使用多元线性回归模型对中国人口增长率进行分析

一、研究背景和目的

在当今时代背景下，“二孩”政策早已全面放开，可在二孩之前，我国早已经进行了为期四十多年的“一孩”政策。因此新的计划生育推行的当今，对已经完成的旧的计划生育政策就需要有一个系统性的评估和诊断。于是也就有了本次研究

本文的“一孩”政策，指的时从1971年到2013年为止的长达四十二年的计划生育政策。在1980年9月，党中央发表《关于控制我国人口增长问题致全体共产党、共青团员的公开信》，提倡一对夫妇只生一个孩子。这项政策推出之后，使得中国自然增长率从1970年的5.8降到1980年的2.24，效果显著。

自然增长率也很大程度上与经济发展等各方面因素相联系，与我们的经济生活息息相关，为了研究这时影响中国自然增长率的原因，分析人口增长，和猜测中国未来的增长趋势，需要建立研究模型。在此文中，采用了多元线性回归模型。

影响自然增长率的因素有很多，大致可以有一下几种：从国民经济上来看，经济增长是人口自然增长的基本源泉；居民消费水平也会一定程度上反应出人口增长率；居民的文化程度，会影响人民的思想，从而影响人口的自然增长率；人口分布，非农业和农业人口的占比和分布也会对人口自然增长率造成影响。

二、模型建立

从国家统计局官网，收集到如下数据信息：



（表1 中国人口自然增长率及相关数据）

根据搜集的数据，为了更加全面的了解人口增长率的影响因素，选择人口自然增长率为解释变量，以此来反映人口增长。选择国民总收入和人均GDP作为精致增长的指标。选择居民消费价格指数增长率作为居民消费水平的代表。鉴于搜集到的数据，暂且考虑这些影响因素。

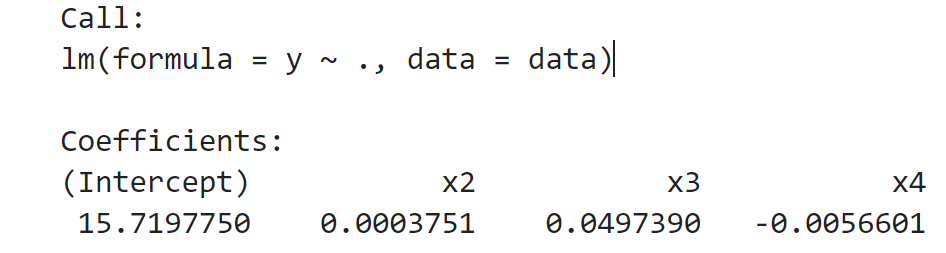
根据以上模型，建立多元线性回归模型：

Y = β1+β2X2+β3X3+β4X4

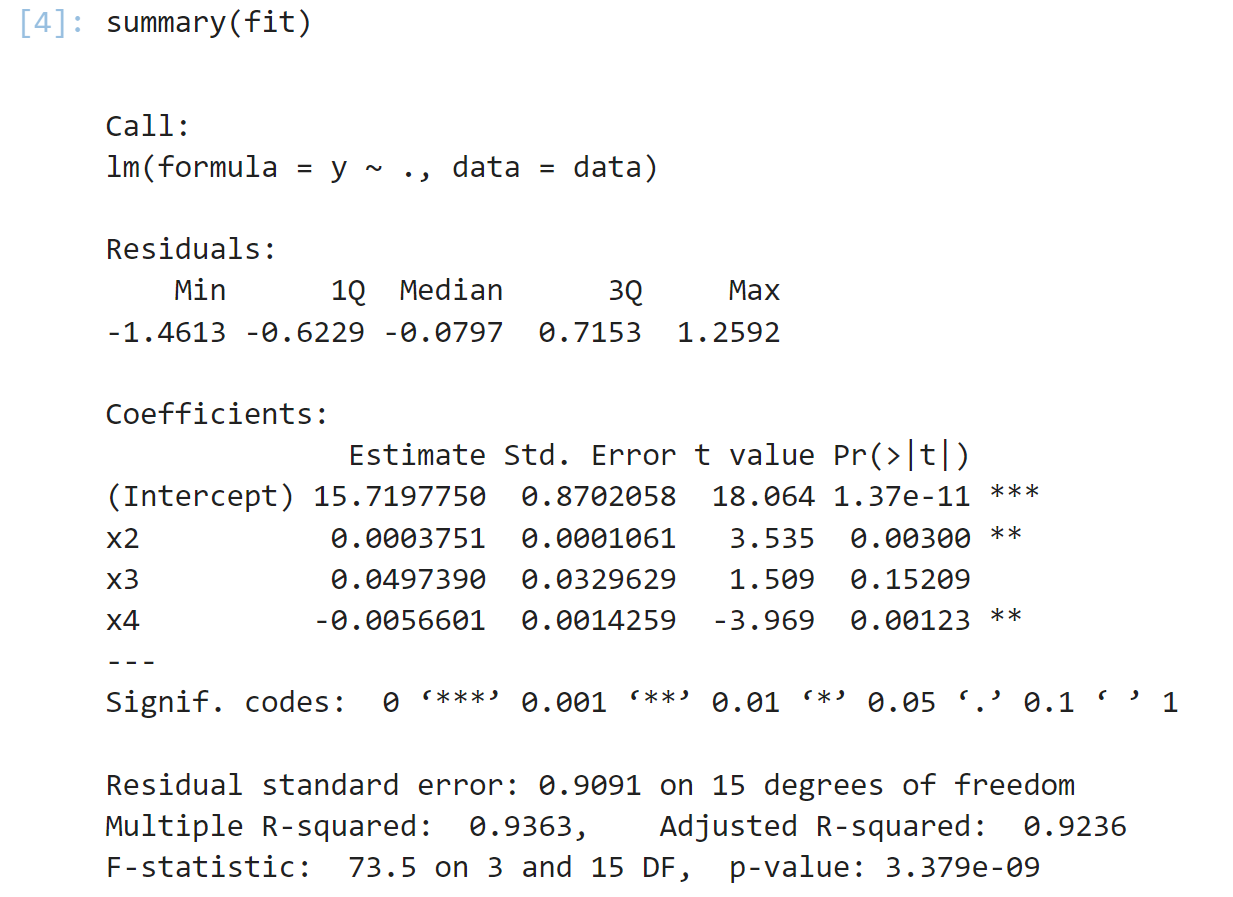
三、模型拟合及参数估计

利用最小二乘估计模型，使用R语言估计步骤如下：

1. 建立相应的文件，将数据文件和R存储在同一目录下，然后使用，data = read.csv("中国增长率及相关数据.csv")函数读取数据，使用colnames(data) = c("y","x2","x3","x4")方法对列名做对应的参数更改。
2. 对于修改好的数据，使用fit=lm(y~.,data=data)对数据进行最小二乘的多元线性回归拟合用summary(fit)进行估计。
3. 估计结果如下：



**（表2）**



**（表3）**

得到的多元回归模型为：

Y = 15.7197750+0.0003751 **X2**+0.0497390 **X3**-0.0056601 **X4**

**S=（0.8702058）（0.0001061）（0.0329629）（0.0014259）**

**t=（18.064） (3.535) (1.509) (-3.969)**

**R^2=0.9363 A-R^2=0.9236**

**F=73.5**

四、模型检验分析

1. 经济意义检验

模型估计结果可以知道，在其他变量不变的情况下，当年国民总收入每增长1亿元，人口增长率增长0.0003751%；在其他变量不变的情况下，当年居民消费价格指数增长率每增长1%，人口增长率增长0.0497390%；在其他条件不变的情况下，当人均GDP每增加1元，人口增长率就会降低0.0056601%。这些理论分析和经验判断一致。

2. 统计检验

拟合优度：由表2中数据可以得到R^2= ，修正的可决系数为A-R^2= ,这说明模型样本拟合的非常好。

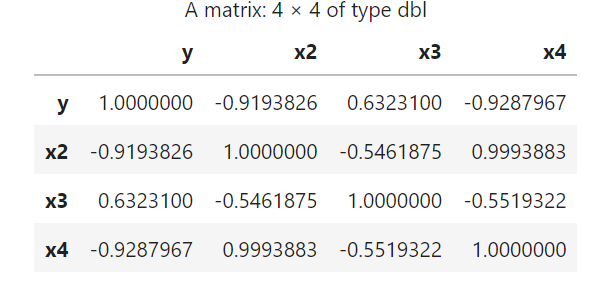
F检验：针对H0：β2=β3=β4=0，给定显著性水平α=0.05，在F分布表中查出自由度为k-1=3和n-k=14的临界值Fα(3,14)=3.34。由表3中得到F=73.5>Fα(3,21)=3.075，应拒绝原假设H0：β2=β3=β4=0，说明回归方程显著，即“国民总收入”、“居民消费价格指数增长率”、“人均GDP”等变量联合起来确实对“人口自然增长率”有显著影响。

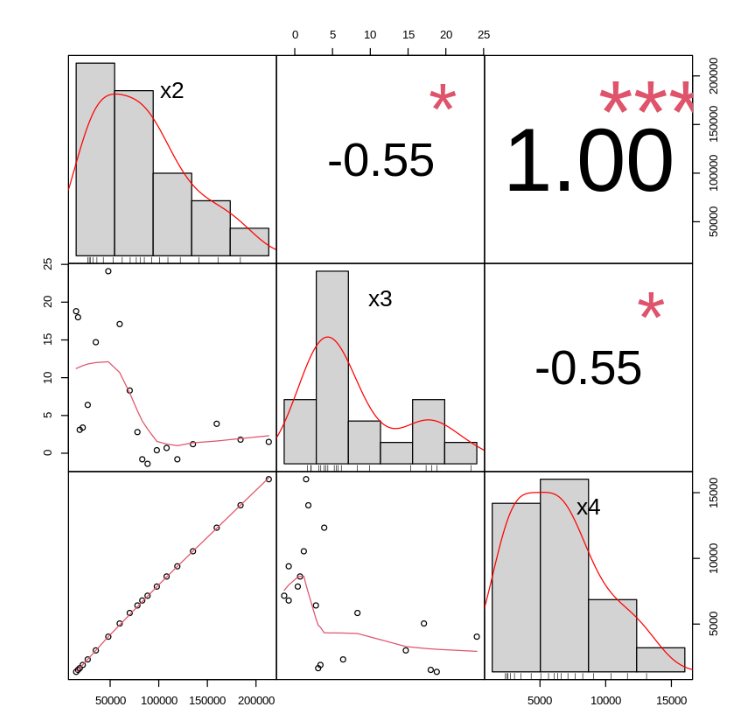
T检验：分别针对H0：βj=0（j=1，2，3，4），给定显著性水平a=0.05，查看t分布表得自由度为n-k=14临界值tα/2(n-k)=2.145。

由表3中的数据可以得到，与β1，β2，β3，β4对应的t统计量分别为17.08010、2.482857、1.412721、-2.884953。除了β3，其绝对值均大于tα/2(n-k)=2.145，这说明分别都应当拒绝H0，也就是说，当在其他解释变量不变的情况下，解释变量“国民总收入”、“人均GDP”分别对被解释变量“人口自然增长率”Y都有显著的影响。

β3的绝对值小于tα/2(n-k)=2.145，这说明接受H0，X3系数对t检验不显著，这说明又存在多重共线性的可能。

所以计算各解释变量的相关系数，选择X2、X3、X4数据，使用PerformanceAnalytics包chart.Correlation(data[,-1], histogram=TRUE, pch=19)来得到相关系数矩阵





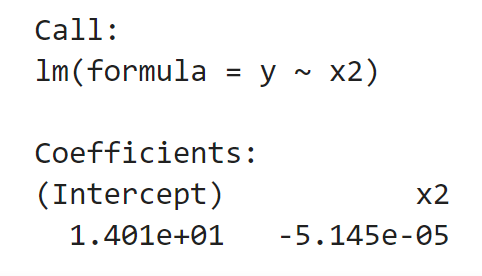
（表4）

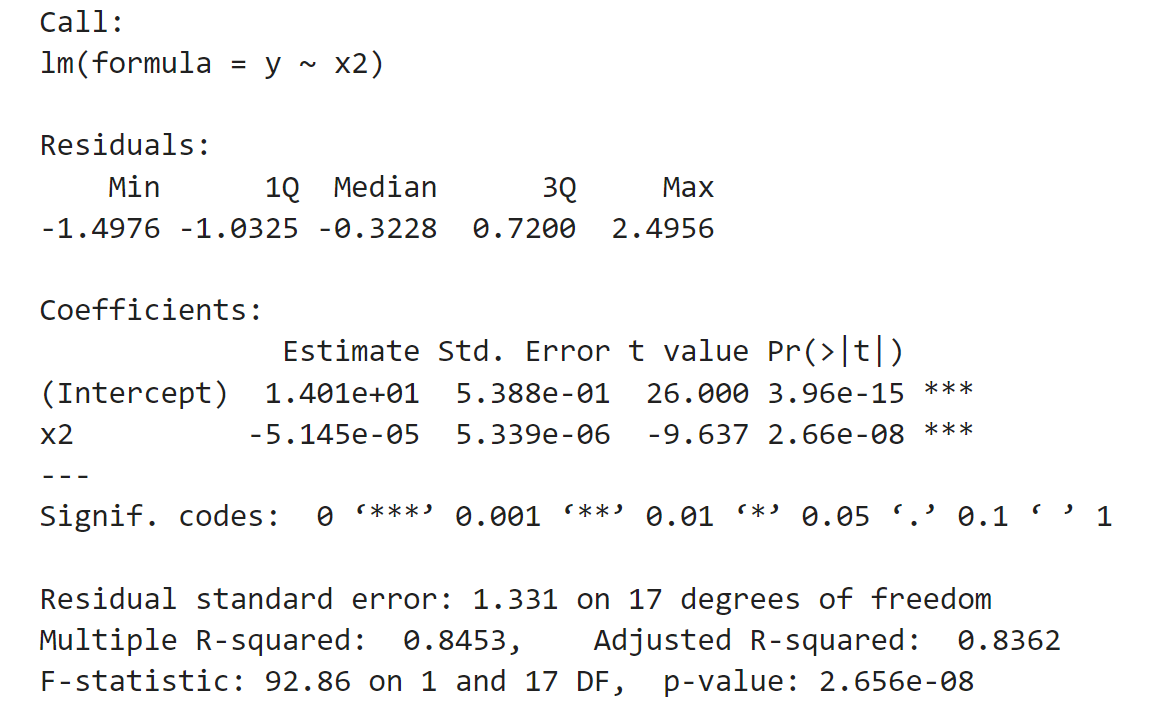
由相关系数矩阵可以看出：各解释变量相互之间的相关系数较高，由此可知确实存在多重共线性的问题。

五、消除多重共线性

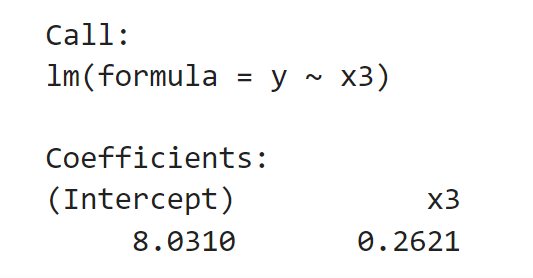
采用逐步回归的方法，去解决多重共线性问题，分别做y对X2，X3，X4的一元回归

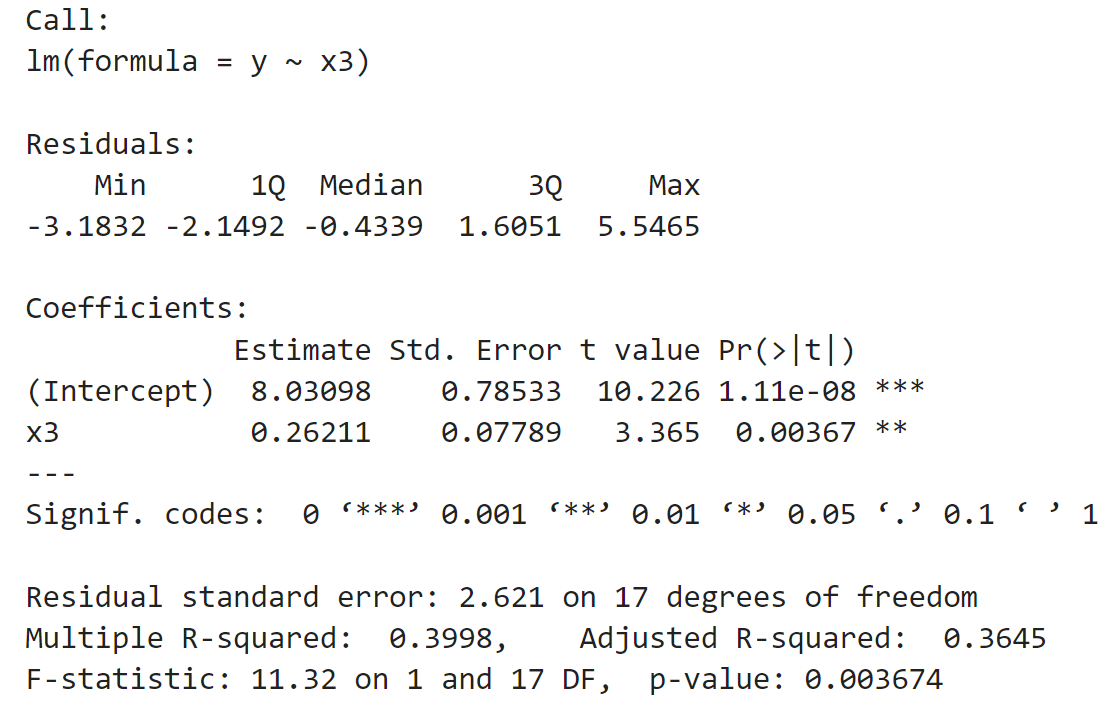
**Y对x2的一元回归**



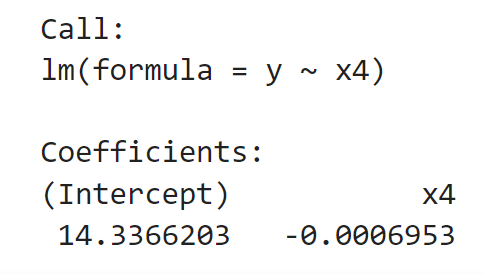


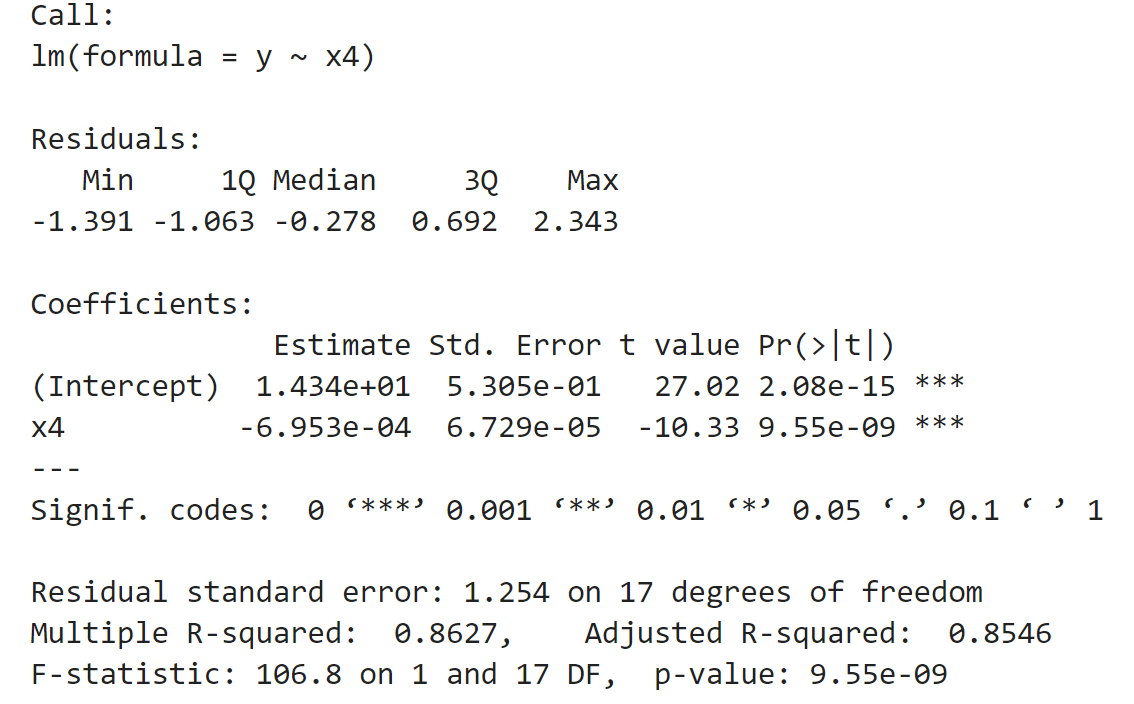
**Y对x3的一元回归**





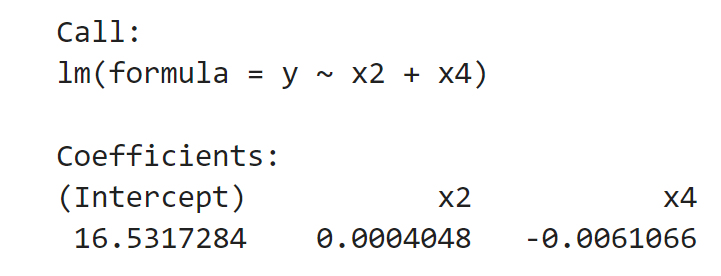
**Y对x4的一元回归**

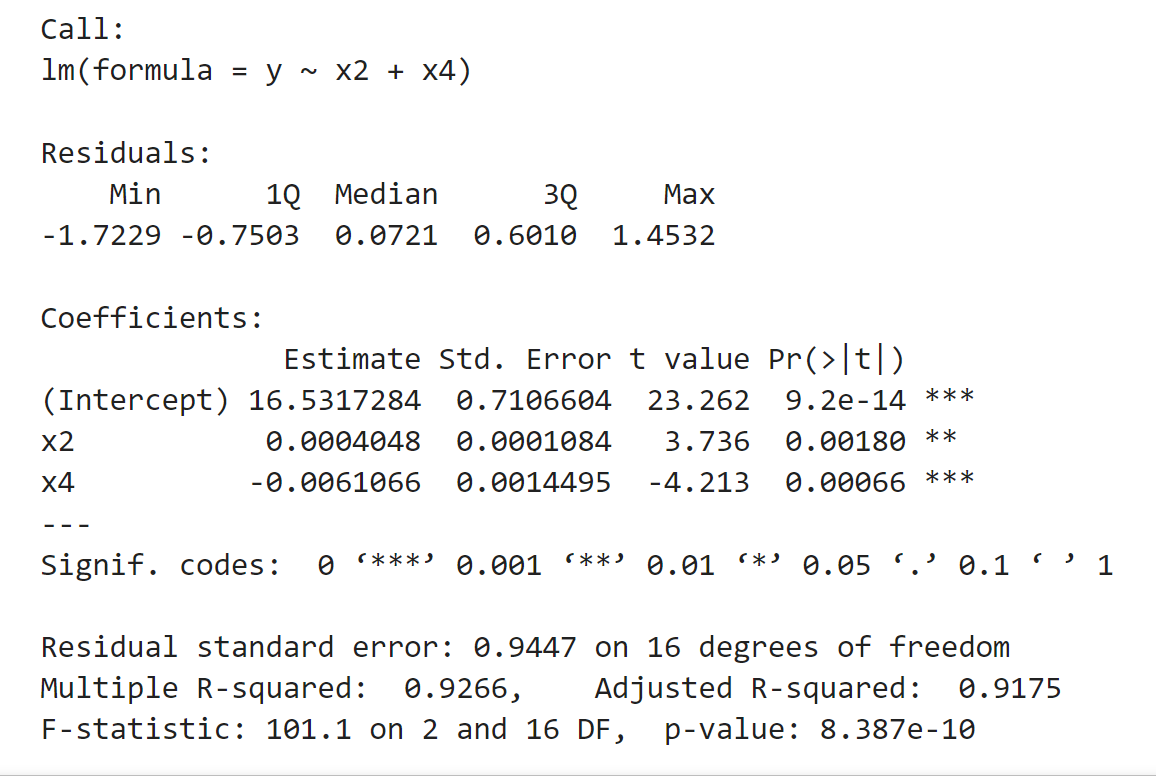




按R^2的大小排序为X4、X2、X3

以X2为基础，顺次加入其他变量逐步回归。首先加入X4的结果为：



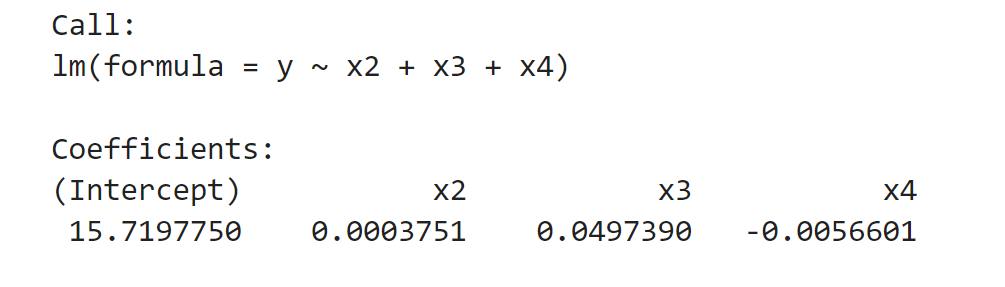


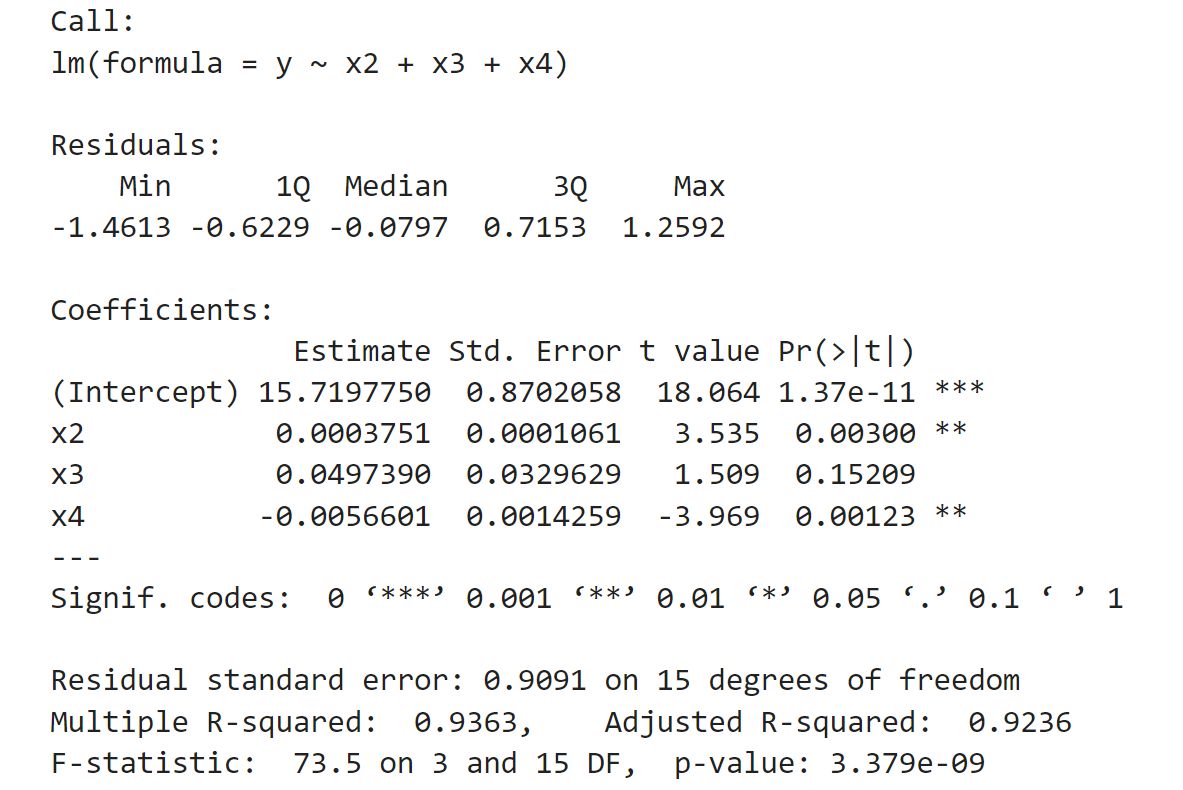
Y=16.5317284+020004048 X2 – 0.0061066 X4

T=（3.736） （-4.213）

R^2=0.9266

当取α=0.05时，tα/2(n-k)=t0.025(18-3)=2.131，X2参数的t检验显著，加入X3回归得





Y=15.7197750 + 0.0003751 X2 + 0.0497390 X3 – 0.0056601 X4

T=(18.064) (3.535) (1.509) (-3.969)

R^2=0.9363 A-R^2=0.9236

F = 73.5

当α=0.05时，tα/2(18-4)=2.145，X3的参数的t检验不显著，予以剔除

所以Y=16.5317284+020004048 X2 – 0.0061066 X4 时消除多重共线性之后的结果。

因此根据研究模型得到结果，在其他变量不变化的情况下，当国民总收入增长1亿元，人口增长率增长0.0003751%；在其他变量不变化的情况下，当人均GDP每增加1元，人口自然增长率就会下降0.0056601%。