

第2章 基本统计学原理

2.1 一般统计指标

1. 最高分、最低分、平均分

这三项指标体现了事物的普遍性与特殊性。通过最高分、最低分这两个极端指标可以知道试卷出题是否超过了哪个极端，或者帮助教师发现该班级中的个别学生，因材施教。平均分体现了该班级考生在这次考试中的平均答题水平，一定程度上代表了该班级考生对该试卷学科知识的掌握程度。

2. 分段统计

分段统计是试卷分析工作中最常用也是最形象直接的指标。通过分段统计，可以一目了然地知道考生在这次考试中的答题情况，例如各分数段的人数和百分比，也可以从中方便地推算出该次考试的及格率、不及格率、优秀率等指标。

3. 标准差

标准差是一种表示分散程度的统计观念，对于学生的考试成绩而言，标准差愈大，表示学生的成绩比较分散，相差较大；标准差愈小，表示学生的成绩比较集中，相差不大。对于试卷分析而言，标准差越大，表示学生的成绩分布比较合理，标准差越小，表示试卷没有把不同的学生区分开。

标准差的简易计算公式：

假设有一组数值 x_1, \dots, x_N （皆为实数），其平均值为：

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

此组数值的标准差为：

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

一个较快求解的方式为：

$$\sigma = \frac{\sqrt{\sum_{i=1}^N x_i^2}}{\sqrt{N \left(\frac{\sum_{i=1}^N x_i}{N} \right)^2}} = \sqrt{\frac{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2}{N^2}}$$

2.2 难度、易度

难度和易度是两个相反的评价指标，求得其一即可。

易度是反映学生完成试题难易程度的指标。一道试题，如果大部分学生都能答对，则该题的易度就大；反之，如果大部分学生都答错，该题的易度就小。答对即通过，因此，易度又称通过率。易度的计算方法通常以答对某题的人数与总人数之比，或者以所有考生解答该题的平均值与该题的满分值之比表示。

易度的计算公式通常为两个：

1. $P=T/N$

$$P = \frac{\text{答对某题的人数}}{\text{参加测验的学生总数}}$$

式中的 P 为该题的易度， N 为参加测验全体学生数， T 为通过该题的学生数。如某试题有 50 名学生参加测验，而答对通过的学生为 40 人，则易度为 0.8。

2. $P = \frac{\bar{X}}{X_m}$

$$P = \frac{\text{学生在某题的得分平均值}}{\text{某题的满分值}}$$

式中 P 为该题易度， \bar{X} 为考生在该题的得分平均值， X_m 为该题的标准分。如某试题学生平均得分为 15 分，而该题满分值为 20 分，则该题的易度为 0.75。

这两个公式中的第一个公式常用于测试题是选择题、填空题和是非判断题的易度计算；第二个公式常用于简答论述题的易度计算，因为考生做这种题时的得

分不是固定的数，无法确定一个通过值，所以取平均数计算更科学合理。同时，从这两个公式也可看出，易度值越大，说明学生的得分越高，试题就越容易；反之，易度值越小，说明学生的得分越低，而试题越难。研究表明，试题过难或过易，即易度值过小或过大都是不当的，一般来说，易度值掌握在 0.5~0.7 之间的试卷是较为合理的。

2.3 区分度

区分度又叫鉴别力，它是测验学生实际水平的区分程度的指标，是一条衡量试题质量的重要标准。区分度高的试题可以把不同水平的学生区别开来；区分度低的试题则不能较好地鉴别学生之间的水平差异。也就是说，区别度高的试题能使学生的测验成绩明显拉开距离；区分度低的试题，可使学生的得分相差不大或没有什么规律。一份好的试卷应该具有良好的区分度，也就是说各个档次的考生应该适当地拉开距离，有所区分，实际水平高的考生应该得高分，实际水平低的考生应该得低分。

试卷的区分度有积极区分度和消极区分度两种。积极区分是指区分的方向与试卷总分的方向一致，而区分的方向与试卷部分的方向不一致的区分是消极区分。

试卷的区分度的取值范围在 -100 至 100 之间。如果区分度是负值，则表示该区分是消极区分；如果区分度为 0，则表示该测试题没有区分；如果区分度是正值，则表示该区分是积极区分。计算区分度的方法有多种，我们采用一种较简单的方法——分组法。分组法是通过比较学生高分组和低分组在该试题上的难度差别来确定区分度大小。具体做法是先将考生的总分由高到低排列，成绩最高的 27% 的考生组成高分组，成绩最低的 27% 的考生组成低分组，然后分别计算高分组和低分组的易度值，分别用 PH 和 PL 表示，则该题的区分度用 D 表示。

公式：
$$D = PH - PL$$

例如，00 计汉本班 45 名考生的数据结构试卷，按成绩排序后计算高分组的易度为 0.8667，低分组的易度为 0.2667，则区分度 $D = 0.8667 - 0.2667 = 0.6$ 。

对于选择题和是非判断题的区分度的计算一般采用上述公式，而对于问答题

可按下面的公式计算:

$$D = \frac{X_H - X_L}{N(H - L)}$$

其中, X_H 为高分组的得分总数, X_L 为低分组的得分总数, N 为参加考试的学生总数的 25%, H 为该题的最高得分, L 为该题的最低得分。对于较大型的问答题, 由于得分情况要比选择题复杂, 可用这个公式计算区分度值。区分度为多少时最合理, 不应该为多少, 我们习惯上按照表 2-1 去判定一份试卷的区分度符合哪个标准。

表 2-1 区分度范围及评价

| 区分度 | 评价 |
|-----------|------------|
| 0.4 以上 | 优秀 |
| 0.30~0.39 | 良好, 如能改进更好 |
| 0.20~0.29 | 尚可, 应该改进 |
| 0.19 以下 | 差, 必须淘汰 |

根据上表可知区分度在 0.19 以下的试卷是不能用的, 在 0.19 以上和 0.29 以下的题则需要修改或淘汰。区分度的最高限值为 100%, 但这是一个不可能达到的理论概念。大量的经验告诉我们一般情况下试卷的区分度在 0.4~0.7 之间为最佳。

2.4 试卷信度

信度根据误差来源分为再测信度、复本信度、内部一致性信度和评分者信度。

再测信度是对同一被试总体进行重复测量而得到的两组数据之间的一致性程度的量化指标; 复本信度要求构建与原测验各方面均相近或等值的平行测验(即复本), 计算被试总体在两次测验上所得分数的皮尔逊积差相关系数, 即得复本信度; 内部一致性信度是指考试内部所有试题的一致程度, 也就是一次考试所测内容的相关程度; 评分者信度考察评分者之间的一致程度。

对于高校专业课考试, 计算再测信度和复本信度的数据无法取得, 所以可以选取内部一致性信度进行评价。

内部一致性信度的计算方法有：分半法、库德—理查森公式及克龙巴赫所创的 α 系数公式。

用“分半法”评估试卷信度时，将涉及如何把一次考试分成两半，可以按奇偶题目分半，也可以随机分半，而采用不同的分法求出的信度系数会有所不同。

α 系数公式：

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum_{i=1}^k S_i^2}{S^2} \right)$$

式中， α ：信度系数， k ：考卷的试题数， S_i^2 ：全体考生第*i*题得分的方差， S^2 ：全体考生的考试总分的方差。

2.5 试卷效度

效度分为构想效度、内容效度、效标关联效度。

构想效度指考试分数能够说明心理学理论的某种结构或特制的程度，主要用于心理测验。而学科测验主要考虑内容效度和效标关联效度。

内容效度指选取的具有代表性的样本组成的考试内容是否能够恰当地代表教学内容总体。到目前为止，还没有一种切实可行的统计方法可以用来合理地估计试题取样的恰当程度，只能由有经验的教师、专家依据考试大纲与相应的双向细目表对每道试题进行比较分析来作出评估。

效标关联效度是指考试结果与效标的相关程度。

根据三种效度的定义，可见在评估高校课程考试的效度时，只能通过计算“效标关联效度”来定量地估计效度。

效标关联效度可用积差相关法求得：

$$r_{xy} = \frac{\sum xy - (\sum x)(\sum y)/n}{\sqrt{[\sum x^2 - (\sum x)^2/n][\sum y^2 - (\sum y)^2/n]}}$$

式中， r_{xy} ：效标关联效度， x ：考试分数， y ：效标分数。

按照教育测量学理论，效标应采用一次公认的客观的同类标准考试成绩。但

对于教师自编试卷，其可靠性和有效性均无法得到保证。

2.6 正态分布的判断

正态分布对试卷分析的作用有两个：一是系统中部分其他统计方法是建立在正态分布基础之上的；二是用于对考试成绩的解释。

正态分布是数理统计中最重要的一个概率分布，多数随机现象都服从或近似服从这个分布。正态分布的图形如图 2-1 所示。

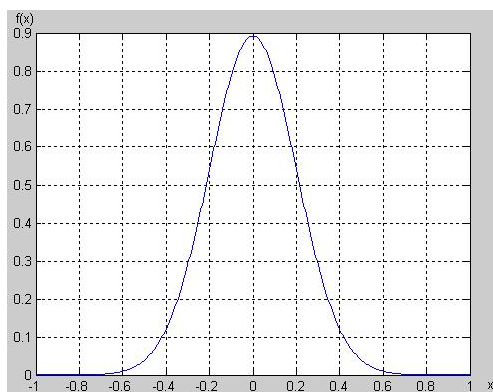


图 2-1 正态分布图

其曲线的数学表达式为：

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

这里 Y 表示变量 X 的高度或纵坐标(即代表 X 点上的无限小区间的频数密度)；

X 表示连续变量的任何一点；

μ 表示平均数；

N 表示总频数；

σ 表示标准差；

π 表示常数，约为 3.14159；

e 表示常数，即自然对数之底，约为 2.71828。

就课程考试成绩来说，在一次符合考生实际水平的考试中，成绩居中等水平的学生占大多数，而成绩特别好和特别差的学生均为少数，也就是说，学生的考分应服从或近似服从正态分布。因此，一份试卷质量高不高、适不适合考生的具体情况，首先应该看考分是否服从或近似服从正态分布，然后再看其他指标是否符合标准。

偏态量和峰态量是用以描述数据分布特征的统计量，在考察数据分布是否服从正态分布时，可以用偏态和峰态来作度量。用原始数据计算偏态量时，其公式为：

$$\alpha_1 = \frac{\sum(x - \bar{x})^3}{N \delta_x^3}$$

其中： α_1 表示偏态量， x 表示原始数据， \bar{x} 表示平均数， δ_x 表示标准差。

当 $\alpha_1 = 0$ 时，表明数据分布呈对称形，符合正态分布。

当 $\alpha_1 > 0$ 时，表明数据分布呈正偏态分布。

当 $\alpha_1 < 0$ 时，表明数据分布呈负偏态分布。

峰态量公式：

$$\alpha_2 = \frac{\sum(x - \bar{x})^4}{N \delta_x^4} - 3$$

其中： α_2 表示峰态量， x 表示原始数据， \bar{x} 表示平均数， δ_x 表示标准差。

当 $\alpha_2 = 0$ 时，表明数据分布呈正峰态。

当 $\alpha_2 > 0$ 时，表明数据分布呈高狭峰。

当 $\alpha_2 < 0$ 时，表明数据分布呈低阔峰。

2.7 自动评语的产生

(1) 正态分布：说明测试结果与学生的实际情况一致，各种难度的题目比例合适。

(2) 正偏态分布：说明试题难度偏高，难度较大的题目比例偏大。呈这种分布的试题有利于把成绩优秀的学生和成绩中等的学生区别开，但不利于将中等程度的学生和成绩较差的学生区分开。

(3) 负偏态分布：说明试题难度偏低，难度较低的题目比例偏大。呈这种分布的试题有利于把成绩较差的学生和成绩中等的学生区别开，但不利于将中等程度的学生和成绩优秀的学生区分开。

(4) 平坡型分布：（低阔峰）说明试题中各种难度的题目比例接近，梯度较大。呈这种分布的试题区分度较高，但分数之间的差异较大。

(5) 陡峭型分布：（高狭峰）说明试题中同等难度的题目较多，梯度偏小。呈这种分布的试题几乎不能将不同程度的学生区分开，分数分布过于集中。