 -0.42  
## se  
## X1 0.78

hist(grades$final, main = "Histogram of final", xlab = "Final Marks", ylab = "Frequencys", col = "Blue")



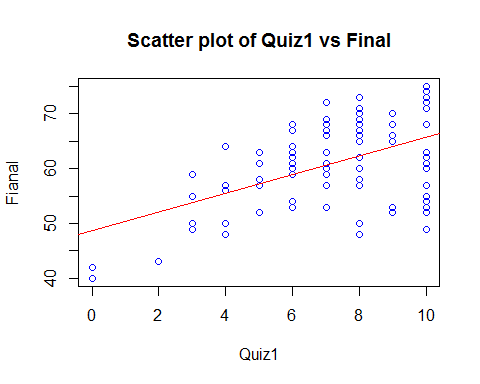
boxplot(grades$final, main = "Box plot of Final", xlab = "Fianl", col = "Blue", horizontal = T)



stem(grades$final)

##   
## The decimal point is at the |  
##   
## 40 | 0  
## 42 | 00  
## 44 |   
## 46 |   
## 48 | 00000  
## 50 | 000  
## 52 | 000000000  
## 54 | 00000  
## 56 | 0000000  
## 58 | 000000  
## 60 | 000000000000  
## 62 | 00000000000000  
## 64 | 000000  
## 66 | 0000000  
## 68 | 0000000000  
## 70 | 000000  
## 72 | 00000  
## 74 | 0000000

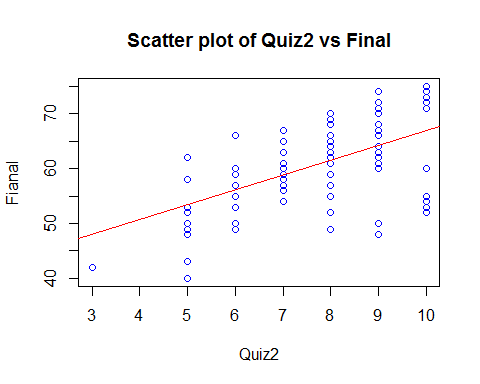
plot(final~quiz1, data = grades,   
 main= "Scatter plot of Quiz1 vs Final",  
 col= 'blue',   
 ylab= 'Fianal', xlab="Quiz1")  
  
# add the regression line / bestfit line  
abline(lm(grades$final~grades$quiz1), col="red")



# Corelation test  
cor.test(grades$final, grades$quiz1)

##   
## Pearson's product-moment correlation  
##   
## data: grades$final and grades$quiz1  
## t = 6.428, df = 103, p-value = 4.094e-09  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3826533 0.6591421  
## sample estimates:  
## cor   
## 0.5350754

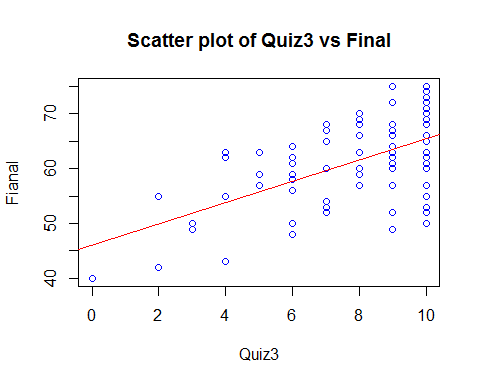
plot(final~quiz2, data = grades,   
 main= "Scatter plot of Quiz2 vs Final",  
 col= 'blue',   
 ylab= 'Fianal', xlab="Quiz2")  
  
# add the regression line / bestfit line  
abline(lm(grades$final~grades$quiz2), col="red")



# Corelation test  
cor.test(grades$final, grades$quiz2)

##   
## Pearson's product-moment correlation  
##   
## data: grades$final and grades$quiz2  
## t = 6.7162, df = 103, p-value = 1.047e-09  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4028080 0.6724089  
## sample estimates:  
## cor   
## 0.5518668

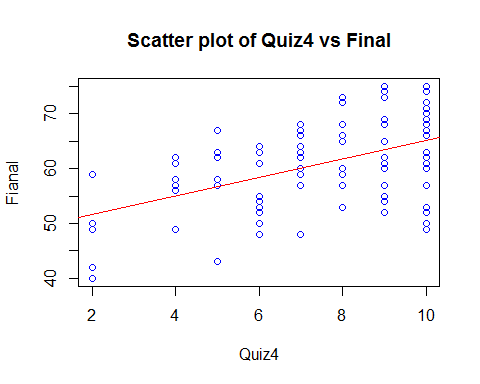
plot(final~quiz3, data = grades,   
 main= "Scatter plot of Quiz3 vs Final",  
 col= 'blue',   
 ylab= 'Fianal', xlab="Quiz3")  
  
# add the regression line / bestfit line  
abline(lm(grades$final~grades$quiz3), col="red")



# Corelation test  
cor.test(grades$final, grades$quiz3)

##   
## Pearson's product-moment correlation  
##   
## data: grades$final and grades$quiz3  
## t = 6.8809, df = 103, p-value = 4.758e-10  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4140461 0.6797318  
## sample estimates:  
## cor   
## 0.5611773

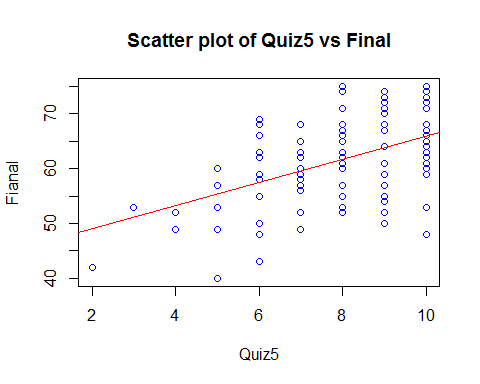
plot(final~quiz4, data = grades,   
 main= "Scatter plot of Quiz4 vs Final",  
 col= 'blue',   
 ylab= 'Fianal', xlab="Quiz4")  
  
# add the regression line / bestfit line  
abline(lm(grades$final~grades$quiz4), col="red")



# Corelation test  
cor.test(grades$final, grades$quiz4)

##   
## Pearson's product-moment correlation  
##   
## data: grades$final and grades$quiz4  
## t = 5.6716, df = 103, p-value = 1.303e-07  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3267180 0.6213986  
## sample estimates:  
## cor   
## 0.4878348

plot(final~quiz5, data = grades,   
 main= "Scatter plot of Quiz5 vs Final",  
 col= 'blue',   
 ylab= 'Fianal', xlab="Quiz5")  
  
# add the regression line / bestfit line  
abline(lm(grades$final~grades$quiz5), col="red")



# Corelation test  
cor.test(grades$final, grades$quiz5)

##   
## Pearson's product-moment correlation  
##   
## data: grades$final and grades$quiz5  
## t = 5.4264, df = 103, p-value = 3.834e-07  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3076485 0.6082107  
## sample estimates:  
## cor   
## 0.4715109

# A regression model for collective set of all the predictors  
model1<-lm(final ~ quiz1+quiz2+quiz3+quiz4+quiz5, data=grades)  
summary(model1)

##   
## Call:  
## lm(formula = final ~ quiz1 + quiz2 + quiz3 + quiz4 + quiz5, data = grades)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.317 -3.641 1.379 4.470 8.887   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 38.7000 3.3600 11.518 <2e-16 \*\*\*  
## quiz1 0.2859 0.5664 0.505 0.615   
## quiz2 0.9422 0.6610 1.425 0.157   
## quiz3 0.9222 0.5724 1.611 0.110   
## quiz4 0.0335 0.5183 0.065 0.949   
## quiz5 0.6993 0.4999 1.399 0.165   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.399 on 99 degrees of freedom  
## Multiple R-squared: 0.3823, Adjusted R-squared: 0.3512   
## F-statistic: 12.26 on 5 and 99 DF, p-value: 2.956e-09

# check the variance factors  
library('faraway')

## Warning: package 'faraway' was built under R version 3.4.2

##   
## Attaching package: 'faraway'

## The following object is masked from 'package:psych':  
##   
## logit

vif(model1)>5

## quiz1 quiz2 quiz3 quiz4 quiz5   
## TRUE FALSE FALSE FALSE FALSE

#Since Quiz1 reported to have a vif value >5, it is descarded from and a   
# new model excluding quiz1 is to be built for further investigations.  
model2<-lm(final ~ quiz2+quiz3+quiz4+quiz5, data=grades)  
summary(model2)

##   
## Call:  
## lm(formula = final ~ quiz2 + quiz3 + quiz4 + quiz5, data = grades)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.778 -3.498 1.213 4.502 9.148   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 38.3404 3.2714 11.720 <2e-16 \*\*\*  
## quiz2 0.9687 0.6565 1.476 0.1432   
## quiz3 1.0729 0.4866 2.205 0.0297 \*   
## quiz4 0.1508 0.4616 0.327 0.7445   
## quiz5 0.7202 0.4963 1.451 0.1499   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.375 on 100 degrees of freedom  
## Multiple R-squared: 0.3808, Adjusted R-squared: 0.356   
## F-statistic: 15.37 on 4 and 100 DF, p-value: 7.85e-10

# check the variance factors  
library('faraway')  
vif(model2)>5

## quiz2 quiz3 quiz4 quiz5   
## FALSE FALSE FALSE FALSE

#From R output of summary(model2) it is evident that slope for   
# predictor variable - quiz2, quiz4, and quiz5 are not significant  
# and hence can be descarded from the model.  
#so, new model is model3 as follows  
model3<-lm(final ~ quiz3, data=grades)  
summary(model3)

##   
## Call:  
## lm(formula = final ~ quiz3, data = grades)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.3759 -4.3759 0.5556 5.6241 11.5556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 46.0613 2.3312 19.759 < 2e-16 \*\*\*  
## quiz3 1.9315 0.2807 6.881 4.76e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.607 on 103 degrees of freedom  
## Multiple R-squared: 0.3149, Adjusted R-squared: 0.3083   
## F-statistic: 47.35 on 1 and 103 DF, p-value: 4.758e-10

# check the variance factors  
library('faraway')  
vif(model3)>5

## quiz3   
## FALSE

#anova test on model to check the F-Values & validation against   
# the F-Statistics for model2 summary results  
anova(model2)

## Analysis of Variance Table  
##   
## Response: final  
## Df Sum Sq Mean Sq F value Pr(>F)   
## quiz2 1 1998.6 1998.56 49.1822 2.803e-10 \*\*\*  
## quiz3 1 410.7 410.66 10.1058 0.001968 \*\*   
## quiz4 1 3.8 3.82 0.0939 0.759916   
## quiz5 1 85.6 85.57 2.1058 0.149872   
## Residuals 100 4063.6 40.64   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#let check Durbin watson statistics valur for model2  
# install.packages('car')  
library('car')

## Warning: package 'car' was built under R version 3.4.2

##   
## Attaching package: 'car'

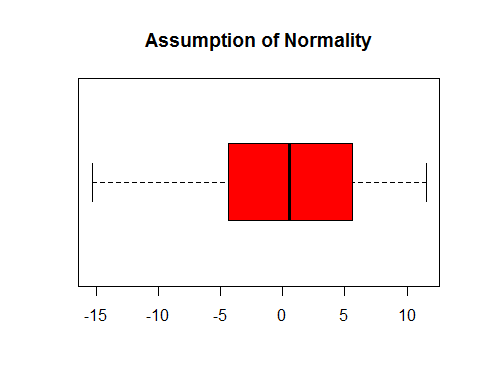
## The following objects are masked from 'package:faraway':  
##   
## logit, vif

## The following object is masked from 'package:psych':  
##   
## logit

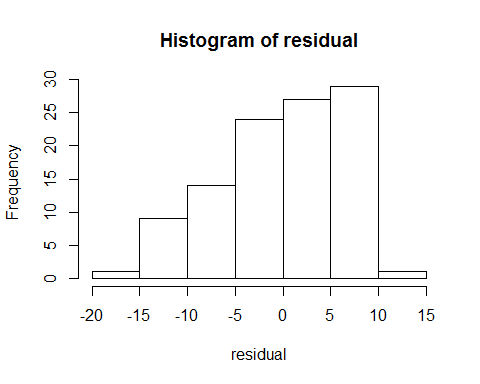
dwt(model2)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.1150608 2.221103 0.226  
## Alternative hypothesis: rho != 0

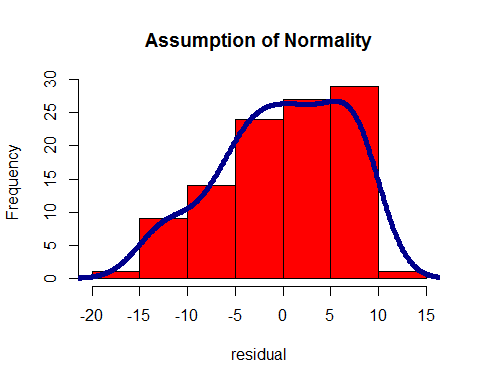
#Q9-Verification of Assumptions - Assumption of Normality  
residual<-residuals(model3)  
boxplot(residual, main = "Assumption of Normality", col = 'red', horizontal = T)



# Histogram with overlay curve  
hstError <- hist(residual)



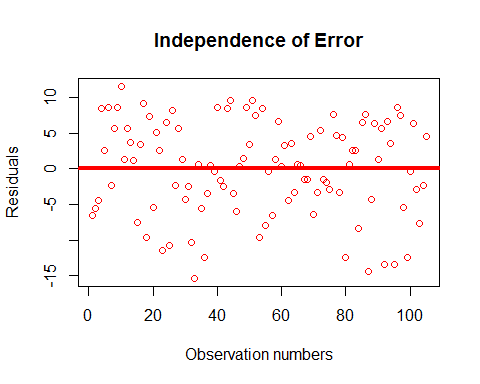
multiplier<-hstError$counts / hstError$density  
densError <- density(residual)  
densError$y<-densError$y \* multiplier[1]  
plot(hstError, main = "Assumption of Normality", col = 'red')  
lines(densError, col = "darkblue", lwd = 5)



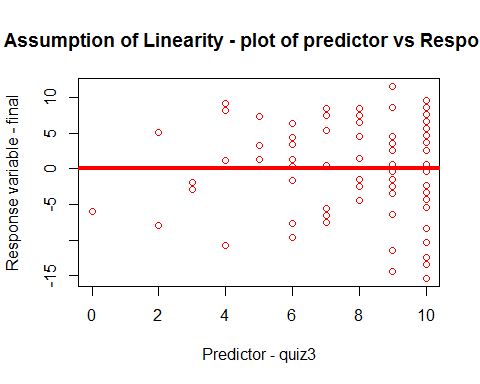
library('psych')  
describe(residual)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 105 0 6.57 0.56 0.41 7.51 -15.38 11.56 26.93 -0.4 -0.73  
## se  
## X1 0.64

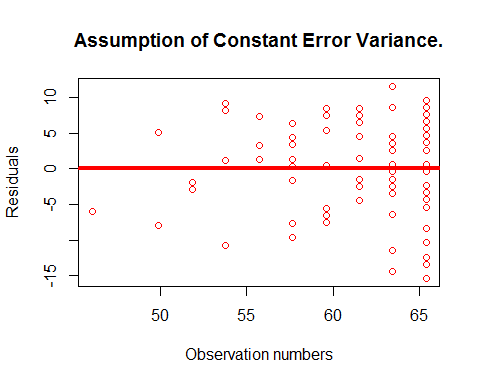
#Q10- Verification of Assumptions - Independence of Error  
library('psych')  
residual<-residuals(model3)  
obsNo<-seq(1:length(residual))  
dsIOE<-data.frame(obsNo, residual)  
plot(dsIOE$obsNo, dsIOE$residual, main= "Independence of Error",  
 col= 'red', ylab= 'Residuals', xlab="Observation numbers")  
abline(h = 0, col = "red", lwd=4)



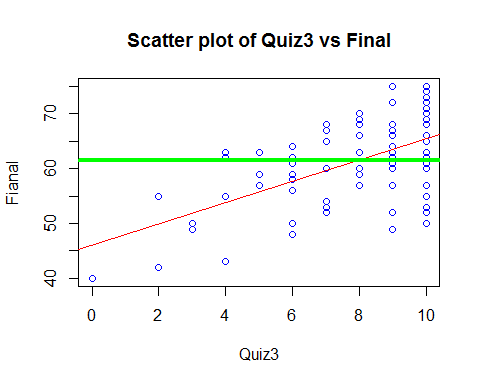
# Q11-Verification of Assumptions - Linearity  
library('psych')  
residual<-residuals(model3)  
plot(grades$quiz3, residual,   
 main= "Assumption of Linearity - plot of predictor vs Response",  
 col= 'red', ylab= 'Response variable - final', xlab="Predictor - quiz3")  
abline(h = 0, col = "red", lwd=4)



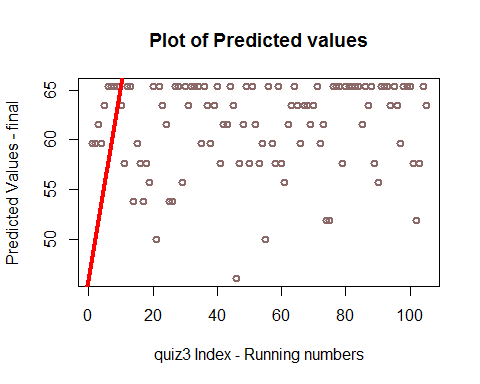
# Q12- Verification of Assumptions - Constant Error Variance  
library('psych')  
residual<-residuals(model3)  
dsPredict<-predict(model3)  
plot(dsPredict, residual, main= "Assumption of Constant Error Variance.",  
 col= 'red', ylab= 'Residuals', xlab="Observation numbers")  
abline(h = 0, col = "red", lwd=4)



#Q13 - Standard error of estimate  
plot(final~quiz3, data = grades, main= "Scatter plot of Quiz3 vs Final",  
 col= 'blue', ylab= 'Fianal', xlab="Quiz3")  
abline(lm(grades$final~grades$quiz3), col="red")  
abline(h=mean(grades$final),col='green',lwd=4)

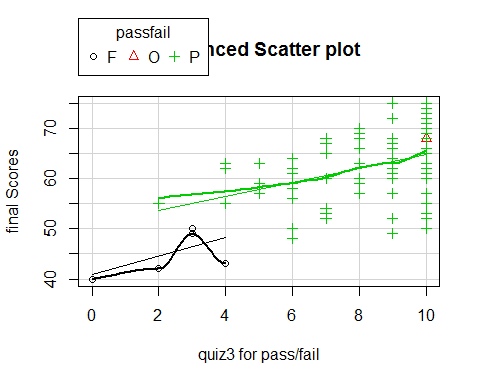


# elaboration of hypothetical values of predictors  
plot(predict(model3),col="rosybrown4",lwd=2,  
 main='Plot of Predicted values',  
 xlab='quiz3 Index - Running numbers',  
 ylab = 'Predicted Values - final')  
abline(model3, col='red', lwd=4)

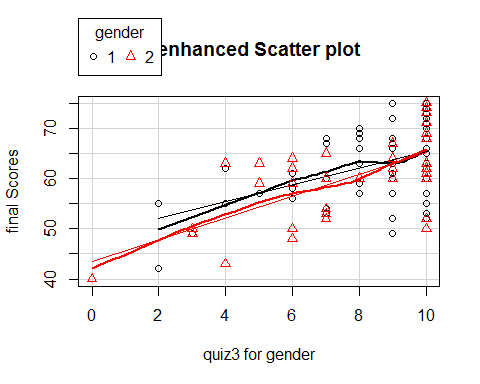


library(car)   
scatterplot(final ~ quiz3|passfail, data=grades,   
 xlab='quiz3 for pass/fail',  
 ylab= 'final Scores',  
 main='enhanced Scatter plot',  
 labels=row.names(grades))

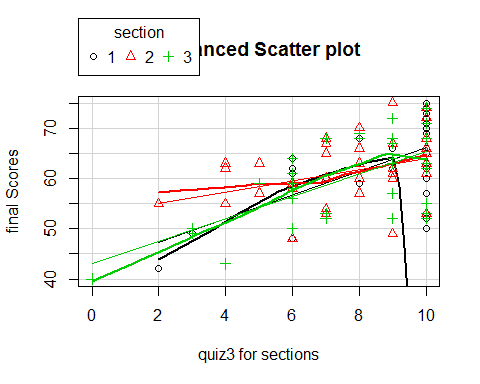
## Warning in smoother(.x[subs], .y[subs], col = col[i], log.x =  
## logged("x"), : could not fit smooth



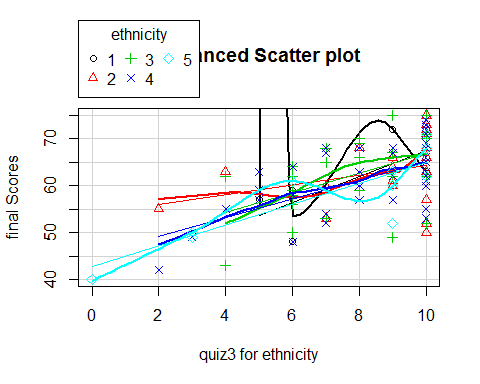
scatterplot(final ~ quiz3|gender, data=grades,   
 xlab='quiz3 for gender',  
 ylab= 'final Scores',  
 main='enhanced Scatter plot',  
 labels=row.names(grades))



scatterplot(final ~ quiz3|section, data=grades,   
 xlab='quiz3 for sections',  
 ylab= 'final Scores',  
 main='enhanced Scatter plot',  
 labels=row.names(grades))



scatterplot(final ~ quiz3|ethnicity, data=grades,   
 xlab='quiz3 for ethnicity',  
 ylab= 'final Scores',  
 main='enhanced Scatter plot',  
 labels=row.names(grades))



## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.