Assignment 12 (S-520)

FNU Anirudh

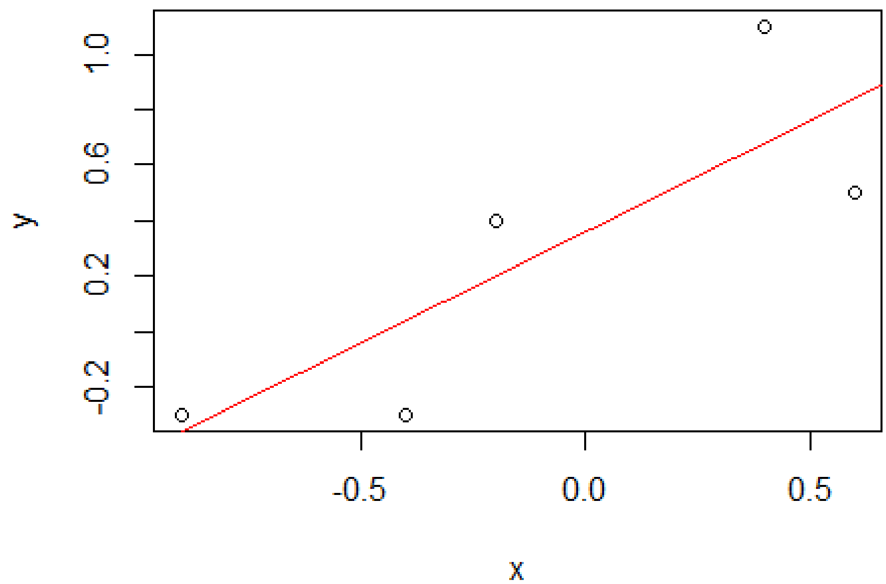
December 10, 2015

Solution 1

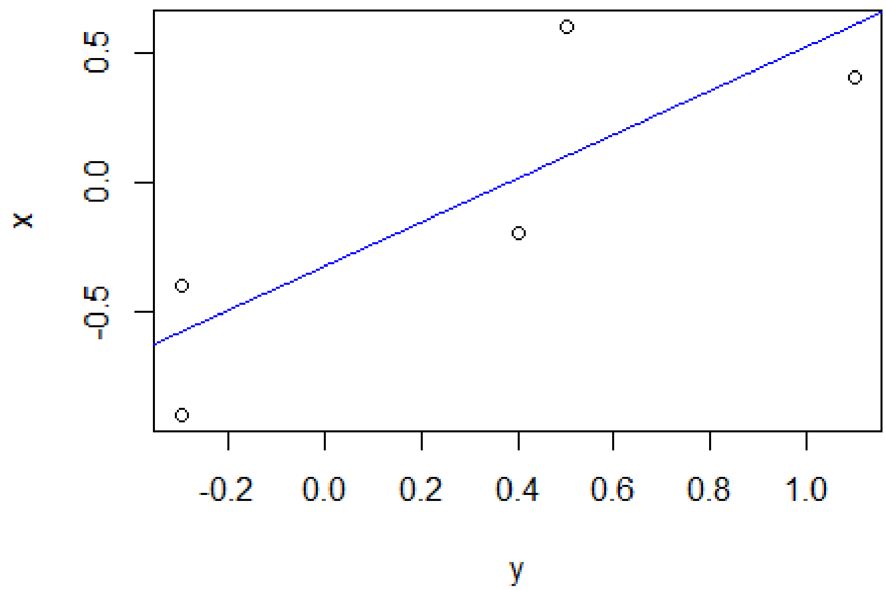
Successfull Flight Maneuver should depend on skill and chance, so for a particular candidate, performance on next flight maneuver will be less than his previous performance because we can't expect chance factor(random) to be equally good on his next manuever even though skills are same.This can be attributed to Regression to Mean.

Solution 2

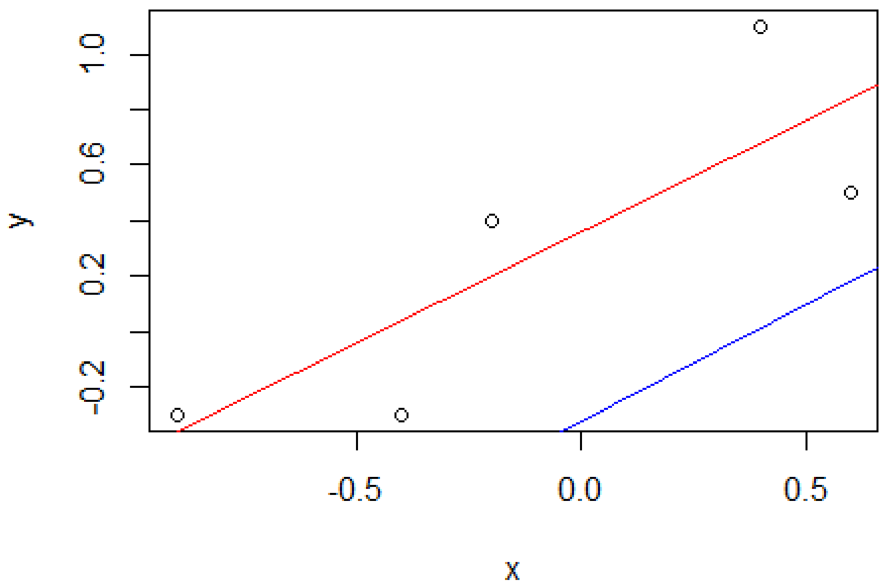
```r
x<- c(-0.2,-0.9,-0.4,0.6,0.4)
y<- c(0.4,-0.3,-0.3,0.5,1.1)
# a) Pearson's Correlation Coefficient is given by
print("Pearson's Correlation Coefficient is =")

## [1] "Pearson's Correlation Coefficient is ="

r=cor(x,y)
r

## [1] 0.8243559

# b)
plot(x,y)
# Predict y from x
slope = r *(sd(y) / sd(x))
intercept = mean(y) - (slope * mean(x))
predictions = intercept + slope*x
SSE = sum((y - predictions)^2)
abline(intercept, slope, col="red")
```



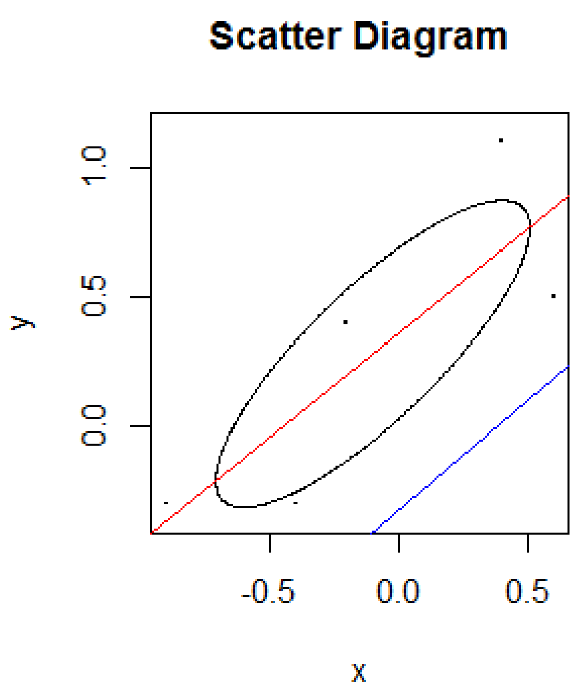

```r
# Equation of line y=intercept + slope*x
# In this case, Equation is y=0.36 + 0.80x
# c)
# Predict x from y
plot(y,x)
slope2 = r *(sd(x) / sd(y))
intercept2 = mean(x) - (slope * mean(y))
predictions2 = intercept + slope*y
SSE2 = sum((x - predictions)^2)
abline(intercept2, slope2, col="blue")
```




```
# Equation of line here is x=intercept1 + slope2* y
# x=-0.32 + 0.84y

# d) Scatter Plot
source('C:/Stats/Assignment 11/binorm.R')
plot(x,y)
abline(intercept, slope, col="red")
abline(intercept2, slope2, col="blue")
```




```
binorm.scatter(cbind(x, y))
abline(intercept, slope, col="red")
abline(intercept2, slope2, col="blue")
```



Solution 3

```
sister<- c(69,64,65,63,65,62,65,64,66,59,62)
brother<- c(71,68,66,67,70,71,70,73,72,65,66)
all<- c(sister,brother)
mb=mean(brother)
ms=mean(sister)
mall=mean(all)
n=length(sister)
sx=sum((sister-ms)^2)/(n-1)
sy=sum((brother-mb)^2)/(n-1)
r=cor(sister,brother)
r2=r^2
# a) The sample coefficient of determination, the proportion of
# variables "explained" by simple linear regression is square of
# correlation coefficent
r2

## [1] 0.3114251

SST = sy*(n-1)
SSR = r2 *SST
SSE=(1-r2)*SST
df1=1
df2=9
msr=r2*SST
mse=SSE/df2
f=(n-2)*r2/(1-r2)
1-pf(f,df1,df2)

## [1] 0.07441681

# b) Ho:beta>0
# Since p-value>0.05, we fail to reject Null Hypothesis hence we
# cannot say that knowing sister's height helps one predict her
# brother's height.
# c)
syy=sqrt(sy)
sxx=sqrt(sx)
beta1=r*(syy/sxx)
gt=qt(0.95,df2)
se=((1-r2)*sy)/((n-2)*sx)
print("Confidence Interval is=")

## [1] "Confidence Interval is="

lower=beta1- (gt*sqrt(se))
upper=beta1+ (gt*sqrt(se))
lower

## [1] 0.05401643
```

```
upper
## [1] 1.127802
# d) Given L=0.1
```

$L^2=0.1$

$L= \beta + q_t \sqrt{MSe/txx}\ - \beta + q_t \sqrt{MSe/txx}\ = 2 \sqrt{MSe/txx}$

Which can be written as

$$\frac{(1-r^2) . s_y^2}{(n-2) . s_x^2} = \left(\frac{0.1}{2 * q_t}\right)^2$$

$$n-2 = \frac{4\ q_t^2 .(1-r^2) .s_y^2}{(0.1)^2\ \ s_x^2}$$

For 95% Confidence Interval

q_t= qt(0.975,df=9) = 2.262157

$$n-2 = \frac{4 * (2.262157)^2 * (1 - 0.311425) * 7.4}{(0.1)^2\ * 6.6}$$

= 1580

We should plan to observe close to 1580 pair of sister-brother.

Solution 4

a) b= cor(test2,test1) * sd(test1)
   a= mean(test2) – b * mean(test1)

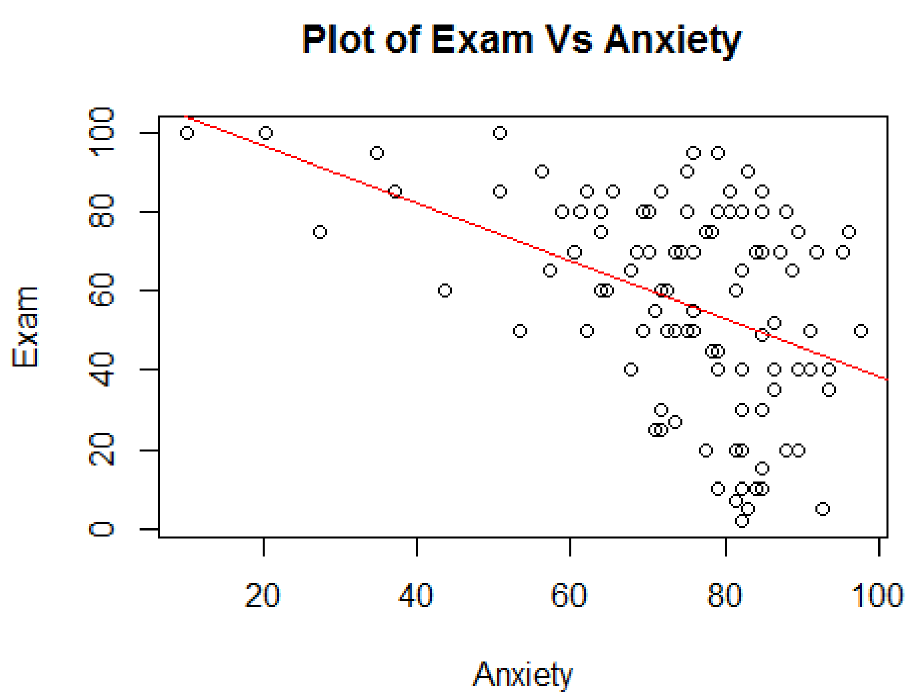
test2= a+ b*test1
So,
test2= 19 + 0.6*80 = 67
Jill should get 67 in Test2 if she scored 80 in first and missed the second one.

b) b = cor(test2, test1) * sd(test1) / sd(test2)
a= mean(test1)- b* mean(test2)
test1= a+btest2
So,
test1= 48.76 + 0.41*76=79.92

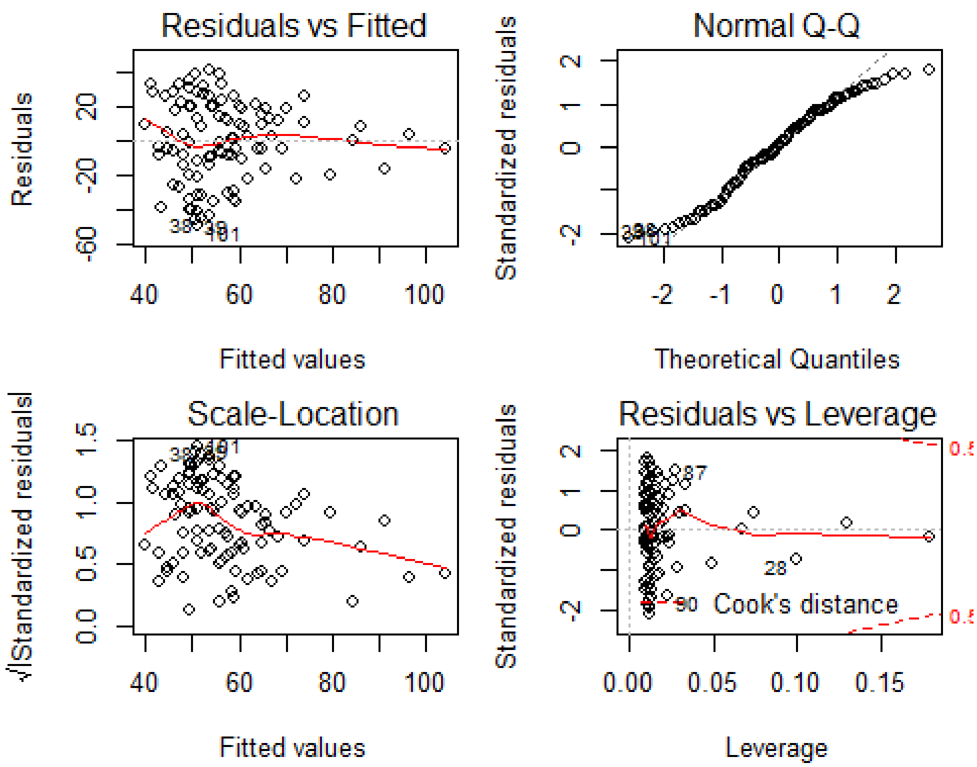
Jack should get close to 80 in test 1 if he scored 76 in test 2 and missed the first one.

Solution 5

```
ext=read.table('C:/Stats/Assignment 12/examanxiety.txt',header = TRUE)
exam<-ext$Exam
anxiety<-ext$Anxiety
b = cor(exam, anxiety) * sd(exam) / sd(anxiety)
a = mean(exam) - b * mean(anxiety)
plot(anxiety, exam,main="Plot of Exam Vs Anxiety",xlab="Anxiety",ylab="Exam")
abline(a, b, col="red")
```



```
# Equation of line is y=a+bx
# Exam=111-0.73* Anxiety

# b)
fit<-lm(Exam~Anxiety,data = ext)
par(mfrow=c(2,2),mar=c(4,4,2,1))
# Plot to check for Homodeskascity, Normality of errors, Independence
# of errors
plot(fit)
```




```
# From Plot, we conculde that out of all assumptions only assumption
# about normality of error seems true.

# c)
# Suppose X and Y are bivariate normal.Then given a specific value
# of X, Y follows normal distribution.
# cor(anxiety,exam)=-0.439 correlation is negative hence we should
# not use bivariate normal and instead fit an ellipse in scatter plot
# ellipse won't fit density but will be cross like shape.
```

Solution 6

```
fit1<-lm(Exam~Anxiety+Revise+Gender,data=ext)
fit2<-lm(Exam~Anxiety+Revise,data=ext)
summary(fit1)
```

```
##
## Call:
## lm(formula = Exam ~ Anxiety + Revise + Gender, data = ext)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.834 -15.029  -1.121  21.845  40.243
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  89.7932    17.8090   5.042 2.08e-06 ***
## Anxiety      -0.5202     0.1973  -2.636  0.00974 **
## Revise        0.2746     0.1703   1.612  0.11009
## GenderMale    0.5587     4.6497   0.120  0.90460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.34 on 99 degrees of freedom
## Multiple R-squared:  0.214,  Adjusted R-squared:  0.1902
## F-statistic: 8.987 on 3 and 99 DF,  p-value: 2.544e-05
```

```
summary(fit2)
```

```
##
## Call:
## lm(formula = Exam ~ Anxiety + Revise, data = ext)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.591 -14.763  -1.397  21.831  40.505
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  90.3422    17.1279   5.275 7.72e-07 ***
## Anxiety      -0.5230     0.1950  -2.682  0.00857 **
## Revise        0.2717     0.1678   1.619  0.10849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.23 on 100 degrees of freedom
## Multiple R-squared:  0.2139, Adjusted R-squared:  0.1982
## F-statistic: 13.61 on 2 and 100 DF,  p-value: 5.931e-06
```

From above we can say that model should be Exam~ Anxiety+Revise since exam scores depend on Revision and Anxiety before the exam, Exam score has nothing to do with gender of person.R^2 value decreases when Gender is included and Residual Standard error increases hence model should be Exam~Anxiety+Revise.