

山东省新高考分数转换模型解析

一、引言

分数模型构建是招生考试制度设计的基础性工作，分数结构一方面建立了对考生进行能力测量的标准，另一方面构建了基础教育知识体系和面向招生专业考生群体的分类。因此，从根本上讲，分数模型决定了考试所建立的基本竞争选拔体系，也决定了基础教育的根本导向。在新高考制度中，建立了‘3+3’的学科组合模式，即共同科目（语文、数学、外语）的‘基础3’，和六科目（物理、化学、生物、地理、历史、政治）中选择的‘加3’，共同科目每科150分，‘加3’科目中每科100分。同时，‘加3’科目实行等级分数制度，将原始分数转换为等级分数，然后再作为录取分数加入总分。等级分数和原始分数需要建立一定的对应关系，应用这种对应关系将原始分数转换到等级分数，也就是分数转换模型。

等级分数是新高考改革中的主要改革点，也是使用高中学业水平考试成绩与全国统一高考成绩融合的举措。等级分数需要进行分数转换模型的设计，理论上存在很多转换方式，由于全国各省的具体情况不同，也需要根据自身情况进行研究，建立符合本省省情的转换模型。在第二批改革试点中，山东省推出了基于等级划分的分段线性转换模型，既建立了

等级划分理念，也保持了分数区分度，比较适用于考生人数多和招生规模大的自身情况。

二、分数与分数转换

按照教育测量的概念，考试分数即是对考生能力评测的结果，反映了考生在测量目标上表现出的水平。测量结果存在于一个可解释的度量空间中，也就是所谓的量表。这个量表中的值对考生来说具有比较直观的解释意义，对考试的应用也十分重要，要考虑使用中的诸多要求，包括解释直观性（直观意义）、可比较性（区分度、精度）、水平呈现（难度）、学习指导（知识点水平）、专业趋向（专业发展潜质）等。要实现分数较好的应用特性，就需要进行分数转换，即将原始分数转换为固定分布形态的量表分数。

如果将考生进行的考试视为统计学中随机试验，则考试结果是考生分数总体的一个抽样过程，考试分数则是随机变量。按照大数定律，当抽样数很大时，这个随机变量应该接近正态分布，所以量表分数大多采用正态分布。

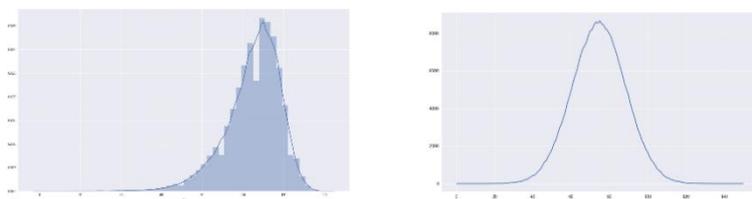


图1 山东省某年某科目高考分数分布（左），标准正态分布（右）

图一中分数分布的统计数据表明该科目分数分布近似正

态分布，呈现一定的负偏态分布形态：min（最小值）：0，max（最大值）：142，mean（中位数）：104.68，variance（方差）：199.31，stdandard deviation（标准差）：14.12，skewness（偏度）：-1.3892，kurtosis（峰度）：4.4581。

将原始分数（卷面的得分）分布转换到正态分布，需要拟合分布密度函数，即对应各个区间的得分频数，按照转换后分数形态计算。转换后的分数设置为固定的均值和方差，形成所谓的标准分数。

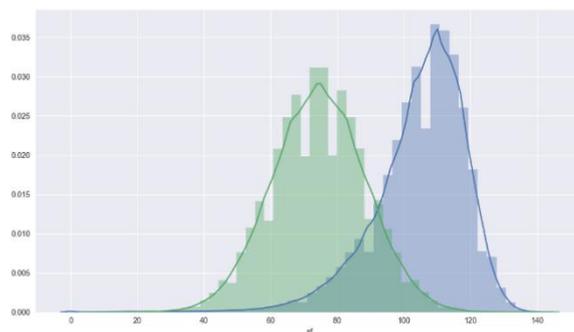


图2 原始分数到正态化分数转换

在分数转换模型中存在几个经常使用的分数定义，也是考试理论中进行分数解释的基本概念。

原始分数（简称原始分）：试卷中各个答题所得分数直接相加得到的卷面分数。具有与考生测试情况直接对应，直接反映答题水平的特点，表现直观，易于解释和理解。目前我国很多大规模考试成绩均采用原始分。原始分对区分度、难度、变化尺度等没有直接信息，不能充分反映测试的信息，

一般为非等距、无原点分数。

等级分数：在某一特定的范围内，对某一特定的人群进行等级划分所给定的分数，叫做等级分数。等级分数的制定一般分为按分数等级和按相对位置来制定。

按分数等级制定是将受试群体的考试分数按照事先规定的分数段标准对应相应的等级。如英国高校入学招生考试即 A-Level 证书考试的成绩采用 A*、A、B、C、D、E、U 等共七个等级，这七个等级分别对应的百分率区间为 90%-100%，80%-89%，70%-79%，60%-69%，50%-59%，40%-49%，40%以下。

按相对位置制定等级分数是根据个体的成绩在同类受试群体中所处的相对位置来确定分数等级，即按照成绩由高到低划分若干等级，规定各等级的人数百分比，再根据个体成绩所属区间确定等级分数。如韩国大学修学能力考试 (CSAT) 成绩报告采用九分制，即等级分数由高到低依次为 9 分，8 分，…，1 分，受试群体的成绩由高到低按比例划分分别为 4%，7%，12%，17%，20%，17%，12%，7%，4%，与 9 分~ 1 分分别对应。这种九分等级制分数也称为标准九分数，标准九分数最早被普遍运用于美国空军的心理测验中。

按此划分规则，还可以定义百分位分数，也称百分位，它标记一个个体分数值在整体中所占百分比位置，即不高于该个体成绩的人数所占的百分比。

标准分数：也称标准分，是一种由原始分推导出来的相

对地位量数，具有预定（一般指正态）分布形态（均值、标准差、标准差区间）的分数量表，同时规定原始分数到标准分数这个量表转换方式。标准分数具有以下常用分数。

Z 分数：符合标准正态分布（均值为 0，标准差为 1）的分数量表。一般的 Z 分数量表使用 8 个标准差，即分值点从 -4 到 +4 分，转换后小于 -4 的对应到 -4，大于 +4 的对应到 +4 分。

T 分数：符合特定正态分布（均值 50，标准差 10，分数区间为 10-90）的分数量表。为了消除小数或负数的情况，对 Z 分数进行线性转换，转到更大的标准分数量表上，即 $T=10Z+50$ ，其中 10 为标准差，50 为平均数。

标准九分数：前面提到的标准九是标准化九分制的简称。它是以 2 为标准差，以 5 为平均数的一个分数量表，即标准九 $=2Z+5$ 。目前的雅思考试（IELTS）和韩国大学修学能力考试（CSAT）等均采用标准九分数。

CEEB 分数：该分数是美国大学入学考试委员会（College Entrance Examination）采用的分数，它采用的是标准差为 100，平均数为 500 的标准分，即 $CEEB=100Z+500$ ，目前 TOEFL、GRE、SAT 等考试均采用 CEEB 分数。

离差智商分数（Deviation IQ）：离差智商是智商的一种形式、是智力发展水平测试指标，由美国心理学家韦克斯勒在智力测验中首创。离差智商的平均数通常为 100，标准

差在韦克斯勒智力量表中为 15,在斯坦福-比内量表中为 16。
离差智商=15Z +100 或 16Z+ 100。

我国大学英语四、六级考试 (CET) 自 2005 年开始使用
标准差为 70, 平均分为 500 的标准分数, 即 $CET = 70Z + 500$

三、分数转换模型

分数转换是将原始分数转为标准量表分数的过程, 是一个根据原始分数总体分布情况和转换后分数要求的计算模型。根据转换后分数的形态和转换质量要求, 一般分为线性转换和非线性转换, 线性转换是直接将原始分数视为正态分布, 使用线性平移缩放的公式, 将原始分数映射到标准分数。

线性转换的公式一般为:

$$Y = \frac{S(x-m)}{s} + M$$

其中 m, s 为原始分数的均值和方差, M, S 为转换后分数的均值和方差。

线性转换公式比较直观简单, 转换过程的计算也比较容易实现。但是, 由于原始分数分布的非正态性和不确定性 (每次考试、每个科目都不相同), 转换后分数的分值点都不一定能够对应到标准分数量表的相应分值, 所以给分数解释带来一定的问题, 一般在正规考试中不会使用。

非线性变换一般采用百分位严格对应的方式, 即将原始分数根据其百分位转换到标准分数相同百分位的分值。这个转换过程需要生成一个转换对应表, 计算过程较为复杂。非

线性转换使得转换后分值具有明确的标准分值意义，即其与正态分布的变量值对应，可以直接导出百分位和标准差间距，是各类大型考试中经常使用的方式。

非线性分数正态化转换过程：

【原始分 f 】 \rightarrow 【原始分百分位 $p(f)$ 】 \rightarrow 【标准分量表分值 $F(p)$ 】

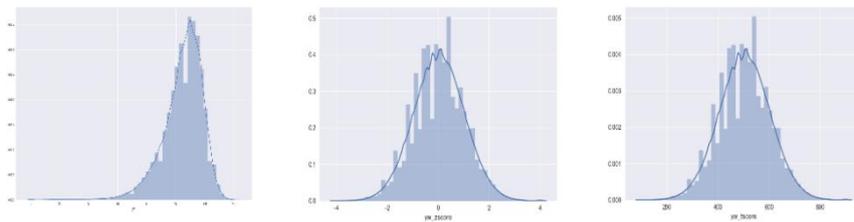


图 3 非线性标准分数转换示例

(从左至右依次为原始分数、Z 分数、T 分数)

如果将线性与非线性方式进行结合，可以设计一类分区间线性转换的模型，即分段线性转换，当这些区间形成划分等级时，也可以称为分级线性转换模型，简称分级线性模型。分级线性模型的具体方式是，将原始分数值分为几个顺序连接且不重叠的区间，每个区间的端点按照非线性转换方式对应到标准分数的分值点，然后各个区间内的分值按照两个端点映射值所形成的线性公式进行转换，这类转换适合等级与连续转换相结合的方式。

分级线性转换的实现过程可以分为以下几个步骤：

(1) 区间划分：根据一定的原则将原始分数划分为若干顺序区间，同时制订转换分数的顺序区间，

(2) 端点映射: $f: \text{原始分数端点 } x_1, x_2, \dots, x_n \xrightarrow{f} \text{标准分数端点 } f(x_1) = y_1, f(x_2) = y_2, \dots, f(x_n) = y_n$ 。

(3) 分 值 转 换 :

$$f(x) = [(f(x_{k+1}) - f(x_k)) / (x_{k+1} - x_k)] * (x - x_k) + f(x_k), x_k < x < x_{k+1}$$

(4) 异常处理: $f(x) = f(p_1), 0 < x < x_1$, $f(x) = f(p_{n-1}), x_n < x \leq$ 最高原始分数。如果设定 x_1 为原始分数最低分, x_n 为原始分数最高分, 则不存在异常处理问题。

在实际应用中, 需要将上述算法转换为分数数据处理过程, 并且要注意分数处理的一些原则, 如零分计算问题, 即是否将零分考虑在计算之中, 缺考考生与实际考试得零分考生的区别等。

山东新高考方案分数转换模型采用了上述分级线性转换模型, 并确定了其中的一些原则, 其分数转换实施方法如下:

步骤一: 制订区间划分原则。山东方案提出了基于正态分布划分考生分数区间的原则, 即根据标准分数量表得到各个区间比例 p_1, p_2, \dots, p_{n-1} , 。按照既定比例将原始分数划分为相应的区间 $p_1: [x_1, x_2], p_2: [x_2, x_3], \dots, p_{n-1}: [x_{n-1}, x_n]$ 。由于标准分数量表一般为固定分布形态 (如正态分布), 所以比例划分对应产生转换分数的区间端点 y_1, y_2, \dots, y_n 。由于在重合点上转换分值是一致的, 所以重合端点不影响其本身的转换计算, 但考虑分值的离散特性, 使用重合端点能够更好地保持模型的连续形态。山东方案确定的具体原始分数区间划分比例是 0.03,

0.07, 0.16, 0.24, 0.24, 0.07, 0.03, 对应的转换分数端点为 20, 30, 40, 50, 60, 70, 80, 90, 100。

步骤二：确定转换公式。由于分级转换的公式是根据原始分数分布形态动态计算得到的，所以，尽管区间划分原则是不变的，但是每次每个科目实施转换的公式是不同的，需要根据具体数据进行推算。在现在的计算技术支持下，这个计算过程可以很容易的实现。首先计算从低分到高分的分数分段统计表（也可以使用相反顺序的统计方式，对应的转换分数也需要顺序相同），其中包括每个分段的累计百分位，当对应到符合预定的累加比例值时，就可以得到原始分数的区间端点，然后分级上面的线性映射方法建立每个区间到转换分数区间的转换公式。

如，历史科目的 0 分是原始分最小端点，其累计百分位应该是 0，转换到端点 20，当计算到 12 分时累计百分位位 0.0298，最接近 0.03，即可以确定 12 分对应到转换分数端点 30。原始分数 0-12 分之间分值的转换公式为：

$$(30-20)/(12-0) * (x - 0) + 20 = 5/6 * x + 20$$

即每个 0-12 的原始分数 x 都可以使用上述公式转换为转换分数。

步骤三：制订总体转换公式，需要考虑保留的小数位、零分处理、作弊分数处理等问题，即针对特定考生的特定分值对公式进行附加的条件约定或改进。

步骤四：使用转换公式对相应科目的所有考生成绩进行转换

分数转换涉及考生评价和录取选拔竞争，是一个高敏感度的数据处理工作，需要考虑分值计算处理、分数解释、分数应用等诸多问题，所以是一个科学计算与社会应用的综合工程，需要进行系统分析、科学设计和严谨实施。

四、分数转换的相关问题

转换分数也称为导出分数，是正规考试项目中经常使用的分数形式。其优点是分数含义明确（可解释性）和分值稳定（标准量表），缺点是隐含了原始答题情况和难度，对重视结果精确性和群体竞争性的考试项目存在结果解释复杂问题。在设计转换模型时，高敏感性和竞争性考试往往涉及较多的应用需求和分数解释问题，需要深度考虑分数特性变化所产生的各种影响。

（1）区分度影响。

如果考试分数的应用是精确性和竞争性的，需要考虑转换过程是否损失分数区分度，即分值的差异性。如，原始分值是 0-100，而转换后为标准 9 段分数，则会损失考生之间分值的差异性，等于加大了分数的尺度，由原来的尺度 1 转为尺度 $100/9$ 。

影响区分度的还有转换量表的方差和分数区间。如果原始分数分布区间为 0-100，转换后量表为 30-90，则区间区

分度会损失 30%。在转换中对区分度影响较大的还有方差(其实也代表着总体信息量), 如果区间没有变化, 而方差缩小了, 区分度也会减少。

(2) 顺序性影响。

考试分数是教育或心理智力行为的水平度量, 当特别注意比较和分辨时, 其顺序关系是十分重要的。如果分数转换过程改变了群体中个体之间的顺序关系, 则该转换一般会受到质疑。这种改变会分为两种情况: 一类是顺序逆转, 在原始分中分数 $a > b$, 转换后的分值 $x(a)$, $y(b)$ 出现 $x(a) < y(b)$; 另一类是顺序消失, 即转换后出现 $x(a) = y(b)$ 的情况。

转换模型的设计一般需要避免顺序逆转, 但很难会避免出现顺序消失情况, 一般是由于分值区间的缩小、小数位的取舍、尺度变大等所造成的。山东新高考方案所使用的转换, 不会出现顺序逆转情况, 由于转换区间的缩小(由 $[0, 100]$ 转换为 $[20, 80]$), 如果不保留小数位, 会出现极少量的顺序消失。

(3) 下移问题。

当使用的标准量表与原始分数处于同一区间或具有重叠区间时, 会产生转换分数与原始分数的比较, 其中较为敏感的是分数下移问题。如果一个分数值在转换后比原始分数小了, 称为分数下移, 反之则称为上移。尽管这种比较没有实质意义, 但大多数转换模型会避免分数下移产生的比较问

题。图 4 中展示了使用分级线性转换模型的转换情况，在原始分数区间 $(28, 74)$ 上出现了下移，即转换分数小于原始分数的情况，而两端分数区间 $([0, 28), (74, 100])$ 都是上移。

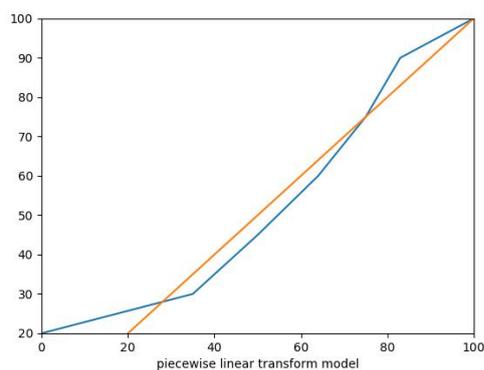


图 4 转换分数上移和下移

(4) 权重影响。

当使用多个科目合成总成绩时，每个科目转换分数量表的制定会影响到该科目的权重问题，即该科目成绩在总成绩中的作用。影响权重的因素主要是转换区间和标准差，当然如果使用的非标准化转换过程（即不遵循正态分布原则转换），则百分位区间的分配就会直接影响到权重问题。权重问题包含了分数分布信息量问题，当方差比例（即方差与分值区间的比例）变化过大时，会直接影响到分数分布的信息度，即拉大或缩减了差异性，使分数在综合分数中的影响度改变。

尽管权重会直接反映一个科目分数的作用，但并非左右总分数差异的直接和唯一的支撑量。总分数方差由各个科目方差和科目间协方差的组成，分值区间并不能直接反映方差，而且各个分数区间的方差影响都会不同。尤其是针对竞争性应用的高考分数，权重只能是一个感性影响量，不能由权重直接得出科目在总分中重要程度的结论，需要进行专门的数据分析和研究。