

参赛密码 \_\_\_\_\_  
(由组委会填写)

## 第十届华为杯全国研究生数学建模竞赛

学 校            山东财经大学

---

参赛队号        10456002

---

队员姓名        1. 董中印

---

                    2. 孙晓琳

---

                    3. 孙晓芹

---

参赛密码 \_\_\_\_\_

(由组委会填写)



## 第十届华为杯全国研究生数学建模竞赛

题目 中等收入定位与人口度量模型研究

### 摘 要：

居民收入分配关系到人民幸福和国家安定，其中中等收入人口比重是反映收入分配格局的重要指标。我国正处于经济转型期和收入分配格局的重要调整期，监控收入分配格局的变化具有重大意义。本文针对中等收入定位与人口度量模型研究，对于不同的问题建立相应的数学模型并求解。

对于问题一，构建新的洛伦兹模型  $L(p)=[1-(1-p)^\alpha e^{-\beta p}][1-(1-p)^\gamma e^{-\eta p}]$ ，与现有的经典洛伦兹模型做比较，拟合精度最优，MSE、MAE 的值均在同类模型中最小，说明所构建的新模型性质优良。

对于问题二，对“收入空间法”和“人口空间法”分别提出了不同的改进策略。针对“收入空间法”中等收入上下限选取的随意性缺陷，本文在收入均值左右一个方差内考虑税收因素确定中等收入，并且建立衡量两极分化程度指标  $\omega$  来比较不同地域不同年份的两极分化程度。对于“人口空间法”，通过人口分组排序的方法确定各个分组的收入状态，针对收入结构的不同情况，定位不同的中等收入，使得对于中等收入的定位和人口度量的研究更具有实用性。

对于问题三，运用改进的收入空间法和人口空间法对 A、B 两个地区不同年份数据进行研究。改进的收入空间法得到 A 地区两年份的中等收入范围分别是 [3500, 9188.4] 和 [4803.1, 12977]，中等收入人口比例分别为 55.45% 和 55.60%，

基尼系数分别为 0.3113、0.3153，收入不平等程度稍有加重；B 地区两年份的中等收入范围分别是 [10708, 23169] 和 [14088, 30369]，中等收入人口比例分别为 59.87% 和 58.09%，基尼系数分别为 0.2553、0.2371，收入两极分化情况有所缓解。横向比较发现，A 地区中等收入人口比例低于 B 地区，收入不平等程度较高，两极分化相对较重。改进的人口空间法得到 A 地区两年份的中等收入状态分别为 0.6264、0.6281，B 地区两年份的中等收入人口状态分别为 0.6083、0.6326。

对于问题四，分析中等收入人群的特征，提出中等收入人群的定义，在此基础上构建了中等人口收入效用指数模型，通过界定效用指数区间来界定中等收入人口，在充分考虑时空差异的基础上，克服了传统方法研究中等收入人口的缺陷。

最后，对全文进行总结，得出中等收入定位和人口度量模型的结论，并提出相关的政策建议。

**关键词：** 中等收入 洛伦兹曲线 收入空间法 人口空间法

## 目录

摘    要.....	1
一、问题重述.....	4
二、问题分析.....	5
三、模型建立与求解.....	6
3.1 问题一模型建立与求解.....	6
3.1.1 新洛伦兹模型构建.....	6
3.1.2 模型的计算与比较.....	9
3.2 收入空间法和人口空间法改进模型.....	10
3.2.1 收入空间法的改进.....	10
3.2.2 改进后收入空间法的检验.....	13
3.2.3 人口空间法的改进.....	15
3.2.4 改进后人口空间法的检验.....	17
3.3 A、B地区两年份的中等收入研究.....	18
3.3.1 基于改进收入空间法的研究.....	18
3.3.2 基于改进人口空间法的研究.....	23
3.4 中等收入人口的新模型及测算.....	27
3.4.1 中等收入人群的特征.....	27
3.4.2 中等收入人群的定义.....	27
3.4.3 确定中等收入人群的原理及经济学意义.....	28
3.4.3 中等收入新模型的建立.....	28
四、结论及政策建议.....	31
4.1 结论.....	31
4.2 政策建议.....	32

## 一、问题重述

居民收入分配关系到广大民众的生活水平，分配公平程度是广泛关注的话题。其中中等收入人口比重是反映收入分配格局的重要指标。我国处于经济转型期，收入分配格局处于重要的调整期，“橄榄型”收入分配格局正处于形成阶段。因此，监控收入分配格局的变化是经济社会发展的重要课题，例如需要回答，与前年比较，去年的收入分配格局改善了吗？改善了多少？可见实际上需要回答三个问题：什么是“橄榄型”收入分配格局？收入分配格局怎样的变化可以称之为改善？改善了多少？直观上，中间部分人口增加，则收入分配格局向好的方向转化。于是基本问题回答什么是中间部分。

一个国家的收入分配可以用统计分布表示。收入分配经验分析说明，收入分配曲线一般是所谓正偏的，即峰值点向左偏，右端拖一个长尾巴，记对应的分布函数为  $F(x)$ ，则  $p = F(x)$  表示收入低于或等于  $x$  的人口比例。由于  $F(m) = 1/2$ ，意味着收入大于或等于平均收入的人口一定不到半数，因此是少数。

记收入低于或等于  $x$  的人口群体拥有收入占总收入的比例为  $L(p)$ ， $L(p)$  称之为收入分配的洛伦兹曲线。经济学界采用所谓的洛伦兹曲线模型  $L(p, \tau)$  拟合数据，使用非线性最小二乘法确定其中参数向量  $\tau$  的估计值  $\hat{\tau}$ ，然后用  $L(p, \hat{\tau}) = \hat{L}(p)$  作为近似的洛伦兹曲线来进行收入分配分析，确定相应的统计密度与分布的估计。也可以使用其他方法(例如多项式、样条函数逼近)来确定洛伦兹曲线，但实践证明使用洛伦兹曲线模型是比较理想的方法之一。经济理论中提出的另一种方法是使用经验分布拟合分组数据而直接形成收入分配的近似分布。

经济理论界考虑取收入落在中位收入  $m$  的一个范围内的人口为中等收入人口，视这种方法为“收入空间法”。这种方法中  $x_l$  与  $x_h$  的取法具有任意性，由于经济进步，通货膨胀等因素的影响，收入的区间是变化的，更多的情形是所有人口的收入都提高了，即全社会的收入区间右移，可见  $x_l$  与  $x_h$  的任意性使纵向比较各年的中等收入人口时出现困难。

另一种方法可以视为“人口空间法”，即选择  $F(m) = 1/2$  邻近的一个范围为中等收入人口，如用 60% 的人口所拥有的收入占总收入的比例来描述中等收入人口的状态，此时中等收入人口的收入范围  $[x_l, x_h]$  当然容易算得，但也存在收入范围变化带来的问题。为此，需要研究中等收入定位与人口度量问题，请你根据表一中给出的分组数据，用数学模型研究给出的问题：

一．构造满足(9)式的新模型  $L(p, \tau)$ ，使得能很好的拟合上述分组数据、反映经济规律。就文献中提出了其他一些模型说明利用这些模型时产生的估计结果优于密度函数的 Kernel 估计法。在现有参考文献中找出至少 10 种模型，与你们

提出的模型进行比较。通过比较，说明你们的模型不差。

二. 研究可否改进上述提到的收入空间法，这时需要研究确定中等收入的范围、中等收入人口的范围的科学方法，以克服中等收入区间取法的任意性；研究可否改进上述提到的人口空间法，例如研究在各年中  $p_1$  与  $p_2$  取不同的值时，纵向比较各年中等收入人口与收入的变动的方法。

三. 利用最后表二至表五所附 A, B 两个地区前后两个不同年份的收入分配分组数据，请研究：(1) 对各地区、各年份的中等收入的数量(或范围)、中等收入人口的数量或范围进行定量描述，说明中等收入人口的变化趋势；(2) 比较两个地区的中等收入人口、收入等变化情况。

四. 除二题中所述方法外，提出中等收入人口的定义、原理及经济学意义，并提出与之相应的中等收入人口的测算方法、模型或指数，说明其经济学意义。

## 二、问题分析

问题一的关键要求要构建一个新的洛伦兹曲线模型，并且检验所构建的模型能够满足题目中所给数据的拟合效果。解决该问题需要通过查阅相关文献，在总结前人研究成果的基础上，构建一个满足洛伦兹曲线性质的新的洛伦兹曲线模型。通过对比分析，新构建的洛伦兹模型要至少不差于甚至优于目前已经存在的大部分洛伦兹曲线模型。

如何克服“收入空间法”在选取中等收入区间时的随意性和“人口空间法”纵向比较的可行性是解决问题二的关键。而“收入空间法”改进的关键是能否找到一种确定中等收入区间端点的方法，使得该区间在选择上具有稳定性，并且随着经济增长、通货膨胀等因素的变化，该区间具有纵向比较性；同理，对于“人口空间法”的改进应主要从解决

问题三的解决应该建立在问题二解决的基础上，运用上述“改进的收入空间法”和“改进的人口空间法”来解决该问题，通过对表二到表五的数据进行定量分析，以对个地区、各年份的中等收入的数量（或范围）、中等收入人口的数量和范围进行定量描述，并分别分析 A、B 两地区中等人口的变化趋势，收入分配情况以及两极分化情况，对比分析 A、B 两地区中等收入人口、收入以及两极分化等情况。

问题四要求提出中等收入人口的定义、原理及经济学意义，这需要对中等收入人口的特征有一个比较全面的了解，这需要查阅相关经济学者对中等收入人口的定义，并结合自己的理解，提出一个更为合理、更为完善的中等人口收入的定义，分析确定中等收入人口的原理，并提出与之相应的中等收入人口的测算方法、

模型或者指数，通过数据检验所建立的模型指数，最后阐明其经济学意义。

### 三、模型建立与求解

#### 3.1 问题一模型建立与求解

##### 3.1.1 新洛伦兹模型构建

假设基于统计分布表示的收入分配的密度函数是  $f(x)$ ，其中  $x$  表示收入，对应的分布函数为  $F(x)$ ，则  $p = F(x)$  表示收入低于或等于  $x$  的人口比例。记收入低于或等于  $x$  的人口群体拥有收入占总收入的比例为  $L(p)$ ，则应有

$$L(p) = \frac{1}{\mu_0} \int_0^x tf(t)dt, p = F(x)$$

$L(p)$  称之为收入分配的洛伦兹曲线。收入分配的洛伦兹曲线  $L(p)$  在收入分配分析中具有重要地位，它表示人口份额等于  $p$  的低收入端拥有的总收入份额，因此  $L(p)$  是定义于  $[0, 1]$  区间上的函数。按经济意义，它应满足如下条件：

$$L(0, \tau) = 0, L(1, \tau) = 1, L'(p, \tau) \geq 0, L''(p, \tau) \geq 0$$

即  $L(p, \tau)$  在  $[0, 1]$  上是凸增函数。

在分析与测算洛伦兹曲线的实际工作中，在只有分组数据可用的条件下，可以先估计收入分配的密度函数，从而得到相应的洛伦兹曲线，或直接估算洛伦兹曲线。国内外学者所提出的模型可以概括为三大类：几何计算法、分布函数法和曲线拟合法。几何分析法，是根据分组数据刻画洛伦兹曲线，利用这一方法不能得到洛伦兹曲线的表达式，只能用来计算基尼系数，但由于在计算分块面积时用直线近似地代替曲线，所估计的基尼系数要小于实际值，尤其在数据点较少时，误差较大；分布函数法，是基于对指标的概率密度函数或概率分布函数的假设，估计其分布参数，然后对洛伦兹曲线进行估计，这类方法较为复杂，同时由于计算收入分配的概率密度的复杂性，很难提出合适的概率函数；曲线拟合法首先假设收入分配服从某一特殊的统计分布函数形式，如对数正态分布、伽马分布、贝塔分布、帕累托分布、威布尔分布等，再据此为洛伦兹曲线选择适当的参数方程直接进行拟合、确定参数，由此估计出洛伦兹曲线。

为了更准确地描述洛伦兹曲线和精确地估计基尼系数，本文通过分析洛伦兹曲线的特性，构建新的洛伦兹曲线模型，对洛伦兹曲线直接进行拟合。根据洛伦兹曲线应满足的性质入手，在查找相关文献的基础上，总结前人所做研究，构建出满足洛伦兹曲线的新模型  $L(p, \tau)$ 。

Chotikapanich (1993) 提出较早的洛伦兹模型：

$$L_c(p) = \frac{e^{\lambda p} - 1}{e^\lambda - 1}, \lambda > 0 \quad (1)$$

Sarabia 等提出了基于经典帕累托分布的洛伦兹模型，其基本形式为：

$$L(p) = 1 - (1 - p)^\beta, \beta \in [0, 1] \quad (2)$$

从 (1) 式出发，他们提出了提出了一族广义帕累托族的洛伦兹曲线模型，具体形式为：

$$L(p) = p^\alpha (1 - (1 - p)^\beta), \alpha \geq 0, \beta \in (0, 1] \quad (3)$$

$$L(p) = p^\alpha (1 - (1 - p)^\beta)^\nu, \beta \in (0, 1] \quad (4)$$

$$L(p) = (1 - (1 - p)^\beta)^\nu, \beta \in (0, 1] \quad (5)$$

另外，基本的洛伦兹经典模型还有：

$$L(p) = 1 - \left(\frac{e^{\lambda(1-p)} - 1}{e^\lambda - 1}\right)^\beta, \beta \in (0, 1], \lambda \in (-\infty, 0) \cup (0, \ln \beta_1^{-1}] \quad (6)$$

$$L(p) = 1 - \left(1 - \frac{e^{\lambda p} - 1}{e^\lambda - 1}\right)^\beta, \beta \in (0, 1], \lambda \in [\ln \beta_2, 0) \cup (0, +\infty) \quad (7)$$

Ogwang 和 Rao (2000) 曾提出用两种混合的方法建立洛伦兹模型：加权积 (weighted product) 及洛伦兹模型的凸组合 (convex combination)。并且得到凸组合模型 (8) 满足洛伦兹曲线性质：

$$L(p) = \delta p^\alpha [1 - (1 - p)^\beta e^{-\gamma p}] + (1 - \delta) \frac{e^{\lambda p} - 1}{e^\lambda - 1}, \delta \in [0, 1] \quad (8)$$

更一般化的经典洛伦兹曲线模型为王祖祥 (2007) 提出的如下的二元参数模型：

$$H_0(p) = 1 - (1 - p)^\beta e^{-\gamma p} \quad (9)$$

(9) 式所表示的二元参数模型作为一个洛伦兹曲线的参数估计模型使用时比基于帕累托的广义洛伦兹模型具有更好的性质。

现在的研究成果已经证明：假定  $L(p)$  为洛伦兹曲线，则对于任意的  $\alpha \geq 0$  和  $\eta \geq 1$ ， $\tilde{L}(p) = p^\alpha L(p)^\eta$  也是洛伦兹曲线。进而，如果对于所有的  $p \in [0, 1]$ ，都有  $\tilde{L}'''(p) \geq 0$ ，则当  $\alpha \geq 0$ ， $\eta \geq 1/2$  且  $\alpha + \eta \geq 1$  时， $\tilde{L}(p)$  也是洛伦兹曲线。

国内外的专家学者在研究洛伦兹模型方面做了大量的工作，王祖祥、Sarabia 等提出了一系列的洛伦兹曲线模型，除上面提到的公式 (9)，如：

$$L(p) = p^\alpha \left( \frac{e^{\lambda p} - 1}{e^\lambda - 1} \right)^\beta, \alpha \geq 0 \quad (10)$$

$$L(p) = p^\alpha \left[ 1 - \left( \frac{e^{\lambda(1-p)} - 1}{e^\lambda - 1} \right)^\beta \right]^{\alpha_1} \left[ 1 - \left( 1 - \frac{e^{\eta p} - 1}{e^\eta - 1} \right)^{\beta_1} \right]^{\alpha_2} \quad (11)$$

$$L(p) = [1 - (1-p)^\beta]^\alpha [1 - (1-p)^\eta]^\lambda \quad (12)$$

$$L(p) = p^\kappa \left\{ \delta p^\alpha [1 - (1-p)^\beta e^{-\gamma p}] + (1-\delta) \frac{e^{\lambda p} - 1}{e^\lambda - 1} \right\}^\rho \quad (13)$$

在研究现有文献基础上, 按照洛伦兹曲线应满足的性质以及推导定理, 本文构建了基于指数成份的洛伦兹模型  $L(p, \tau)$ , 形式如下:

$$L(p) = [1 - (1-p)^\alpha e^{-\beta p}] [1 - (1-p)^\gamma e^{-\eta p}] \quad (14)$$

其中,  $0 \leq \alpha + \beta \leq 1$ ,  $0 \leq \gamma + \eta \leq 1$ 。

可以证明, 该模型满足洛伦兹模型的一般定义:

$$L(0) = 0;$$

$$L(1) = 1;$$

$$\begin{aligned} L'(p) &= [\alpha(1-p)^{\alpha-1} e^{-\beta p}] [1 - (1-p)^\gamma e^{-\eta p}] + \\ & [1 - (1-p)^\alpha e^{-\beta p}] [\gamma(1-p)^{\gamma-1} e^{-\eta p} + \eta p(1-p)^\gamma e^{-\eta p}] \end{aligned}$$

在满足条件  $\alpha + \beta \geq 0$ ,  $\gamma + \eta \geq 0$  的情况下, 当  $p \in [0, 1]$  时, 满足  $L'(p) \geq 0$ 。

$$\begin{aligned} L''(p) &= [\alpha(1-\alpha)(1-p)^{\alpha-1} e^{-\beta p} - \alpha\beta p(1-p)^{\alpha-1} e^{-\beta p} + \beta(1-p)^\alpha e^{-\beta p} \\ & - \alpha\beta p(1-p)^{\alpha-1} e^{-\beta p} - \beta^2 p^2(1-p)^\alpha e^{-\beta p}] [1 - (1-p)^\gamma e^{-\eta p}] \\ & + 2[\alpha(1-p)^{\alpha-1} e^{-\beta p} + \beta p(1-p)^\alpha e^{-\beta p}] [\gamma(1-p)^{\gamma-1} e^{-\eta p} \\ & + \eta p(1-p)^\gamma e^{-\eta p}] [\gamma(1-\gamma)(1-p)^{\gamma-1} e^{-\eta p} \\ & - \gamma\eta p(1-p)^{\gamma-1} e^{-\eta p} + \eta(1-p)^\gamma e^{-\eta p} \\ & - \gamma\eta p(1-p)^{\gamma-1} e^{-\eta p} - \eta^2 p^2(1-p)^\gamma e^{-\eta p}] [1 - (1-p)^\alpha e^{-\beta p}] \\ & = [(\alpha + \beta) - (\alpha + \beta)^2 - \beta^2 p^2(1-p)] (1-p)^{\alpha-1} e^{-\beta p} \\ & + [(\gamma + \eta) - (\gamma + \eta)^2 - \eta^2 p^2(1-p)] (1-p)^{\gamma-1} e^{-\eta p} \end{aligned}$$

当满足条件  $\alpha + \beta \leq 1$ ,  $\gamma + \eta \leq 1$  的情况下, 当  $p \in [0, 1]$  时,  $L''(p) \geq 0$ 。

综上所述, 新构建的洛伦兹模型模型(14)式满足洛伦兹曲线的定义及性质,

可以用来拟合题目中给出的数据，并且与现有经典洛伦兹模型做出比较。

### 3.1.2 模型的计算与比较

有关收入与人口的数据一般情况下可以得到收入人口分布的分组数据，这种数据的完整形式为  $(p_i, x_i/\mu)$ ,  $i=1,2,\dots,n$  或者  $(p_i, L_i)$ ,  $i=1,2,\dots,n$ ，其中  $x_i$  是收入区间点，满足  $0 \leq x_1 < x_2 < \dots < x_n < x_{n+1}$ ，通常  $x_{n+1}$  理解为充分大的正数。 $n$  通常不大，例如  $n=10$ 。表一中提供这样描述的数据。经济学界采用所谓的洛伦兹曲线模型  $L(p, \tau)$  拟合分组数据其中  $\tau$  是一组参数，使用非线性最小二乘法求解

$\min \sum_{i=1}^n (L(p_i, \tau) - L_i)^2$  确定其中参数向量  $\tau$  的估计值  $\hat{\tau}$ ，然后用  $L(p, \hat{\tau}) = \hat{L}(p)$  作为近似的洛伦兹曲线来进行收入分配分析。

本文采用无约束非线性最小二乘法来拟合表一中数据，并且对新构建模型和相关文献中找到的经典洛伦兹模型进行比较。为了检验模型拟合的效果，本文采用如下三个标准：

均方误差 (MSE, mean squared error) :  $\frac{1}{n} \sum_{i=1}^n [L(p_i, \hat{\tau}) - L_i]^2$

平均绝对误差 (MAE, mean absolute error) :  $\frac{1}{n} \sum_{i=1}^n |L(p_i, \hat{\tau}) - L_i|$

最大绝对误差 (MAS, maximum absolute error)  $\max_{1 \leq i \leq n} |L(p_i, \hat{\tau}) - L_i|$

针对题目附表中给出的数据，分别采用不同洛伦兹模型，拟合结果如表 1。

表 1 新构建洛伦兹模型与经典洛伦兹模型拟合结果比较

模型	估计参数	MSE	MAE	MAS
(1)	$\lambda=2.689$	0.01706	0.0141	0.0359
(2)	$\beta=0.4641$	0.05694	0.0496	0.0774
(3)	$\alpha=0.8031, \beta=0.6794$	0.003261	0.0027	0.0055
(4)	$\alpha=1.392, \beta=0.5181, \nu=0.4779$	0.002024	0.0016	0.0036
(5)	$\beta=0.782, \nu=1.763$	0.004543	0.0038	0.0080
(6)	$\lambda = -2.423, \beta = 0.9162$	0.01629	0.0140	0.0217
(7)	$\lambda = 2.423, \beta = 0.9163$	0.01629	0.0140	0.0218
(8)	$\delta = 1.023, \alpha = 0.7516, \beta = 0.6459$ $\gamma = 0.08411, \lambda = -2.789$	0.001366	9.6866e-004	0.0021
(9)	$\beta = 0.8495, \gamma = -0.7048$	0.01089	0.0092	0.0177

(10)	$\alpha=0.6923, \lambda=3.045, \beta=0.6466$	0.01733	0.0113	0.0411
(12)	$\alpha = 1.494, \beta = 0.7148,$ $\eta = 2.217, \lambda = 0.523$	0.0004288	3.1525e-004	7.8636e-004
构建模型	$\alpha=0.7119, \beta=-0.3813,$ $\gamma=2.299, \eta=0.3372$	0.000409	2.8566e-004	8.2201e-004

注释：模型标号与文中模型公式的标号一致，为简化文章不再把模型表达式再次罗列到结果中。

通过比较可以发现，新构建的模型就本组数据拟合效果很好，均方误差 (MSE) 为 0.000409、平均绝对误差 (MAE) 为 0.000286、最大绝对误差 (MAS) 为 0.000822，误差为所比较模型中绝大多数模型的最小值，并且拟合优度几乎为 1。由此得出新构建的洛伦兹模型拟合的误差远小于现有的大部分洛伦兹模型的拟合误差，该模型是一个很好的洛伦兹模型。

### 3.2 收入空间法和人口空间法改进模型

针对中等收入的测算可以分解成以下四步：第一，选择“空间”；第二，定义中等收入；第三，确定中等收入区间；第四，计算数据。首先，最基础的工作就是中等收入定义及“空间”的选择，当前最常用的选择方法是“收入空间法”，这里的收入包括工资、年终奖以及其他的福利费。同时，中等收入也可以定义在“人口空间”。第二步要给出中等收入的大致定义。多数的学者选择收入中位数作为收入空间的中等收入依据，在人口空间中表示为人口的百分之五十。另外的一个中等收入标准是平均收入，尽管在现实中一半的人达不到平均收入水平，这仍然不失为一种研究方法。第三步需要选择中位数附近的区间，以此来定义中等收入。最后一步是对中等收入的数据估算，通过计算相关的指数来反映所研究的相关问题。但是对于“收入空间法”和“人口空间法”在研究中等收入的问题上依然存在一些缺陷，本文提出以下的改进措施。

#### 3.2.1 收入空间法的改进

中等收入人口比重是反映和衡量收入分配格局的重要指标。在理想的橄榄型的收入分配格局下，收入差距不大，社会消费旺盛，人民生活水平高，社会稳定。目前经济理论界划定中等收入人群较常用的方法有“收入空间法”和“人口空间法”。其中“收入空间法”取收入落在中位收入  $m$  的一个范围内的人口为中等收入人口，如图 1 所示为：

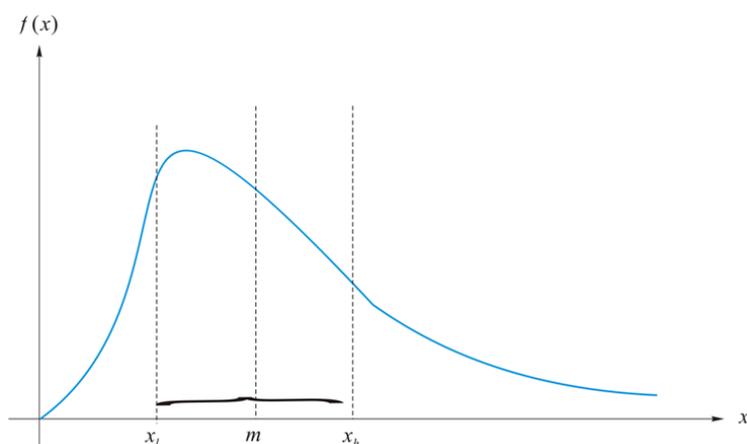


图 1 收入分配概率密度示意图

取其中收入属于  $(x_l, x_h)$  中的人口为中等收入人口，这时中等收入人口比例  $M$  显然等于  $F(x_h) - F(x_l)$ 。显然，这种方法中  $x_l$  和  $x_h$  的取法具有任意性，由于经济进步，通货膨胀等因素的影响，收入的区间是变化的，更多的情形是所有人口的收入都提高了，即全社会的收入区间右移，可见  $x_l$  与  $x_h$  的任意性使纵向比较各年的中等收入人口时出现困难。因此，构建一种合理的动态区间来刻画中等收入人群，观测其变化趋势具有极大的现实意义。下面针对“收入空间法”的不足，本文的解决思路为：构建了一个综合多因素的区间模型，统计分析表明，距平均值一个标准差以内的范围，在正态分布中，此范围所占比率为全部数值的 68%，虽然收入分密度函数不服从正态分布，而是正偏的、右端尾部较长的分布曲线，但是用  $[\mu - \sigma, \mu + \sigma]$  能够有效的包含中间区域部分，且该部分的收入差距不大。这在区间端点的选择上能够克服“收入空间法”在选取上下限时的随意性。本文在此区间的基础上，通过加入税收因素，最终把该区间的修正区间作为中等收入人群的收入区间。

### 3.2.1.1 模型假设

1. 假设能够根据所获得的区域收入数据得到较为准确的收入均值和标准差；
2. 若所得数据是分组数据，那么假设在每一小区间内人口及收入是均匀分布。

### 3.2.1.2 符号说明

1.  $x_l$  为所确定收入空间的下限；
2.  $x_h$  为所确定收入空间的上限；
3.  $\mu$  为收入均值；
4.  $\sigma$  为收入的标准差；

5.  $\omega$  为两极分化衡量指标。

### 3.2.1.3 多因素收入空间模型的构建

本文所构建的模型是在“收入空间法”的基础之上,通过综合考虑税收因素、数据的方差和均值等特征量,基于均值附近左右一个标准差范围内所包含的概率密度较高的特性,来确定收入空间的范围,以消除收入空间法在区间取值上的任意性。中等收入模型所确立的区间为 $[x_l, x_h]$ ,其中上下限所要满足的条件为:

$$\begin{cases} x_l = \max \{3500, \mu - \sigma\} & (15) \\ x_h = \min \left\{ \left(2 - \frac{\mu - x_l}{\sigma}\right)\sigma + \mu, \mu + 1.2\sigma \right\} & (16) \end{cases}$$

其中,式(15)为综合考虑了税收因素的模型区间的下限;式(16)是所选取区间的上限,该上限使得在选取中等收入人群时,具有灵活变动性,这样就克服了区间选取的任意性,确定了在一个合理的相对固定的区间内满足人口收入差距不大、收入处于中等范围内的人口比例。

其中 3500 为我国 2011 年公布的个人所得税起征点,个人所得税的征集一方面是为了调整我国居民收入的贫富差距,另一方面,税收是在保障了居民生活水平的前提下起征。因此,将个人所得税起征点作为确定中等收入人群下限的考量因素,可以充分的排除目前生活水平相对较低的人群。

一个国家的收入分配可以用统计分布表示,图 2 是一系列收入分配的密度函数。收入分配经验分析说明,收入分配曲线一般是所谓正偏的,即峰值点向左偏,右端拖一个长尾巴,如图 2 所示:

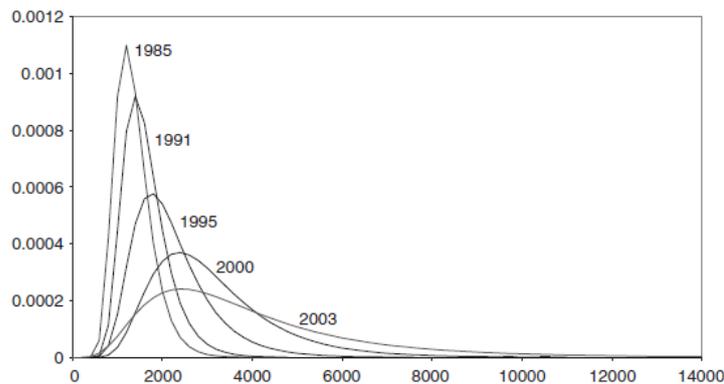


图 2 收入分配的密度函数 (横轴为收入, 纵轴为概率)

为了比较不同地区和不同年份收入分配的两极分化程度,本文创新性的构建了一个指标  $\omega$ , 以衡量居民收入两极分化的大小, 该指标越大, 说明收入分配

的两极分化程度越高，该指标越小，说明收入分配两极分化的程度越低。该指标需要满足的经济规律方面的要求概括为：

在其他情况不变的条件下，当 $\sigma$ 扩大时， $\omega$ 减小，这说明随着贫富收入离散程度的加大，两极分化程度具有增大的趋势；在其他情况不变的条件下，当 $p$ 增大时， $\omega$ 减小，这说明随着中等收入区间人口比例的增加，两极分化程度减弱；在其他情况不变的条件下，当 $\mu$ 增大时， $\omega$ 减小，这说明随着人均收入水平的提高，若贫富收入离散程度没有改变，则两极分化程度减轻了。本文根据该指标的性质，将其定义为：

$$\omega = 1 - \frac{\mu p}{2\sigma} \quad (17)$$

$\omega$ 是一个在区间 $(-\infty, 1]$ 上变动的指标，当收入分配完全是均匀分配时，不存在两极分化的情况，此时， $\omega = -\infty$ ；而当中间收入人群为零时，即形成完全的两极分化时， $\omega = 1$ ，为此本文考虑另一种特殊的情形，即中间收入人口所占比例为50%，变异系数为0.5时， $\omega = 0.5$ 。自此，本文界定 $\omega \in (-\infty, 0.5]$ 时，两极分化程度比较轻，而当 $\omega \in (0.5, 1]$ 时，两极分化程度比较高。

该指数克服了该模型区间选取中存在长度不一致的情形。

本文构建的多因素收入空间模型的优越性体现在：

1. 区间的选择有效排除了生活水平相对较低的人群；
2. 该区间的选择具有稳定性，克服了收入空间选择的随意性，并且，该区间随着每年实际收入分布的不同而有所变化，克服了由于经济增长等因素变化造成的纵向不可比较性；
3. 若中间部分比前一年隆起得更高，那么该区间所包含的的比率将增大，因此，可认为中等收入人口扩大了；若两边人口扩大了，则该区间所包含的比率将减少，可认为中等收入人口下降了，本区间确定的中等收入人群能够很好的与现实情况吻合。
4. 该区间的上下限不是固定不变的，而是随着每年收入状况发生改变，通过指标 $\omega$ 的定义，可以有效比较不同年份以及不同地区之间两极分化的程度。

### 3.2.2 改进后收入空间法的检验

本部分运用表一中给出的收入分配的分组数据对多因素收入空间模型进行

检验分析，计算得到该分组数据的密度函数曲线（图3）和洛伦兹曲线（图4）。

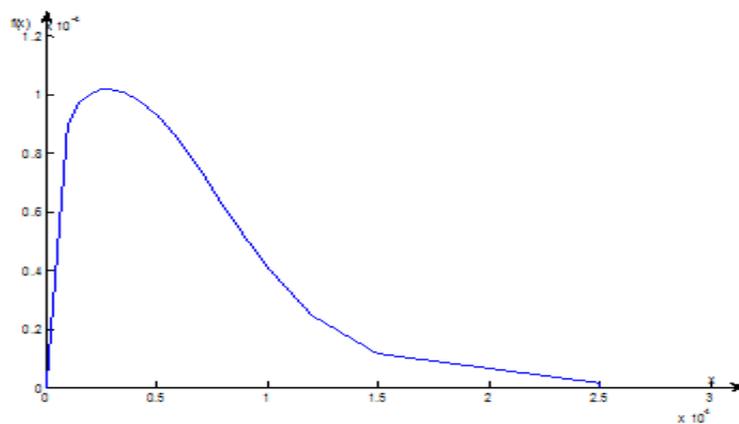


图3 密度函数曲线

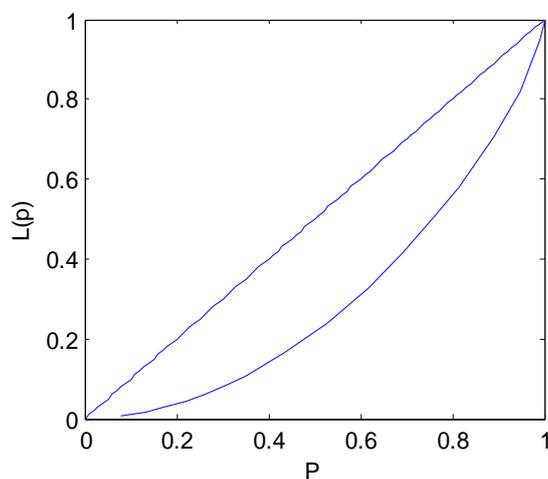


图4 洛伦兹曲线

中等收入人口分布区间上下限的界定：

1. 均值。根据前面模型假设，对于分组数据，每一小段数据服从均匀分布，可以得到：

$$\mu = \sum_{j=0}^{16} \frac{(x_j + x_{j+1})}{2} f_j = 6535.5$$

2. 标准差。根据（15）式中所得到的均值，运用方差求值公式，可到该分组数据的标准差：

$$\begin{aligned} \sigma &= \sqrt{\sum_{j=0}^{16} \left( \frac{x_j + x_{j+1}}{2} - \mu \right)^2 f_j + (x_{18} - \mu)^2 f_{17}} \\ &= 4981.4 \end{aligned}$$

3. 中等收入区间的确定。对于中等收入区间的下限，根据上面（15）式，由于

$$3500 > \mu - \sigma = 1554.1$$

因此可以确定中等收入区间的下限为 3500.

根据（16）式

$$x_h = \min \left\{ \left( 2 - \frac{\mu - x_l}{\sigma} \right) \sigma + \mu, \mu + 1.2\sigma \right\}$$

$$= 12513$$

因此可以确定中等收入区间为[3500,12513]，通过计算可以得到人口概率区间为[0.309,0.9]，中间人口所占的比例为 0.5910。为了进一步分析两极分化问题，可计算得到的中等收入化指标  $\omega$ ：

$$\omega = 1 - \frac{\mu p}{2\sigma} = 0.6231$$

根据该模型的结果可以看出，该区域收入分配分组数据中间人口所占的比例为 59.10%，两极分化指标为  $0.6231 \in (0.5, 1]$ ，该区域的中等化水平相对来说较好，但是两极分化程度比较明显，总体上还没有达到理想的“橄榄形”。究其原因，主要是因为该地区贫困人口的数量所占总人口的比例较高，从而导致了中等收入人群的数量所占人口的比例较低。另外，该结果还表明，该组数据所反映的区域高收入人群数量较少，收入分配格局大体呈现了“不对称的橄榄形”，该区域应该着重于减少贫困人口的数量，增加中等收入人口的比重。

### 3.2.3 人口空间法的改进

经济理论界研究中等收入问题较常用的另一种方法为“人口空间法”，即选择  $F(m)=1/2$  邻近的一个范围为中等收入人口，比如取范围  $p_1$  到  $p_2$ （其中  $p_2 = p_1 + p$ ， $p$  为中等人口所占的比例，按照经济学定义及不同地域不同年份已经给出）。再用此  $p_1$  到  $p_2$  的人口所拥有的收入占总收入的比例来描述中等收入

人口的状态： $S = L(p_2) - L(p_1) = \frac{1}{\mu} \int_{p_1}^{p_2} F^{-1}(p) dp$ （其中  $\mu = \int_0^1 F^{-1}(p) dp$  为平均收入），

此时中等收入人口的收入范围  $[x_l, x_h]$  容易算得。

人口空间法较之于收入空间法更容易解释和比较，因为经济发展、收入增加

导致所有人口的收入都右移时，总是取中间的一定比例进行纵向比较似乎总是可行的。但是当收入范围变大收入分配发生变化时，低端人口收入下降，高端收入人口收入增加，直观上两极分化扩大了，也即中等收入人口应该是下降了，但按人口空间法，中间固定比例的人口拥有的总收入比例可能固定不变。同时，当收入分布中两极分化的情况出现变化时，假设两极分化严重，即用密度函数表示时出现严重右偏且厚尾，也即中间部分空洞化，但是按照“人口空间法”所确定中等收入人口所占有收入的比重可能不会出现变化，体现不出人口分化导致中等收入人口的减少。一系列由人口空间法研究中等收入的问题都与实际的经济直观不符。

综合人口空间法研究中等收入问题时所存在的缺陷，可以归结为不能够体现出收入出现分化（包括社会整体收入分化和中等收入内部出现分化），而导致得出的中等收入定位和人口度量不够合理。在研究改进人口空间法的问题上，本文总结“人口空间法”之所以出现上述问题主要是因为收入的分布函数  $F(x)$  不同导致中等。不同的分布函数  $F(x)$  有不同的概率密度函数  $f(x)$ ，体现在分布函数的图像上即表现为不同的形态，而采用“人口空间法”取范围  $p_1$  到  $p_2$ （其中  $p_2 = p_1 + p$ ， $p$  为中等人口所占的比例）。再用此  $p_1$  到  $p_2$  的人口所拥有的收入占总收入的比例来描述中等收入人口的状态，进而计算中等收入人口的收入范围  $[x_l, x_h]$ 。

综上所述，本文认为，针对人口空间法我们应做的有意义的改进应该是中等收入的定义上，也就是  $p_1$ 、 $p_2$  的选取问题。目前国内外的研究中没有统一标准。同时， $p_1$ 、 $p_2$  的选取有着重要的现实意义，它不仅是用来描述中等收入人口所拥有的收入占总收入的比例，更能直接影响到中等收入人口的收入范围界定。因此我们认为针对“人口空间法”的改进应该是提出较好选择的  $p_1$ ， $p_2$  依据。对于题目中举出的选择  $p_1 = 20\%$  到  $p_2 = 80\%$ ，用此中间 60% 的人口所拥有的收入占总收入的比例来描述中等收入人口的状态。根据我国人口收入的实际情况以及题目中给定的数据，可以看出低收入人群所占比重偏大，如果按照  $p_1 = 20\%$ 、 $p_2 = 80\%$  中间 60% 的人群表示中等收入人口不能很好的描述我国的收入分布情况。在此我们提出应该把中等收入的概念加以推广，不但要研究中间收入人口的状态，还要研究各个收入等级的状，本文提出对人口空间法采取按人口分组排序的方法进行改进。

针对人口空间法，在选取分组时，可以考虑  $n$  个分组，求出不同组内的人口收入状态，对于中等人口的比较可以从对应分组做同一地区不同年份的纵向

比较和不同地区同一年份的横向比较。改进的具体做法为：取范围  $p_i$  到  $p_{i+1}$ （其中  $p_{i+1} = p_i + \Delta p, i=1,2,3,\dots,n; n$  表示分组数）。用此  $p_i$  到  $p_{i+1}$  的人口所拥有的收入占总收入的比例来描述不同人口分组的收入状态：

$$S_i = L(p_{i+1}) - L(p_i) = \frac{1}{\mu} \int_{p_i}^{p_{i+1}} F^{-1}(p) dp \quad (18)$$

（其中  $\mu = \int_0^1 F^{-1}(p) dp$  为平均收入），

根据不同地区的整体收入状态，我们可以确定出该地区内各个人口分组的收入情况。此时通过确定中等收入的区间来研究该地区的中等收入问题，计算出中等收入人口的收入范围  $[x_l, x_h]$ 。

### 3.2.4 改进后人口空间法的检验

根据题目中表一所给出的数据，来检验改进后的人口空间模型。

选取分组数为 20，即  $n=20$ ，这样把人口分成 20 个等间距的区间，并且依次按照收入增加排列。由公式（18）可以求出各个人口分组对应的收入状况，如表 2。

表 2 基于分组排序的“人口空间法”收入状态表

$p_i$	$p_{i+1}$	收入状态	$p_i$	$p_{i+1}$	收入状态
0%	5%	$S_1=0.0022$	50%	55%	$S_{11}=0.0451$
5%	10%	$S_2=0.0065$	55%	60%	$S_{12}=0.0494$
10%	15%	$S_3=0.0109$	60%	65%	$S_{13}=0.0540$
15%	20%	$S_4=0.0153$	65%	70%	$S_{14}=0.0588$
20%	25%	$S_5=0.0196$	70%	75%	$S_{15}=0.0640$
25%	30%	$S_6=0.0239$	75%	80%	$S_{16}=0.0701$
30%	35%	$S_7=0.0282$	80%	85%	$S_{17}=0.0777$
35%	40%	$S_8=0.0324$	85%	90%	$S_{18}=0.088$
40%	45%	$S_9=0.0366$	90%	95%	$S_{19}=0.1054$
45%	50%	$S_{10}=0.0408$	95%	100%	$S_{20}=0.1709$

按照中等收入人口的不同定义，由“人口空间法”测算的中等收入人口所占的收入比例也不一样。中等收入群体的界定应该具有综合性、相对性、动态性等几个方面，对于不同的社会区域和时间，中等收入的标准是不一样的，这需要根据所研究对象所处的社会状态而定。对于题目中给定的数据，由于不同机构、不同学者的立足点不同，采用的指标和划分标准不同，得出的中等收入标准不尽相同。这里为了检验改进的“人口空间法”，本文在参考相关文献的基础上，根据

表一数据拟合的概率密度函数分别选取 20%—80%、25%—85%、30%—90%、25%—80%、30%—85%、35%—90%的人口作为中等收入人群，分别计算各个中等收入定义下中等收入群体的状态，结果如表 3。

表 3 改进“人口空间法”不同中等收入定义下的检验结果

按“人口空间”中等收入划分标准	20%—80%	25%—85%	30%—90%	25%—80%	30%—85%	35%—90%
低收入人口所占比重	0.0349	0.0545	0.0784	0.0545	0.0784	0.1066
中等收入人口所占比重	0.5229	0.581	0.6452	0.5033	0.5571	0.617
高收入人口所占比重	0.4421	0.3644	0.2763	0.4421	0.3644	0.2763

由上表可以看出题目中所给数据所描述的收入分配结构中，中等收入所占社会总收入的 55%左右，低收入人口总收入所占比例不足 10%，高收入人口所占的收入比例约为 45%。按照 20%—80%来确定的中等收入的标准得到的收入分配状态如图（5）

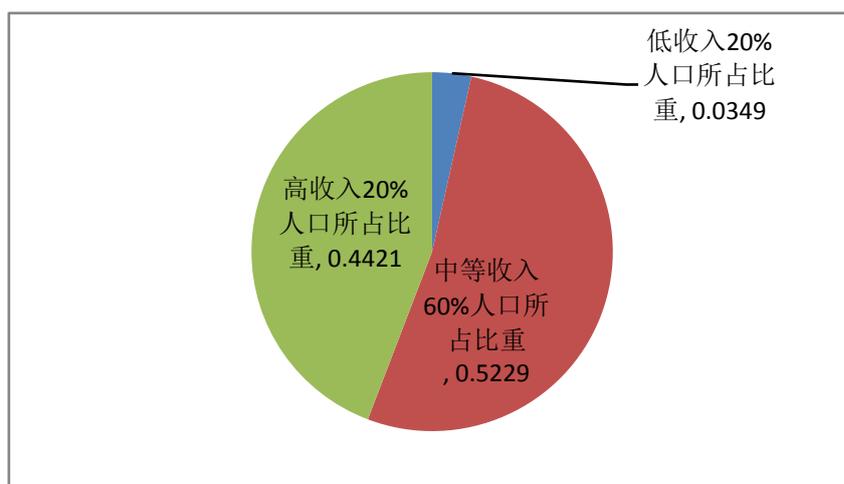


图 5 20%—80%来确定的中等收入的标准等到的收入分配状态

此时通过确定中等收入的区间来研究该地区的中等收入问题，计算出中等收入人口的收入范围[2331,9104]。

### 3.3 A、B 地区两年份的中等收入研究

针对问题三的要求，本文分别利用第二问改进的两种方法方法来解决问题：第一种，改进的收入空间法；第二种，改进后的人口空间法。并且计算出 A、B 地区两年份的基尼系数，综合评价 A、B 地区两年份的中等收入。

#### 3.3.1 基于改进收入空间法的研究

该部分运用改进的收入空间法，即“多因素收入空间法”，分别对  $A$ ， $B$  两个地区的两个不同年份的分组数据进行分析。

### 3.3.1.1 计量结果

1. 根据题目中表二至表五所给数据，可以得到两地区不同年份的均值和方差如下：

表 4 A、B 地区不同年份收入的基本统计量

地区	年份	均值	标准差
$A$	1	6281.34	2844.2
	2	8890.21	4087.1
$B$	1	16938.46	6230.8
	2	22228.53	8140.3

从上表中所给出的结果可以看出， $A$ ， $B$  两地区的居民收入均值都随着时间而又所增加，标准差也相应的有所增加，并且地区  $A$  的变动幅度要高于地区  $B$  的变动幅度。

2. 根据多因素收入空间法的计算公式：

$$\begin{cases} x_l = \max \{3500, \mu - \sigma\} \\ x_h = \min \left\{ \left(2 - \frac{\mu - x_l}{\sigma}\right)\sigma + \mu, \mu + 1.2\sigma \right\} \end{cases}$$

得到两地区不同年份的中等收入空间如下：

$$A1 = [3500, 9188.4]$$

$$A2 = [4803.1, 12977]$$

$$B1 = [10708, 23169]$$

$$B2 = [14088, 30369]$$

中等收入人口比例空间。由前面假设 (2)，分组数据中每一小区间上的数是均匀分布的，根据这个假设，通过计算，可以得到中等收入空间所对应的人口概率空间分别为：

$$P_{A1} = [0.2599, 0.8144]$$

$$P_{A2} = [0.2563, 0.8123]$$

$$P_{B1} = [0.2281, 0.8268]$$

$$P_{B2} = [0.2287, 0.8096]$$

中等收入人口比例。两地区每个年份中等收入区间所占的人口比例为：

$$p_{A1} = 0.8144 - 0.2599 = 0.5545$$

$$p_{A2} = 0.8123 - 0.2563 = 0.5560$$

$$p_{B1} = 0.8268 - 0.2281 = 0.5987$$

$$p_{B2} = 0.8096 - 0.2287 = 0.5809$$

从上述结果可以看出，随着年份二相对于年份一，*A*，*B*两地区的中等收入区间都有所增大，但是该区间内的人口比例变化并不是很大，说明，随着收入的较大幅度的增加，中等收入区间的人口比例在发生较为缓慢的变化。

6. 两极分化指标。根据两极分化指标公式：

$$\omega = 1 - \frac{\mu p}{2\sigma}$$

可以得到两地区每个年份中等收入化指标为：

$$\omega_{A1} = 0.224$$

$$\omega_{A2} = 0.3553$$

$$\omega_{B1} = 0.3553$$

$$\omega_{B2} = 0.3771$$

把上述计算结果总结到表格中，见表 5：

表 5 A、B 地区两年份收入空间法结果

	$\mu$	$\sigma$	中等收入区间	人口区间	人口比例	$\omega$	
A	年份一	6281.34	2844.2	[3500, 9188.4]	[0.2599, 0.8144]	0.5545	0.224
	年份二	8890.21	4087.1	[4803.1, 12977]	[0.2563, 0.8123]	0.5560	0.3553
B	年份一	16938.46	6230.8	[10708, 23169]	[0.2281, 0.8268]	0.5987	0.3553
	年份二	22228.53	8140.3	[14088, 30369]	[0.2287, 0.8096]	0.5809	0.3771

根据两地区两个年份所给的数据，拟合了各自的收入密度函数曲线（图 6）和洛伦兹曲线（图 7）。

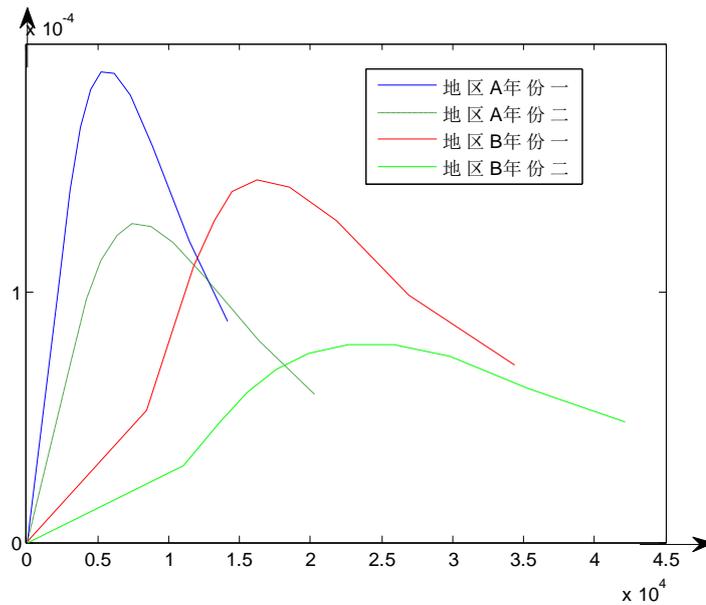


图 6 A, B 两地区不同年份的收入密度曲线 (横坐标为收入, 单位: 元, 纵坐标为概率)

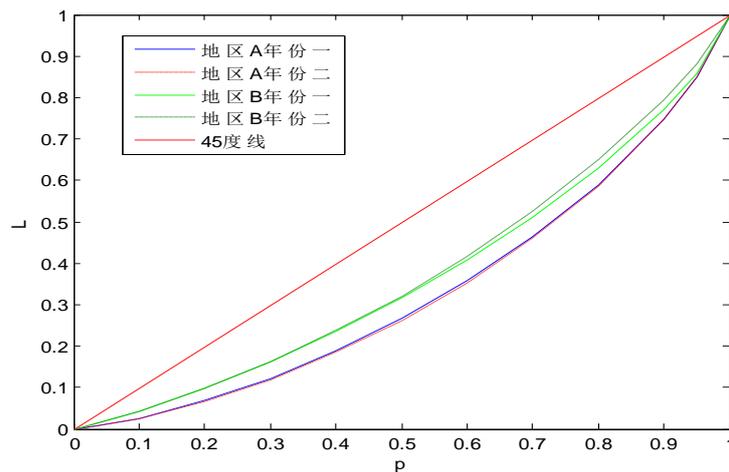


图 7 A, B 两地区不同年份的洛伦兹曲线

根据本文建立的洛伦兹曲线模型, 计算得到两地区两年份的基尼系数, 见表 6。

表 6 A、B 两地区不同年份的基尼系数

A		B	
年份一	年份二	年份三	年份四
0.3113	0.3153	0.2553	0.2371

### 3.3.1.2 结果解释

1. A 地区的特征及不同年份的比较。从表 1 的数据中可以看出, 相对于年份一, 年份二的平均收入有了明显的提高, 这表明随着经济的发展, A 地区的居民收入水平得到了很大的改善, 根据题目中所给出的数据, 可以算出, 该地区的平

均收入增长率为 41.53%，但是，分析数据可以发现，高收入人群的收入增长率要明显高于低收入人群的收入增长率，粗略计算，低收入人群的收入增长率大约为 37%，而高收入人群的收入增长率大约为 42%。同时，根据所给出的数据可以计算得出该地区不同年份的标准差有较大的差别，表现为，年份二的标准差要远高于年份一的标准差，说明，随着经济的增长，A 地区的收入差距拉大了，这也可以从两极分化指标看出，虽然，该地区两年分的两极分化指标都处于比较合理的范围，但是年份二的两极分化指标 0.3553，比年份一的两极分化指标 0.224 扩大了，这说明，A 地区随着经济的发展两极分化程度加重了，从表 6 可以看出，年份二的基尼系数要高于年份一的基尼系数，说明该地区的收入不平等现象加重了。

利用“改进的收入空间法”可以得出 A 地区两个年份的中等收入区间，从结果中可以得出，随着经济的发展，年份二的中等收入区间较年份一的收入区间有所增大，且区间内的人口比例有所增加，从人口区间可以看出，中等收入人口比例的增加主要是由一部分低收入人群向中等收入人群的转移造成的，同时，中等收入人群向高等收入人群略有转移。总体来看，A 地区的收入分配结构大体满足“橄榄型”，但随着经济的发展，A 地区两级分化程度加深，收入不平等现象有所加重，贫困人口的比例还相对较高，因此，今后应努力促进并扶持低收入人群往中等收入人群发展，并且预防两极分化的加深，促进社会的贫富差距的缩小。

2. B 地区的特征及不同年份的比较。从表 1 的数据中可以看出，相对于年份一，年份二的平均收入有了明显的提高，这表明随着经济的发展，B 地区的居民收入水平得到了很大的改善，根据题目中所给出的数据，可以算出，该地区的平均收入增长率为 31.23%，但是，分析数据可以发现，中间收入人群部分收入增长率要明显高于低收入人群的收入增长率，粗略计算，低收入人群的收入增长率大约为 31%，高收入人群的收入增长率大约为 27%，而中等收入人群的收入增长率大约为 39%。通过计算，该地区的两极分化程度都处于较为合理的水平，年份二较年份一有了一定程度的改善，两极分化指标由 0.3553 增加到 0.3771。同时，从表 6 可以看出，该地区的基尼系数有所减小，说明该地区的收入分配状况随着经济的发展得到了一定程度的改善。

利用“改进的收入空间法”可以得出 B 地区两个年份的中等收入区间，从结果中可以得出，随着经济的发展，年份二的中等收入区间较年份一的收入区间有所增加，但是区间内的人口比例有所减少，从人口比例区间可以看出，中等收入

人口比例的增加主要是由于中等收入人群收入的高增长率致使其向高等收入人群的转移造成的。总体来看，B地区的收入分配结构大体满足“橄榄型”，随着经济的发展，B地区两级分化和收入不平等现象有所改善（从上图7中可以明显看出年份二的基尼系数较年份一的基尼系数有所减小）。今后发展应继续改善两极分化的状况，促使社会贫富向更理想的“橄榄型”发展。

3. A、B地区之间的比较。从所计算的结果中可以看出，A地区收入水平与B地区比较要较低，其中等收入人口比例也比较低，大约低了3-4个百分点，说明经济较发达的地区，其中等收入人口比例要较高。从两极分化指标来看，在年份一，B地区的两极分化指数要明显高于A地区，说明B地区的两极分化程度要比A地区的两极分化程度严重，但是随着经济发展水平的提高，到年份二时，B地区较之A地区的两极分化程度有所减轻。另外，B地区的基尼系数要明显低于A地区的基尼系数，说明B地区的收入分配要优于A地区的收入分配格局。同时从两地区的收入增长率可以看出，A地区的收入增长率要明显高于B地区的收入增长率，平均增长率相差了10.3%。可能的原因是，在经济发展水平较低时，更加重视经济的快速发展，而对收入分配问题有所忽略；但是当经济发展水平较高时，收入分配问题成为关注的重点。政府应该在提高经济发展水平的同时，要预防收入分配格局的恶化，而当经济发展到较高水平时，政府应在发展经济的同时，考虑优化社会的收入分配结构，以实现共同富裕的目标。

### 3.3.2 基于改进人口空间法的研究

改进的人口空间法，在选取分组时，可以考虑 $n$ 个分组，求出不同组内的人口收入状态，对于中等人口的比较可以从对应分组做同一地区不同年份的纵向比较和不同地区同一年份的横向比较。改进的具体做法为：取范围 $p_i$ 到 $p_{i+1}$ （其中 $p_{i+1} = p_i + \Delta p, i = 1, 2, 3, \dots, n; n$ 表示分组数）。用此 $p_i$ 到 $p_{i+1}$ 的人口所拥有的收入占总收入的比例来描述不同人口分组的收入状态，根据不同地区的整体收入状态，我们可以确定出该地区内各个人口分组的收入情况。此时通过确定中等收入的区间来研究该地区的中等收入问题，计算出中等收入人口的收入范围 $[x_l, x_h]$ 。

#### 3.3.2.1 A地区两年份的中等收入变化

采用改进人口空间法对题目中表二和表三中A两地区前后两个不同年份的

收入分组数据进行研究，得到表 7 结果。

表 7 A 地区两年份基于分组排序的“人口空间法”收入状态表

年份一		年份二		
人口区间	收入区间	收入状态	收入区间	收入状态
[0%, 10%]	[0.00, 2228.28]	0.0250	[0.00, 3081.27]	0.0241
[20%, 20%]	[2228.28, 3066.03]	0.0423	[3081.27, 4199.72]	0.0410
[20%, 30%]	[3066.03, 3790.18]	0.0548	[4199.72, 5272.06]	0.0536
[30%, 40%]	[3790.18, 4519.24]	0.0661	[5272.06, 6383.72]	0.0656
[40%, 50%]	[4519.24, 5254.75]	0.0781	[6383.72, 7461.83]	0.0780
[50%, 60%]	[5254.75, 6166.38]	0.0906	[7461.83, 8751.34]	0.0909
[60%, 70%]	[6166.38, 7273.48]	0.1062	[8751.34, 10294.02]	0.1069
[70%, 80%]	[7273.48, 8813.52]	0.1270	[10294.02, 12500.51]	0.1264
[80%, 90%]	[8813.52, 11424.93]	0.1584	[12500.51, 16362.67]	0.1603
[90%, 95%]	[11424.93, 14171.91]	0.1008	[16362.67, 20288.83]	0.1020

按照人口空间法，我们按照人口的收入排序定义 30%至 90%的人口为中等收入人口，则中等收入人口的状态即定义为  $S = L(0.9) - L(0.3) = \frac{1}{\mu} \int_{0.3}^{0.9} F^{-1}(p) dp$ 。对

于 A 地区两个年份分别计算其中等收入状态，得到年份一中等收入人口的收入比重为 0.6264，年份二中等收入人口的收入比重为 0.6281。A 地区两年份的收入结构如图 8。

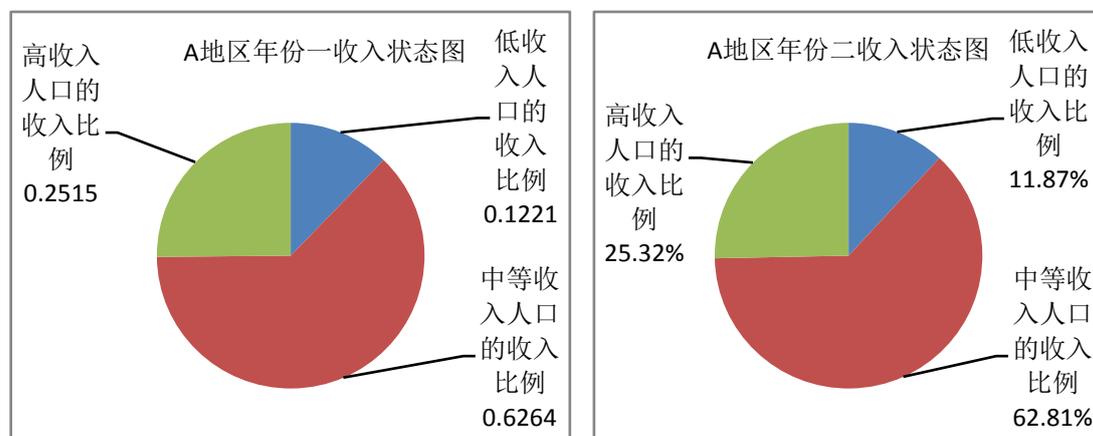


图 8 A 地区两年份的收入结构

对比 A 地区两年的收入结构，我们可以发现，就“人口空间法”定义下的中等收入人口的收入状态未出现较大变化，但是，年份二的中等收入人口所占收入比例相对于年份一提高了 0.27%。中等收入人口的财富状况出现了改善。一方面说明 A 地区的收入结构变化不大，另一方面中等收入人口的收入比重约为 63%，所占比重较大，可以看出 A 地区的收入结构满足中间大、两头小的“橄榄形”结构。

计算出 A 地区两年份中等收入的范围分别为 [3790.18, 11424.93] 和 [5272.06, 16362.67], 可见中等收入的收入水平提高了, 其中中等收入的下限提高了 39.1%, 上限提高了 43.22%, 中等收入上下限差距由 7634.8 元提高到了 11091 元, 说明按照“人口空间法”测算的中等收入的范围拉大了 45.27%。

### 3.3.2.2 B 地区两年份的中等收入变化

采用改进人口空间法对题目中表四和表五中 B 两地区前后两个不同年份的收入分组数据进行研究, 得到表 8 结果:

表 8 B 地区两年份基于分组排序的“人口空间法”收入状态表

年份一		年份二		
人口区间	收入区间	收入状态	收入区间	收入状态
[0%, 10%]	[0.00, 8465.55]	0.0427	[0.00, 11062.50]	0.0411
[20%, 20%]	[8465.55, 10293.33]	0.0551	[11062.50, 13531.18]	0.0559
[20%, 30%]	[10293.33, 11770.00]	0.0652	[13531.18, 15472.69]	0.0652
[30%, 40%]	[11770.00, 13173.47]	0.0737	[15472.69, 17599.77]	0.0747
[40%, 50%]	[13173.47, 14422.27]	0.0813	[17599.77, 19814.62]	0.0841
[50%, 60%]	[14422.27, 16246.88]	0.0904	[19814.62, 22681.13]	0.0953
[60%, 70%]	[16246.88, 18510.11]	0.1024	[22681.13, 25818.75]	0.1086
[70%, 80%]	[18510.11, 21794.50]	0.1182	[25818.75, 29848.37]	0.1250
[80%, 90%]	[21794.50, 26918.59]	0.1423	[29848.37, 35288.50]	0.1449
[90%, 95%]	[26918.59, 34375.61]	0.0883	[35288.50, 42150.00]	0.0856

对于 B 地区, 我们同样按照人口的收入排序定义 30%至 90%的人口为中等收入人口, 则中等收入人口的状态即定义为  $S = L(0.9) - L(0.3) = \frac{1}{\mu} \int_{0.3}^{0.9} F^{-1}(p) dp$ 。计

算其中等收入状态, 得到 B 地区年份一中等收入人口的收入比重为 0.6264, 年份二中等收入人口的收入比重为 0.6281。

对于 B 地区两个年份分别计算其中等收入状态, 得到年份一中等收入人口的收入比重为 0.6083, 年份二中等收入人口的收入比重为 0.6326。B 地区两年份的收入结构如图 9。

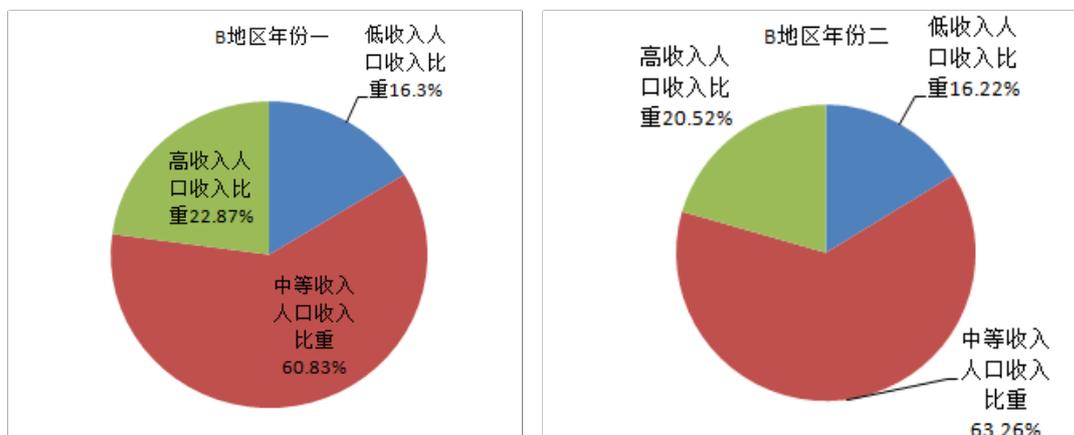


图 9 B 地区两年份的收入结构

对比 B 地区两年的收入结构，我们可以发现，就“人口空间法”定义下的中等收入人口的收入状态未出现较大变化，但是，中等收入人口所占的收入比重有 60.83%提高到了 63.26%，年份二的中等收入人口所占收入比例相对于年份一提高了 3.99%。中等收入人口的财富状况出现了改善。这与年份二的收入密度函数相对年份一的概率密度函数相一致，说明 B 地区年份二相对于年份一中等收入情况出现了较大改善。另外，B 地区的中等收入人口的收入比重约为 60%，所占比重较大，可以看出 B 地区的收入结构满足中间大、两头小的“橄榄形”结构。

计算出 B 地区两年份中等收入的范围分别为 [11770, 26918.59] 和 [15472.69, 35288.50]，可见中等收入的收入水平提高了，其中中等收入的下限提高了 31.46%，上限提高了 31.09%，中等收入上下限差距由 15149 元提高到了 19816 元，说明按照“人口空间法”测算的中等收入的范围拉大了 30.81%。

### 3.3.2.3 A、B 两地区对比分析

由 A、B 地区两年份基于分组排序的“人口空间法”收入状态以及定义出的中等收入可以横向比较 A、B 两地区的中等收入情况。根据 3.2.1 和 3.2.2 的计算结果，总结得到两地区的中等收入对比分析表，见表 9。

表 9 A、B 两地区中等收入对比分析

地区	A 地区		B 地区	
年份	年份一	年份二	年份一	年份二
平均收入(元)	6281.34	8890.21	16938.46	22228.53
中等收入状态	0.6264	0.6281	0.6083	0.6326
中等收入范围	[3790.1, 11424.9]	[5272.0, 16362.6]	[11770, 26918.5]	[15472.6, 35288.5]

从上表的对比结果可以看出，A 地区的平均收入水平低于 B 地区的平均收入水平，A、B 两地区的经济发展水平存在明显的差异，B 地区相对于 A 地区整体的收入水平高。同时，A 地区在两个年份中中等收入人口的收入状态变化较小，越为 63.%，改善幅度较小，年份二相对于年份一改善 0.27%；而 B 地区在两个年份中中等收入状态变化稍大，由年份一的 60.83%提高到了 63.26%，改善了 3.99%，说明 B 地区较之于 A 地区中等收入人口所占收入的比重改善更明显，也就是说 B 地区经济相对发达，同时中等收入的状态改善更显著。

### 3.4 中等收入人口的新模型及测算

党的十六大报告中提出要“以共同富裕为目标，扩大中等收入者比重，提高低收入者收入水平”，中等收入人口的提法顿时成为社会关注的焦点，而党的十八大报告中继续强调“收入分配差距缩小，中等收入群体持续扩大”，因此，以中等收入比重作为探究收入分配格局变化的指标是衡量经济社会发展的重要课题，对于理解扩大中等收入者比重的必要性和落实扩大中等收入者比重的战略措施具有重要的意义。

经济理论界将中等收入人口定义为中位收入附近的人口，但严格来讲，中等收入也不是一个单纯的经济学的概念，是一个集资产状况、生活质量、职业特征、区域特征等为一体的多元化概念，对于中等收入的界定应该更多的结合中等收入人群的特征。

#### 3.4.1 中等收入人群的特征

(1) 中等收入人群是一个综合的概念，它应该受多种因素的影响，而不能仅仅考虑每年居民收入的变动，最基本的影响因素应该包含居民所生活的城市以及区域的经济发展水平、居民消费水平以及居民可支配财产；

(2) 中等收入人群是一个相对的概念，中等收入必然对应着低收入和高收入，由于不同国家和地区的居民收入具有很大的差别，中等收入人群的收入区间具有较大的差异性，因此，对于国家的不同地区，不能进行一刀切，而只能从追求一个综合考虑多种因素的模型出发，具体测算出不同地区的中等收入人群所占的比例；

(3) 中等收入人群应是代表了一类处于特定生活状态的人，这种状态的人应该类似小康水平，这类人的生活特征应该包括：不为衣食住行所困扰，赚取的收入以及所拥有的流动资产可与自己的负债相抵还能有结余以供自己追求生活或者精神上更高层次的消费，但是这种更高层次的消费不是奢侈。

根据上述中等人群的特征，本文提出了中等人口的定义、确定中等收入人群的原理以及其所包含的经济学意义。最后本文给出了与之相应的中等收入测算模型。

#### 3.4.2 中等收入人群的定义

中等收入人群是指，处于社会某种生活状态的一类人群，该类人群所获得的

收入能够使其在所生活的地域内，除满足基本的生活需要和个人长期负债以外，还有剩余能力追求更高层次的生活和精神的消费，而这种更高层次的消费是在一定程度内生活的改善，是享受而不是奢侈，该类人群处于整个社会中间群体中相对集中的部分。

### 3.4.3 确定中等收入人群的原理及经济学意义

在开始介绍具体原理和经济学意义时，首先需要定义一个名词概念：效用指数。效用指数是指居民在一定区域范围内从所获得的收入中所获得的效用的满足感。这种满足感随着收入的增加而增加，但是增加的速度越来越慢。

本文通过效用指数来界定中等收入人群，效用指数是通过定义在居民收入上的效用指数模型测算出来的，该效用指数受不同区域的经济发展水平和居民消费水平等因素的影响。本文划定效用指数在区间[0.5,0.9]内的人群为中等收入人群。这种界定方法具有其合理性，因为中等收入是一个综合的概念，单纯的从收入和人口比例方面来划分中等收入不仅具有较大的随意性，而且所划分出的中等人群可能存在生活水平并不高、消费需求并不旺盛等缺点。因此，通过效用指数空间来界定中等收入人群，不仅仅从居民收入出发，而且还综合考虑了多种经济因素，从而使得该方法所确定出来的中等收入人群所达到的生活状态更加理想。

### 3.4.3 中等收入新模型的建立

恩格尔系数是食品支出总额占个人消费支出总额的比重。19世纪德国统计学家恩格尔根据统计资料，对消费结构的变化得出一个规律：一个家庭收入越少，家庭收入中（或总支出中）用来购买实物的支出所占的比例就越大，随着家庭收入的增加，家庭收入中（或总支出中）用来购买食物的支出比例则会下降。推而广之，一个国家（地区）越穷，每个家庭的平均收入中（或平均支出中）用于购买实物的支出所占比例就越大，随着国家（地区）的赋予，这个比例呈下降趋势。联合国根据恩格尔系数的大小，对世界各国的生活水平有一个划分标准，即一个国家平均家庭恩格尔系数大于60%为贫困；50%–60%为温饱；40%–50%为小康；30%–40%属于相对富裕；20%–30%为富足；20%以下为极其富裕。

本文构建了效用指数模型来反应居民从收入中所获的满足感，通过界定效用指数空间来确定中等收入人口区间。由于相同的收入在不同区域间所获得的效用会因为区域间的经济发展水平不同而有所差别，因此，本文通过恩格尔系数定义修正项，来反映不同区域对效用指数的影响。

### 3.4.3.1 模型假设

假设一个地区的恩格尔系数和平均消费水平能够完全反映该地区的经济发展水平和消费物价水平。

### 3.4.3.2 符号说明

1.  $U$  为居民收入的效用指数；
2.  $A$  为居民收入修正项；
3.  $EC$  为所研究区域的平均消费水平；
4.  $I$  为居民收入
5.  $\alpha$  为恩格尔系数

### 3.4.3.3 模型建立

本文建立的效用指数函数如下所示：

$$U(A, I) = 1 - e^{-\frac{I}{A}}$$

其中  $U(A, I)$  为居民收入的效用指数， $U \in [0, 1]$ ； $I$  为居民收入； $A$  为修正项，它受地域因素的影响。

该修正项需满足的性质为：

1. 平均消费水平  $EC$  越高，修正项  $A$  越大，表示同等收入水平下，在平均消费水平较高地域的居民从收入中所获得的效用越小；
2. 恩格尔系数越高，修正项  $A$  越小，在越贫困的地区，居民从同等收入水平中所获得的效用越大。

因此将修正项  $A$  定义为：

$$A = f(EC, \alpha) = (2 - \alpha)EC, \quad 0 < \alpha < 1$$

因此，效用指数模型可以确定为：

$$U(EC, \alpha, I) = 1 - e^{-\frac{I}{(2-\alpha)EC}}$$

该模型满足的性质有：

$$U(0) = 0$$

$$U(+\infty) = 1$$

$$U'' = -\left[\frac{1}{(2-\alpha)EC}\right]^2 e^{-\frac{I}{(2-\alpha)EC}} \leq 0$$

不同地区的恩格尔系数大小不同，考虑下面的特殊情况，当某居民的收入为 0 时，该人从收入中所获得效用为 0；当某居民收入为无穷大是，该人从收入中所获得的效用为 1；当某个区域的恩格尔系数为 50%时，当收入等于该地区的平均消费时，效用指数为 0.4866；根据上述结果做相应参考，本文将中等收入人群的效用区间界定为[0.5, 0.9]，表示从收入中获得的满意度为 50%到 90%的一类人。

#### 3.4.3.4 模型检验

针对题目中所给出的表一进行检验，由于所给表中缺乏相关数据，因此我们在此对数据做一些相关的假设（此假设并不影响对结果的分析）。假设平均消费水平为 3600 元/月，该地区的恩格尔系数为 50%。

由此可以通过效用函数得到中等收入人群的收入区间：

$$0.5 \leq 1 - e^{-\frac{I}{(2-0.5)3600}} \leq 0.9$$

$$0.1 \leq e^{-\frac{I}{5400}} \leq 0.5$$

$$3742.7 \leq I \leq 12434$$

因此可以确定中等收入人群为[3742.7,12434]，中等收入人群所占比例区间为[0.3289,0.8991]，中等收入人群所占的比例为 57.02%。

#### 3.4.3.5 模型评价

该模型将区域发展水平作为确定中等收入人群的重要考量因素，运用恩格尔系数和平均消费水平作为修正因素，确定的收入区间更具有合理性；通过定义效用函数来确定效用空间，将居民的效用做为确定中等收入人群的指标，可以有效避免收入空间法和人口空间法单纯从数字方面界定中等收入人群的缺陷。但是该模型还有许多地方可以改进，例如效用区间的界定还需要找到更加合理的依据。由于时间的限制，不能通过数据检验来确定该模型的通用性和优越性。

## 四、结论及政策建议

### 4.1 结论

本文重点研究中等收入定位与人口度量模型，题目中所给的问题成为我们研究的一条主线，首先通过研究问题一，构建出新的洛伦兹曲线模型，为下文收入分布的拟合等奠定了基础；其次，对于中等收入人口的界定，题目中提出了收入空间法和人口空间法，因存在缺陷，使得模型的改进势在必行，根据题目三所给数据，利用改进的收入空间模型和人口空间法进行定量描述，使得改进的方法得到了检验；最后根据以上模型的建立，结合自身的价值判断，提出符合经济规律的中等收入人口，提出中等收入人口的新定义。通过建模研究，本文得出以下结论：

第一，为了更好的拟合洛伦兹曲线，学者们不断的对洛伦兹模型进行改进和创新，在研究了国内外学者的成果后，本文在满足洛伦兹曲线性质的前提下，构建了基于指数成份的洛伦兹模型： $L(p)=[1-(1-p)^\alpha e^{-\beta p}][1-(1-p)^\gamma e^{-\eta p}]$ ，利用题中所给数据，将新洛伦兹模型与文中提及的其他模型进行拟合对比，在所有比较模型中，本文构建的新模型拟合的误差远小于现有的大部分洛伦兹模型的拟合误差，并且拟合优度几乎为 1，充分证明了本文构建的新模型是一个较好的洛伦兹模型。

第二，收入空间法和人口空间法存在的区间界定随意性的缺陷，对于收入空间法，本文本着动态的原则，加入税收因素，将个人所得税起征点作为确定中等收入人群下限的考量因素，建立了多因素收入空间动态模型，同时引入判定两极分化程度的指标  $\omega$ ，为两极分化现象的分析提供依据。通过表一所给数据进行模型验证时，利用收入空间法确定的中等收入区间为[3500,12513]，中间人口所占的比例为 0.5910，与中国现行的收入分配格局大体一致，并且计算出的两极分化指标为 0.6231，两极分化程度比较明显，总体上还没有达到理想的“橄榄形”（只针对所给数据的分析）。而对于人口空间法，本文主要按照人口分组排序来改进，提出较好选择的  $p_1$ ， $p_2$  依据，即增加分组数量。根据表一数据拟合的概率密度函数分别选取 20%—80%、25%—85%、30%—90%、25%—80%、30%—85%、35%—90%的人口作为中等收入人群，中等收入所占社会总收入的 55%左右，这与现实情况的偏离在可接受的范围内。

第三，本文通过改进的收入空间法和人口空间法对 A、B 地区两个年份的中等收入情况进行定量描述。基于收入空间法，A 地区收入水平与 B 地区比较要较低，其中等收入人口比例也比较低，大约低了 3-4 个百分点，说明经济较发达的地区，其中等收入人口比例要较高。基于人口空间法，A 地区在两个年份中中等收入人口的收入状态变化较小，约为 63. %，改善幅度较小，年份二相对于年份一改善 0. 27%；而 B 地区在两个年份中中等收入状态变化稍大，由年份一的 60. 83% 提高到了 63. 26%，改善了 3. 99%，说明 B 地区较之于 A 地区中等收入人口所占收入的比重改善更明显，也就是说 B 地区经济相对发达，同时中等收入的状态改善更显著，这与收入空间法得到的结论一致。

第四，从效用函数的角度出发，对中等收入人口进行新的界定为，处于社会某种生活状态的一类人群，该类人群所获得的收入能够使其在所生活的地域内，除满足基本的生活需要和个人长期负债以外，还有剩余能力追求更高层次的生活和精神的消费，而这种更高层次的消费是在一定程度内生活的改善，是享受而不是奢侈。本文提出了效用指数模型，来反映居民从收入中所获的满足感，通过界定效用指数空间来确定中等收入人口区间。经过表一的数据进行验证，得到的中等收入人口比例为 57. 02%，与其他方法所得结论相差不大，从而证明了此种方法的可行性。

## 4. 2 政策建议

优化收入分配格局是国家宏观调控的重要目标之一，而中等收入人口比重则可以作为反映收入分配格局的重要指标。一般来讲，这一人口比重越大，意味着收入分配结构越合理，“橄榄型”收入分配格局被誉为理想的收入分配格局。在这种收入分配格局下，收入差距不大，社会消费旺盛，人民生活水平高，社会稳定。我国正处于经济转型期，“橄榄型”收入分配格局正处于形成阶段，继续优化我国收入分配格局，整体提高居民收入，对于促进我国经济发展，实现 2020 的收入倍增计划具有重要的战略意义。

从宏观上讲，优化我国收入分配格局，深入收入分配制度改革，提高中等收入人群比重首先要依靠政府职能，继续完善我国社会保障体系，统筹城乡发展，最大限度地降低中等以下收入人群的比重，改善社会福利制度，增加就业机会，实现低收入人群向中等收入人群的跨越。其次，优化我国税收结构，根据我国现行分配格局，建立合理完善的税收体系，适当的提高税收起征点，减少中等及以

下收入人群的税收压力。最后，继续坚持“稳增长、调结构、促改革”的主线，实施经济发展战略，促进经济稳步增长，调整产业结构，以发展第三产业为动力，结合我国国情，将劳动密集型和技术、资本密集型产业结合，加快产业创新，普遍提高全国居民的收入。

本文针对我国现行收入分配格局，结合定量分析结果，提出政策建议。通过定量分析结果发现，高增长率往往伴随着两极分化程度的加重现象的发生，因此在追求经济高增长率的同时，防范两极分化的工作应该同步进行，避免掉进“中等收入陷阱”中；当经济发展到较高水平时，两极分化现象会有所改善，此时，政府应抓住机遇，合理优化社会分配格局，使较高的经济发展水平真正成为实现共同富裕的动力。

## 参考文献

- [1] Chotikapanich, D., D. S. P. Rao, K. K. Tang. Estimating income inequality in China using grouped data and the generalized Beta distribution. *The Review of Income and Wealth*[J]53:127-47, 2007.
- [2] Foster, J. E., M. C. Wolfson. Polarization and the decline of the middle class: Canada and the U.S. *Journal of Economic Inequality* [J] 8:247-273, 2009.
- [3] Wang, Z. X., Y-K Ng, R. Smyth. A general method for creating Lorenz curves. *The Review of Income and Wealth*[J]57:561-582, 2011.
- [4] Wang, Z. X., R. Smyth. A hybrid method for creating Lorenz curves with an application to measuring world income inequality[J]1-48, 2013.
- [5] Sen, A. .Poverty: An ordinal approach to measurement. *Econometrica*[J]44: 219-232, 1976.
- [6] Ogowang, T., U. L. Gouranga Rao. Hybrid models of the Lorenz curve. *Economics Letters*[J]69:39-44, 2000.
- [7] Sarabia, J. M., E. Castillo, M. Pascual, M. Sarabia. Mixture Lorenz curves. *Economics Letters*[J]89:748-756, 2005.
- [8] 陈家鼎, 陈奇志. 关于洛伦兹曲线和基尼系数的统计推断. *应用数学学报*[J]34: 385-400, 2011.
- [9] 王祖祥. 几种著名经济评价指标及构造方法. *东岳论丛*[J]3: 31-34, 2000.
- [10] 龙莹. 中国中等收入群体规模动态变迁与收入两极分化:统计描述与测算. *财贸研究*[J]2: 92-99, 2012.
- [11] 徐建华, 陈承明, 安翔, 对中等收入的界定研究, *上海统计*. 8: 12-14, 2003.
- [12] 庄健, 张永光. 基尼系数和中等收入群体比重的关联性分析. *数量经济技术经济研究*[J]4:145-152, 2007.
- [13] 宏观经济研究院经济和社会发展研究所课题组. 中等收入者的概念和划分标准. *宏观经济研究*[J]5:53-55, 2004.