死亡率预测

模型选择：

本文采用了经典的Lee\_Carter模型对人口的死亡率进行建模并预测，该模型是由Lee 和Carter在1992年提出同时考虑了年龄因素与时间因素对死亡率的影响，其具体表达式如下：

其中为x年龄组，t时间内的中心死亡率，为时间相关变量反映了时间对死亡率的影响， 、为和年龄组相关的变量，反映了不同年龄组对死亡率的影响，反映了不同年龄组对时间变化的敏感程度，为随机误差项。为了保证模型有唯一解，引入了两个约束条件：

本文采用采用极大似然估计的方法对模型中三个参数进行估计，可以认为某时间、某年龄组的死亡人数满足泊松分布：

其中为t时刻、x年龄组的暴露人口，泊松分布参数。

由此可得在任意时刻、任意年龄组的Lee\_Carter模型的极大似然函数为：

其中C为任意常数。通过参数迭代的方法极大化似然函数就可得到各个参数的估计值。

本文采用了联合国1986-2020年的死亡率数据对中国从0至100+年龄段的数据对死亡率进行估计，得到的各个参数的估计值随年龄组或时间的变化如下图所示：

图表

描述已自动生成

从图中可以看出参数的变化与真实死亡率随年龄先上升后下降的趋势是一致的，参数的值先急剧下降到10岁左右的达到最低点，然后随年龄缓慢上升。参数随年龄的变化则较为剧烈，在0到40岁左右有一个剧烈的摆动期，60至70岁左右缓慢上升，70岁左右后伴随小幅波动下降。而参数随时间的变化则整体呈下降趋势，存在一定的随机扰动，这与建国以来随着经济社会的发展死亡率逐渐下降的趋势吻合。

下图展示了模型拟合的不同时刻不同年龄组的残差分布情况，可以看出整体而言较高年龄段的死亡率拟合程度较好，而0至30岁左右的年龄组的拟合度波动较大，这可能与参数在0到40岁左右的剧烈摆动有关，对比现有文献的结果较少出现此类剧烈波动的现象，说明模型仍有很大的进步空间。

图表, 散点图

描述已自动生成

*图表, 直方图

描述已自动生成图表, 直方图

描述已自动生成*

以上两幅图展示了模型拟合的1986-2020年间女性不同年龄组死亡率的变化与预测的未来50年女性不同年龄组死亡率的变化，下图则展示了对k值的预测及其置信区间：

图表

描述已自动生成

数据包络分析

数据包络分析是针对具有多指标输入和多指标输出系统的较为有效的评价方法。本文希望通过将南京市2010到2021年间政府对社会福利、教育、医疗卫生等民生相关领域的投入作为输入变量，将出生率、新增结婚人数、人口自然增长率作为输出变量，每一年度作为一个决策单元（Decision Making Units，DMU）对南京市目前的民生政策对人口变化的影响进行评价。

本文采用的市数据包络分析中的BCC模型，设有n个DMU，每个DMU有m种输入和s种输出，表示第j个DMU的第i种输入量，表示第j个DMU的第r 个输出量，表示第i种投入的权值，表示第r种输出的权值。

令，， ，

则第j个DMU的效率评价指数为：

通过Charnes – Cooper变换：，可以得到如下线性规划问题：

经对偶变换后得到：

求解该模型得到如下效益分析表：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 决策单元 | 技术效益 | 规模效益 | 综合效益 | 松弛变量S- | 松弛变量S+ | 有效性 |
| 2010 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | DEA强有效 |
| 2011 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | DEA强有效 |
| 2012 | 1.000 | 0.921 | 0.921 | 6175.139 | 5276.584 | 非DEA有效 |
| 2013 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | DEA强有效 |
| 2014 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | DEA强有效 |
| 2015 | 0.828 | 0.954 | 0.790 | 1434.408 | 1.069 | 非DEA有效 |
| 2016 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | DEA强有效 |
| 2017 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | DEA强有效 |
| 2018 | 0.867 | 0.999 | 0.866 | 62458.775 | 1.992 | 非DEA有效 |
| 2019 | 0.765 | 0.995 | 0.762 | 85108.170 | 1.994 | 非DEA有效 |
| 2020 | 0.718 | 0.921 | 0.662 | 59523.601 | 3.175 | 非DEA有效 |
| 2021 | 0.845 | 0.733 | 0.619 | 65976.585 | 4.412 | 非DEA有效 |

其中，综合技术效益(overall efficiency, OE)反映的是决策单元在一定（最优规模时）投入要素的生产效率，是对决策单元的资源配置能力、资源使用效率等多方面能力的综合衡量与评价，值等于1时，代表该决策单元的投入与产出结构合理，相对效益最优；值大于1时，代表该决策单元的投入与产出结构处于超级效益模式；值小于1时，代表该决策单元的投入与产出结构不合理，相对效益未能达到最优，可能存在不同程度的投入冗余和产出不足，其值为技术效益\*规模效益。

技术效益(technical efficiency, TE)反映的是由于管理和技术等因素影响的生产效率，其值等于1时，代表投入要素得到了充分利用，在给定投入组合的情况下，实现了产出最大化。

规模效益(scale efficiency, SE)反映的是由于规模因素影响的生产效率，通常结合规模报酬表进行分析其值等于1时，代表规模效率有效（规模报酬不变），也就是规模适宜，已达到最优的状态；若规模报酬递增（并非其值递增递减或者小于0大于0），代表服务规模过小，需要扩大规模以增加规模效益；若规模报酬递减（并非其值递增递减或者小于0大于0），代表服务规模过大，存在规模过度扩张风险。

松驰变量S-（差额变数）指为达到目标效率可以减少的投入量，即非 DEA 有效单元的实际值和目标值之差，松驰变量S+（超额变数）指为达到目标效率可以增加的产出量，即非 DEA 有效地区的目标值和实际值之差。

有效性分析结合综合效益指标，S-和S+共3个指标，可判断DEA有效性，如果综合效益=1且S-与S+均为0，则“DEA强有效”，如果综合效益为1但S-或S+大于0，则“DEA弱有效”，如果综合效益<1则为“非DEA有效”。

图表

描述已自动生成

可以看出2015年以前政策的效益持续处于较高水平，2015年突然下降这与2015年全面二孩政策实施的时间点相吻合，可能的原因是二孩政策导致的生育率的突然上升与相对连贯的民生政策产生了矛盾导致政策的综合效益下降。而经过2016、2017年的短暂回升后，综合效益持续下降。结合具体的政策变化和生育率变化情况针对2015年后的结果本文提出如下两种猜想：

猜想1：现有政策投入不足或者政策实施的技术细节仍有较大的提升空间，无法将生育率等增长的潜力完全激发出来。要想获得取得更好的效果需要增大投入。

猜想2：现有二孩政策在2015-2017年间引起的生育率提升是过去几十年来计划生育政策压抑的生育意愿得到释放的结果，长期来看并不能有效的提高生育率。需要改善现有的政策策略。

在如下两种猜想前提下进行进一步的分析，以下两表展示了对松弛变量的分析：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 决策单元 | 松驰变量S-分析 | | | | | | | | |  | | 投入冗余率 | | | | | | |
| 社会福利 | | 教育 | 卫生健康 | | 固定资产投资 | | 房屋竣工面积 | | 汇总 | | 社会福利 | 教育 | | 卫生健康 | 固定资产投资 | | 房屋竣工面积 |
| 2010 | 0.000 | | 0.000 | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | 0.000 | | 0.000 | 0.000 | | 0.000 |
| 2011 | 0.000 | | 0.000 | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | 0.000 | | 0.000 | 0.000 | | 0.000 |
| 2012 | 5850.642 | | 14.507 | 0.000 | | 19.012 | | 290.978 | | 6175.000 | | 0.164 | 0.115 | | 0.000 | 0.004 | | 0.171 |
| 2013 | 0.000 | | 0.000 | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | 0.000 | | 0.000 | 0.000 | | 0.000 |
| 2014 | 0.000 | | 0.000 | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | 0.000 | | 0.000 | 0.000 | | 0.000 |
| 2015 | 1431.219 | | 3.189 | 0.000 | | 0.000 | | 0.000 | | 1434.000 | | 0.028 | 0.018 | | 0.000 | 0.000 | | 0.000 |
| 2016 | 0.000 | | 0.000 | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | 0.000 | | 0.000 | 0.000 | | 0.000 |
| 2017 | 0.000 | | 0.000 | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | 0.000 | | 0.000 | 0.000 | | 0.000 |
| 2018 | 62431.826 | | 24.192 | 2.757 | | 0.000 | | 0.000 | | 62459.000 | | 0.463 | 0.096 | | 0.029 | 0.000 | | 0.000 |
| 2019 | 84898.312 | | 69.838 | 27.454 | | 0.000 | | 112.566 | | 85108.000 | | 0.514 | 0.242 | | 0.234 | 0.000 | | 0.071 |
| 2020 | 59474.197 | | 37.435 | 11.970 | | 0.000 | | 0.000 | | 59524.000 | | 0.374 | 0.122 | | 0.099 | 0.000 | | 0.000 |
| 2021 | 65858.758 | | 76.910 | 40.916 | | 0.000 | | 0.000 | | 65977.000 | | 0.394 | 0.238 | | 0.272 | 0.000 | | 0.000 |
| 决策单元 | | 松驰变量S+分析 | | | | | | |  | | 产出不足率 | | | | | | | |
| 结婚登记（对） | | | 自然增长率 | | 出生率 | | 汇总 | | 结婚登记（对） | | | 自然增长率 | | | 出生率 | |
| 2010 | | 0.000 | | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | | 0.000 | | | 0.000 | |
| 2011 | | 0.000 | | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | | 0.000 | | | 0.000 | |
| 2012 | | 5276.584 | | | 0.000 | | 0.000 | | 5277.000 | | 0.068 | | | 0.000 | | | 0.000 | |
| 2013 | | 0.000 | | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | | 0.000 | | | 0.000 | |
| 2014 | | 0.000 | | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | | 0.000 | | | 0.000 | |
| 2015 | | 0.000 | | | 0.000 | | 1.069 | | 1.000 | | 0.000 | | | 0.000 | | | 0.103 | |
| 2016 | | 0.000 | | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | | 0.000 | | | 0.000 | |
| 2017 | | 0.000 | | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | | 0.000 | | | 0.000 | |
| 2018 | | 0.000 | | | 0.578 | | 1.413 | | 2.000 | | 0.000 | | | 0.107 | | | 0.124 | |
| 2019 | | 0.000 | | | 0.000 | | 1.994 | | 2.000 | | 0.000 | | | 0.000 | | | 0.208 | |
| 2020 | | 0.000 | | | 1.494 | | 1.681 | | 3.000 | | 0.000 | | | 0.428 | | | 0.177 | |
| 2021 | | 0.000 | | | 3.013 | | 1.399 | | 4.000 | | 0.000 | | | 2.450 | | | 0.193 | |

可以看出2015年后政策效益的整体倾向于投入冗余，这表明了目前政策投入的增加对刺激生育率等与人口增长相关的指标增加是相对乏力的，进一步的增大投入并不能有效的刺激生育增加。针对计划生育政策，政府仍然需要积极的调整政策策略以促进人口结构的健康发展。

面板数据回归

通过整合江苏省各个市2010 – 2021年的与人口数、GDP、医院数、普通高等院校数等数据，获得了与人口预测相关的面板数据，对其进行面板数据分析。

模型选择：

面板数据模型一般有三种形式可以选择：混合估计模型（POOL）、固定效应模型（FE）、随机效应模型（RE），可以通过如下三种统计检验进行选择：

F检验用于FE模型和POOL模型选择对比 ，P值小于0.05意味着FE模型更优，反之则使用POOL模型。

Breusch-Pagan检验用于RE模型和POOL模型选择对比，P值小于0.05意味着RE模型更优，反之则使用POOL模型。

Hausman检验用于FE模型和RE模型选择对比，P值小于0.05意味着FE模型更优，反之则使用RE模型。

|  |  |  |  |
| --- | --- | --- | --- |
| 检验类型 | 统计量 | P | 结论 |
| F检验 | 92.356 | 0.000\*\*\* | FE模型 |
| Breusch-Pagan检验 | 324.127 | 0.000\*\*\* | RE模型 |
| Hausman检验 | -8.523 | 1.000 | RE模型 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | |

根据检验结构选择随机效应模型。在随机效应模型中，个体特征被视为随机变量，其变异被建模为一个误差项。这个误差项表示个体特征的随机差异，它是与时间无关的随机效应。因此，在随机效应模型中，个体之间的差异被认为是随机的，并且在个体间的变异中考虑了随机误差项。

随机效应模型的回归方程可以写成以下形式：

Y\_it = β\_0 + β\_1X\_1\_it + β\_2X\_2\_it + ... + β\_k\*X\_k\_it + α\_i + u\_it

其中，

Y\_it 是第 i 个个体在时间 t 上的因变量观测值。

X\_1\_it, X\_2\_it, ..., X\_k\_it 是第 i 个个体在时间 t 上的 k 个自变量观测值。

β\_0, β\_1, β\_2, ..., β\_k 是自变量的系数，表示自变量对因变量的影响。

α\_i 是个体 i 的随机效应，表示个体 i 的特征对因变量的影响，它是一个随机变量。

u\_it 是模型的误差项，表示未被解释的随机误差。

个体 i 的随机效应 α\_i 被认为是一个符合某种分布（通常是正态分布）的随机变量，并且它在所有个体之间是独立且与时间无关的。在拟合随机效应模型时，通常通过最大似然估计等方法来估计参数。

模型的拟合结果如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| RE模型 | | | | | | |
| 变量 | 系数 | 标准误差 | t | P | R² | F |
| const | 428.45 | 43.937 | 9.751 | 0.000\*\*\* | within=0.417 between=0.272 overall=0.28 | F=24.836 P=0.000\*\*\* |
| deposit | 0 | 0 | 2.079 | 0.039\*\* |
| GDP | 0 | 0 | 0.517 | 0.606 |
| school | 0.35 | 1.438 | 0.244 | 0.808 |
| hospital | 0.43 | 0.138 | 3.113 | 0.002\*\*\* |
| 因变量：population | | | | | | |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | |

RE随机模型的F检验结果显示，显著性P值为0.000\*\*\*，水平上呈现显著性，拒绝原假设，因此模型是有效的。