**实验报告**

|  |  |  |
| --- | --- | --- |
| **课程名称**： **应用回归分析实验** | **学期： 2023年秋季学期** | **成绩**： |
| **指导教师： 陈丹** | **学生姓名**： **枫叶** | **学生学号**： |
| **实验名称**： **多元回归分析** | | |
| **实验编号**：3 | **实验日期**： | **实验学时**： |
| **学院： 数学与统计学院** | **专业： 统计学** | **年级**： **2021级** |

1. **实验目的**

掌握多元回归分析应用

1. **使用环境**

R语言

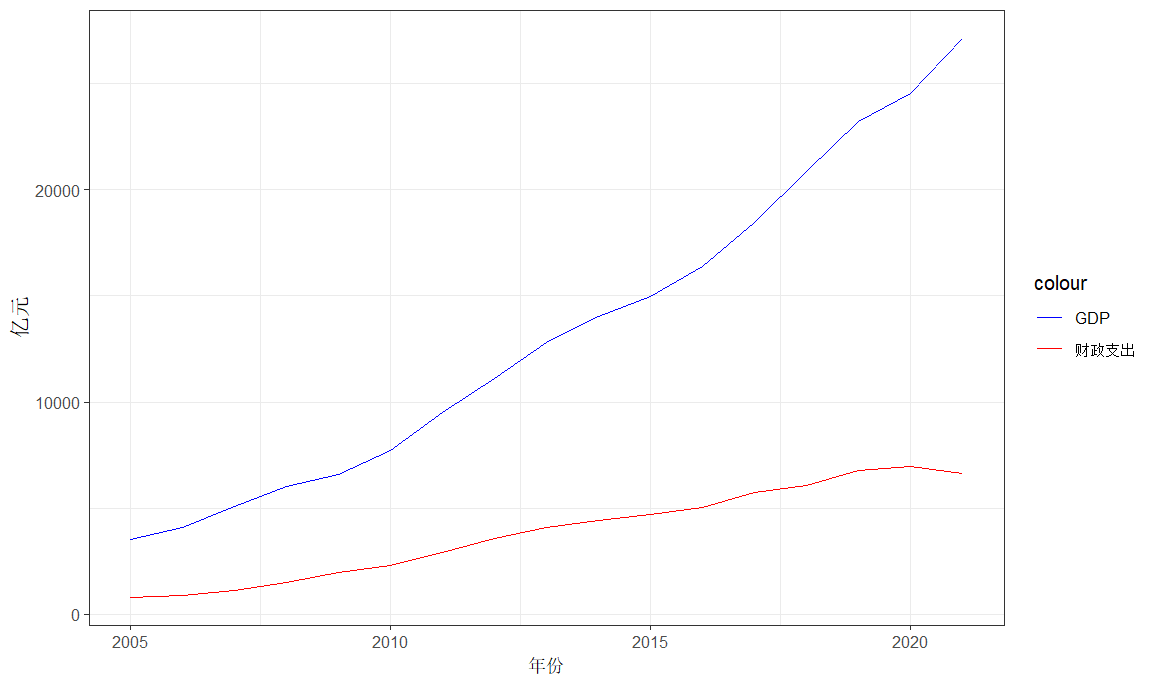
**三、实验内容**

多元回归分析流程

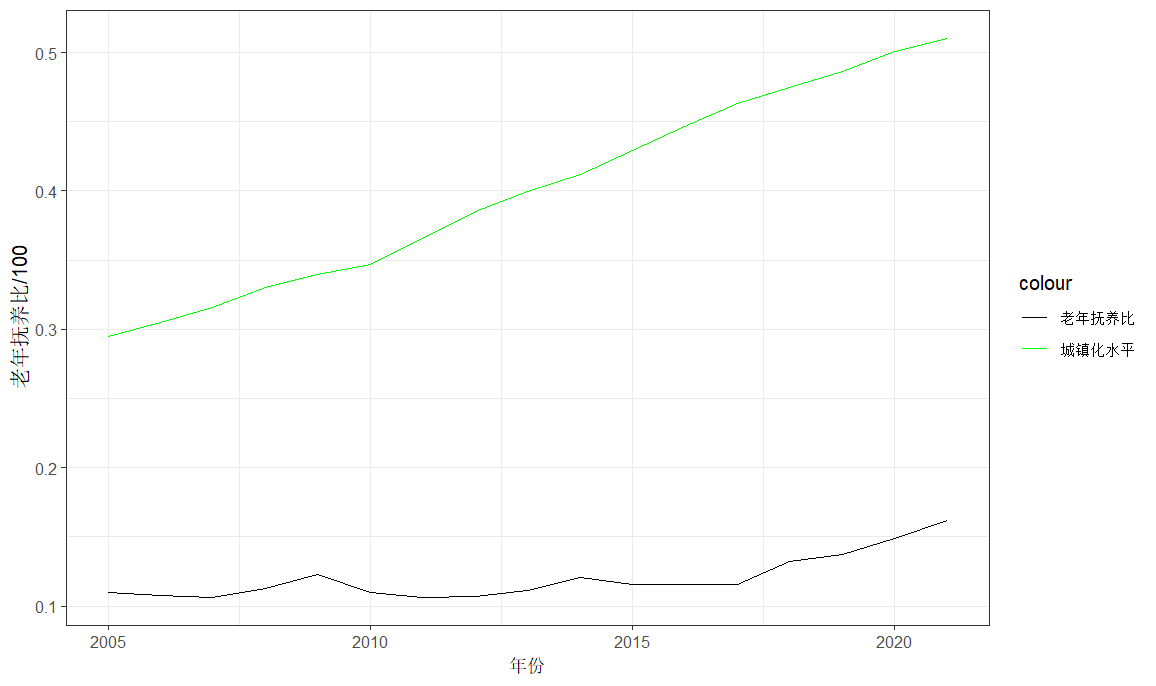
1. **算法介绍及结果**

**数据说明**：数据获取自国家统计局网站，时间跨度为2005-2021,变量包括云南省经济增长（以GDP取对数处理）,政府干预（以财政支出取对数处理），城镇化水平（以城镇人口占总人口表示），老年抚养比（65岁以上人口数比劳动年龄人口数），所有价格均以2005年为基期，按GDP平减指数折算

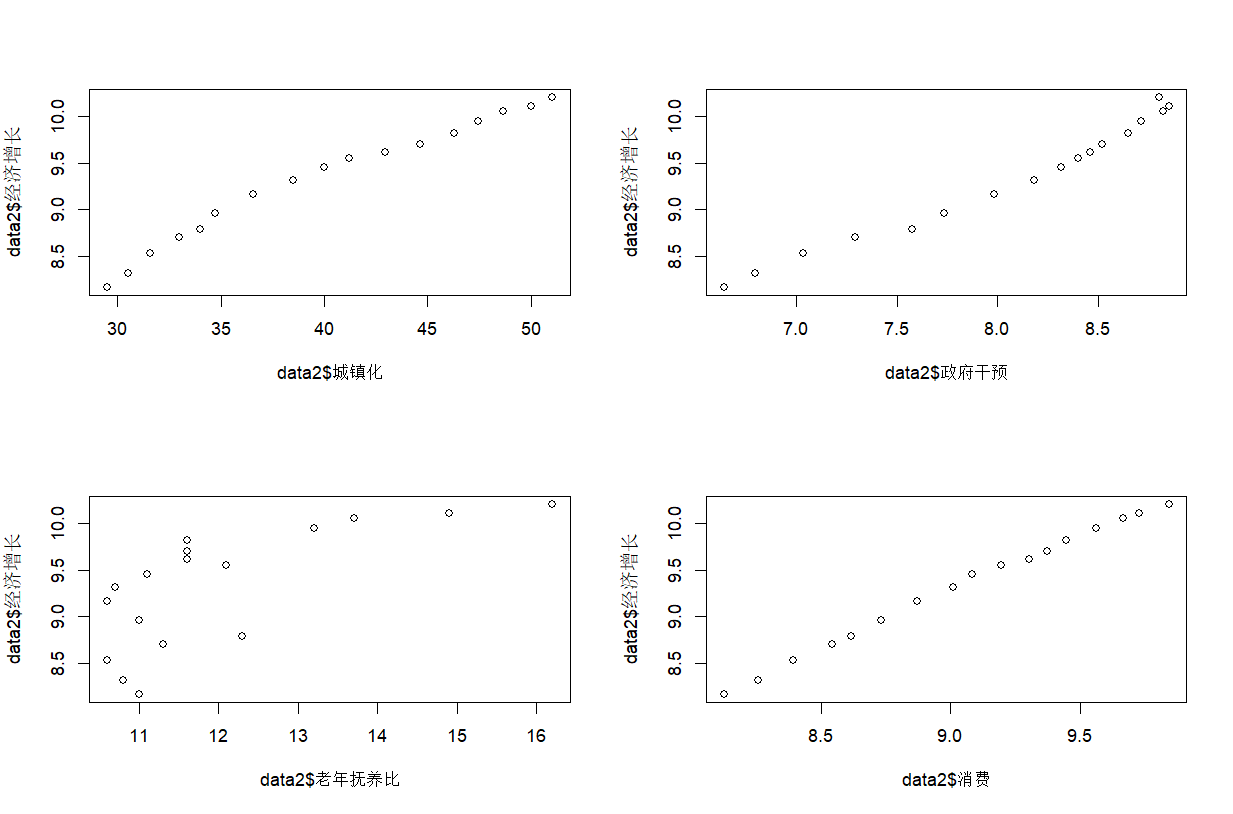
**1.描述性统计**



从上图来看，GDP始终保持较高增速，财政支出的增速相对较低，且在2020年出现下降趋势



从上图来看，城镇化水平在不断提升，这意味着云南人口持续向城镇集中，农村空心化问题可能加剧，而老年抚养比有起有落，在2017年开始持续上升

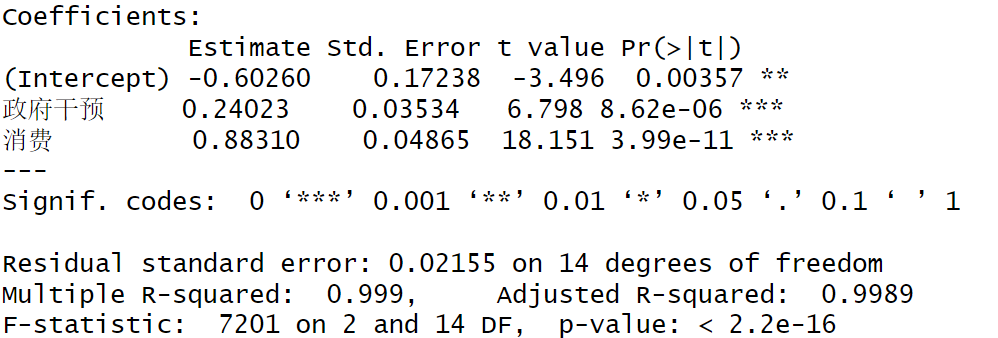


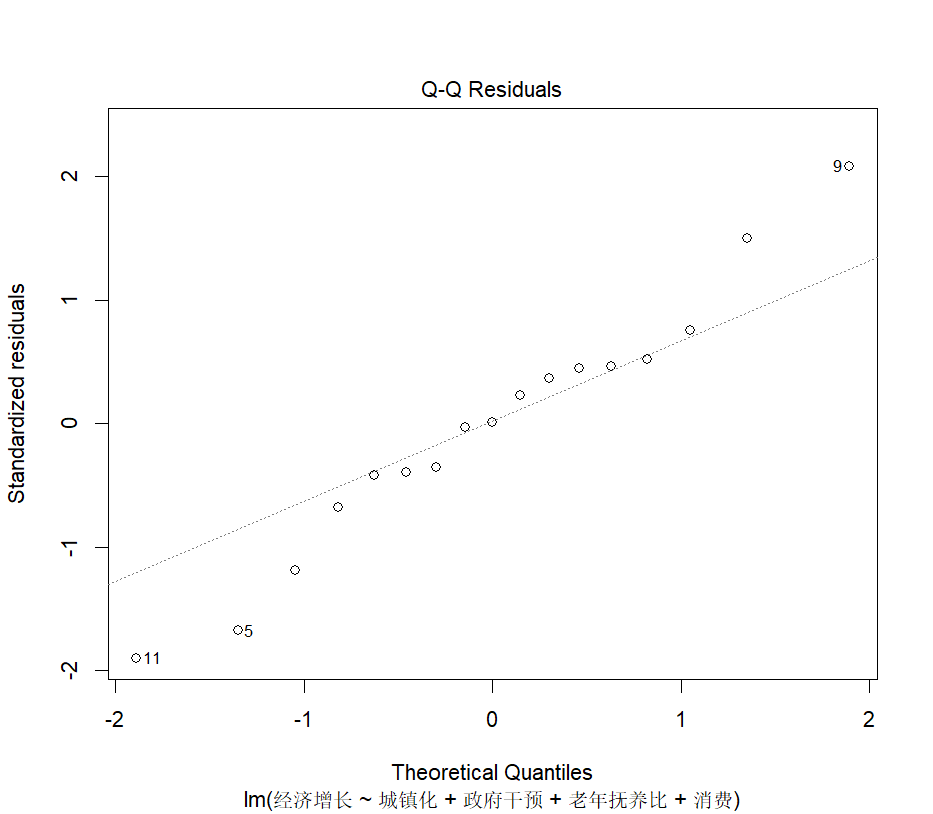
从散点图可以看出，城镇化水平、政府干预和人均消费极有可能与经济增长存在正相关的线性关系，而老年抚养比与经济增长无明显的关系

**2.回归分析**

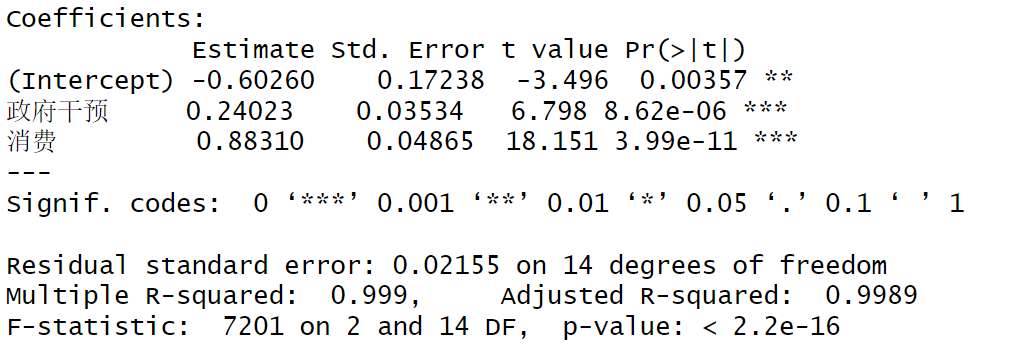
建立如下回归模型

其中GDP表示国民生产总值，Fin表示财政支出，con表示人均消费，odr表示老年抚养比，urb表示城镇化水平，i表示2005年至2020年的各个观测值，求得回归结果如下





作出残差QQ图，其并不呈一条直线，这表明数据可能不服从正态分布，但shapiro检验p值为0.96，故我们接受残差服从正态分布的假设，下面进一步使用AIC准则筛选显著变量，筛选后的回归结果如下，斜率系数均在0.001显著性水平上显著，修正为0.9989，这表明模型拟合效果较好，大部分波动都被自变量解释了



最终模型为

经济学意义上来说，0.24意味着财政支出增加1%，GDP增加24%，0.88意味着人均消费增加1%，GDP增加88%，这一意义显然是有违常理的，但由于这里仅出于练习目的应用多元回归，故不作深究。

**3.方差分析**

作出方差分析表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 方差来源 | 自由度 | 平方和 | 均方 | F值 | P值 |
| 政府干预 | 1 | 6.5372 | 6.5372 | 14071.71 | <2.2e-16 |
| 消费 | 1 | 0.1531 | 0.1531 | 329.45 | 3.989e-11 |
| 残差 | 14 | 0.0065 | 0.0005 |  |  |
| 总 | 16 | 6.6968 |  |  |  |

注意到绝大部分波动是由政府干预所解释的，这符合我国政府主导经济发展的情况，但这并不能证明存在因果关系，也可能是GDP高使得政府可以有更大的开支 ，此外这里用政府干预一词其实并不恰当，单纯的支出度量不能反映政府干预程度，应当用财政支出与GDP的比值或其他相对值来度量，这里暂且不作改进

**4.预测**

用2022年数据预测结果为27226.88，0.95显著性水平区间估计为(26373.77,28107.58)，真实值为28954.20亿元，预测结果有所低估

**代码**

|  |
| --- |
| **library(tidyverse)**  **library(ggplot2)**  **library(readxl)**  **data <- read\_excel("D:/预删除文件夹/大三上/应用回归分析/task3数据.xlsx")**  **data2 <- data %>%**  **mutate(经济增长=log(GDP),**  **财政支出=log(财政支出),**  **城镇化=城镇人口/总人口\*100,**  **消费=log(人均消费支出)) %>%**  **select(年份,经济增长,城镇化,政府干预,老年抚养比,消费)**  **#描述统计**  **layout(matrix(c(1,1)))**  **ggplot(data,aes(x=年份))+**  **geom\_line(aes(y=GDP,col="blue")) +**  **geom\_line(aes(y=财政支出,col="red"))+**  **scale\_color\_manual(values=c("blue", "red"),**  **labels = c("GDP", "财政支出"))+**  **scale\_y\_continuous(name = "亿元") +**  **theme\_bw()**  **ggplot(data,aes(x=年份))+**  **geom\_line(aes(y=老年抚养比/100,col="black")) +**  **geom\_line(aes(y=城镇人口/总人口,col="green")) +**  **scale\_color\_manual(values=c("black","green"),**  **labels = c("老年抚养比", "城镇化水平")) +**  **theme\_bw()**  **par(mfrow=c(2,2))**  **plot(data2$城镇化,data2$经济增长)**  **plot(data2$财政支出,data2$经济增长)**  **plot(data2$老年抚养比,data2$经济增长)**  **plot(data2$消费,data2$经济增长)**  **model <- lm(data=data2,经济增长 ~ 城镇化+财政支出+老年抚养比+消费)**  **summary(model)**  **plot(model,2)**  **shapiro.test(model$residuals)**  **model2 <- step(model)**  **summary(model2)**  **anova(model2)**  **new <- data.frame(政府干预=log(6699.74),消费=log(18951))**  **exp(predict(model2,newdata = new,interval = "confidence"))** |

**六、参考文献**

**七、教师评语**