David's Musings: On The Good Judgment Project, and on being the 365,625th most famous person in history.

The Good Judgment Project (GJP) has roots in Phil Tetlock's study "Expert Political Judgment", which may be best known for its conclusion that the "expert" forecasters he studied were often hard-pressed to do better than the proverbial dart-throwing chimp. Tetlock and colleagues believe that forecasting tournaments are the best way to compare forecasting ability; and that participants can improve their forecasting skills through a combination of training and practice, with frequent feedback on their accuracy. Combining training and practice with what GJP's research suggests is a stable trait of forecasting skill seems to produce the phenomenon that GJP calls "superforecasters". These have been so accurate that they even outperformed the forecasts of intelligence analysts who have access to classified information. (Extracted from the (public) Project blog http://goodjudgmentproject.com/blog/ with minor edits.)

Partly for my own interest, and partly to have material for my "Probability in the Real World" course, I am participating in this GJP. Participants in teams are asked to assess the probability (as of today) of specified geopolitical events happening before a specified deadline. For instance ``Before 1 May 2014, will China confiscate the catch or equipment of any foreign fishing vessels in the South China Sea for failing to obtain prior permission to enter those waters?" Of course you are not supposed to just guess an answer -- rather, you are supposed to search for relevant news and analysis by other people, and then (like a jury in a trial) assess and discuss this evidence to make your judgment. And, of course, you update probabilities as news (or no news) appears.

How is this relevant to an undergraduate Statistics course? For a start, there's the practical issue of how one should "score" the accuracy of probability assessments in general, and those changing over time in particular; and the philosophical point that one can indeed judge relative accuracy of different forecasters, but not their absolute accuracy. It also turns out, via a kind of statistical detective story examining the nuances of the GJP's scoring rules, that one could actually "game the system" by announcing dishonest probabilities under some circumstances, but I won't publicly say how to do so.

If you join the GJP there is a lengthy orientation, including tests of your "cognitive style", of your background factual knowledge of obscure geopolitics, and your ability to assess your own level of knowledge (as you might guess, most people are over-confident).

And a briefing on cognitive biases, in the spirit of Kahneman's *Thinking, Fast and Slow*. All are quite fascinating, to me personally.

Changing (at first sight) topics, when Andrew Gelman writes "This book is a guaranteed argument-starter. I found something to argue with on nearly every page" then I couldn't resist looking at the book: "Who's Bigger?: Where Historical Figures Really Rank" by Steven Skiena and Charles Ward. They say they have taken all people, dead or alive, with Wikipedia entries (about 700,000) and ranked them in four overlapping ways (Significance, Fame, Celebrity, Gravitas) using statistical analyses based on underlying data such as Wikipedia page length, PageRank applied to Wikipedia cross-references, and news frequency. Much of the book consists of chapters on different categories of people -- Modern World Leaders, Sports Players, Performing Arts, etc. -- naming and briefly discussing the top-ranked and some surprisingly low-ranked individuals.

For a comparison, Google Ngrams is a cool tool without pretensions to be more than what it is -- "a graph showing how (relatively frequently) given phrases have occurred in a corpus of books over the selected years". It has many fun uses -- for instance, to discover whether writers treat the word "data" as singular or plural, just check the relative frequencies of "data are" versus "data is" -- but the many potential misuses are clearly the responsibility of the user, not the tool provider.

So the "Who's Bigger" project is potentially interesting to me as an analogous tool, because they have a website whoisbigger.com where they claim that "for every person in Wikipedia" you can enter their name and find a page with their ranking. Sounds very interesting. To check it out, I went to Google Scholar to find there the 5 most highly cited authors tagged with "label:probability", and typed these into the whoisbigger.com search. Of these 5, only Richard A. Davis doesn't have a Wikipedia entry, so was not under consideration; Frank Kelly and David Freedman are taken to be different people with those names; Terrence Fine and "Paul Erdos" return no page. Somewhat later, I discovered that cutting-and-pasting the exact Hungarian accent for "Paul Erdos" from Wikipedia does fetch a page identifying the correct person -- but with no understandable data. Persevering, no variant of "David A. Freedman" or " David Freedman (statistician)" worked, though finally "Frank Kelly (mathematician)" identified the correct person and ranked him as 95,823 in Fame. It does better for famous historical figures -- ranking Jacob Bernoulli as 14,401 and Andrey

Kolmogorov as 5,177 sounds reasonable. But from this limited foray I would regard their rankings, outside the top few thousands, as absurdly incomplete and unreliable. If only they had modeled the project on Google Ngrams, and put more effort into making the website actually do what it claims, with proper name disambiguation, and less into their own *We interpret Stephen King as the Charles Dickens of our time* style of commentary.

As the authors write, their analysis *treats people as memes ...... there are several forces acting on our collective memory to determine which figures get preserved for posterity*. And in many ways they are aware of the defects of such analysis: of using the English language Wikipedia, that Wikipedia entries over-represent contemporary people, for instance.

To me their key claim, and their justification for the project, is that *our rankings show an excellent correlation with published rankings by human experts, and correlate better with these experts than they do among themselves*. That's interesting to me, because it suggests projects for my undergraduate course. Repeat

such a comparison for people in some category that interests you. Or look at historical figures and see if their elaborate analyses seem better than simply taking the length of the article in the final printed Encyclopedia Britannica.

This spotlights a certain conceptual circularity in the project. Wikipedia is after all the product of a crowd of contributors, and the length of a particular article already is influenced by this crowd's consensus opinion of the subject's importance. That this can be used as a broader consensus measure of significance is hardly insightful. But the key claim above is another contribution to the long-running "wisdom of crowds versus experts" debate, as was our opening quote from the GJP.

David Aldous,
Berkeley.

Editor's Note: This is the ninth installment of a regular opinion   column.

Past Conferences, Meetings and Workshops

ISOSS Silver Jubilee Celebration Meeting, 22 November 2013, Lahore, Pakistan

From left to right: Professors Munir Ahmed, Abdus Salam Hirai, Shahjahan Khan, Javed Siddiqi, and Muhammad Hanif Mian,

On 22 November 2013, a large number of statisticians as well as many patrons and supporters of statistics gathered at the ISOSS House in Lahore, Pakistan, to celebrate the 25 years of professional services of the Islamic Countries Society of Statistical Sciences (ISOSS). The Society was established in Lahore, Pakistan in 1988 in the First Islamic Countries Conference on Statistical Sciences (ICCS-1) to promote statistical sciences and their diverse applications in various sectors of lives and societies in the Islamic

countries. Its membership is open to statisticians of all orientations.

The founding President of ISOSS, Professor Munir Ahmad, gave a brief background and history of ISOSS and its landmark contributions, emphasizing the need for the creation of provincial and local Societies of statisticians in Pakistan to organize national conferences and promote other statistical activities. The founding Secretary General of ISOSS, Professor Akhlaq Ahmad, and two most senior statisticians of Pakistan, Professor