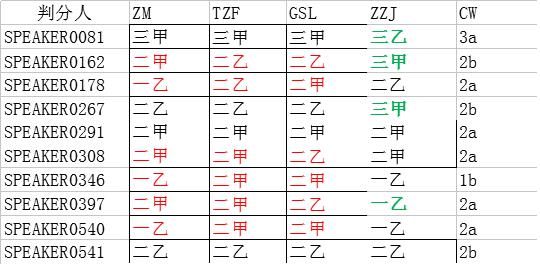
普通话水平评判的一致性与正确率评价

# 问题描述



# 数值处理

为通过数值指标对普通话评判水平进行量化比较，需要对普通话的使用能力进行数值化处理，以方便计算。

普通话水平数字化处理：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 普通话等级 | 一乙 | 二甲 | 二乙 | 三甲 | 三乙 |
| 分值 | 4 | 3 | 2 | 1 | 0 |

每个人对十个人的评分可以看作一个中的向量，例如ZM人的评价结果为：

而被评判人的真实成绩为：

所有的数值结果依法转化如下：

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 对象 | ZM | 评分 | TZF | 评分 | GSL | 评分 | ZZJ | 评分 | CW | 评分 |
| 1 | 三甲 | 1 | 三甲 | 1 | 三甲 | 1 | 三乙 | 0 | 三甲 | 1 |
| 2 | 二甲 | 3 | 二乙 | 2 | 二乙 | 2 | 三甲 | 1 | 二乙 | 2 |
| 3 | 一乙 | 4 | 二乙 | 2 | 二甲 | 3 | 二乙 | 2 | 二甲 | 3 |
| 4 | 二乙 | 2 | 二乙 | 2 | 二乙 | 2 | 三甲 | 1 | 二乙 | 2 |
| 5 | 二甲 | 3 | 二甲 | 3 | 二甲 | 3 | 二甲 | 3 | 二甲 | 3 |
| 6 | 二甲 | 3 | 二甲 | 3 | 二乙 | 2 | 二甲 | 3 | 二甲 | 3 |
| 7 | 一乙 | 4 | 二甲 | 3 | 二甲 | 3 | 一乙 | 4 | 一乙 | 4 |
| 8 | 二甲 | 3 | 二甲 | 3 | 二乙 | 2 | 一乙 | 4 | 二甲 | 3 |
| 9 | 一乙 | 4 | 二甲 | 3 | 二甲 | 3 | 一乙 | 4 | 二甲 | 3 |
| 10 | 二乙 | 2 | 二乙 | 2 | 二乙 | 2 | 二乙 | 2 | 二乙 | 2 |

# 正确率计算

每一次评价的结果要么正确要么错误，一个比较直接的正确率计算为一个人评价的正确次数比上总次数，可得到正确率如下表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 人 | ZM | TZF | GSL | ZZJ |
| 正确率 | 0.7 | 0.8 | 0.7 | 0.4 |

# 一致性评价

通常情况下，我们希望一个人的对普通话评价的好坏程度，应该和真实的好坏程度相一致。因为真实的好坏程度应该与大多数人对普通话的评价标准一致。也就是说，一个人如果一致性很好，那么如果他认为一段话说地很好，那么真实情况，那条普通话应该真的好，反之亦然。因此，在衡量一致性的同时，我们希望一个人的评价结果和真实情况相近。一个数学的表达就是，如果ZM的评价一致性程度很好，那么与这两个向量的距离应该很近，也就是：

应该较小。（上述距离选取为欧式距离）

计算结果：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 人 | ZM | TZF | GSL | ZZJ |
| 距离值 | 1.7321 | 1.4142 | 1.7321 | 2.4495 |

因为在考虑评价值和真实值之间一致性大小的时候。可以发现TZF的评价与大多数人（真实情况）较为吻合。

# 评价者间的一致性

为了探究两个评价者间的一致性，我们分别求出他们两两间的距离如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ZM | TZF | GSL | ZZJ |
| ZM |  | 2.6458 | 2.4495 | 3.3166 |
| TZF | 2.6458 |  | 1.7321 | 2.4495 |
| GSL | 2.4495 | 1.7321 |  | 3.3166 |
| ZZL | 3.3166 | 2.4495 | 3.3166 |  |

# 评价倾向

ZM的评价倾向可以如下计算为：

我们希望上述值更接近0，这样说明评价更加客观。

因为正确率和一致性是更为重要的指标，其中比较明显的是ZM 的上述指标为3，而他总共也就错了三个位置，我们认为ZM 的评价较为手松。

# 结论

TZF 具有较高的评价的正确性和一致性。同时GSL与TZF的看法较为一致，两人的观点更为相近。同时ZM 的评价存在一定过高的估计被评价人的真实水平的倾向。ZZJ因为具有较高的错误率，同时和其他三人的看法都存在较大差异，我们认为ZZJ 的评价是缺乏正确性的。