VIII. Beyond "THE RASCH MODEL"

All models are wrong. Some are useful. G.E.P.Box
Models must be used but must never be believed. Martin Bradbury Wilk

The Basic Ideas

We have thus far occupied ourselves entirely with the basic, familiar form of the Rasch model. I justify this fixation in two ways. First, it is the simplest and the form that is most used and second, it contains the kernel $(\beta_\nu - \delta_i)$ for pretty much everything else. It is the mathematical equivalent of a person throwing a dart at a balloon. Scoring is very simple; either you hit it or you don't and they know if you did or not. The likelihood of the person hitting the target depends only on the skill of the person and the "elusiveness" of the target. If there is one *The Rasch Model*, this is it.

This simple difference between skill (*ability*) and elusiveness (*difficulty*) lumps together a multitude of factors that might, or might not, influence the rate of success at bursting balloons: size, distance, wind, gravity, gender, age, noise, dart quality, or darter's state of sobriety. Leaving out factors does not give us leave to ignore them. Some we might want to prove don't matter so that we can continue to ignore them or they might be the very things we want to analyze with our measures. For example, does proficiency at bursting balloons with darts vary with the gender or age of the person and is it amenable to instruction? Other factors like wind, lighting, or decibel level, are just annoyances that we might choose to deal with by controlling the conditions so that they are the same for everyone or that we might choose to tinker with to better tailor the task to the participant.

These extraneous factors may very well affect the difficulty we find for the agent and the measure we get for the object, but successful measurement depends on them not *interacting*, in an analysis of variance sense, with either, which means that the person who is best at throwing darts on a calm day is also best on a windy day. The difficulty ordering of the balloons must not depend on, say, the sobriety of the dart throwers and the ordering of the dart throwers must not depend on, say, the color of the balloons[1]. But we must establish this (or make a persuasive argument) for every situation. That was the last chapter; in this chapter we are expanding our repertoire of Rasch models to deal with more situations.

Rather than treating each balloon as a unique "fixed effect" and estimating a difficulty specific to it, there may be other types of effects for which it is more effective and certainly more parsimonious to represent the difficulty as a composite, i.e., linear combination of more basic factors like size, distance, drafts. With estimates of the relevant effects in hand, we would have a good sense of the difficulty of any target we might face in the future. This is the idea behind Fischer's (1973) *Linear Logistic Test Model* (LLTM), which dominates the Viennese school and has been almost totally absent in Chicago.

Let me switch to a baseball pitcher attempting to throw a baseball through the strike zone. This is still a dichotomous item (ball or strike) but scoring is more problematic than, Did the balloon burst? To deal with the scoring, we bring in a judge to arbitrate between the pitcher and the strike

---

[1] While this may seem like a throw-away but there could be people who have trouble hitting red balloons against green backgrounds.

zone. The pitcher's task is to throw the ball in the strike zone[2] and the umpire's task[3] is to determine whether or not the pitcher succeeded[4]. Whether or not a pitch is ruled a strike is determined by the proficiency of the pitcher, the location of the strike zone, and the harshness of the judge[5]. This is Linacre's *Many-Facet model* (Linacre, 1994, 2004). In this example, "many" is three.

Our baseball game thus far as had the batter as a rather passive observer. The batter's task is to hit the ball and advance as many bases as possible. A mathematically-endowed manager might expect that, for any 100 batters, 68 will get zero bases, 23 one base[6], 5 two bases, 1 three bases, and 3 four bases. The situation involves the proficiency of the batter, the difficulty of hitting the pitch, and the difficulty of getting from one base to the next. Andrich's *Rating Scale Model* (Andrich, 1978).

I could just as well have said, in the last paragraph, that this is the essence of the Masters *Partial Credit Model* (Wright and Masters, 1982) rather than the Andrich Rating Scale Model (*RSM*). It could be handled with either. But if we imagine a rather odd game[7] where the bases are moved periodically, say sometimes spaced at 90 feet, sometimes at 60 feet, sometimes at 150 feet, the Rating Scale Model would not apply. Because the different arrangements change the likelihoods of the task scores (0, 1, 2, 3, or 4), we need different parameters in the model for each arrangement of the bases. Then this is the Partial Credit Model (*PCM*).

With either *RSM* or *PCM*, we could, and probably should, include impartial umpires on the bases to arbitrate the score in disputed situations. This suggests adding Linacre to Andrich and Masters giving us the *Many-Facet Rating Scale* and *Many-Facet Partial Credit Models*. Continuing right along, we can add Fischer's (1973) linear logistic test model (*LLTM*) on top of any of these models and decompose the task difficulties and the referee's harshness into fundamental components, perhaps using the pitcher's velocity, variety of pitches, distance and height of the outfield walls, and age and experience.

Finally, Rasch (1960) started with the Poisson model circa 1950 with his original problem in reading remediation, for seconds needed to read a passage or for errors made in the process. Andrich (1973) used it for errors in written essays. It could also be appropriate for points scored in almost any game. The Poisson can be viewed as a limiting case of the binomial (see Wright, 2003 and Andrich, 1988) where the probability of any particular error becomes small (i.e., $\beta_v - \delta_i$ large positively) enough that the δ_i and the probabilities are all essentially equal.

These are a few of my favorite models and I have avoided the math about as long as possible.

---

[2] Officially the strike zone is the 3-dimensional region above home plate and between the batter's shoulders and knees. The locations of the knees and shoulders vary with the height and stance of the batter and opinion of the umpire..

[3] The umpire is viewing the strike zone from a slightly oblique angle and has a few hundredths of a second to determine if the pitch passes through any part of the zone. In addition to the problem in psychophysics, where the batter's knees start or shoulders stop is somewhat subjective.

[4] This may or may not be what the pitcher is actually trying to do on a given pitch but I am making a measurement analogy not coaching baseball.

[5] And a lot of other rules that I'm not concerned with at the moment.

[6] I have lumped walks in with singles, but they don't have quite the same value, although in either case the batter reaches first base. And I am ignoring errors and fielder's choice.

[7] Maybe not that odd. Outside the bases, the dimensions of the field are less rigorously controlled, which can influence the likelihood of multi-base hits.