第九小组应用回归报告

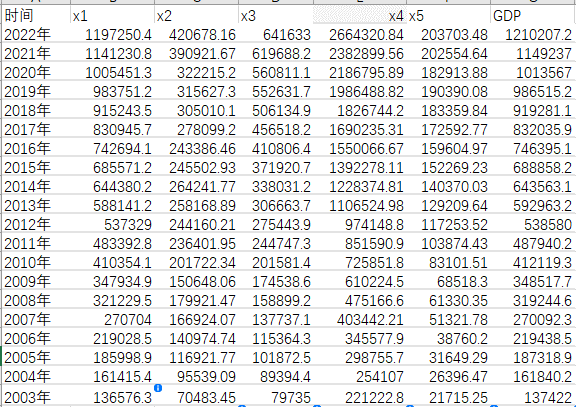
组长：谭铭瑞 组员：施治强，陈艺铭，高冬情，萧海澜

**负责部分**

|  |  |
| --- | --- |
| 谭铭瑞 | 报告整理，自相关性检验 |
| 施治强 | 数据寻找，逐步回归 |
| 陈艺铭 | 多重共线性，岭回归 |
| 高冬情，萧海澜 | PPT制作与演讲 |

2023-06-07

#数据导入  
library(readxl)  
datagz <- read\_excel("C:/Users/Administrator/Desktop/广州市GDP与第一第二第三产业生产总值1(3).xlsx")  
View(datagz)



#国民总收入x1,进出口总额x2,最终消费支出x3,货币和准货币(M2)供应量(亿元)x4,财政收入x5,GDP  
#1.逐步回归  
lm\_ <- lm(GDP~x1+x2+x3+x4+x5,data = datagz)  
lm\_step <- step(lm\_,direction = "both")

## Start: AIC=300.49  
## GDP ~ x1 + x2 + x3 + x4 + x5  
##   
## Df Sum of Sq RSS AIC  
## - x3 1 4178 36774012 298.49  
## - x5 1 532450 37302284 298.78  
## <none> 36769835 300.49  
## - x2 1 5222830 41992664 301.15  
## - x4 1 31759328 68529163 310.94  
## - x1 1 650413254 687183088 357.05  
##   
## Step: AIC=298.49  
## GDP ~ x1 + x2 + x4 + x5  
##   
## Df Sum of Sq RSS AIC  
## - x5 1 544913 37318926 296.79  
## <none> 36774012 298.49  
## + x3 1 4178 36769835 300.49  
## - x2 1 18768794 55542807 304.74  
## - x4 1 32778466 69552478 309.24  
## - x1 1 3035227723 3072001736 385.00  
##   
## Step: AIC=296.79  
## GDP ~ x1 + x2 + x4  
##   
## Df Sum of Sq RSS AIC  
## <none> 37318926 296.79  
## + x5 1 544913 36774012 298.49  
## + x3 1 16642 37302284 298.78  
## - x2 1 23383396 60702322 304.52  
## - x4 1 62037246 99356171 314.37  
## - x1 1 7099134172 7136453098 399.85

summary(lm\_step)

##   
## Call:  
## lm(formula = GDP ~ x1 + x2 + x4, data = datagz)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3335.3 -564.5 48.1 674.4 2882.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.260e+02 1.457e+03 -0.155 0.87869   
## x1 9.206e-01 1.669e-02 55.169 < 2e-16 \*\*\*  
## x2 5.449e-02 1.721e-02 3.166 0.00599 \*\*   
## x4 3.216e-02 6.236e-03 5.157 9.54e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1527 on 16 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 3.15e+05 on 3 and 16 DF, p-value: < 2.2e-16

#结果为选择x1,x2,x4  
#自相关性检验  
lmgz = lm(GDP~x1+x2+x4,data = datagz)   
#检测自相关性  
#计算dw  
library(lmtest)

## 载入需要的程辑包：zoo

##   
## 载入程辑包：'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

dwtest(lmgz,alternative = "two.side")

##   
## Durbin-Watson test  
##   
## data: lmgz  
## DW = 2.4977, p-value = 0.574  
## alternative hypothesis: true autocorrelation is not 0

#k=4,n=20时dl = 1,du=1.68此时dw=2.49 处于(4-du,4-dl)区间，无法判断自相关  
#3.多重共线性检验  
#检测多重共线性  
library(car)

## 载入需要的程辑包：carData

vif(lm\_step)

## x1 x2 x4   
## 258.91510 21.12597 186.82357

#剔除x1  
lm\_data1<-lm(lm(GDP ~ x2 + x4 , data = datagz))  
summary(lm\_data1)

##   
## Call:  
## lm(formula = lm(GDP ~ x2 + x4, data = datagz))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -57099 -7502 2439 5986 36146   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.638e+04 1.913e+04 0.856 0.40370   
## x2 6.644e-01 1.769e-01 3.755 0.00158 \*\*   
## x4 3.646e-01 2.156e-02 16.910 4.56e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20490 on 17 degrees of freedom  
## Multiple R-squared: 0.9968, Adjusted R-squared: 0.9964   
## F-statistic: 2617 on 2 and 17 DF, p-value: < 2.2e-16

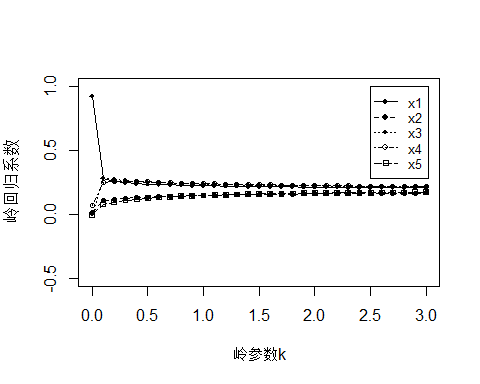
vif(lm\_data1)

## x2 x4   
## 12.40772 12.40772

#此时发现无法消除多重共线性  
#转而使用岭回归  
  
#4.岭回归  
datas<-data.frame(scale(datagz[,2:7]))  
library(MASS)  
ridge\_data<-lm.ridge(GDP~x1+x2+x3+x4+x5-1,data = datas,lambda = seq(0,3,0.1))  
beta<-coef(ridge\_data)  
beta

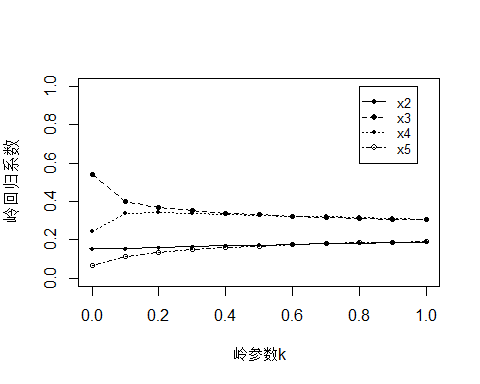
## x1 x2 x3 x4 x5  
## 0.0 0.9246766 0.01385426 -0.00155624 0.06702152 -0.003453516  
## 0.1 0.2822488 0.11152213 0.28430790 0.24871767 0.077711277  
## 0.2 0.2580820 0.11915393 0.27200448 0.25856685 0.096095992  
## 0.3 0.2475371 0.12474222 0.26366516 0.25867509 0.108524402  
## 0.4 0.2410736 0.12937939 0.25745238 0.25645702 0.117960695  
## 0.5 0.2364877 0.13336475 0.25251281 0.25362465 0.125476581  
## 0.6 0.2329647 0.13684707 0.24842577 0.25071529 0.131633012  
## 0.7 0.2301217 0.13992156 0.24495290 0.24791687 0.136775014  
## 0.8 0.2277491 0.14265670 0.24194473 0.24529271 0.141133816  
## 0.9 0.2257207 0.14510474 0.23930062 0.24285724 0.144872829  
## 1.0 0.2239541 0.14730699 0.23694915 0.24060518 0.148111852  
## 1.1 0.2223929 0.14929683 0.23483769 0.23852362 0.150941161  
## 1.2 0.2209967 0.15110167 0.23292624 0.23659731 0.153430279  
## 1.3 0.2197355 0.15274431 0.23118367 0.23481098 0.155633718  
## 1.4 0.2185867 0.15424390 0.22958530 0.23315034 0.157594873  
## 1.5 0.2175325 0.15561664 0.22811122 0.23160239 0.159348745  
## 1.6 0.2165590 0.15687637 0.22674517 0.23015554 0.160923894  
## 1.7 0.2156549 0.15803500 0.22547369 0.22879953 0.162343860  
## 1.8 0.2148110 0.15910278 0.22428555 0.22752528 0.163628226  
## 1.9 0.2140199 0.16008866 0.22317128 0.22632482 0.164793418  
## 2.0 0.2132752 0.16100042 0.22212280 0.22519111 0.165853319  
## 2.1 0.2125715 0.16184491 0.22113325 0.22411793 0.166819743  
## 2.2 0.2119044 0.16262814 0.22019666 0.22309979 0.167702805  
## 2.3 0.2112701 0.16335545 0.21930789 0.22213182 0.168511221  
## 2.4 0.2106652 0.16403156 0.21846246 0.22120971 0.169252535  
## 2.5 0.2100869 0.16466070 0.21765643 0.22032960 0.169933316  
## 2.6 0.2095327 0.16524663 0.21688633 0.21948805 0.170559306  
## 2.7 0.2090003 0.16579272 0.21614913 0.21868200 0.171135551  
## 2.8 0.2084880 0.16630202 0.21544209 0.21790868 0.171666497  
## 2.9 0.2079939 0.16677725 0.21476280 0.21716561 0.172156081  
## 3.0 0.2075167 0.16722090 0.21410911 0.21645055 0.172607801

k<-ridge\_data$lambda  
plot(k,k,type = "n",xlab ="岭参数k" , ylab = "岭回归系数", ylim=c(-0.5,1))  
linetype<-c(1:5)  
char<-c(18:25)  
for(i in 1:5)  
 lines(k,beta[,i],type="o",lty=linetype[i],pch=char[i],cex=0.75)  
legend(2.5,1,inset = 0.5,legend = c("x1","x2","x3","x4","x5"),cex=0.8,pch = char, lty = linetype)

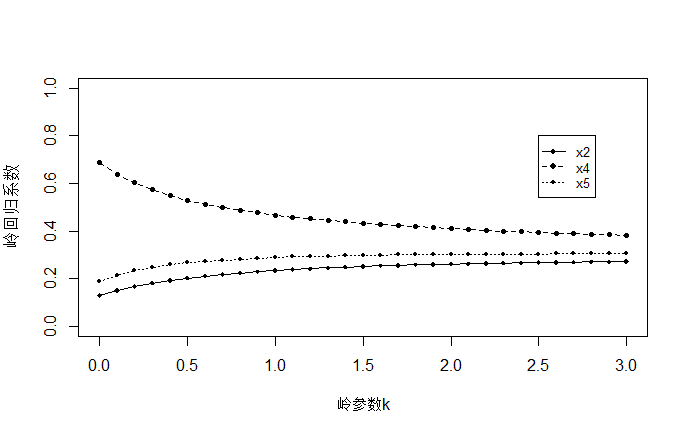


#由图可发现，x1与x4一个上升的同时伴随着另一个下降，说明两者具有多重共线性，二者选择其一删除，从经济角度思考，我们决定删除x1,。

而其他的变量x2x3x5在k>0.5后趋于稳定，且变化幅度不大，绝对值也不小,选择保留。  
#剔除x1



#此时发现x2与x3一个升起伴随着另一个下降，说明其具有多重共线性，选择其一删除即可，我们选择删除x3



#可以看到此时的岭迹图是比较健康的

#可以看出，岭参数k大于2时，岭参数取值基本稳定，不妨选取k=2

y=0.1679685x2 + 0.6026944x4 + 0.2341634x5

可以看到我们的因变量GDP与进出口总额x2,货币和准货币(M2)供应量(亿元)x4,财政收入x5有正相关。

跟我们的模型，我们发现，

当进出口总额增加时，说明对外贸易量增加，进而使GDP增加。

当货币供应量增加，说明政府介入市场，流入市场货币增加，刺激经济活动，导致GDP增加

当财政收入增加，说明政府在一定时期内的总收入增加如税收等，可用于投资或支出，刺激生产和就业，从而提高GDP