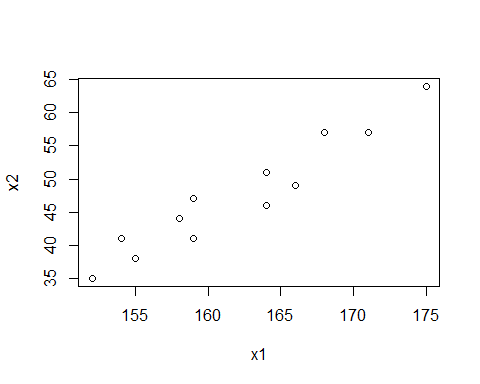
Rtest\_210330

ae

2021/3/30

## 直观分析：图示法

x1=c(171,175,159,155,152,158,154,164,168,166,159,164)  
x2=c(57,64,41,38,35,44,41,51,57,49,47,46)  
plot(x1,x2)



## 用离均差乘积和计算相关系数

lxy<-function(x,y)   
sum(x\*y)-sum(x)\*sum(y)/length(x)  
r=lxy(x1,x2)/sqrt(lxy(x1,x1)\*lxy(x2,x2))  
r

## [1] 0.9593031

## 计算相关系数r的t值

n=length(x1)#向量的长度  
tr=r/sqrt((1-r^2)/(n-2))#相关系数假设检验t统计量  
tr

## [1] 10.74298

## 计算t值和P值，作结论

cor.test(x1,x2)#相关系数假设检验

##   
## Pearson's product-moment correlation  
##   
## data: x1 and x2  
## t = 10.743, df = 10, p-value = 8.21e-07  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.8574875 0.9888163  
## sample estimates:  
## cor   
## 0.9593031

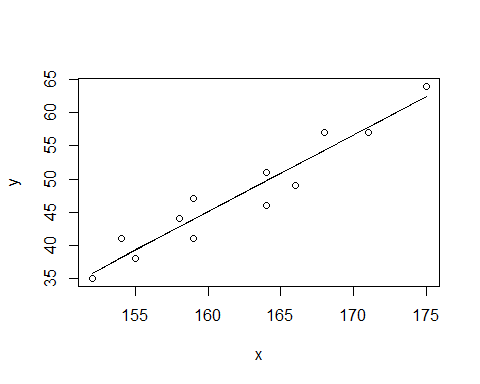
## 建立直线回归方程

x=x1#自变量,数据来自例2.2  
y=x2#因变量,数据来自例2.2  
b=lxy(x,y)/lxy(x,x)#线性回归方程斜率  
a=mean(y)-b\*mean(x)#线性回归方程截距  
c(a=a,b=b)#显示线性回归方程估计值

## a b   
## -140.36436 1.15906

## 散点图

plot(x,y)#做散点图  
lines(x,a+b\*x)#添加估计方程线

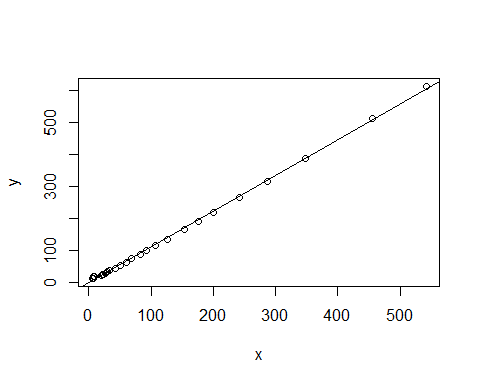


## 例 4-3

library(xlsx)

## Warning: package 'xlsx' was built under R version 4.0.4

yX = read.xlsx("./mvstats5.xlsx","d4.3",encoding="UTF-8")  
fm=lm(y~x,data=yX)  
plot(y~x,data=yX)#做散点图  
abline(fm)#添加回归线



anova(fm)#模型方差分析

## Analysis of Variance Table  
##   
## Response: y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## x 1 712077 712077 27428 < 2.2e-16 \*\*\*  
## Residuals 29 753 26   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(fm)#回归系数t检验

##   
## Call:  
## lm(formula = y ~ x, data = yX)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.630 -3.692 -1.535 5.338 11.432   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.19656 1.16125 -1.03 0.311   
## x 1.11623 0.00674 165.61 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.095 on 29 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.9989, Adjusted R-squared: 0.9989   
## F-statistic: 2.743e+04 on 1 and 29 DF, p-value: < 2.2e-16

## 例 4-4

library(xlsx)  
yX = read.xlsx("./mvstats5.xlsx","d4.4",encoding="UTF-8")  
fm=lm(y~x1+x2+x3+x4,data=yX)  
fm

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3 + x4, data = yX)  
##   
## Coefficients:  
## (Intercept) x1 x2 x3 x4   
## 23.5321088 -0.0033866 1.1641150 0.0002919 -0.0437416

#标准化偏回归系数  
coef.sd<-function(fm){  
b=fm$coef;b  
si=apply(fm$model,2,sd);si  
bs=b[-1]\*si[-1]/si[1]  
bs}  
  
coef.sd(fm)

## x1 x2 x3 x4   
## -0.0174513678 1.0423522972 0.0009628564 -0.0371053994

summary(fm) #多元数据相关系数t检验

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3 + x4, data = yX)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.0229 -2.1354 0.3297 1.2639 6.9690   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 23.5321088 4.5990714 5.117 2.47e-05 \*\*\*  
## x1 -0.0033866 0.0080749 -0.419 0.678   
## x2 1.1641150 0.0404889 28.751 < 2e-16 \*\*\*  
## x3 0.0002919 0.0085527 0.034 0.973   
## x4 -0.0437416 0.0092638 -4.722 7.00e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.79 on 26 degrees of freedom  
## Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997   
## F-statistic: 2.289e+04 on 4 and 26 DF, p-value: < 2.2e-16

模型方程为： y^ = 23.532109 - 0.003387 \* x1 + 1.164115 \* x2 + 0.000292 \* x3 - 0.043742 \* x4

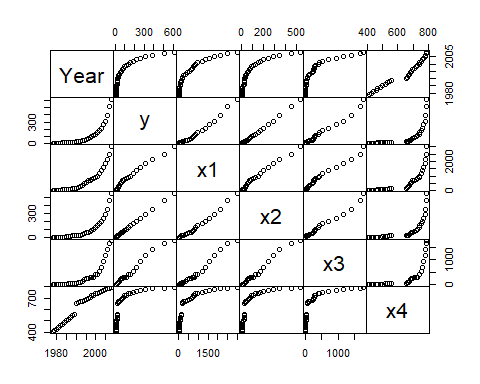
由方差结果可见，模型的F为2.29e+04，P<0.0001，该模型是有意义的

# 例 4.5

# 两个变量的相关系数  
cor(yX) #多元数据相关系数矩阵

## Year y x1 x2 x3 x4  
## Year 1.0000000 0.8141349 0.8842687 0.8308056 0.8234883 0.9742470  
## y 0.8141349 1.0000000 0.9871498 0.9994718 0.9912053 0.6956619  
## x1 0.8842687 0.9871498 1.0000000 0.9907018 0.9867664 0.7818066  
## x2 0.8308056 0.9994718 0.9907018 1.0000000 0.9917094 0.7154297  
## x3 0.8234883 0.9912053 0.9867664 0.9917094 1.0000000 0.7073820  
## x4 0.9742470 0.6956619 0.7818066 0.7154297 0.7073820 1.0000000

plot(yX,gap = 0) #多元数据散点图



# 计算相关系数的t值和P值  
source('msaR.r') #调用自定义函数  
msa.cor.test(yX) #多元数据相关系数检验

## corr test:   
## Year y x1 x2 x3 x4  
## Year 0.000 0.000 0.000 0.000 0.000 0  
## y 7.550 0.000 0.000 0.000 0.000 0  
## x1 10.197 33.267 0.000 0.000 0.000 0  
## x2 8.039 165.614 39.214 0.000 0.000 0  
## x3 7.817 40.336 32.772 41.560 0.000 0  
## x4 23.268 5.215 6.752 5.514 5.389 0  
## lower is t value, upper is p value

以上结果可以看出，财政收入和国内生产总值及其税收、出口贸易总额、经济活动人口之间的关系都非常密切，t>0.6,P<0.001，财政收入与税收之间的关系最为密切，t=0.9995,P<0.001