

《实用回归分析》实验报告

****

**实验一 一元、多元线性回归**

学生姓名 丁浩洋 201813230001

李睿星 20171323017

杨 钦 201813870052

李子豪 201813870017

庞浩哲 201813870023

学 院 数学与统计学院

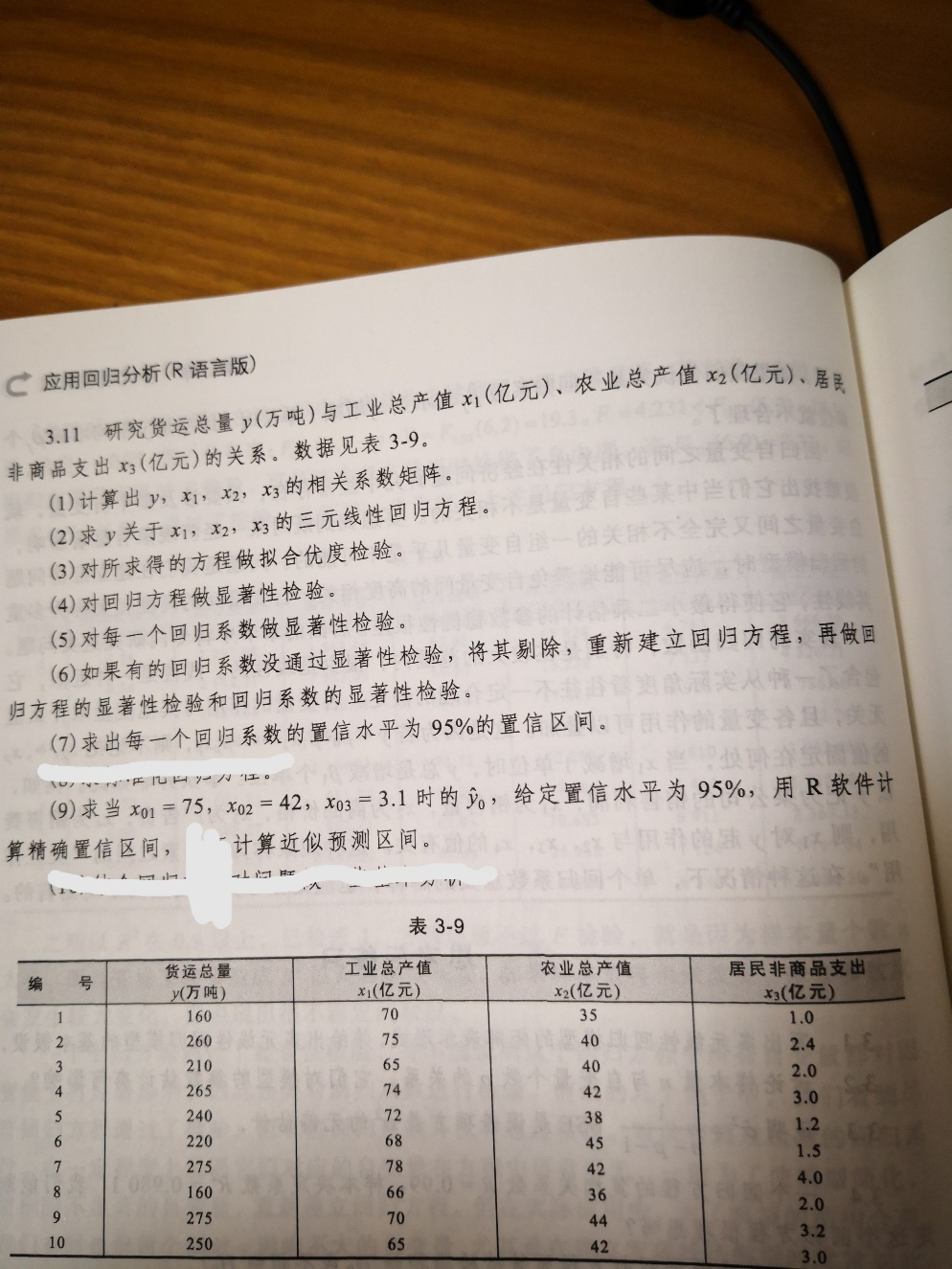
专 业 统计学

指导教师 吴香华

**二Ｏ二Ｏ年三月二十日**

**实习目的：**掌握R语言的安装和基本操作，建立线性回归分析模型、估计参数、进行统计检验和预测，学会分析软件结果。

**实习内容：**



**一、R语言程序及实现步骤（由rmarkdown文件显示）**

# 研究货运总量y(万吨)与工业总产值x1(亿元)、农业总产值x2(亿元)、居民非商品支出x3(亿元)的关系  
y<-c(160,260,210,265,240,220,275,160,275,250) #货运总量y万吨  
x1<-c(70,75,65,74,72,68,78,66,70,65) #工业总产值x1亿元  
x2<-c(35,40,40,42,38,45,42,36,44,42) #农业总产值x2亿元  
x3<-c(1.0,2.4,2.0,3.0,1.2,1.5,4.0,2.0,3.2,3.0) #居民非商品支出x3亿元  
data1<-data.frame(y,x1,x2,x3,stringsAsFactors=FALSE)  
data1

## y x1 x2 x3  
## 1 160 70 35 1.0  
## 2 260 75 40 2.4  
## 3 210 65 40 2.0  
## 4 265 74 42 3.0  
## 5 240 72 38 1.2  
## 6 220 68 45 1.5  
## 7 275 78 42 4.0  
## 8 160 66 36 2.0  
## 9 275 70 44 3.2  
## 10 250 65 42 3.0

# (1)计算出y关于x1,x2,x3的相关系数矩阵。  
class1<-cor(data1)  
class1

## y x1 x2 x3  
## y 1.0000000 0.5556527 0.7306199 0.7235354  
## x1 0.5556527 1.0000000 0.1129513 0.3983870  
## x2 0.7306199 0.1129513 1.0000000 0.5474739  
## x3 0.7235354 0.3983870 0.5474739 1.0000000

# (2)求y关于x1,x2,x3的三元线性回归方程  
equation=lm(y~x1+x2+x3)  
equation

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3)  
##   
## Coefficients:  
## (Intercept) x1 x2 x3   
## -348.280 3.754 7.101 12.447

# (3)对所求得的方程做拟合优度检验  
summary(equation)

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25.198 -17.035 2.627 11.677 33.225   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -348.280 176.459 -1.974 0.0959 .  
## x1 3.754 1.933 1.942 0.1002   
## x2 7.101 2.880 2.465 0.0488 \*  
## x3 12.447 10.569 1.178 0.2835   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.44 on 6 degrees of freedom  
## Multiple R-squared: 0.8055, Adjusted R-squared: 0.7083   
## F-statistic: 8.283 on 3 and 6 DF, p-value: 0.01487

# (4)对回归方程做显著性检验  
anova(equation)

## Analysis of Variance Table  
##   
## Response: y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## x1 1 5234.1 5234.1 9.5248 0.021491 \*   
## x2 1 7659.1 7659.1 13.9378 0.009698 \*\*  
## x3 1 762.2 762.2 1.3870 0.283510   
## Residuals 6 3297.1 549.5   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# (5)对每一个回归系数做显著性检验。  
  
cor.test(y,x1)

##   
## Pearson's product-moment correlation  
##   
## data: y and x1  
## t = 1.8903, df = 8, p-value = 0.09538  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1137799 0.8780797  
## sample estimates:  
## cor   
## 0.5556527

cor.test(y,x2)

##   
## Pearson's product-moment correlation  
##   
## data: y and x2  
## t = 3.0266, df = 8, p-value = 0.01639  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.1870312 0.9316643  
## sample estimates:  
## cor   
## 0.7306199

cor.test(y,x3)

##   
## Pearson's product-moment correlation  
##   
## data: y and x3  
## t = 2.9647, df = 8, p-value = 0.01802  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.1724866 0.9296522  
## sample estimates:  
## cor   
## 0.7235354

# (6)如果有的回归系数没通过显著性检验，将其剔除，重新建立回归方程，再做回归方程的显著性检验和回归系数的显著性检验。  
  
newequation=lm(y~x2+x3)  
newequation;

##   
## Call:  
## lm(formula = y ~ x2 + x3)  
##   
## Coefficients:  
## (Intercept) x2 x3   
## -72.771 6.334 20.755

#显示详细检验结果  
summary(newequation)

##   
## Call:  
## lm(formula = y ~ x2 + x3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.779 -11.510 -3.418 7.754 47.155   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -72.771 123.944 -0.587 0.576  
## x2 6.334 3.371 1.879 0.102  
## x3 20.755 11.418 1.818 0.112  
##   
## Residual standard error: 27.69 on 7 degrees of freedom  
## Multiple R-squared: 0.6833, Adjusted R-squared: 0.5928   
## F-statistic: 7.551 on 2 and 7 DF, p-value: 0.01788

#相关系数检验  
cor.test(y,x2)

##   
## Pearson's product-moment correlation  
##   
## data: y and x2  
## t = 3.0266, df = 8, p-value = 0.01639  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.1870312 0.9316643  
## sample estimates:  
## cor   
## 0.7306199

cor.test(y,x3)

##   
## Pearson's product-moment correlation  
##   
## data: y and x3  
## t = 2.9647, df = 8, p-value = 0.01802  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.1724866 0.9296522  
## sample estimates:  
## cor   
## 0.7235354

# (7)求出每一个回归系数的置信水平为95%的置信区间。  
confint(equation)

## 2.5 % 97.5 %  
## (Intercept) -780.0603287 83.499990  
## x1 -0.9766149 8.484688  
## x2 0.0529179 14.148507  
## x3 -13.4147488 38.309689

# (9)求当x1=75,x2=42,x3=3.1时的y,给定置信水平为95%，用R软件计算精确置信区间，计算近似预测区间。  
  
new<-data.frame(x1=75,x2=42,x3=3.1)   
yconf<-predict(equation,new,interval = "confidence",level = 0.95)#置信区间  
yconf

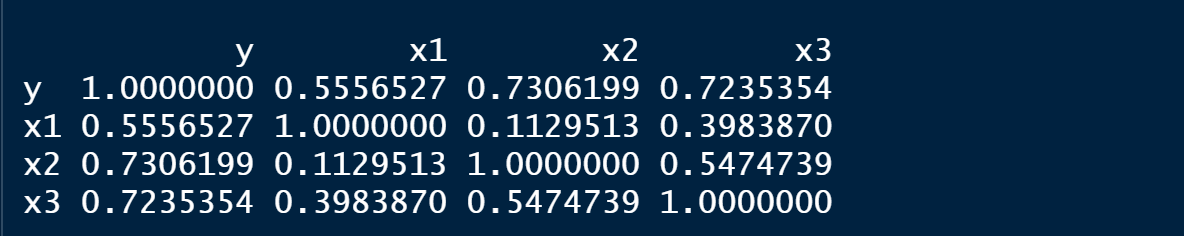
## fit lwr upr  
## 1 270.0897 241.6338 298.5456

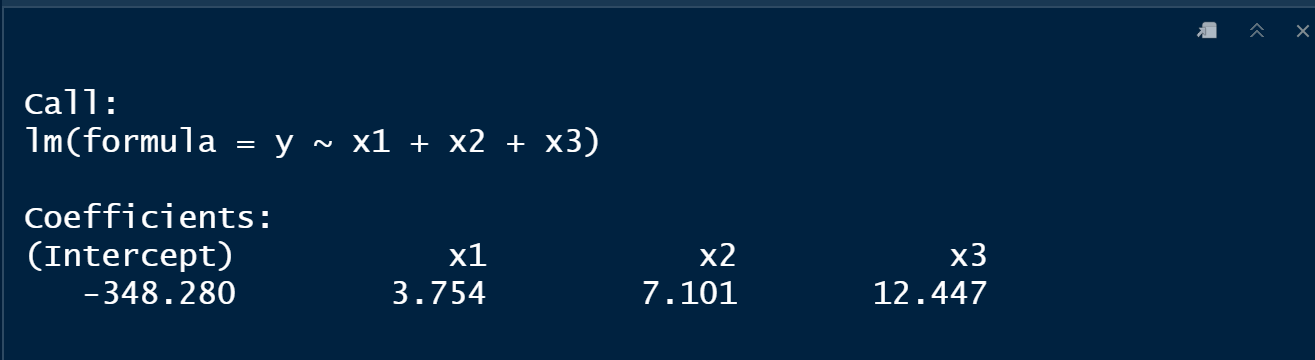
ypred<-predict(equation,new,interval = "prediction",level = 0.95)#预测区间  
ypred

## fit lwr upr  
## 1 270.0897 206.059 334.1204

**二、运行结果分析**

（1）相关系数矩阵如图所示：

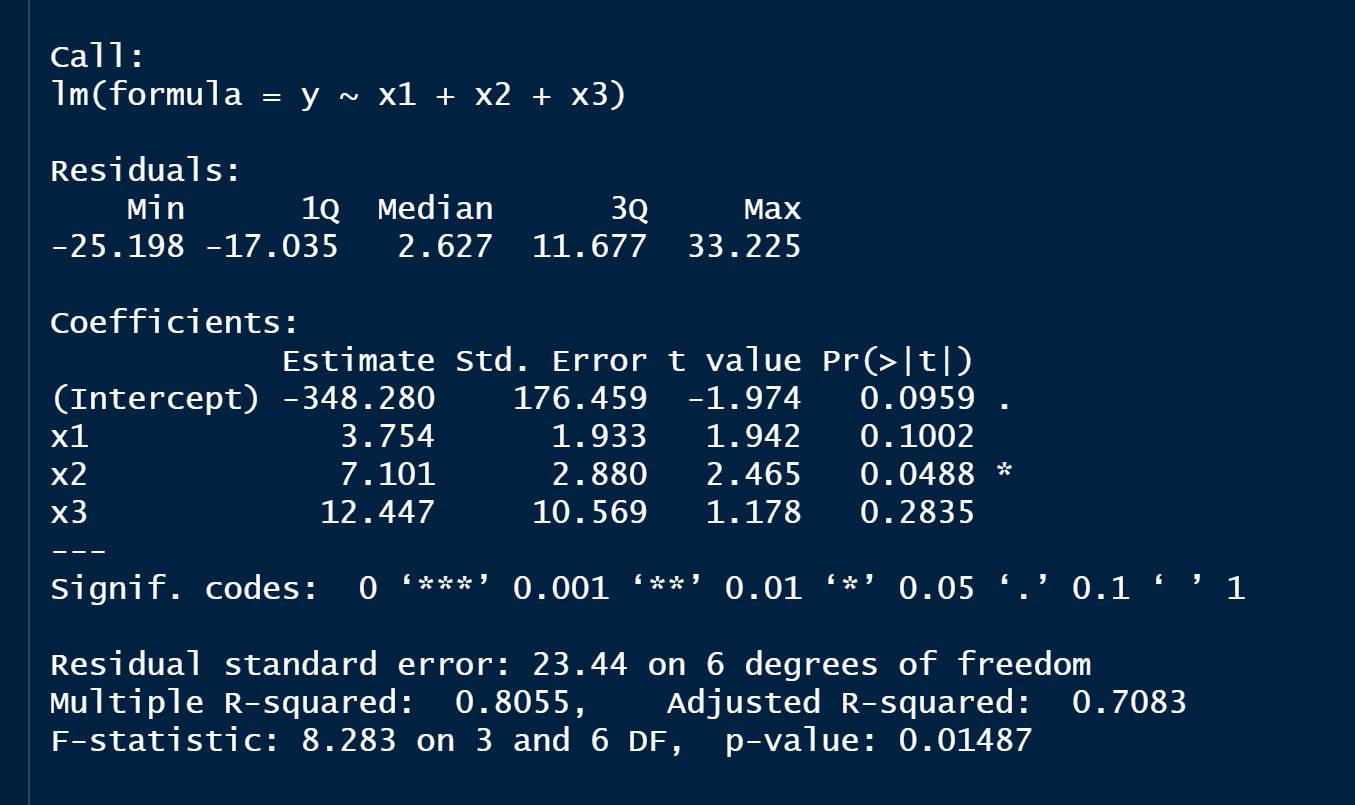


（2） 由

故故

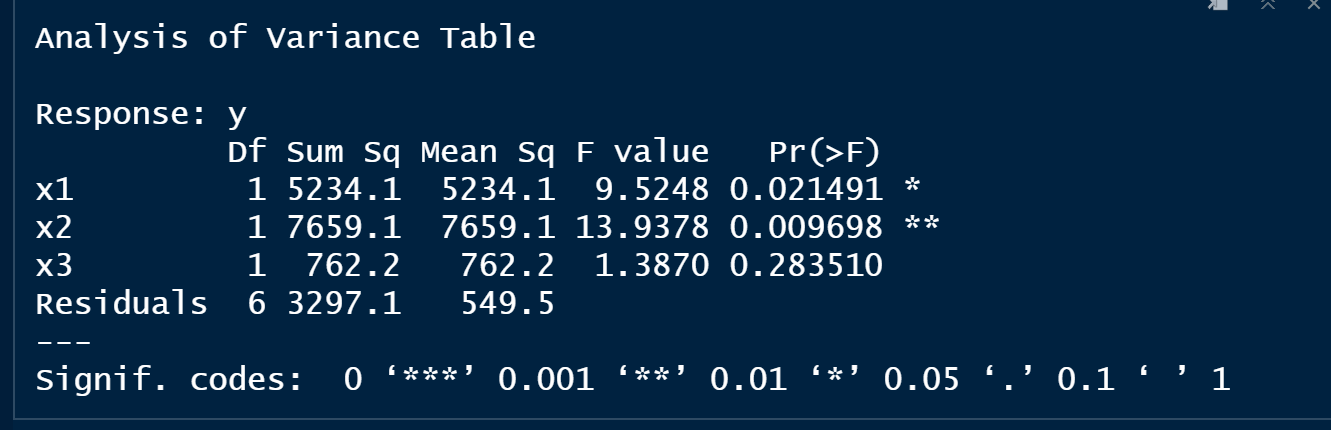
**由图可知回归方程为：y=3.754\*x1+7.101\*x2+12.447\*x3-348.280**

（3）



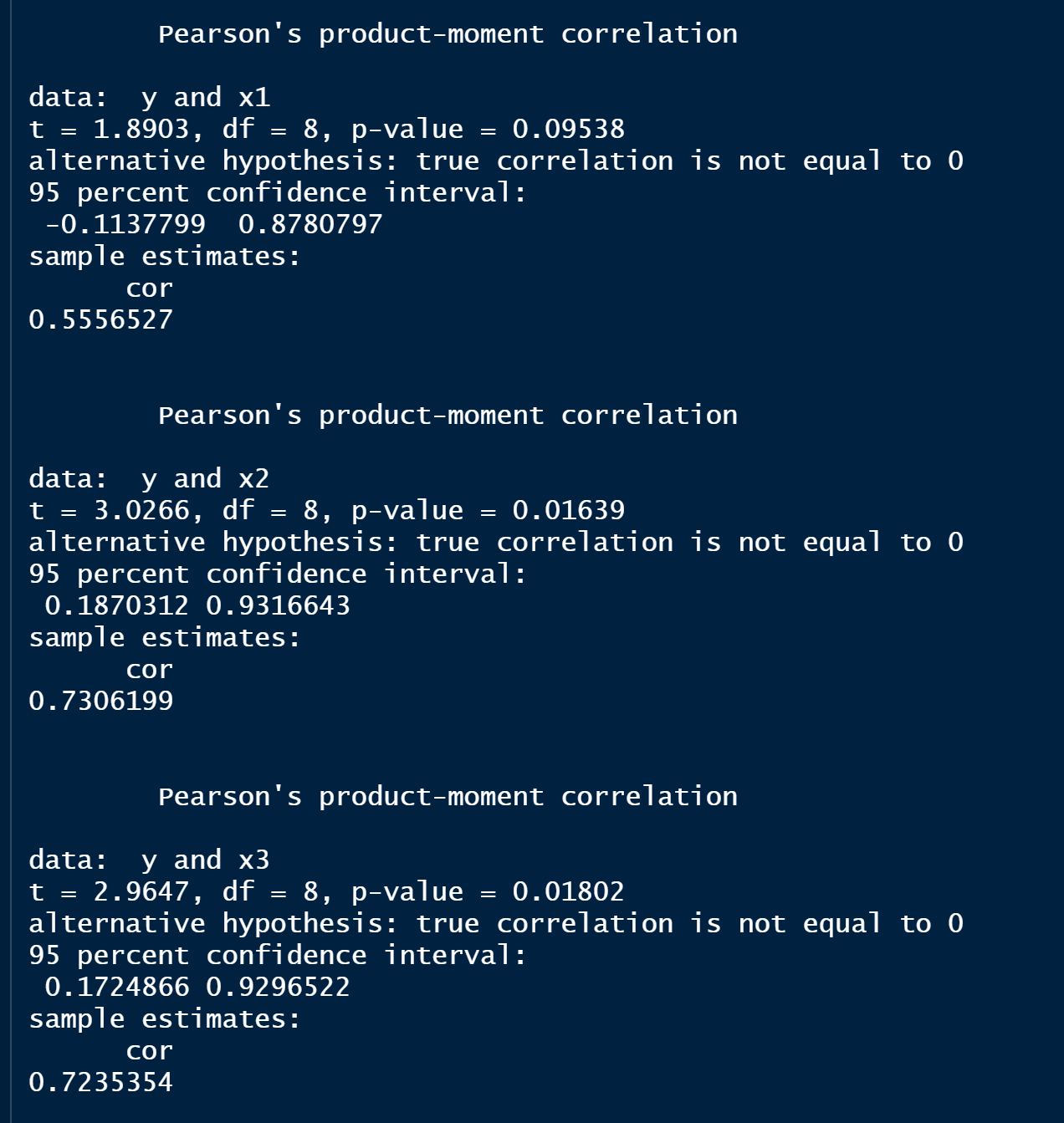
**拟合优度为0.8055，修正后的拟合优度为0.7083接近于1，说明拟合效果较好。**

（4）

****

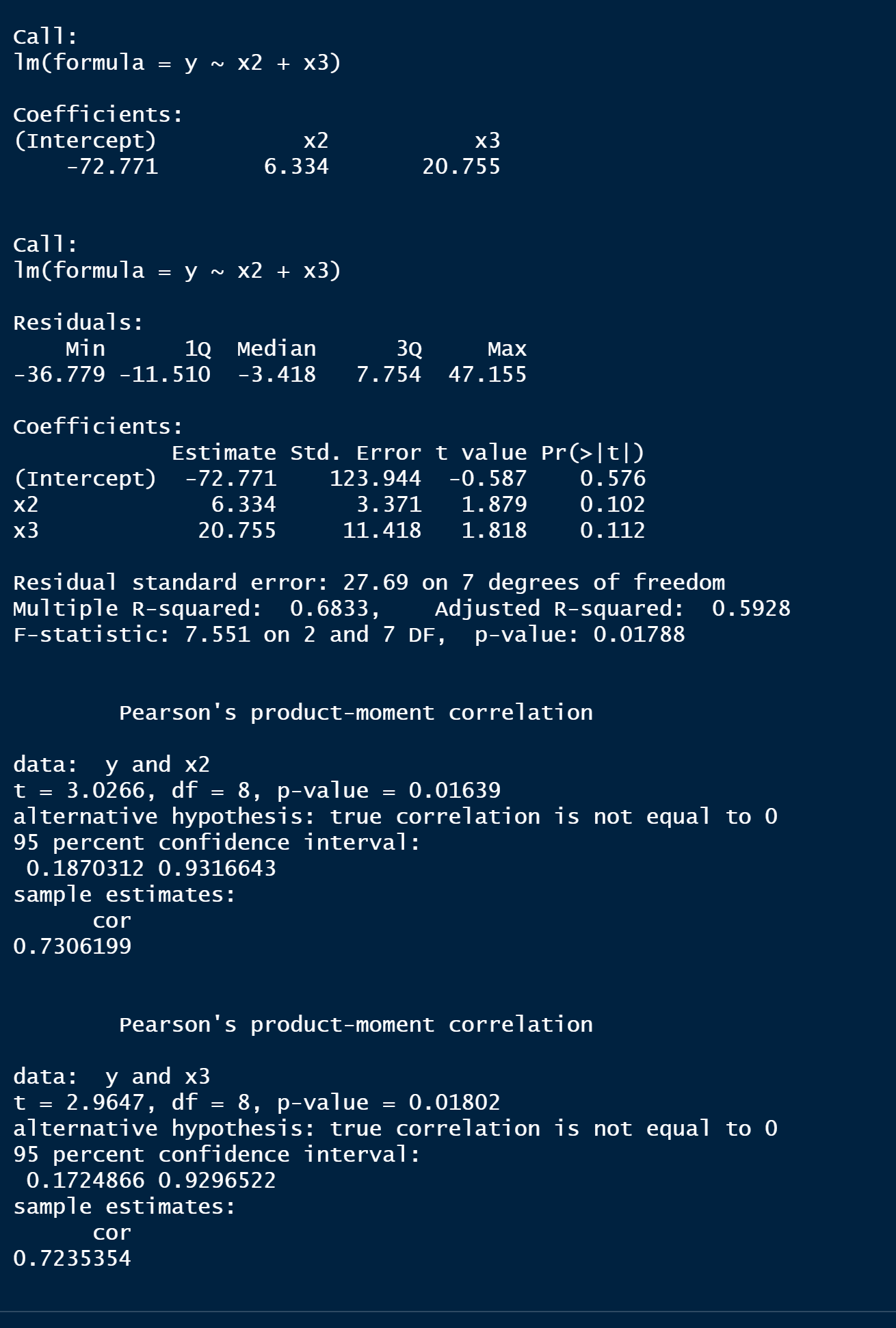
**x3p值为0.283510，偏大，所以x3数在5%显著性水平下不显著，由第（3）问中分析可知，整体函数平均值为0.01487，在5%水平下显著**

（5）.对每一个回归系数做显著性检验

****

**在5%的检验中，结果显示,x1的p值为0.09538大于0.05,相关系数为0.5556527,应将其剔除。x2的p值为0.01639小于0.05,相关系数为0.7306199 。x3的p值为0.7306199小于0.05,相关系数为0.7235354。所以,x2与x3跟y相关性显著，将其保留**

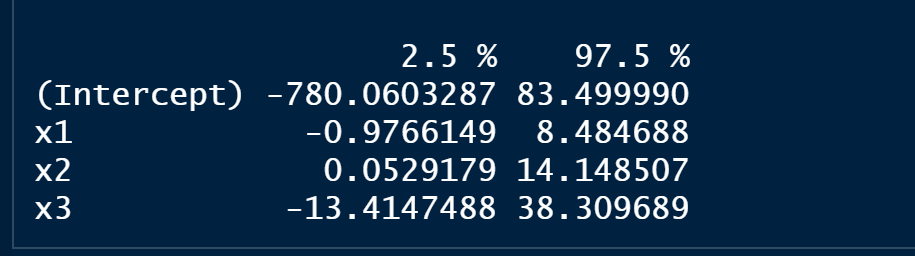
(6)如果有的回归系数没通过显著性检验，将其剔除，重新建立回归方程，再做回归方程的显著性检验和回归系数的显著性检验



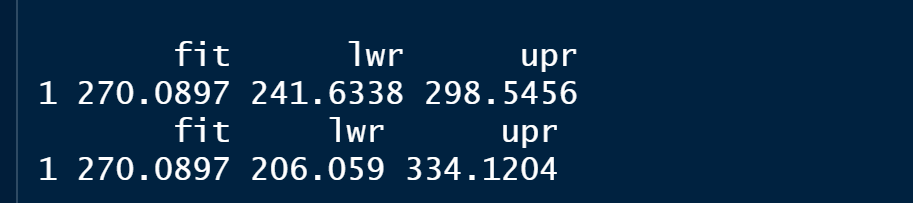
**根据第5题的检验结果，剔除变量x1,调用lm函数重新建立回归方程；调用summary函数和cor.test函数分别进行回归方程的显著性检验和回归系数的显著性检验。结果显示：重建后方程的拟合优度为0.6833，调整后的拟合优度为0.5928，；F统计量的值为7.551，p值仅为0.01788；且x2、x3关于y的相关系数分别为0.7306和0.7235354**

1. 求出每一个回归系数的置信水平为95%的置信区间

**结果如图所示：**



(8)求当x1=75,x2=42,x3=3.1时的y,给定置信水平为95%，用R软件计算精确置信区间，计算近似预测区间



**置信区间为(241.6338,298.5456)，预测值y为270.0897;**

**预测区间为(206.059,335.1204)，预测值y为270.0897**

故

故