**我国居民消费水平影响因素及其统计研究**

**姓名**

**编号 专业班级 学号**

**摘要**：我国国内消费水平发展较慢是一个待解决的问题。本文采用近22年的国内消费水平指数、商品房销售价格、GDP等数据，通过相关分析、主成分分析、回归分析方法，分析影响居民消费水平的因素，探究居民消费的内在动力。根据回归方程，我国社会发展状况、居民收入水平越高，我国居民消费水平越高。我国物价水平、居民贫困程度越低，我国居民的消费水平越高，根据此结论本文提出了建议。本文的研究结果对今后的我国居民消费水平影响因素研究及刺激我国消费水平增长有所帮助。

**关键词**：消费水平；相关分析；主成分分析；回归分析

1 **问题提出与分析**

我国经济增长依赖于投资、出口与消费这“三驾马车”,投资与出口稳步增长,但国内消费水平发展较慢,内需不足一直是我国经济的短板。[1]本文采用近22年的国内消费水平指数、商品房销售价格、GDP等数据，采用相关分析、主成分分析、回归分析方法，分析影响居民消费水平的因素，探究居民消费的内在动力。

研究我国居民消费水平影响因素，首先应该从政府网站获取数据，再通过相关分析、主成分分析、回归分析，建立模型，得到影响因素与消费水平的关系，并对模型进行检验，提出建议。

**2 数学模型、原理及算法**

**2.1 相关分析**

本文采用*Pearson*相关矩阵描述变量间的相关性。设是*p*维总体，获得*n*组样本数据

样本中第*j*维的均值为



第*j，k*维数据的协方差为（*j，k*相等时表示该维数据的方差）



第*j，k*维数据的相关系数



是无量纲的量，总有.

观测数据的*Pearson*相关矩阵为



其中越接近1，两变量之间的线性相关密切程度越高,反之越低。[2]

**2.2 主成分分析**

主成分分析是一种统计方法，通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量，转换后的变量叫主成分。步骤如下：

*Step1:*指标数据标准化。

*Step2:*原变量相关性检验。若通过相关性检验得出变量间相关度较大的结论,则原变量适合做主成分分析,否则不适合做主成分分析。

*Step3:*确定提取主成分的个数并提取主成分。确定提取主成分的个数的常用方法是选择特征值大于或等于1的主成分作为主成分，也可规定提取主成分个数。

*Step4:*确定主成分表达式。用原指标的线性组合表示主成分。

*Step5:*主成分命名，确定主成分的含义。

**2.3 回归分析**

**2.3.1 回归参数估计**

线性回归模型的矩阵形式为



其中Y称为观测变量，X为设计矩阵，并假定它是列满秩的。Y和X由样本数据决定，是已知的，是待估计的回归参数向量，是不可观测的随机误差向量。

的最小二乘估计即选择使误差项的平方和





达到最小，其中,为此,对的偏导数应为零，得



即



将此方程组写成矩阵形式，称为正规方程



由于，故存在，解正规方程即得回归系数的最小二乘估计为



将代入模型并略去误差项，得到经验回归方程



**2.3.2 回归分析的步骤**

一个完整的回归分析应包括以下步骤：

*Step1:*相关分析。分析变量与因变量之间是否存在相关性,若存在,则有继续做回归分析的意义。

*Step2:*线性回归模型建立。

*Step3:*模型检验。其中包括拟合优度检验、线性回归关系检验。

若拟合优度大，能够反映的因变量总变化量的比例大，拟合程度高，线性关系显著。

拟合优度的定义为



其中表示因变量的拟合值，表示因变量的平均值，表示因变量的观测值。

线性回归关系的假设检验方法如下：



根据离差平方和的分解式以及各项的意义，构造检验统计量



其中,MSR为均方回归,MSE为均方残差。若计算出F的值大于给定的值,则拒绝原假设,认为因变量与自变量的线性关系显著;否则认为不显著。

*Step4:*系数检验。以对自变量的系数作显著性检验为例，其他自变量检验过程相同，进行假设检验如下：



基于系数的最小二乘估计以及以下性质

(1)线性回归模型中各自变量系数服从正态分布；

(2)；

(3)与SSE相互独立。

构造检验统计量



其中是矩阵的主对角线上的第1个元素，是的标准差的估计，它是的主对角线上的第1个元素的算术平方根。若计算出*t*的绝对值大于给定的值，则拒绝原假设，认为自变量对因变量的影响显著。

*Step5:*多重共线性检验，检验解释变量间是否存在相关关系。多重共线性是指线性回归模型中的解释变量之间由于存在精确相关关系或高度相关关系而使模型估计失真或难以估计准确。

多重共线性可通过VIF或容差判断。VIF为方差膨胀主成分,是指解释变量之间存在多重共线性时的方差与不存在多重共线性时的方差之比,其对应共线性存在情况如表1所示。VIF的倒数是容差*tolerance,*其对应共线性存在情况如表2所示。

表1 VIF值对应共线性存在情况

|  |  |  |  |
| --- | --- | --- | --- |
| VIF | 0<VIF<10 | 10≤VIF<100 | VIF≥100 |
| 多重共线性 | 无 | 较强 | 严重 |

表2 容差对应共线性存在情况

|  |  |  |  |
| --- | --- | --- | --- |
| 容差 | >0.1 | 0.01≤*tolerance* ≤0.1 | ≤0.01 |
| 多重共线性 | 无 | 较强 | 严重 |

*Step6:*DW检验,误差的自相关性的假设检验：

构造检验统计量

其中*w*为误差项与因变量的相关系数；若D的值趋于4或趋于0，则拒绝原假设，若D的值趋于2，则接受原假设。

*Step7:*残差分析。分为残差正态性检验与残差图分析。

残差正态性检验首先计算学生化残差。

若假设误差向量,则残差向量，其中，因此



其中为H的主对角线上的第*i*个元素。

，即一般情况下，的方差不等。将标准化，再以代替，则得到所谓的学生化残差



残差正态性的频率检验：

残差正态性检验的基本思想是将学生化残差落在一定范围内的频率与标准正态分布在相应范围内的概率作比较，若二者差异较大，则认为残差不服从正态分布，否则无理由拒绝误差项独立同正态分布的假设。

残差图是指以残差为纵坐标，以任何其他有关量的值为横坐标的散点图，本文将横坐标取为因变量的拟合值。若线性回归关系正确且误差向量服从正态分布，则因变量的拟合值与残差向量相互独立，这时残差图中的点应该大致分布在一个水平的带状区域内，且不呈现任何明显的趋势。

**3 实验数据收集及预处理**

**3.1 数据来源**

从*eps*数据库与国家统计局选取了近22年 (1997—2018年)的全国居民消费水平指数、商品房价格、GDP、全国总人口数、居民储蓄存款余额、商品零售价格指数、居民消费价格指数、固定资产投资价格指数、城镇居民家庭可支配收入、城镇家庭恩格尔系数、农村家庭恩格尔系数、货币供应量（M2）、就业人员平均实际工资指数、年末城镇登记失业率的数据。

商品房价格、商品零售价格指数、居民消费价格指数、固定资产投资价格指数直接反应物价水平,GDP是国民经济核算的一项核心指标,居民储蓄存款余额、城镇居民家庭可支配收入、城镇家庭恩格尔系数、农村家庭恩格尔系数、货币供应量（M2）、就业人员平均实际工资指数、年末城镇登记失业率直接反应我国居民的购买力水平。

**3.2 数据处理**

由于各个指标间量级差异大，因此先对数据进行标准化处理，处理后的数据量级相同，便于后续分析。

**4 基于统计方法的数据分析**

**4.1相关分析**

13个指标的*Pearson*相关矩阵如表3所示。

表3 各指标间的*Pearson*相关矩阵

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 1 | 0.99 | 0.94 | 0.99 | 0.73 | 0.67 | 0.42 | 0.99 | -0.66 | -0.89 | 0.99 | -0.28 | 0.58 |
|  |  | 1 | 0.92 | 1.00 | 0.72 | 0.67 | 0.43 | 1.00 | -0.64 | -0.88 | 1.00 | -0.29 | 0.56 |
|  |  |  | 1 | 0.93 | 0.76 | 0.69 | 0.54 | 0.94 | -0.86 | -0.98 | 0.91 | -0.03 | 0.80 |
|  |  |  |  | 1 | 0.70 | 0.64 | 0.40 | 1.00 | -0.65 | -0.89 | 1.00 | -0.28 | 0.57 |
|  |  |  |  |  | 1 | 0.98 | 0.88 | 0.73 | -0.47 | -0.67 | 0.68 | -0.35 | 0.66 |
|  |  |  |  |  |  | 1 | 0.91 | 0.68 | -0.40 | -0.59 | 0.62 | -0.37 | 0.58 |
|  |  |  |  |  |  |  | 1 | 0.44 | -0.35 | -0.43 | 0.38 | -0.16 | 0.58 |
|  |  |  |  |  |  |  |  | 1 | -0.67 | -0.90 | 1.00 | -0.26 | 0.59 |
|  |  |  |  |  |  |  |  |  | 1 | 0.91 | -0.63 | -0.37 | -0.84 |
|  |  |  |  |  |  |  |  |  |  | 1 | -0.88 | -0.06 | -0.81 |
|  |  |  |  |  |  |  |  |  |  |  | 1 | -0.29 | 0.54 |
|  |  |  |  |  |  |  |  |  |  |  |  | 1 | 0.18 |
|  |  |  |  |  |  |  |  |  |  |  |  |  | 1 |

其中，表示商品房价格，表示GDP，表示全国总人口数，表示居民储蓄存款余额，表示商品零售价格指数，表示居民消费价格指数，表示固定资产投资价格指数，表示城镇居民家庭可支配收入，表示城镇家庭恩格尔系数，表示农村家庭恩格尔系数，表示货币供应量(M2),表示就业人员平均实际工资指数，表示年末城镇登记失业率。

通过表3可知，指标间相关程度较大，且指标数较多，因此不适合直接用原指标做回归分析。

**4.2 主成分分析**

通过相关分析知，原指标间的相关程度较大，因此适合做主成分分析。

本文首先用主成分法提取了五个主成分，得到其累积特征值方差百分比为99.4%，主成分如下：



下对五个主成分的含义作出解释。

第一主成分在商品房价格、GDP、全国总人口、居民储蓄存款余额、城镇居民家庭可支配收入、农村居民家庭恩格尔系数、货币供应量（M2）占比较大，因此其含义是我国社会发展状况。

第二主成分在城镇家庭恩格尔系数、就业人员平均实际工资指数、居民消费价格指数、商品零售价格指数、固定资产投资价格指数占比较大，因此其含义是我国物价水平。

第三主成分在固定资产投资价格指数、居民消费价格指数、商品零售价格指数、居民储蓄存款余额、年末城镇登记失业率占比较大，因此其含义是我国居民的生活成本。

第四主成分在就业人员平均实际工资指数、年末城镇登记失业率、固定资产投资价格指数、城镇家庭恩格尔系数占比较大，因此其含义是我国居民收入水平。

第五主成分在城镇家庭恩格尔系数、农村家庭恩格尔系数、年末城镇登记失业率占比较大，因此其含义是我国居民贫困程度。

**4.3 回归分析**

**4.3.1 相关分析**

我国消费水平指数*y*与五个主成分的*Pearson*相关矩阵如表4所示。

表4 *Pearson*相关矩阵

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *y* | *f1* | *f2* | *f3* | *f4* | *f5* |
| *y* | 1 | 0.67 | -0.42 | 0.21 | 0.42 | -0.31 |
| *f1* |  | 1 | 0 | 0 | 0 | 0 |
| *f2* |  |  | 1 | 0 | 0 | 0 |
| *f3* |  |  |  | 1 | 0 | 0 |
| *f4* |  |  |  |  | 1 | 0 |
| *f5* |  |  |  |  |  | 1 |

由表4知，我国消费水平指数*y*与五个主成分间相关性较大，因此适合做回归分析。

**4.3.2 最小二乘法估计回归参数**

以通过主成分分析提取得到的主成分作为自变量，我国居民消费水平指数（以1997年为100）的标准化后的数据作为因变量，建立回归模型。

通过最小二乘法，得到回归参数如表5所示。

表5 回归参数

|  |  |
| --- | --- |
| 自变量 | 回归参数 |
| 常量 |  |
| 第一主成分*f1* | 0.665 |
| 第二主成分*f2* | -0.418 |
| 第三主成分*f3* | 0.212 |
| 第四主成分*f4* | 0.420 |
| 第五主成分*f5* | -0.312 |

**4.3.3 模型检验**

根据SPSS的计算，模型拟合优度为93.6%，拟合度较高，提取出的主成分能较大程度地描述我国居民消费水平的变化。

计算出的模型的F值及其显著性如表6所示。

表6 ANOVA计算结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 平方和 | 自由度 | 均方 | F值 | 显著性 |
| Sr | 14.971 | 5 | 2.994 | 31.999 | 0.000 |
| Se | 1.029 | 11 | 0.094 |  |  |
| 总计 | 16.000 | 16 |  |  |  |

根据表6，模型的显著性为0.000，因此拒绝原假设，认为我国居民消费水平关于各主成分存在显著的线性回归关系。

**4.3.4 系数检验**

通过SPSS计算回归参数的显著性如表7所示。

根据表7，回归参数在显著水平下，除常数项外，其余各回归参数均不显著为零。

表7 回归参数的显著性

|  |  |
| --- | --- |
| 自变量 | 显著性 |
| 常量 | 1.000 |
| 第一主成分*f1* | 0.000 |
| 第二主成分*f2* | 0.000 |
| 第三主成分*f3* | 0.018 |
| 第四主成分*f4* | 0.000 |
| 第五主成分*f5* | 0.002 |

**4.3.5 多重共线性检验**

由于通过主成分提取后，各主成分间不存在相关关系，因此不存在多重共线性，不再进行多重共线性检验。

**4.3.6 DW检验**

根据SPSS的计算，检验统计量D=1.530，较接近2，误差项间不存在明显的自相关关系。

**4.3.7 残差分析**

将学生化残差的实测累积概率与正态分布的累积概率（即预期累积概率）对比，作出学生化残差的正态P-P图如图1所示。



图1 学生化残差的正态P-P图

由图1可知，残差可能有某种规律性。

以横坐标为因变量的拟合值作出残差图。如图2所示。



图2 横坐标为因变量的拟合值的残差图

由图2可知，因变量的拟合值与残差向量不呈现明显的趋势，因变量拟合值与残差向量不存在明显的相关关系。

综合以上分析，本文建立的回归方程



具有一定的实际意义。

**5 结果分析**

根据回归方程，我国社会发展状况、我国居民生活成本、我国居民收入水平越高，我国居民消费水平越高，我国物价水平、居民贫困程度越低，我国居民的消费水平越高。这一结论也符合常识。

线性回归模型在拟合优度、模型检验、系数检验、多重共线性检验都表现较好，但根据残差分析，残差可能仍然有某些规律性。这是由于消费水平由很多因素决定，很难做到将影响消费水平的全部指标都包含进本文的研究中，因此今后的研究可在本文的基础上，加以考虑其他因素，例如交通的发展可能导致旅游消费增加。

**6 建议**

第一，提高我国综合国力。“国家”，先有国，后有家，只有国家富强起来了，人民才能获得生活的保障，才能使消费环境稳定。

第二, 提高居民的人均可支配收入, 包括改善工资薪酬制度, 以及适度提高最低工资水平，同时促进居民工资的稳步提高。鼓励居民获取工资以外的其他收入, 以增加居民的总体收入, 可以在一定程度上改善居民消费水平。[3]

第三, 政府应时刻关注物价水平对居民消费的影响程度, 调控物价指数,以防物价过高对居民的消费水平造成严重的影响, 降低居民消费水平。

第四, 坚决打赢脱贫攻坚战，坚持精准扶贫、精准脱贫。强化党政一把手负总责的责任制，坚持大扶贫格局，重点攻克深度贫困地区脱贫任务，增强国民购买力。

**参考文献**

[1] 俞琴.居民消费水平影响因素分析[J].当代经济,2017(19):151-153.

[2] 梅长林，范金城. 数据分析方法[M].第一版.北京:高等教育出版社，2006.

[3] 严玲.关于居民消费水平影响因素分析[J].商,2016(02):108.

[4] eps 数据库 http://olap.epsnet.com.cn/ 2018.12.1.

[5] 国家统计局 http://data.stats.gov.cn/ 2018.12.1