Multiple Regression

**Chinki**

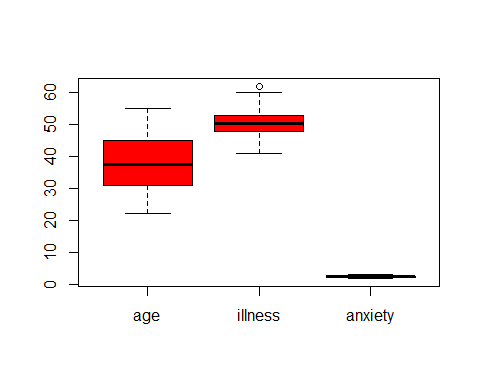
**May 4, 2017**

A hospital administrator wished to study the relation between patientsatisfaction (Y) and patient's age(X1 in years), severity of illness (X2 as an index)and anxiety (X3 as an index).

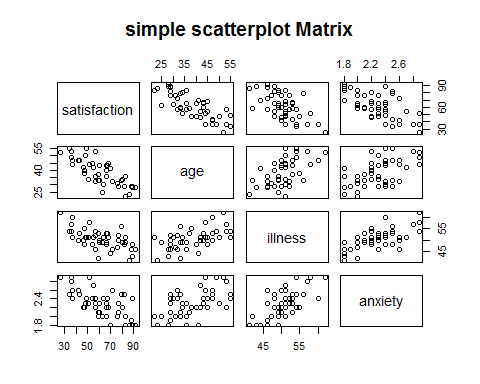
Sourse of the data: Applied linear statistical models

#Reading the data and giving headers  
hospital\_record=read.table("CH06PR15.txt")  
colnames(hospital\_record)=c("satisfaction","age","illness","anxiety")

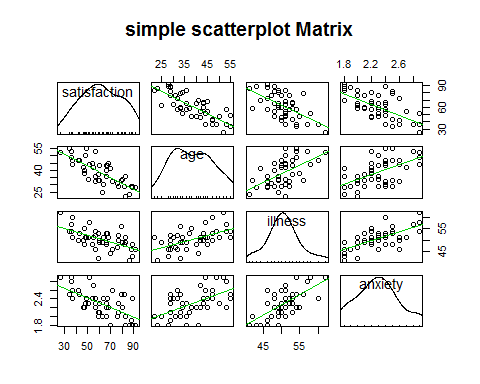
#Creating boxplots  
boxplot(hospital\_record[,-1],col="red")

 Boxplot of age is symmetric and boxplot of illness has one outlier.

#Scatterplot   
pairs(~satisfaction+age+illness+anxiety,data = hospital\_record,main="simple scatterplot Matrix")



library("car")  
scatterplotMatrix(~satisfaction+age+illness+anxiety,data = hospital\_record,main="simple scatterplot Matrix",ellipse=("FALSE"),smooth=F)



sactterplot of satisfaction and age,anxiety & illness is showing a liner trande.

# Correlation matrix for data set  
cor(hospital\_record[,1:4])

## satisfaction age illness anxiety  
## satisfaction 1.0000000 -0.7867555 -0.6029417 -0.6445910  
## age -0.7867555 1.0000000 0.5679505 0.5696775  
## illness -0.6029417 0.5679505 1.0000000 0.6705287  
## anxiety -0.6445910 0.5696775 0.6705287 1.0000000

#Correlation coeficent with p values  
library(psych)

##   
## Attaching package: 'psych'

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

corr.test(hospital\_record,use = "pairwise",y=NULL,method = "pearson",adjust = "holm",alpha=0.5)

## Call:corr.test(x = hospital\_record, y = NULL, use = "pairwise", method = "pearson",   
## adjust = "holm", alpha = 0.5)  
## Correlation matrix   
## satisfaction age illness anxiety  
## satisfaction 1.00 -0.79 -0.60 -0.64  
## age -0.79 1.00 0.57 0.57  
## illness -0.60 0.57 1.00 0.67  
## anxiety -0.64 0.57 0.67 1.00  
## Sample Size   
## [1] 46  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## satisfaction age illness anxiety  
## satisfaction 0 0 0 0  
## age 0 0 0 0  
## illness 0 0 0 0  
## anxiety 0 0 0 0  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

#Runnig regression model  
result=lm(hospital\_record$satisfaction~hospital\_record$age+hospital\_record$illness+hospital\_record$anxiety)

summary(result)

##   
## Call:  
## lm(formula = hospital\_record$satisfaction ~ hospital\_record$age +   
## hospital\_record$illness + hospital\_record$anxiety)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.3524 -6.4230 0.5196 8.3715 17.1601   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 158.4913 18.1259 8.744 5.26e-11 \*\*\*  
## hospital\_record$age -1.1416 0.2148 -5.315 3.81e-06 \*\*\*  
## hospital\_record$illness -0.4420 0.4920 -0.898 0.3741   
## hospital\_record$anxiety -13.4702 7.0997 -1.897 0.0647 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.06 on 42 degrees of freedom  
## Multiple R-squared: 0.6822, Adjusted R-squared: 0.6595   
## F-statistic: 30.05 on 3 and 42 DF, p-value: 1.542e-10

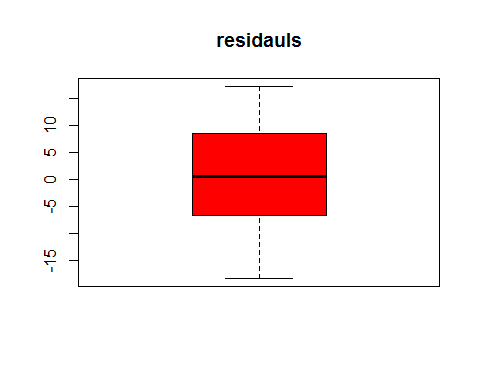
**Conclusion**

* The regression function indicate that the mean of satisfaction is expected to decrease by 1.14 per year change in age ,holding the illness and anxiety is constant.
* The regression function idicate that the mean of satisfaction is expected to decrease by 0.44 per unit change in the illness , holding other xs constant.
* The regression function indicate that the mean of satisfaction is expected to decrease by 13.47 per unit change in anxiety.
* We should reject H0 as p value<0.05.so age ,illness ,anixiety have no effect on satisfactory.
* illness (0.37) is not significant.
* R square is 66%.
* F statistics is a good indicator of whether there is a relationship between our predictor and the response variables. Inference in the multiple regression setting is typically performed in a number of steps. We begin by testing whether the explanatory variables collectively have an effect on the response variable, The null hypothesis (H0 : There is no relationship between predictor and the responce variable). As pvalue <0.05, conclude to reject null.so there is realtionship between predictor and responce.

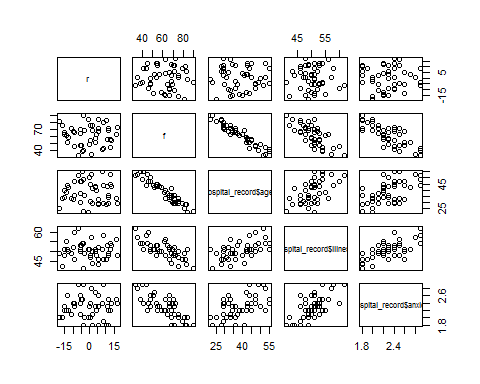
H0: β1= β2= β3=0

Ha: at least one is not equal to 0.

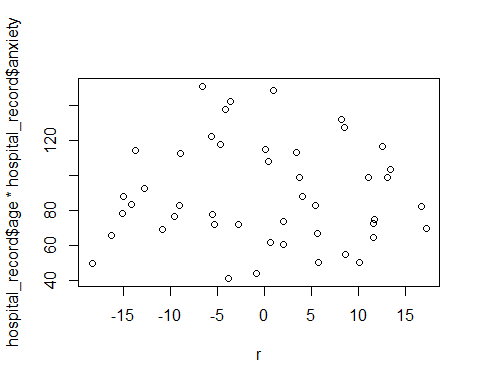
# Boxplot of residuals   
r=residuals(result)  
boxplot(r,col="red",main="residauls")

 No outliers are present in the data of residuals.

#Plot of residuals and fitted value,predictors,interaction  
f=fitted.values(result)  
pairs(r~f+hospital\_record$age+hospital\_record$illness+hospital\_record$anxiety)



plot(r,hospital\_record$age\*hospital\_record$anxiety)



**Constancy of error variance.**

group1=ifelse(hospital\_record$age>40,c("1"),c("2"))  
group1

## [1] "1" "2" "2" "1" "2" "1" "1" "1" "1" "2" "2" "1" "2" "2" "1" "2" "2"  
## [18] "2" "2" "1" "2" "1" "1" "2" "1" "1" "2" "2" "2" "1" "2" "2" "2" "2"  
## [35] "1" "2" "2" "2" "1" "1" "1" "2" "1" "1" "2" "2"

#Levene test  
levene.test(r,group1)

## Warning: 'levene.test' is deprecated.  
## Use 'leveneTest' instead.  
## See help("Deprecated") and help("car-deprecated").

## Warning in leveneTest.default(...): group1 coerced to factor.

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 1 2.342 0.1331  
## 44

Fail to reject H0, and conclude has constant variance.

group2=ifelse(hospital\_record$illness>2.4,c("1",c("2")))  
group2

## [1] "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1"  
## [18] "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2"  
## [35] "1" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2"

levene.test(r,group2)

## Warning: 'levene.test' is deprecated.  
## Use 'leveneTest' instead.  
## See help("Deprecated") and help("car-deprecated").

## Warning in leveneTest.default(...): group2 coerced to factor.

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 1 2.0984 0.1545  
## 44

As p value is 0.15 fail to reject H0 and conclude constat variance.

**Lack of fit The full model is treating it as a two-way ANOVA and the reduced model is the linear regression with two predictors.**

#Reduce model  
reduce=lm(hospital\_record$satisfaction~hospital\_record$age+hospital\_record$anxiety)  
full=lm(hospital\_record$satisfaction~factor(hospital\_record$age)+factor(hospital\_record$anxiety)+factor(hospital\_record$illness))  
anova(reduce,full)

## Analysis of Variance Table  
##   
## Model 1: hospital\_record$satisfaction ~ hospital\_record$age + hospital\_record$anxiety  
## Model 2: hospital\_record$satisfaction ~ factor(hospital\_record$age) +   
## factor(hospital\_record$anxiety) + factor(hospital\_record$illness)  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 43 4330.5   
## 2 1 180.5 42 4150 0.5474 0.8163

We fail to reject null H0: Y= β0 +β1X1+β2X2 +β3 X3 at 95% level of significant. It apperas that the variable age, illness & anxiety do not contribute significant information to satisfaction whan age ,illness & anxiety has been taken into underconsideration.

**Comparing the model ( Dropping illness )**

reduce=lm(hospital\_record$satisfaction~hospital\_record$age+hospital\_record$anxiety)  
full=lm(hospital\_record$satisfaction~hospital\_record$age+hospital\_record$anxiety+hospital\_record$illness)  
anova(reduce,full)

## Analysis of Variance Table  
##   
## Model 1: hospital\_record$satisfaction ~ hospital\_record$age + hospital\_record$anxiety  
## Model 2: hospital\_record$satisfaction ~ hospital\_record$age + hospital\_record$anxiety +   
## hospital\_record$illness  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 43 4330.5   
## 2 42 4248.8 1 81.659 0.8072 0.3741

We fail to reject reject null hypothesis, H0: β2=0 at the 5% level of significance. It appears that the variables illness do not contribute significant information to the satisfaction once the age and anxiety have been taken into consideration

#An interval estimate of the mean   
Result1=lm(satisfaction~age+illness+anxiety,data = hospital\_record)  
predict(Result1,data.frame(age=25,illness=45,anxiety=2.2),interval="confidence",levels=0.90)

## fit lwr upr  
## 1 80.4264 73.88088 86.97192

A 90% confidence interval is given by (73.88 ,86.97).