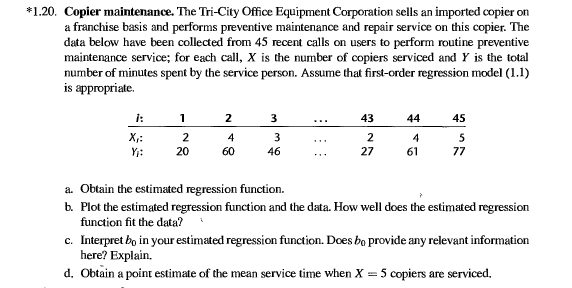
Question 1.

Pg 186 3.4

Pg 64 for original

a.



[http://www.stat.columbia.edu/~fwood/Teaching/w4315/Fall2010/homework\_4/homework\_4\_solution](qq://txfile/#)

<http://www.ams.sunysb.edu/~xing/AMS578/Documents/hw2_sol.pdf>

[http://www.math.mun.ca/~aoyet/stat3521/sam](qq://txfile/#)

(a)

> copier=read.table("copier.txt", header=F)

> colnames(copier)=c("y","x")

> copier

y x

1 20 2

2 60 4

3 46 3

4 41 2

5 12 1

6 137 10

7 68 5

8 89 5

9 4 1

10 32 2

11 144 9

12 156 10

13 93 6

14 36 3

15 72 4

16 100 8

17 105 7

18 131 8

19 127 10

20 57 4

21 66 5

22 101 7

23 109 7

24 74 5

25 134 9

26 112 7

27 18 2

28 73 5

29 111 7

30 96 6

31 123 8

32 90 5

33 20 2

34 28 2

35 3 1

36 57 4

37 86 5

38 132 9

39 112 7

40 27 1

41 131 9

42 34 2

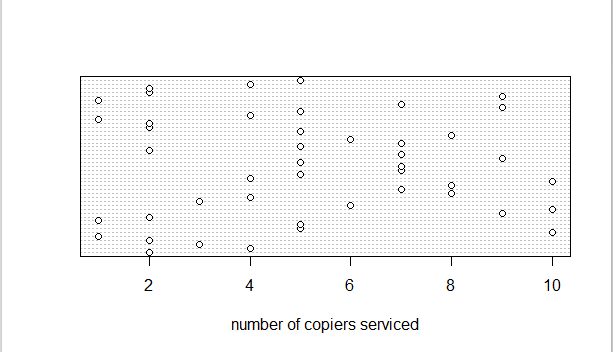
43 27 2

44 61 4

45 77 5

attach(copier)

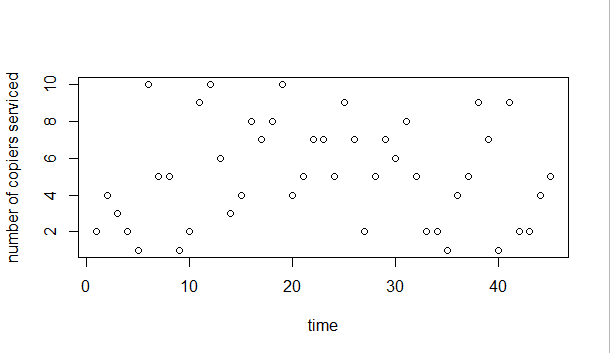
dotchart(x, xlab="number of copiers serviced")



We can see from the plot that x spreads from 1 to 10, and the distribution seems normal. There is no outlying case with respect to this variable.

(b)

> plot(x, xlab="time", ylab="number of copiers serviced")



The plot shows that there is no correlation between time and the number of copiers serviced.

(c)

> xbar=mean(x)

> ybar=mean(y)

> ssx=sum((x-xbar)^2)

> spxy=sum((y-ybar)\*(x-xbar))

> b1=spxy/ssx

> b1

[1] 15.03525

> b0=ybar-b1\*xbar

> b0

[1] -0.5801567

> yhat=b0+b1\*x

> residual=y-yhat

> stem(residual)

The decimal point is 1 digit(s) to the right of the |

-2 | 30

-1 |

-1 | 3110

-0 | 99997

-0 | 44333222111

0 | 001123334

0 | 5666779

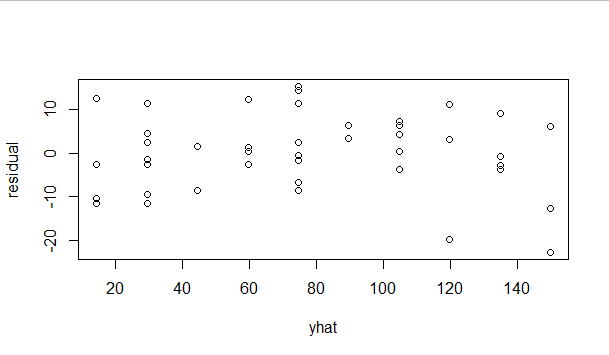
1 | 112234

1 | 5

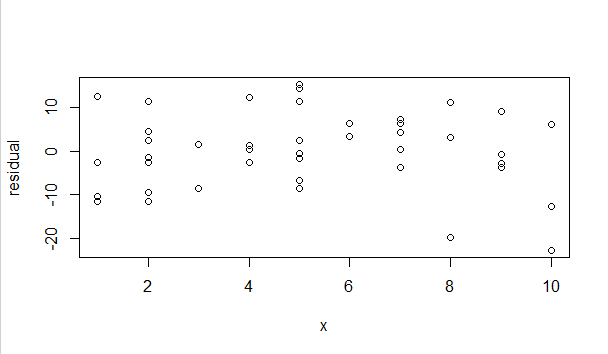
We see that the residuals range from -23 to 15; there are two outliers -20 and -30 if we take those out the residuals range from -13 to 15.

(d)

> plot(yhat,residual)



> plot(x,residual)



Therefore the plots provide the same information. From these plots we can see that the regression

Function is linear and the error variance is constant.

(e)

> rankofres=rank(residual)

> n=length(y)

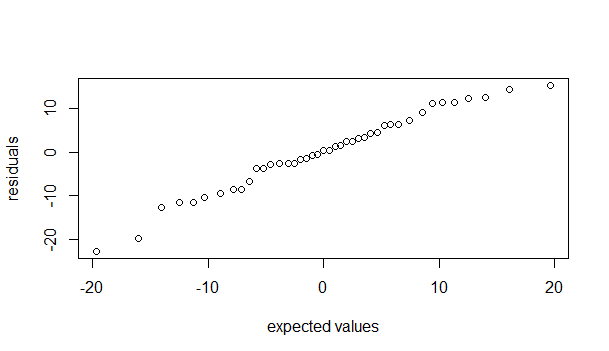
> SSE=sum(residual^2)

> MSE=SSE/(n-2)

> zscore=qnorm((rankofres-0.375)/(n+0.25))

> expres=zscore\*sqrt(MSE)

> plot(expres,residual,xlab="expected values",ylab="residuals")



> Sxy=sum((expres-mean(expres))\*(residual-mean(residual)))

> Sxx=sum((expres-mean(expres))^2)

> Syy=sum((residual-mean(residual))^2)

> r=Sxy/sqrt(Sxx\*Syy)

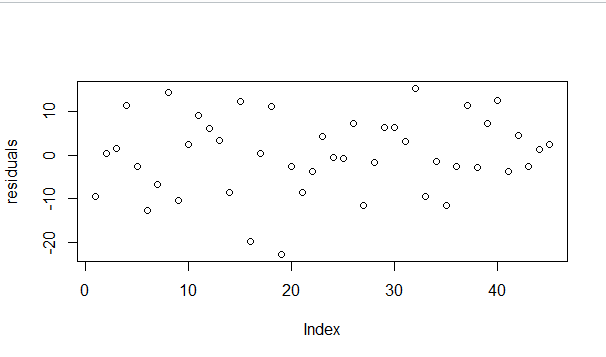
> r

[1] 0.9892079

Therefore the coefficient of correlation is 0.9892079. From table B.6, we see that the critical value is between 0.977 and 0.981 when n=45 which is smaller than the coefficient of correlation; therefore assumption of normality is reasonable.

(f)

> plot(residual,ylab="residuals")



Therefore the error terms are not correlated over time.

(g)

Null hypothesis - H0: error variance is constant.

Alternative hypothesis - Ha: error variances is not constant.

Decision rule: if X^2(BP)>X^2(0.95, 1)=3.841 we reject the null hypothesis;

If X^2(BP)<3.841 we accept the null hypothesis.

> plot(residual,ylab="residuals")

> e2=residual^2

> spe2x=sum((e2-mean(e2))\*(x-mean(x)))

> ssx=sum((x-mean(x))^2)

> b1e2=spe2x/ssx

> SSE=sum(e2)

> SSRsr=b1e2^2\*ssx

> chi2BP=(SSRsr/2)/((SSE/n)^2)

> chi2BP

[1] 1.31468

Conclusion:

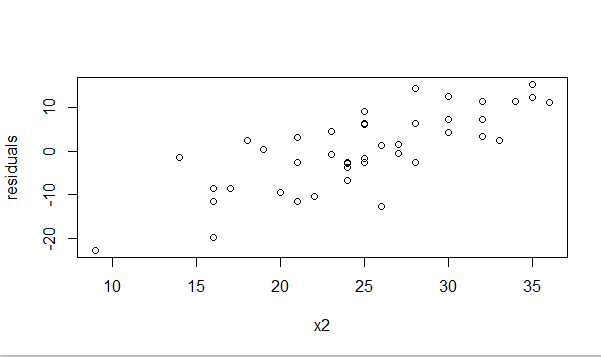
X^2(BP)=1.31468<3.841.

Therefore we accept the null hypothesis that error variance is constant.

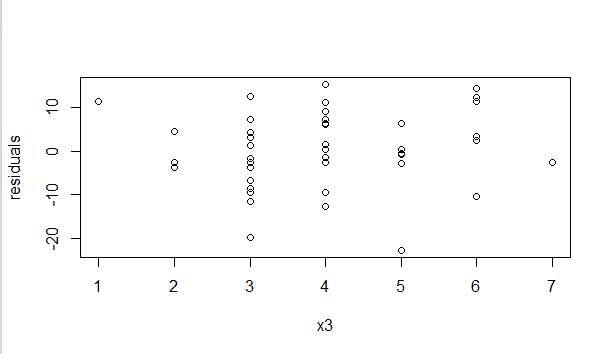
(h)

> copier1=read.table("copier1.txt",header=F)

> attach(copier1)



> plot(V4,residual,xlab="x3",ylab="residuals")



From the plot of x2 we see that there is a positive linear correlation between x2 and residuals but we see no correlation from the second plot. Therefore the model can be included by including x2

2.

> electr=read.table("electr.txt",header=F)

> electr

V1 V2

1 2 3.2

2 3 2.9

3 4 -1.7

4 5 -2.0

5 6 -2.3

6 7 -1.2

7 8 -0.9

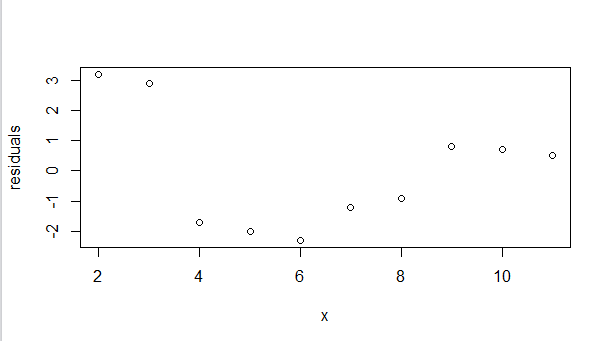
8 9 0.8

9 10 0.7

10 11 0.5

> attach(electr)

> plot(V1,V2,xlab="x",ylab="residuals")



The problem is that there is a pattern for the distribution of the residuals here therefore error variance is not constant; and yes a transformation might alleviate the problem

3. (a)

> drug=read.table("drug.txt",header=F)

> drug

V1 V2

1 -1 0.5

2 0 2.1

3 1 -3.4

4 -1 0.3

5 0 -1.7

6 1 4.2

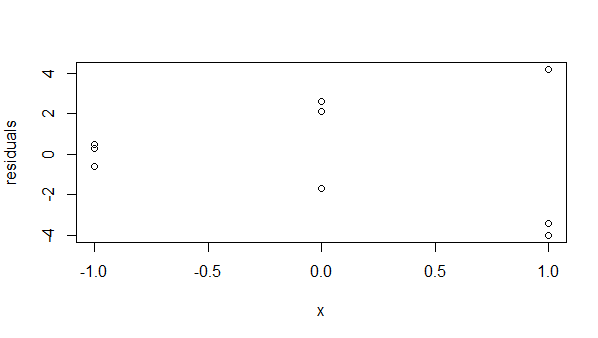
7 -1 -0.6

8 0 2.6

9 1 -4.0

> attach(drug)

> plot(V1,V2,xlab="x",ylab="residuals")



From the plot we see that x gets larger when the residuals get larger and when the residuals get smaller; therefore we conclude that error variance is not constant.

(b)

Null hypothesis - H0: error variance is constant.

Alternative hypothesis - Ha: error variances is not constant.

Decision rule: if X^2(BP)>X^2(0.95, 1)=3.841 we reject the null hypothesis;

If X^2(BP)<3.841 we accept the null hypothesis.

> x=V1

> e=V2

> e2=e^2

> spe2x=sum((e2-mean(e2))\*(x-mean(x)))

> ssx=sum((x-mean(x))^2)

> b1e2=spe2x/ssx

> SSE=sum(e2)

> SSRsr=b1e2^2\*ssx

> chi2BP=(SSRsr/2)/(SSE/length(e))^2

> chi2BP

[1] 3.717924

Since 3.717924<3.841, we accept the null hypothesis; this does not support my preliminary findings in part (a).

4.

4.3 (a)

> copier=read.table("copier.txt", header=F)

> colnames(copier)=c("y","x")

> attach(copier)

[1] 5.111111

From the handwritten sheet attached we see the b0 and b1 tend to err in opposite directions.

(b)

CI for :



> SSE=sum(resi^2)

> n=length(x)

> MSE=SSE/(n-2)

> s2b0=MSE\*((1/n)+xbar^2/sum((x-xbar)^2))

> sb0=sqrt(s2b0)

> t=qt((1-0.05/4), n-2)

> CI0=c(b0-t\*sb0, b0+t\*sb0)

> CI0

[1] -7.092642 5.932329

CI for :



> s2b1=MSE/sum((x-xbar)^2)

> sb1=sqrt(s2b1)

> CI1=c(b1-t\*sb1, b1+t\*sb1)

> CI1

[1] 13.91322 16.15728

(c) yes , since 0 lies within CI of b0 and 14 lies within CI of b1.

4.7

(b)

Scheffe:

> SP2=2\*qf(1-0.1,2,n-2)

> SP=sqrt(SP2)

> SP

[1] 2.204725

Bonferroni:

> BP=qt(1-0.1/(2\*2),n-2)

> BP

[1] 2.016692

SP>BP, therefore Scheffe Procedure will provide tighter prediction limits.

(c)

4 copiers:

> yhat1=b0+b1\*4

> s2pred=(1+1/n+(4-xbar)^2/sum((x-xbar)^2))\*MSE

> spred=sqrt(s2pred)

> CIy1=c(yhat1-BP\*spred,yhat1+BP\*spred)

> CIy1

[1] 41.35419 77.76748

Therefore the 90% CI for 4 copiers is (41.35419, 77.76748).

7 copiers:

> yhat2=b0+b1\*7

> s2pred2=(1+1/n+(7-xbar)^2/sum((x-xbar)^2))\*MSE

> spred2=sqrt(s2pred2)

> CIy2=c(yhat2-BP\*spred2,yhat2+BP\*spred2)

> CIy2

[1] 86.39922 122.93394

Therefore the 90% CI for 7 copiers is (86.39922, 122.93394).

5.

(a)

> typo=read.table("typo.txt",header=F)

> attach(typo)

> x=V1

> y=V2

> spxy=sum(x\*y)

> sxx=sum(x^2)

> b1=spxy/sxx

> b1

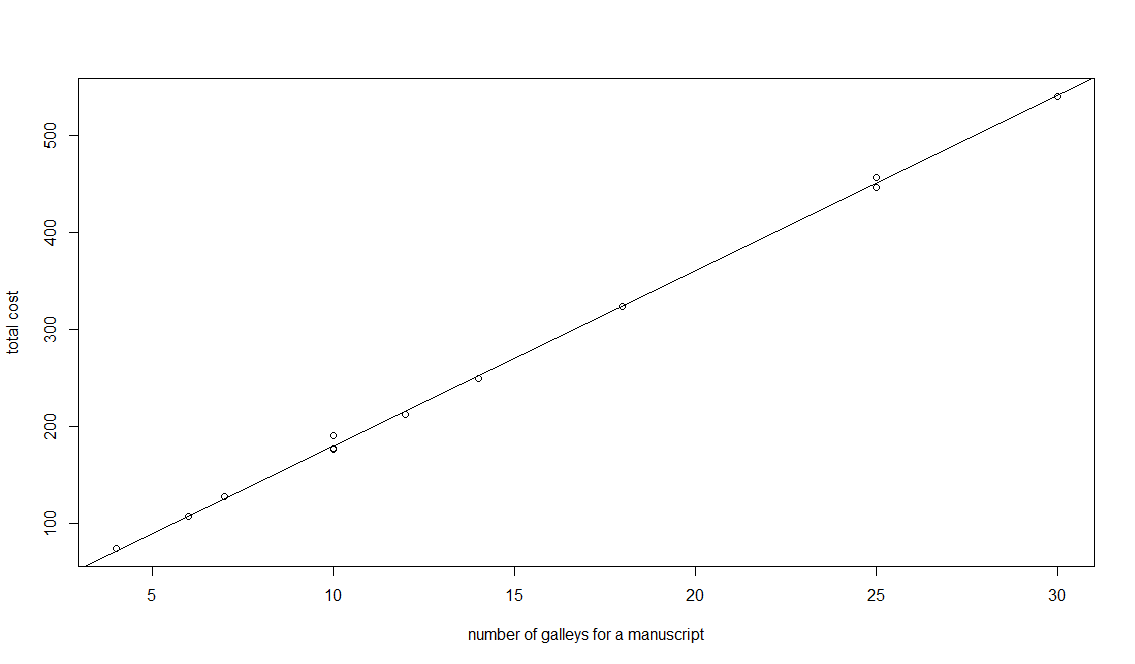
[1] 18.0283

Therefore, the estimated regression function is: Y=18.0283X.

(b)

> plot(x,y,xlab="number of galleys for a manuscript",ylab="total cost")

> abline(a=0,b=b1)



Since all the dots are really close to the line we conclude that the linear regression function through the origin appears to provide a good fit here.

(c)

Null hypothesis - H0: =18.0283 (obtained from part a)



Alternative hypothesis - Ha: not equal to 18.0283.



Decision rule: if t<-t(0.99, 1) or t>t(0.99,1) we reject the null hypothesis;

If -t(0.99, 1)<t<t(0.99, 1), we accept the null hypothesis.

where by checking the table we see t(0.99, 1)=2.718; (df=12-1=11)

> n=12

> yhat=b1\*x

> residual=y-yhat

> SSE=sum(residual^2)

> MSE=SSE/(n-1)

> s2b1=MSE/sum(x^2)

> sb1=sqrt(s2b1)

> t=(17.5-18.0283)/sb1

> t

[1] -6.646642

-6.646642<-2.718

t<-t(0.99,1) therefore we reject the null hypothesis. This standard should be revised.

(d)

> ts=2.718

> y1=b1\*10

> s2pred=(1+10^2/sum(x^2))\*MSE

> spred=sqrt(s2pred)

> CIy1=c(y1-ts\*spred,y1+ts\*spred)

> CIy1

[1] 167.8445 192.7216

Therefore the 98% prediction interval is (167.8445, 192.7216).

6.

5.4

YTY:

> flavor=read.table("flavor.txt", header=F)

> attach(flavor)

> n=length(x)

> X=V1

> Y=V2

> > UnitVec <- rep(1, n)

> X <- cbind(UnitVec, x)

> colnames(X) <- c("1", "x")

> Y <- matrix(y, ncol = 1)

> YY <- t(Y) %\*% Y

> YY

[,1]

[1,] 503.77

XTX:

> XX <- t(X) %\*% X

> XX

1 x

1 5 0

x 0 160

XTY:

> XY <- t(X) %\*% Y

> XY

[,1]

1 49.7

x -39.2

5.12

(XTX)^(-1):

> XXIN=solve(XX)

> XXIN

1 x

1 0.2 0.00000

x 0.0 0.00625

5.23

(a)

Vector of estimated regression coefficients:

> b <- solve(XX) %\*% XY

> b

[,1]

1 9.940

x -0.245

Vector of residuals:

> e <- Y - yhat

> e

[,1]

[1,] -0.18

[2,] 0.04

[3,] 0.26

[4,] 0.08

[5,] -0.20

SSR:

> J <- matrix(1, nrow = n, ncol = n)

> SSR <- t(b) %\*% t(X) %\*% Y - (t(Y) %\*% J %\*% Y)/n

> SSR

[,1]

[1,] 9.604

SSE:

> SSE <- t(Y) %\*% Y - t(b) %\*% t(X) %\*% Y

> SSE

[,1]

[1,] 0.148

Estimated variance-covariance:

> MSE <- sum(e^2)/3

> s2b <- MSE \* solve(t(X) %\*% X)

> s2b

1 x

1 0.009866667 0.0000000000

x 0.000000000 0.0003083333

Point estimate of E(Yh) when Xh=-6:

> Xh <- -6

> Xh <- cbind(1,Xh)

> Yhhat <- Xh%\*%b

> Yhhat

[,1]

[1,] 11.41

Estimated variance of Yhath when Xh=-6:

> s2Yhhat <- MSE\*Xh%\*%solve(t(X)%\*%X)%\*%t(Xh)

> s2Yhhat

[,1]

[1,] 0.02096667

(b)

We see that the values are symmetric; their mean is 0.

(c)

> H <- X %\*% solve(t(X) %\*% X) %\*% t(X)

> H

[,1] [,2] [,3] [,4] [,5]

[1,] 0.6 0.4 0.2 0.0 -0.2

[2,] 0.4 0.3 0.2 0.1 0.0

[3,] 0.2 0.2 0.2 0.2 0.2

[4,] 0.0 0.1 0.2 0.3 0.4

[5,] -0.2 0.0 0.2 0.4 0.6

(d)

> I <- matrix(0, nrow = 5, ncol = 5)

> diag(I) <- 1

> s2e <- (I - H) \* MSE

> s2e

[,1] [,2] [,3] [,4] [,5]

[1,] 0.019733333 -0.019733333 -0.009866667 0.000000000 0.009866667

[2,] -0.019733333 0.034533333 -0.009866667 -0.004933333 0.000000000

[3,] -0.009866667 -0.009866667 0.039466667 -0.009866667 -0.009866667

[4,] 0.000000000 -0.004933333 -0.009866667 0.034533333 -0.019733333

[5,] 0.009866667 0.000000000 -0.009866667 -0.019733333 0.019733333

7.

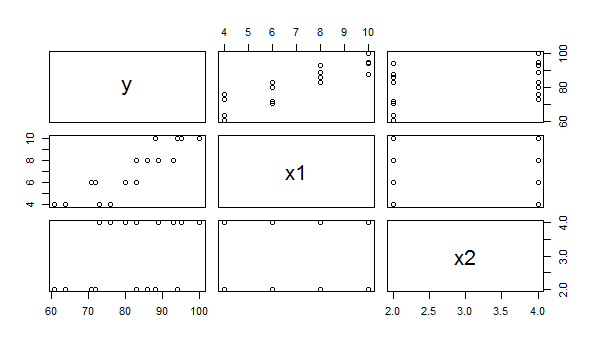
6.5

(a)

> brand=read.table("brand.txt",header=F)

> colnames(brand)=c("y","x1","x2")

> pairs(brand)



> cor(brand)

y x1 x2

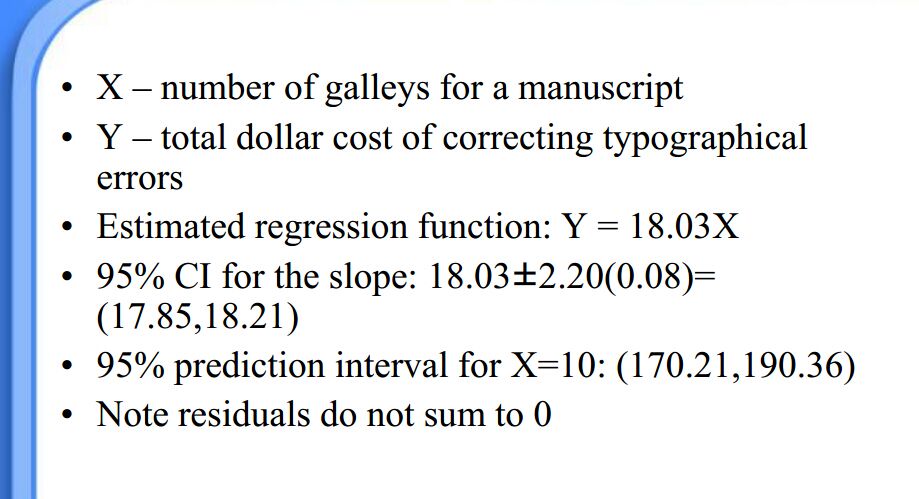
y 1.0000000 0.8923929 0.3945807

x1 0.8923929 1.0000000 0.0000000

x2 0.3945807 0.0000000 1.0000000

Based on these we can tell that there is a strong positive correlation between x1 and y, and a strong positive relation between y and x2. However there is no correlation between x1 and x2.

(b)



http://www.ams.sunysb.edu/~xing/AMS578/Documents/hw3\_sol.pdf