UNIVERSITY OF TORONTO CUP

Statistical Correction Scoring Guidelines

This method is based on the principles and techniques described by Cheng and Farmer[1], and further refined by Alex Marchi of McGill University. We are grateful for their contributions to the scoring model for this competition.

Overview

The purpose of employing this statistical correction is to encourage fairness and consistency across the board for all competitors. The model reflects variation in judging styles and scoring distributions, and is blind to the characteristics of any competitor.

Typically, a competition compares the scores of individual competitors against the scores of other competitors. With this correction, competitors are now comparing the difference between their score and the average score assigned by the judge who assigned it — reflecting how that particular oralist performed compared to other oralists that same judge evaluated. This allows competitors to detach themselves from worrying about "difficult" vs. "easy" judges, and promotes fair scoring without unduly constraining justices with lengthy rubrics and instructions.

Given the number of competitors that each justice will evaluate, and the total number of competitors and justices, it is extremely unlikely that this correction will result in significant changes to the rankings of teams — especially not those near the top. Teams are equally likely to have their scores increased as well as decreased, and an increase in one round may be matched by a decrease in another round. Any increases or decreases reflect actual differences in judging and are justifiable changes that promote fair and consistent scoring.

Corrected scores will be used to generate Day 1 rankings and determine which 8 teams move on to the elimination rounds. This correction method will not be used for Day 2, as there are too few competitors to effectively perform such methods. We are confident that the larger panels in those rounds will result in consistent judging for all competitors.

In the interest of disclosure, we will be providing all competitors with a summary of both their assigned scores and their adjusted scores after the competition closes.

---

[1] Cheng, E.K., Farmer, S.J. 2013. A Normalized Scoring Model for Law School Competitions. Green Bag 16(2D): 377-394. http://www.greenbag.org/v16n4/v16n4_articles_cheng_and_farmer.pdf

Technical Explanation

Definitions

You have i oralists and j judges.

Assume that scores given by judges fall within a normal distribution (commonly known as a "bell curve," where both extremely low scores and extremely high scores are rare).

Define μ to be the mean score assigned by all judges.
Define σ to be the standard deviation (how spread out the scores are; the average distance from the mean) of the scores assigned by all judges.
Define μ_j to be the mean score given by judge j.
Define σ_j to be the standard deviation of the scores given by judge j.

Scoring adjustment

Let i_{kj} be the raw score given to the i^{th} oralist in the k^{th} round by the j^{th} judge.

The "Marchi score" (M) is the normalized difference between the raw score and the average score given by that particular judge.

For the i^{th} oralist, we define the Marchi score as:
$M = (i_{kj} - \mu_j) / \sigma_j$

The "adjusted score" (A) is a comparison between the i^{th} oralist's judge and all other judges. If the average score assigned by that judge was lower than the average score assigned by all judges, the oralist's score for this round would go up, and vice versa.

For the i^{th} oralist, we define the adjusted score as:
$A = (M * \sigma) + \mu$

Example

The μ score of all judges is 80 with a σ of 10.
Judge Judy assigns a μ_j score of 70 with a σ_j of 10.
Jordan receives a raw score of i_{kj} = 60 in Round 1 when judged by Judy.

M = (60 - 70) / 10 = -1
A = (-1 * 10) + 80 = 70

Jordan's adjusted score in Round 1 is 70 (compared to his raw score, 60) to account for the lower score that Judy typically assigns compared to the overall average score across judges.

If this score increase seems large, remember that Jordan may encounter approximately eight judges over the course of four rounds (two per round), and that his partner may experience the same correction, potentially in the other direction. In order to realize an actual 10 point increase in Jordan's overall score, he would have to consistently encounter judges who assign lower scores than average. The adjustment Jordan receives in Round 1 would likely be countered by an opposite adjustment when he encounters a judge that assigns higher scores than average.