北京航空航天大學
BEIHANG UNIVERSITY

国内生产总值指数与各行业生产总值指数的关系

姓名：张振宇

学号：BY1706104

班级：A 班

指导老师：韦卫

2018 年 1 月 12 日

目录

摘要

国内生产总值 GDP 是核算体系中一个重要的综合性统计指标，也是中国新国民经济核算体系中的核心指标，反映了一国（或地区）的经济实力和市场规模。本文基于 78 年以来中国近 40 年 GDP 与各行业的增值数据，采用主成分分析和多元线性回归的方法进行了统计分析，得出了较好的结论。

关键字：国内生产总值指数，主成分分析，多元线性回归，R

1. 引言

国内生产总值（以下或简称 GDP）是按市场价格计算的一个国家（或地区）所有常住单位在一定时期内生产活动的最终成果。国内生产总值有三种表现形态，即价值形态、收入形态和产品形态。从价值形态看，它是所有常住单位在一定时期内生产的全部货物和服务价值超过同期投入的全部非固定资产货物和服务价值的差额，即所有常住单位的增加值之和；从收入形态看，它是所有常住单位在一定时期内创造并分配给常住单位和非常住单位的初次收入之和；从产品形态看，它是所有常住单位在一定时期内所出产的最终使用的货物和服务价值减去货物和服务进口价值。在实际核算中，国内生产总值有三种计算方法，即生产法、收入法和支出法。三种方法分别从不同方面反映国内生产总值及其构成，理论上计算结果相同。

国内生产总值 GDP 是核算体系中一个重要的综合性统计指标，也是中国新国民经济核算体系中的核心指标，反映了一国（或地区）的经济实力和市场规模。一个国家或地区的经济究竟处于增长抑或衰退阶段，从这个数字的变化便可以观察到。一般而言，GDP 公布的形式不外乎两种，以总额和百分比率为计算单位。当 GDP 的增长数字处于正数时，即显示该地区经济处于扩张阶段；反之，如果处于负数，即表示该地区的经济进入衰退时期了。国内生产总值是指一定时间内所生产的商品与劳务的总量乘以“货币价格”或“市价”而得到的数字，即名义国内生产总值，而名义国内生产总值增长率等于实际国内生产总值增长率与通货膨胀率之和。因此，即使总产量没有增加，仅价格水平上升，名义国内生产总值仍然是会上升的。在价格上涨的情况下，国内生产总值的上升只是一种假象，有实质性影响的还是实际国内生产总值变化率，所以使用国内生产总值这个指标时，还必须通过 GDP 缩减指数，对名义国内生产总值做出调整，从而精确地反映产出的实际变动。因此，一个季度 GDP 缩减指数的增加，便足以表明当季的通货膨胀状况。如果 GDP 缩减指数大幅度地增加，便会对经济产生负面影响，同时也是货币供给紧缩、利率上升、进而外汇汇率上升的先兆。

国内生产总值是反映常住单位生产活动成果的指标。常住单位是指在一国经济领土内具有经济利益中心的经济单位。经济领土是指由一国政府控制或拥有的地理领土，也就是在本国的地理范围基础上，还应包括该国驻外使领馆、科研站和援助机构等，并相应地扣除外国驻本国的上述机构（国际机构不属于任何国家的常住单位，但其雇员则属于所在国家的常住居民）。经济利益中心是指某一单位或个人在一国经济领土内拥有一定活动场所，从事一定的生产和消费活动，并持续经营或居住一年以上的单位或个人，一个机构或个人只能有一个经济利益中心。一般就机构（单位）而言，不论其资产和管理归属哪个国家控制，只要符合上述标准，该机构在所在国就具有了经济利益中心。就个人而言，不论其国籍属于哪个国家，只要符合上述标准，该居民在所在国就具有经济利益中心。因为常住单位的概念严格地规定了一个国家的经济主体范围，所以其对于确定国内生产总值的计算口径，明确国内与国外的核算界限以及各种交易量的范围都具有重要意义。

本论文将基于改革开放以来近 40 年的 GDP 数据增值和行业的数据增值进行统计分析，得出相关结论。

2. 统计方法介绍

2.1. 主成分分析

主成分分析（principal component analysis, PCA）是一种降维技术，把多个变量化为能够反映原始变量大部分信息的少数几个主成分。

设 X 有 p 个变量，为 n×p 阶矩阵，即 n 个样本的 p 维向量。首先对 X 的 p 个变量寻找正规化线性组合，使它的方差达到最大，这个新的变量称为第一主成分，抽取第一主成分后，第二主成分的抽取方法与第一主成分一样，依次类推，直到各主成分累积方差达到总方差的一定阈值。

主成分分析的计算步骤：

假设样本观测数据矩阵为：X=(x1,x2,x3,...xp)，xi 为 n 个样本在第 i 个属性上的观测值，是一个列向量：

1. 对原始数据标准化处理（0 均值化处理）。
2. 计算样本相关系数矩阵。
3. 计算协方差矩阵的特征值和特征向量。
4. 选择重要的主成分，并写出主成分表达式。
5. 计算主成分得分。
6. 根据主成分得分的数据，做进一步的统计分析。

主成分分析可以得到 p 个主成分。但是，由于各个主成分的方差是递减的，包含的信息量也是递减的，所以实际分析时，一般不是选取 p 个主成分，而是根据各个主成分累计贡献率的大小选取前 k 个主成分，这里贡献率就是指某个主成分的方差占全部方差的比重，实际也就是某个特征值占全部特征值总和的比重。贡献率越大，说明该主成分所包含的原始变量的信息越强。主成分个数 k 的选取，主要根据主成分的累积贡献率来决定，即一般要求累计贡献率达到 85%以上，这样才能保证综合变量能包括原始变量的绝大多数信息。

此外，在实际应用中，选择了重要的主成分后，还要注意主成分实际含义解释。主成分分析中一个很关键的问题是如何给主成分赋予新的意义，给出合理的解释。一般而言，这个解释是根据主成分表达式的系数结合定性分析来进行的。主成分是原来变量的线性组合，在这个线性组合中个变量的系数有大有小，有正有负，有的大小相当，因而不能简单地认为这个主成分是某个原变量的属性的作用，线性组合中各变量系数的绝对值大者表明该主成分主要综合了绝对值大的变量，有几个变量系数大小相当时，应认为这一主成分是这几个变量的总和，这几个变量综合在一起应赋予怎样的实际意义，这要结合具体实际问题和专业，给出恰当的解释，进而才能达到深刻分析的目的。

2.2. 多元线性回归

社会经济现象的变化往往受到多个因素的影响，因此，一般要进行多元回归分析，我们把包括两个或两个以上自变量的回归称为多元线性回归。

多元线性回归的基本原理和基本计算过程与一元线性回归相同，但由于自变量个数多，计算相当麻烦，一般在实际中应用时都要借助统计软件。这里只介绍多元线性回归的一些基

本问题。

但由于各个自变量的单位可能不一样，比如说一个消费水平的关系式中，工资水平、受教育程度、职业、地区、家庭负担等等因素都会影响到消费水平，而这些影响因素（自变量）的单位显然是不同的，因此自变量前系数的大小并不能说明该因素的重要程度，更简单地来说，同样工资收入，如果用元为单位就比用百元为单位所得的回归系数要小，但是工资水平对消费的影响程度并没有变，所以得想办法将各个自变量化到统一的单位上来。具体而言，就是将所有变量包括因变量都先转化为标准分，再进行线性回归，此时得到的回归系数就能反映对应自变量的重要程度。这时的回归方程称为标准回归方程，回归系数称为标准回归系数，表示如下。

$$Y = X\beta + \varepsilon$$

3. 实验过程

3.1. 数据准备

3.1.1. 数据介绍

数据来自国家统计局国家数据官方网站（http://data.stats.gov.cn/easyquery.htm?cn=C01），主要包括自 1978 年以来国家生成总值增值以及各行业的增长数据，如下表所示。其中，数据以 excel 形式保存在本地。

数据库：年度数据										
变量编号	**y**	x1	x2	x3	x4	x5	x6	x7	x8	x9
时间	国内生产总值指数(上年=100)	农林牧渔业增加值指数(上年=100)	工业增加值指数(上年=100)	建筑业增加值指数(上年=100)	批发和零售业增加值指数(上年=100)	交通运输、仓储和邮政业增加值指数(上年=100)	住宿和餐饮业增加值指数(上年=100)	金融业增加值指数(上年=100)	房地产业增加值指数(上年=100)	其他行业增加值指数(上年=100)
1978年	111.7	104.1	116.4	99.5	123.1	108.9	118.1	110.1	105.7	111.2
1979年	107.6	106.1	108.7	102	108.7	108.3	111.1	98	104.1	110.1
1980年	107.8	98.5	112.6	126.6	98.1	104.3	103.9	107.3	107.9	114.8
1981年	105.1	107	101.7	103.2	129.5	101.9	117.5	104.7	96.5	107.4
1982	109	111.5	105.8	103.4	99.3	111.4	131.6	143.1	109.1	113.3

年										
1983年	110.8	108.3	109.7	117	121.2	109.5	119.4	126.5	105.2	112
1984年	115.2	112.9	114.8	110.8	124.7	114.9	108.1	130.7	127.7	115.3
1985年	113.4	101.8	118	122.1	133.5	113.8	106.3	117.1	125	111.5
1986年	108.9	103.3	109.6	115.8	109.4	113.9	115.6	130.2	125.9	103
1987年	111.7	104.7	113.1	117.8	114.7	109.6	109.7	122.6	129.3	110.3
1988年	111.2	102.5	115.1	108	111.8	112.5	125.1	120.2	112.7	109.1
1989年	104.2	103.1	105	91.6	89.3	104.2	109.9	125.8	115.9	104.8
1990年	103.9	107.3	103.4	101.2	94.7	108.3	103.5	102.2	106.2	103.7
1991年	109.3	102.4	114.3	109.6	105.2	110.6	108.2	102.8	112	115.4
1992年	114.2	104.7	121	121	110.5	110.1	127	106.5	126.6	111.5
1993年	113.9	104.7	120	118	108.6	112.5	108.2	111.3	110.8	116.7
1994年	113	104	118.8	113.6	108.2	108.5	127.1	109.7	112	112.6
1995年	111	105	114	112.4	108.2	111	110.2	108.8	112.4	110.3
1996年	109.9	105.1	112.5	108.5	107.6	111	106.8	107.9	104	112.7
1997年	109.2	103.5	111.3	102.6	108.8	109.2	110.9	109	104.1	115.8
1998年	107.8	103.5	108.9	109	106.5	110.6	111.1	105.1	107.7	109.6
1999年	107.7	102.8	108.6	104.3	108.7	112.2	107.7	105.4	105.9	111.4
2000年	108.5	102.4	109.9	105.7	109.4	108.6	109.3	107	107.1	113.1
2001年	108.3	102.8	108.7	106.8	109.1	108.8	107.6	107	111	112.9
2002年	109.1	102.9	110	108.8	108.8	107.1	112.1	107.5	109.9	113.7

2003年	110	102.5	112.8	112.1	109.9	106.1	112.4	107.4	109.8	110.8
2004年	110.1	106.3	111.6	108.2	106.6	114.5	112.3	104.7	105.9	112.7
2005年	111.4	105.2	111.6	116	113	111.2	112.3	114.1	112.2	112
2006年	112.7	105	112.9	117.2	119.5	110	112.6	123.7	115.5	110.8
2007年	114.2	103.7	114.9	116.2	120.2	111.8	109.6	125.8	124.4	111.5
2008年	109.7	105.4	110	109.5	115.9	107.3	109.6	112.1	101	111.1
2009年	109.4	104.2	109.1	118.9	111.9	103.4	103.8	116.4	111.8	108.5
2010年	110.6	104.3	112.6	113.8	114.6	109.5	108.2	108.9	107.5	108
2011年	109.5	104.3	110.9	109.7	112.5	109.7	105.1	107.7	107.4	109.6
2012年	107.9	104.5	108.1	109.8	110.3	106.1	106.5	109.4	104.7	108.1
2013年	107.8	104	107.7	109.7	110.5	106.6	103.9	110.6	107.2	107.5
2014年	107.3	104.2	107	109.1	109.7	106.5	105.8	109.9	102	108.5
2015年	106.9	104	106	106.8	106.1	104.1	106.2	116	103.2	109.3
2016年	106.7	103.5	106	106.6	106.7	106.5	106.9	105.7	108.6	109.3

3.1.2. 数据预处理

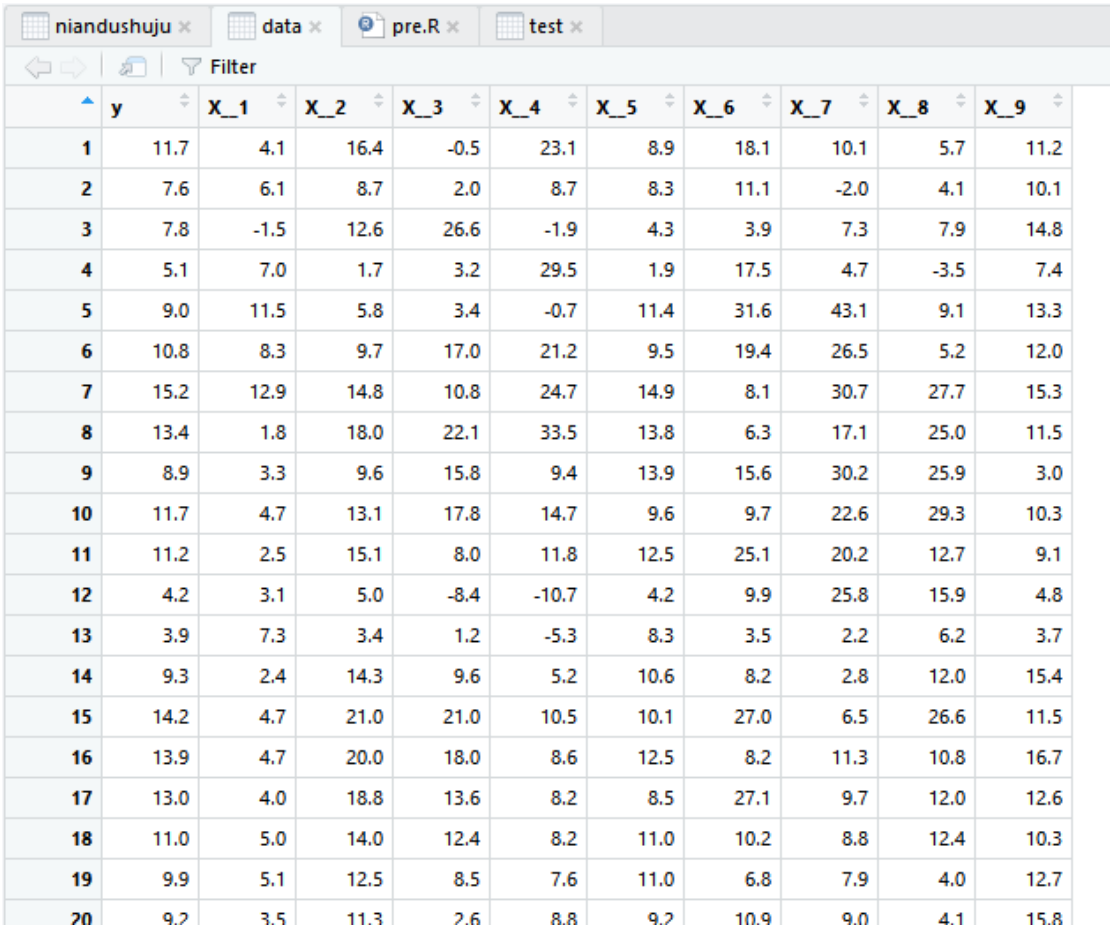
从 excel 中导入数据，并删除冗余信息，命令如下。

数据预处理：

```
library(readxl)
niandushuju <- read_excel("niandushuju.xls")
View(niandushuju)
data <- niandushuju[42,]
niandushuju <- niandushuju[-42,]
niandushuju <- niandushuju[-1,]
names(niandushuju)[1] <- "y"
```

```
xlabel <- niandushuju[1,]
niandushuju
data <- niandushuju[-1,]
View(data)
names(data)
names(data)[2:11] <- names(data[1:10])
names(data)[1] <- "year"
data <- data[-1]
data <- as.data.frame(data)
data <- as.matrix(data)
data=apply(data,2,as.numeric)
data <- data-100
data <- as.data.frame(data)
savehistory("D:/zhenyu/Desktop/test.Rhistory")
```

数据集中的数据局部：



niandushuju × | data × | pre.R × | test ×

Filter

	y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	11.7	4.1	16.4	-0.5	23.1	8.9	18.1	10.1	5.7	11.2
2	7.6	6.1	8.7	2.0	8.7	8.3	11.1	-2.0	4.1	10.1
3	7.8	-1.5	12.6	26.6	-1.9	4.3	3.9	7.3	7.9	14.8
4	5.1	7.0	1.7	3.2	29.5	1.9	17.5	4.7	-3.5	7.4
5	9.0	11.5	5.8	3.4	-0.7	11.4	31.6	43.1	9.1	13.3
6	10.8	8.3	9.7	17.0	21.2	9.5	19.4	26.5	5.2	12.0
7	15.2	12.9	14.8	10.8	24.7	14.9	8.1	30.7	27.7	15.3
8	13.4	1.8	18.0	22.1	33.5	13.8	6.3	17.1	25.0	11.5
9	8.9	3.3	9.6	15.8	9.4	13.9	15.6	30.2	25.9	3.0
10	11.7	4.7	13.1	17.8	14.7	9.6	9.7	22.6	29.3	10.3
11	11.2	2.5	15.1	8.0	11.8	12.5	25.1	20.2	12.7	9.1
12	4.2	3.1	5.0	-8.4	-10.7	4.2	9.9	25.8	15.9	4.8
13	3.9	7.3	3.4	1.2	-5.3	8.3	3.5	2.2	6.2	3.7
14	9.3	2.4	14.3	9.6	5.2	10.6	8.2	2.8	12.0	15.4
15	14.2	4.7	21.0	21.0	10.5	10.1	27.0	6.5	26.6	11.5
16	13.9	4.7	20.0	18.0	8.6	12.5	8.2	11.3	10.8	16.7
17	13.0	4.0	18.8	13.6	8.2	8.5	27.1	9.7	12.0	12.6
18	11.0	5.0	14.0	12.4	8.2	11.0	10.2	8.8	12.4	10.3
19	9.9	5.1	12.5	8.5	7.6	11.0	6.8	7.9	4.0	12.7
20	9.2	3.5	11.3	2.6	8.8	9.2	10.9	9.0	4.1	15.8

如图所示，整理好的数据存储在 data 这一 data.frame 数据结构中。

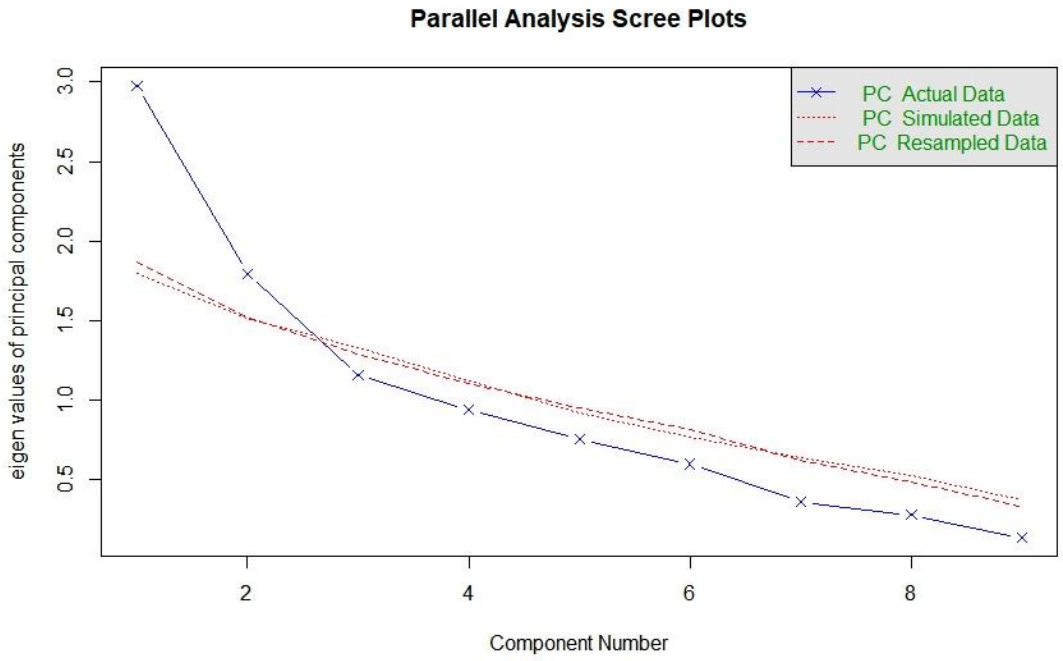
3.2. 主成分分析

3.2.1. 主成分个数的选取

利用处理好的数据，进行主成分分析。

首先确定主成分的个数。代码如下：

```
library(psych)
fa.parallel(data[,-1],fa="pc")
```



如上图所示，根据 Kaiser-Harris 准则（大于 1 的），应选取三个主成分；根据平行分析模拟的结果，应选择两个主成分。

接下来，首先选择两个主成分进行了分析，代码如下：

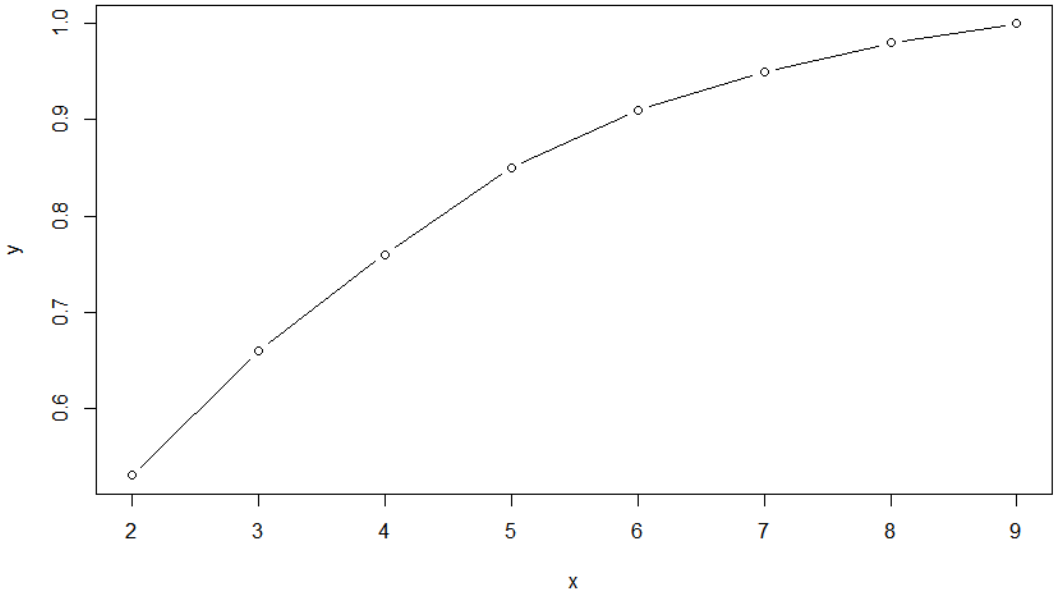
```
principal(data[,-1],nfactors=2,rotate = "none")
```

结果如图所示，即覆盖率仅为 53%，过低，因此需要选取更多的主成分。

```
> principal(data[,-1],nfactors=2,rotate = "none")
Principal Components Analysis
Call: principal(r = data[, -1], nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
      PC1   PC2   h2   u2 com
X__1 0.18  0.78 0.64 0.36 1.1
X__2 0.81 -0.41 0.83 0.17 1.5
X__3 0.63 -0.45 0.60 0.40 1.8
X__4 0.48 -0.02 0.24 0.76 1.0
X__5 0.72  0.13 0.54 0.46 1.1
X__6 0.38  0.49 0.38 0.62 1.9
X__7 0.47  0.66 0.66 0.34 1.8
X__8 0.74  0.06 0.55 0.45 1.0
X__9 0.47 -0.32 0.32 0.68 1.8

                       PC1  PC2
SS loadings           2.98 1.79
Proportion Var        0.33 0.20
Cumulative Var        0.33 0.53
Proportion Explained  0.62 0.38
Cumulative Proportion 0.62 1.00

Mean item complexity =  1.4
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.12
 with the empirical chi square  43.64  with prob <  0.0011

Fit based upon off diagonal values = 0.83
```

类似地，分别选取主成分个数如下表所示，得到覆盖率。

主成分个数	2	3	4	5	6	7	8	9
覆盖率	0.53	0.66	0.76	0.85	0.91	0.95	0.98	1



3.2.2. 分析与解释

考虑降维和效果的折衷，选五个主成分。参数结果如下图所示。

```
Principal Components Analysis
Call: principal(r = data[, -1], nfactors = 5, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
      PC1   PC2   PC3   PC4   PC5   h2    u2 com
X__1 0.18  0.78  0.28  0.29 -0.22 0.86 0.145 1.9
X__2 0.81 -0.41  0.16 -0.16  0.10 0.89 0.113 1.7
X__3 0.63 -0.45 -0.31  0.16  0.18 0.76 0.243 2.7
X__4 0.48 -0.02  0.03  0.80  0.17 0.91 0.087 1.8
X__5 0.72  0.13  0.15 -0.15 -0.47 0.81 0.187 2.0
X__6 0.38  0.49  0.32 -0.30  0.62 0.96 0.041 3.8
X__7 0.47  0.66 -0.34 -0.05  0.04 0.78 0.221 2.4
X__8 0.74  0.06 -0.52 -0.20 -0.12 0.88 0.124 2.0
X__9 0.47 -0.32  0.67 -0.05 -0.12 0.78 0.218 2.4

                       PC1  PC2  PC3  PC4  PC5
SS loadings           2.98 1.79 1.16 0.94 0.75
Proportion Var        0.33 0.20 0.13 0.10 0.08
Cumulative Var        0.33 0.53 0.66 0.76 0.85
Proportion Explained  0.39 0.24 0.15 0.12 0.10
Cumulative Proportion 0.39 0.63 0.78 0.90 1.00

Mean item complexity =  2.3
Test of the hypothesis that 5 components are sufficient.

The root mean square of the residuals (RMSR) is  0.07
 with the empirical chi square  13.21  with prob <  0.00028

Fit based upon off diagonal values = 0.95
```

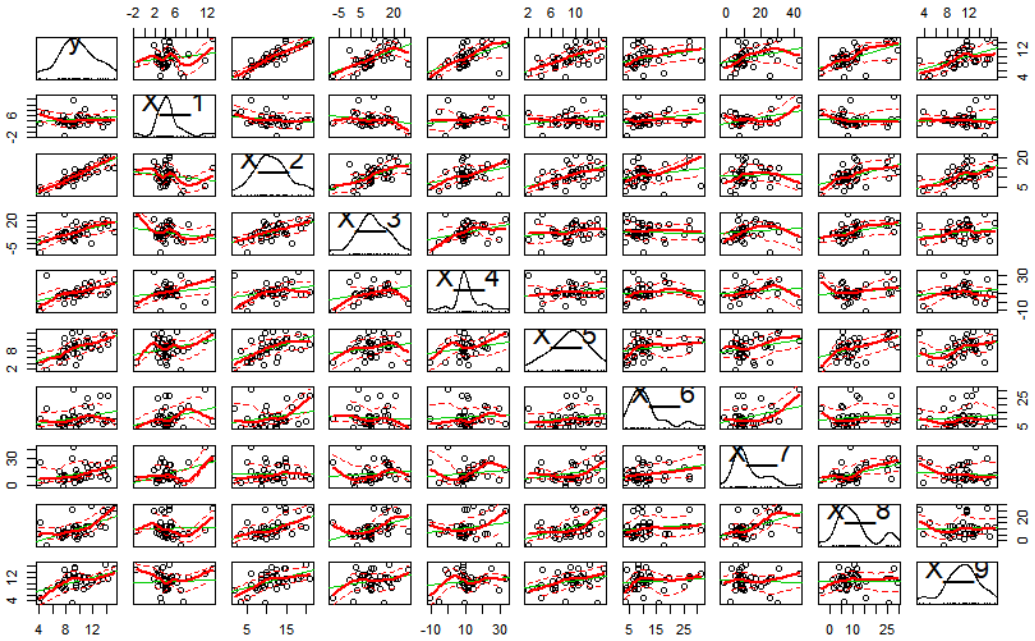
有以上结果，可以得出以下结论：第一，由于第一主成分和第二主成分体量相当，所以其指标具有同等地位的的指导意义，因此，x1（工业）和 x2（农业）此消彼长；第二，x8（房地产）在我国经济增长中扮演重要历史角色。

3.3. 多元线性回归

3.3.1. 散点图矩阵

绘制散点图矩阵，特别关注 y 变量与其他变量之间的线性关系。代码如下：

```
libraray(car)
scatterplotMatrix(data)
```



如图所示，可观察 x5，x6，x7 线性关系较弱，其他变量则线性关系较强。

3.3.2. 模型拟合

虽然 x5，x6，x7 线性关系较弱，此处仍然将其纳入模型，进行建模，期望通过后续数据分析来进行筛选。

```
fit <- lm(y~X__1+X__2+X__3+X__4+X__5+X__6+X__7+X__8+X__9,data)
summary(fit)
```

```
> summary(fit)

Call:
lm(formula = y ~ X__1 + X__2 + X__3 + X__4 + X__5 + X__6 + X__7 +
    X__8 + X__9, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.24583 -0.30913  0.07791  0.30892  1.30726

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.04449    0.41557   2.513 0.017765 *
X__1         0.20691    0.05071   4.080 0.000322 ***
X__2         0.43950    0.04448   9.881 8.65e-11 ***
X__3         0.03927    0.01875   2.095 0.045045 *
X__4         0.06363    0.01269   5.015 2.44e-05 ***
X__5         0.03351    0.04144   0.809 0.425237
X__6        -0.01082    0.01723  -0.628 0.535009
X__7         0.04637    0.01481   3.130 0.003961 **
X__8         0.01071    0.02019   0.531 0.599739
X__9         0.07508    0.04034   1.861 0.072888 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5823 on 29 degrees of freedom
Multiple R-squared:  0.9642,    Adjusted R-squared:  0.9531
F-statistic: 86.77 on 9 and 29 DF,  p-value: < 2.2e-16
```

由上图可知，整体回归效果显著，p 值无限小，拒绝原假设。对于当个回归系数，x1，x2，x4 效果特别显著，x7 效果显著，x3 和常数项也有较好的线性性质，在 α =0.5 时拒绝原假设。R^2 为 0.9642。

3.3.3. 模型调整

基于上小节结果，针对模型进行调整，删除未通过假设检验的项目。结果如下：

```
> fit <- lm(y~X__1+X__2+X__3+X__4+X__7,data)
> summary(fit)

Call:
lm(formula = y ~ X__1 + X__2 + X__3 + X__4 + X__7, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.08419 -0.34735  0.07057  0.26872  1.45273

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.50633    0.34697   4.341 0.000126 ***
X__1         0.22979    0.04702   4.887 2.57e-05 ***
X__2         0.48651    0.02732  17.806  < 2e-16 ***
X__3         0.04275    0.01812   2.359 0.024403 *
X__4         0.06161    0.01283   4.800 3.32e-05 ***
X__7         0.04669    0.01161   4.023 0.000315 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5952 on 33 degrees of freedom
Multiple R-squared:  0.9574,    Adjusted R-squared:  0.951
F-statistic: 148.4 on 5 and 33 DF,  p-value: < 2.2e-16
```

```
> fit <- lm(y~X__1+X__2+X__4+X__7,data)
> summary(fit)

Call:
lm(formula = y ~ X__1 + X__2 + X__4 + X__7, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.31087 -0.25230  0.06388  0.22310  1.41118

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.58721    0.36771   4.316 0.000130 ***
X__1         0.19894    0.04810   4.136 0.000219 ***
X__2         0.51689    0.02567  20.139  < 2e-16 ***
X__4         0.06905    0.01325   5.212 9.12e-06 ***
X__7         0.05306    0.01202   4.414 9.74e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6339 on 34 degrees of freedom
Multiple R-squared:  0.9502,    Adjusted R-squared:  0.9444
F-statistic: 162.3 on 4 and 34 DF,  p-value: < 2.2e-16
```

```
> fit <- lm(y~X__1+X__2+X__4,data)
> summary(fit)

Call:
lm(formula = y ~ X__1 + X__2 + X__4, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.62247 -0.32418 -0.01647  0.37316  2.12119

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.72982    0.45280   3.820 0.000523 ***
X__1         0.28838    0.05392   5.348 5.59e-06 ***
X__2         0.52749    0.03159  16.699  < 2e-16 ***
X__4         0.07042    0.01637   4.301 0.000130 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7836 on 35 degrees of freedom
Multiple R-squared:  0.9217,    Adjusted R-squared:  0.915
F-statistic: 137.4 on 3 and 35 DF,  p-value: < 2.2e-16
```

模型	全模型	未通过 α =0.5	未通过 α =0.01	只保留最显著
R^2	0.9642	0.9574	0.9502	0.9217

由上可知，国内生产总值增长基本有工业，农业和批发和零售业带动，受该三者的影响，金融对国内生产总值亦有有利影响。

4. 结论

主成分分析和多元线性回归分别从不同角度分析了国民生产总值的影响因素，两者的结

论有共同的地方，亦有个性的因子；两者均分析得到工业和农业对国家的影响，pca 更得出了房地产的结论，多元回归分析则发掘出金融对 GDP 的影响，取得了较好的结果，与直观认知相同。

参考文献

[1]孙海燕，周梦，李卫国，冯伟. 《数理统计》. 北京航空航天大学出版社，2014.10.

[2]Kabacoff,R. I.R 语言实战.北京:人民邮电出版社,2013.