

**数理统计大作业（一）**

**国民幸福感线性回归分析**

|  |  |
| --- | --- |
| 学 院： | 计算机学院 |
| 学 号： | ZY2006109 |
| 姓 名： | 姬轶 |
| 班 号： | 23班 |
| 序 号： | 172 |

北京航空航天大学

**2020 年12月**

摘 要

提升国民幸福感日渐成为我国的一个主要目标，那么，哪些因素可以作为评判一个国家国民幸福感的指标呢？

本文以 19 年的世界幸福报告做分析，探究国民幸福感与国家各项数据之间的关系。我们以官方给出的国民幸福感作为因变量；探究其与人均GDP、社会支持度、健康预期寿命、生活选择自由度、国家腐败程度 5 项指标的相对值之间的线性相关性，并将以上指标作为自变量。

本文利用了统计软件 SPSS ，使用逐步回归法进行线性回归，最终筛选出了与因变量相关性最大的几个因素，并确定了国民幸福感与这些因素之间的最优回归方程。

关键词：国民幸福感，多元线性回归，逐步回归法，SPSS

目录

[一、 引言 1](#_Toc59728206)

[1.1. 研究背景 1](#_Toc59728207)

[1.2. 研究内容 1](#_Toc59728208)

[二、 多元线性回归分析 1](#_Toc59728209)

[2.1. 多元线性回归 1](#_Toc59728210)

[2.2. 变量与数据集 2](#_Toc59728211)

[2.3. 步进条件 3](#_Toc59728212)

[三、 问题建模 3](#_Toc59728213)

[3.1. 逐步回归法 3](#_Toc59728214)

[3.2. 回归模型 6](#_Toc59728215)

[四、 结论与展望 6](#_Toc59728216)

[参考文献 7](#_Toc59728217)

# 引言

## 研究背景

《世界幸福报告》是对全球幸福状况的里程碑式调查。第一份报告于2012年发布，第二份报告于2013年发布，第三份报告于2015年发布，第四份报告于2016年更新。3月20日，在庆祝国际幸福日的活动上，联合国发布了2017年世界幸福度，按幸福度排名155个国家。随着政府，组织和公民社会越来越多地使用幸福指标来指导其决策，该报告继续获得全球认可。经济，心理学，调查分析，国家统计，卫生，公共政策等领域的领先专家描述了如何有效地使用幸福感测度来评估国家的进步。

幸福分数和排名使用盖洛普世界民意调查的数据。分数基于对民意调查提出的主要生活评估问题的答案[1]。这个问题称为Cantril阶梯，要求被调查者考虑一个阶梯，该阶梯的最长寿命对他们来说是10，而最糟糕的寿命是0，并以此等级来评价他们当前的寿命。分数来自2013-2016年全国代表性样本，并使用盖洛普权重使估算具有代表性。幸福分数后面的各列估算了五个因素（经济生产，社会支持，预期寿命，自由度和腐败度）中每个因素在何种程度上有助于使每个国家的生活评估高于反乌托邦，一个假设的国家，其五个因素的价值均等于世界最低的国家平均值。它们对每个国家/地区的总分没有影响，但是确实解释了为什么某些国家的排名高于其他国家。

## 研究内容

本文的研究内容是建立合适的模型，对国民幸福度进行研究。国民幸福度受多种因素影响，本文使用多元线性回归模型对国民幸福度进行建模。本文通过逐步回归法，以较小的代价将最合适的变量引入模型，从而构建出最优回归方程。

# 多元线性回归分析

## 多元线性回归

因为印象国民幸福感的因素远不止一个，若使用一元线性回归进行分析会有非常大的局限性，且拟合出来的模型无法反映多方面的影响，所以需要使用多元线性回归分析，多元线性回归，即对于因变量 与 n 个普通自变量有关，满足：

其中是与 无关的未知参数，称为回归系数，而称为回归因子， 是不可观测的随机变量， 为经过逐步回归法后得到的模型中变量个数[2]。

## 变量与数据集

《世界幸福报告》是对全球幸福状况的里程碑式调查。随着政府，组织和公民社会越来越多地使用幸福指标来指导其决策，该报告继续获得全球认可。经济，心理学，调查分析，国家统计，卫生，公共政策等领域的领先专家描述了如何有效地使用幸福感测度来评估国家的进步，因此，选取人均GDP、社会支持度、健康预期寿命、生活选择自由度、国家腐败程度 5 各因素作为考虑。

表2-1是2019年《世界幸福报告》中相应得分与因素评估。

表2-1 2019年全球幸福分数数据节选（前20名）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 国家 | 幸福感指数 | 人均国内生产总值差值 | 社会支持度差值 | 健康预期寿命差值 | 做生活选择的自由度差值 | 腐败程度差值 |
| Finland | 7.769 | 1.34 | 1.587 | 0.986 | 0.596 | 0.393 |
| Denmark | 7.6 | 1.383 | 1.573 | 0.996 | 0.592 | 0.41 |
| Norway | 7.554 | 1.488 | 1.582 | 1.028 | 0.603 | 0.341 |
| Iceland | 7.494 | 1.38 | 1.624 | 1.026 | 0.591 | 0.118 |
| Netherlands | 7.488 | 1.396 | 1.522 | 0.999 | 0.557 | 0.298 |
| Switzerland | 7.48 | 1.452 | 1.526 | 1.052 | 0.572 | 0.343 |
| Sweden | 7.343 | 1.387 | 1.487 | 1.009 | 0.574 | 0.373 |
| New Zealand | 7.307 | 1.303 | 1.557 | 1.026 | 0.585 | 0.38 |
| Canada | 7.278 | 1.365 | 1.505 | 1.039 | 0.584 | 0.308 |
| Austria | 7.246 | 1.376 | 1.475 | 1.016 | 0.532 | 0.226 |
| Australia | 7.228 | 1.372 | 1.548 | 1.036 | 0.557 | 0.29 |
| Costa Rica | 7.167 | 1.034 | 1.441 | 0.963 | 0.558 | 0.093 |
| Israel | 7.139 | 1.276 | 1.455 | 1.029 | 0.371 | 0.082 |
| Luxembourg | 7.09 | 1.609 | 1.479 | 1.012 | 0.526 | 0.316 |
| United Kingdom | 7.054 | 1.333 | 1.538 | 0.996 | 0.45 | 0.278 |
| Ireland | 7.021 | 1.499 | 1.553 | 0.999 | 0.516 | 0.31 |
| Germany | 6.985 | 1.373 | 1.454 | 0.987 | 0.495 | 0.265 |
| Belgium | 6.923 | 1.356 | 1.504 | 0.986 | 0.473 | 0.21 |
| United States | 6.892 | 1.433 | 1.457 | 0.874 | 0.454 | 0.128 |
| Czech Republic | 6.852 | 1.269 | 1.487 | 0.92 | 0.457 | 0.036 |

由此，可以为变量定义符号，如下表所示：

表2-2 各变量说明

|  |  |
| --- | --- |
| 变量名 | 符号 |
| 幸福度 |  |
| 人均GDP |  |
| 社会支持度 |  |
| 健康预期寿命 |  |
| 生活选择自由度 |  |
| 国家腐败程度 |  |

## 步进条件

本次回归分析中因变量 为国民幸福分数，变量集为表2-2中，输入方法使用步进，逐步回归法中使用的和取 SPSS 的默认值 。

# 问题建模

## 逐步回归法

逐步回归法是从一个自变量开始，视自变量对 的显著程度，从大到小地依次逐个引入回归方程，并且根据后续引入变量而对之前的变量显著性的影响儿决定是否将其除去。每一步引入新的变量都要进行 F 值检验。这个过程反复进行至所有变量都引入模型为止，如图3-1所示。



图3-1 逐步回归法引入变量过程

图3-2为模型的摘要信息，可以看到通过逐步回归法每引入一个变量，模型的 R 方都会增大，标准差都会减小。



图3-2 模型摘要

模型的方差分析如图 3-3 所示，因为所有模型的显著性检验值均小于 0.001，故可以认为每个模型引入的变量均对因变量有显著的影响；此外可以看出模型每引入一个变量，残差的均方误差均减小



图3-3 方差分析

相关的模型系数如图 3-4 所示：



图3-4 模型的回归系数

图 3-5 中的的回归标准化残差中，可以看出所有的观测残差点均匀的落在预测残差直线的两侧，可以得出残差的正态性假设基本成立。

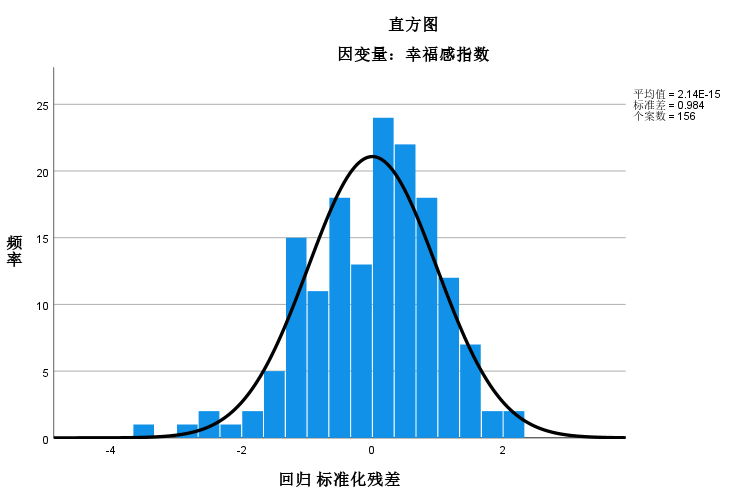


图 3-5 幸福感指数的回归标准化残差

## 回归模型

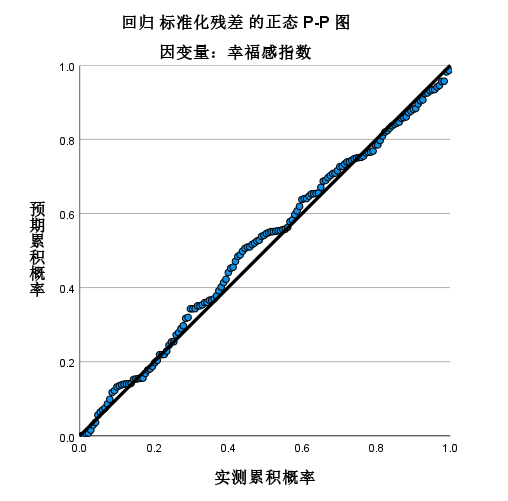


图 3-6 幸福感指数的正态 P-P 图

根据图 3-6 可得，国民幸福感的回归方程为：

# 结论与展望

根据 3.2 节中所述的模型的不足之处，可以细化数据的统计粒度，如按月统计相应的数据，从而提高样本数量，更好的反映各变量对国民幸福感的影响程度的动态变化。或者可以根据年份引入修正因子，手动将这种动态变化引起的偏差控制在一个合理的范围。

# 参考文献

[1] 《2019年全球幸福报告》，2019

[2] 孙海燕，周梦，李卫国，冯伟. 数理统计[M]. 北京：北京航空航天大学应用数学与系统 科学学院，2015: P160-P171