**SES指标的计算方法**

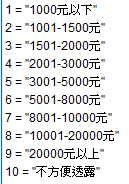
SES的合成已经有大量前人文献给出说明，下面以常用问卷对人口学问题设置的选项为例详细说明SES指标合成步骤，本文主要参考国内对SES有丰富研究经验的学者任春荣教授的研究 [1] 。

1. 收入、职业、教育程度指标的预处理

首先要保证研究总体在收入、职业、教育程度三个指标上均有有效填答（事先进行用户清洗，缺失值剔除）。之后分别对三个指标进行由低到高的等级排列：

* **收入处理方法（以CNNIC收入分类为例）（Z5）：**

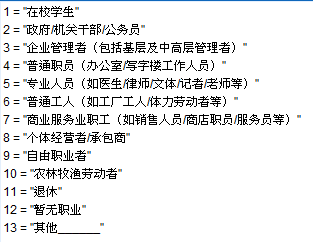
Z5.您的家庭月收入（从各种途径得到的全部收入总和，包括零花钱）为？



将收入选择10（不方便透露）的用户剔除

* **职业处理方法（以常见的职业分类为例）（Z4.1）：**

Z4您目前的职业是？



将职业选项合并，按照等级从地到高排列（Z4.1）：

11/12/13→1 其他

1 →2 学生

10/9/8 →3 个体经营者、自由职业、农林牧渔

7/6 →4 普通蓝领

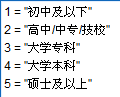
5/4 →5 普通白领

3/2 →6 高级白领

注意：目前常见的研究论文在计算个体SES时将在校与在职被试分开研究[2]或改为计算家庭SES，但实际项目中很少有家庭全部成员职业、学历的统计，研究也经常涉及在职和在校两类人群，因此这里采用待验证的折中方法：将学生的职业等级赋值为2（仅高于退休、无业和其他），此方法参考流行病学对在职在校群体的研究[3]。

* **学历题划分方法（以常见的学历分类为例）（Z3）：**

Z3.到目前为止，您的最高学历(包括在读)是？



学历选项已经是按照等级由低到高排列的，不需要进一步处理

1. 3种SES合成方法
2. 简单加和

简单加和法是将问题选项所代表的社会阶层由低到高排列并赋值（见第一部分），然后将作答选项加和。即：

**SES = Z5 + Z4.1 + Z3**

1. 加权均数

加权均数法是在简单加和方法基础上，以样本群体在各选项的分布比例为权重，将作答选项乘权重后加和，这个方法严重受样本分布影响（因此不推荐使用）。

**SES = a \* Z5 + b \* Z4.1 + c \* Z3**

(a/b/c为根据样本分布计算出的收入、职业、学历的权重)

1. 因子分析

因子分析法首先确定主因子个数，计算研究群体在各因子上的得分。若主因子个数多于一个，需要对主因子二次综合，以主因子负荷（β1、β2、β3）为权重，乘以各因子分，加和后除以第一个因子的特征值（εf），即获得每个研究对象的SES得分。

**SES = (β1 \* Z5 ＋ β2 \* Z4.1 + β3 \* Z3) / εf**

正文内保险大调研和cdc美妆数据的验算选择简单加和法得出，因为这是三种算法中将业务指标分离得最好的。但此选择不能代表所有商业研究案例，建议对SES指标感兴趣的同学具体情况具体分析，选择方法进行用户分类前进行三种尝试，。

参考文献：

[1] 任春荣. (2010). 学生家庭社会经济地位(ses)的测量技术. *教育學報,6(5)*, 77-82.

[2] 李春玲. (2005). 当代中国社会的声望分层——职业声望与社会经济地位指数测量. *社会学研究(2)*, 74-102.

[3] 董碧蓉, 葛宁, & 刘关键. (2001). 社会经济地位、环境因素及个人行为对肺结核发病危险的研究. *中华流行病学杂志*, 22(2), 102-104.